Opis predmetu zákazky Príloha č. 1

[Obsah](#_Toc110518766)

[1 Referencie 4](#_Toc110518767)

[1.1 Zoznam použitých skratiek a pojmov 4](#_Toc110518768)

[1.2 Používané skratky odborov NBS 4](#_Toc110518769)

[2 Predmet verejného obstarávania 6](#_Toc110518770)

[2.1 Úvod 6](#_Toc110518771)

[2.2 Rozsah dodávky 6](#_Toc110518772)

[2.3 Realizácia riešenia NBS DWH 7](#_Toc110518773)

[3 Opis predmetu zákazky 8](#_Toc110518774)

[3.1 Východisková situácia 8](#_Toc110518775)

[3.1.1 Procesy a systémy pre spracovanie dát 8](#_Toc110518776)

[3.1.2 Súčasná architektúra aplikačnej vrstvy práce s dátami 13](#_Toc110518777)

[3.1.3 Pilot DWH 21](#_Toc110518778)

[3.2 Technické požiadavky 26](#_Toc110518779)

[3.2.1 Úvod 26](#_Toc110518780)

[3.2.2 Požadovaná architektúra aplikačnej vrstvy práce s dátami 26](#_Toc110518781)

[3.2.3 Logická architektúra 27](#_Toc110518782)

[3.2.4 Fyzická architektúra 28](#_Toc110518783)

[3.2.5 Aplikačná architektúra 32](#_Toc110518784)

[3.2.6 Metodika vývoja 39](#_Toc110518785)

[3.2.7 Relačná časť NBS DWH 42](#_Toc110518786)

[3.2.8 DataLake časť NBS DWH 48](#_Toc110518787)

[3.2.9 Informačná bezpečnosť 50](#_Toc110518788)

[3.2.10 Iniciálny load 51](#_Toc110518789)

[3.2.11 Migrácia dát z projektu Pilot DWH 52](#_Toc110518790)

[3.2.12 Metodika fyzického vývoja a inštalácie na produkciu 52](#_Toc110518791)

[3.2.13 Nástroje na podporu prevádzky 53](#_Toc110518792)

[3.2.14 Nástroje pre Data Governance a riešenia Master Data Managamentu 65](#_Toc110518793)

[3.3 Biznisové požiadavky 65](#_Toc110518794)

[3.3.1 Zdrojové (vstupné) systémy 65](#_Toc110518795)

[3.3.2 Reportingová oblasť 69](#_Toc110518796)

[3.3.3 Výstupné dátové oblasti 73](#_Toc110518797)

[3.3.4 Funkčné požiadavky 76](#_Toc110518798)

[3.4 Požiadavky na organizáciu a výstupy projektu 77](#_Toc110518799)

[3.4.1 Projektové riadenie 77](#_Toc110518800)

[3.4.2 Prístup k projektu a RACI matica 77](#_Toc110518801)

[3.4.3 Pravidlá pre riadenie komunikácie 78](#_Toc110518802)

[3.4.4 Projektový plán 82](#_Toc110518803)

[3.4.5 Požadované výstupy projektu 84](#_Toc110518804)

[3.4.6 Akceptačné kritériá 90](#_Toc110518805)

[3.4.7 Riziká projektu 90](#_Toc110518806)

[4 Opis predmetu zákazky pre služby podpory prevádzky a rozvoja 91](#_Toc110518807)

[4.1 Úvod 91](#_Toc110518808)

[4.2 Požiadavky na služby aplikačnej podpory prevádzky NBS DWH 91](#_Toc110518809)

[4.2.1 Zoznam činností vykonávaných v rámci Služieb podpory prevádzky 91](#_Toc110518810)

[4.2.2 Parametre kvality poskytovanej služby podpory prevádzky 92](#_Toc110518811)

[4.3 Požiadavky na služby rozvoja systému 93](#_Toc110518812)

[4.3.1 Požiadavky na realizáciu zmien na základe požiadaviek verejného obstarávateľa 93](#_Toc110518813)

[4.3.2 Parametre kvality poskytovanej služby rozvoja systému 93](#_Toc110518814)

[4.3.3 Servisné služby čiastkového plnenia 93](#_Toc110518815)

[5 Prílohy 94](#_Toc110518816)

[5.1 A – Zdrojové systémy NBS DWH 94](#_Toc110518817)

[5.2 B – DataMarty NBS DWH 94](#_Toc110518818)

[5.3 C – Metodika k predpisu NBS č.20/2020 o projektovom riadení 94](#_Toc110518819)

[5.4 D – RACI matica 94](#_Toc110518820)

[5.5 E – Projektový plán 94](#_Toc110518821)

[5.6 F - Zmluva o prevádzke, údržbe a rozvoji systému 94](#_Toc110518822)

[5.7 G -Špecifikácia Servisných služieb čiastočného plnenia 94](#_Toc110518823)

# Referencie

## Zoznam použitých skratiek a pojmov

|  |  |
| --- | --- |
| Skratka / Pojem | Vysvetlenie |
| API | Aplikačný interface |
| BI | Business intelligence. Skratka a teminus technikus zväčša používaný pre definovanie rámca reportingu a analytických nástrojov. |
| DL = Data Lake | Časť NBS DWH založená na BigData technológiách pre spracovanie neštruktúrovaných dát |
| DWH | Dátový sklad |
| ETL | Extraction – Transformation – Load, proces spracovania a transformácie dát do DHW |
| HW | Hardware-Hardvér |
| NBS | Národná banka Slovenska |
| NBS DWH | Celobankový Ddátový sklad NBS. Pozostáva z dvoch častí:  1.Relačnej časti  2. DataLake |
| ODI | Oracle Data Integrátor |
| OUT | Výstupná vrstva relačnej časti NBS DWH |
| PL | Prezentačná vrstva relačnej časti NBS DWH |
| RBÚZ | Register bankových úverov a záruk |
| RS | Register subjektov |
| SW | Software-Softvér |
| ŠZP | Štatistický zberový portál |
| TL | Transakčná vrstva relačnej časti NBS DWH |
| UID | Jedinečný prístupový identifikátor/užívateľské meno |
| VO | Verejné obstarávanie |
| Pilot DWH | Pilotný projekt DWH |

Tabuľka 3: Zoznam použitých skratiek a pojmov

## Používané skratky odborov NBS

| Skratka odboru | Názov odboru |
| --- | --- |
| OBO | Odbor bankových obchodov |
| ODB | Odbor dohľadu nad bankovníctvom |
| ODK | Odbor dohľadu nad kapitálovým trhom |
| OEM | Odbor ekonomických a menových analýz |
| OFI | Odbor finančných technológií a inovácií |
| OFR | Odbor finančného riadenia |
| OFS | Úsek dohľadu a ochrany finančného spotrebiteľa |
| OIA | Odbor interného auditu |
| OIT | Odbor informačných technológií |
| OPD | Odbor dohľadu nad poisťovníctvom a dôchodkovým sporením |
| OPH | Odbor riadenia peňažnej hotovosti |
| OPM | Odbor finančnej stability |
| OPS | Odbor platobných systémov |
| ORR | Odbor riadenia rizika |
| OST | Odbor štatistiky |
| OVM | Odbor výskumu |
| OVO | Odbor vysporiadania bankových obchodov |

Tabuľka 4: Skratky odborov

# Predmet verejného obstarávania

## Úvod

Predmetom verejného obstarávania je výber uchádzača, ktorý v súčinnosti s verejnýcm obstarávateľaom (ďalej aj ako „NBS“) dodá riešenie centrálneho celobankového dátového skladu (ďalej aj ako „NBS DWH“). Riešenie novej dátovej integrácie je navrhnuté ako jeden dátový ekosystém v kombinácii dvoch platforiem:

* DataLake
* Relačná časť

Uchádzač musí vyvinúť/dodať nové komponenty dátového skladu, ktoré budú:

* Integrovať dáta z existujúcich informačných systémov NBS a z externých systémov
* Poskytovať operatívny reporting, meranie výkonnosti a podporu analytických procesov nad dátami pre potreby jednotlivých organizačných zložiek spoločnosti.
* Podporovať prevádzku systému
* Umožnia samostatný rozvoj NBS DWH pracovníkmi IT a ďalších organizačných zložiek spoločnosti

Ďalej musí uchádzač v maximálnej miere vyžiť výsledky projektu Pilot DWH, ktorý mal za úlohu otestovať možnosti niektorých technológií. Časti projektu Pilot DWH, ktoré je nutné prepoužiť sú:

* Technickú infraštruktúru – Oracle DB, MS SQL/TOLAP a PowerBI
* Know-how pri extrakcii dát zo ŠZP a RBUZ
* Oracle Data Integrator ako ETL nástroj pre NBS DWH ako výsledok interného výberu.

## Rozsah dodávky

Súčasťou dodávky uchádzača je splnenie a dodávka nasledovných úloh:

1. Vypracovať detailnú analýzu požiadaviek
2. Vypracovať návrh dátového modelu na základe detailnej analýzy biznisových požiadaviek
3. Dodať riešenie, ktoré bude v prípade spracovania osobných údajov pokrývať celý životný cyklus administrácie osobných údajov od evidencie až po ich likvidáciu (vrátane vymazania osobných údajov na základe žiadosti dotknutej osoby).
4. Realizácia integrácie na zdrojové systémy, vrátane pravidelného nahrávania dát do dátového skladu (kap. 3.3.1)
5. Vypracovať návrh kapacitného plánu
6. Dodať odporučenia konfigurácie DataLake a dodávka všetkých potrebných licencií tejto časti.
7. Inštalácia SW zabezpečenia (DataLake časť DWH, Frameworky, ETL nástroja).
8. Prevádzka a podpora SW zabezpečenia (DataLake časť DWH, Frameworky) počas trvania projektu.
9. Vypracovať návrh a implemenovať nové Knowledge Modulov ODI
10. Implementovať dátový model a dodať inštalačné scripty
11. Implementovať ETL úlohy a dodať inštalačné nastavenia
12. Implementovať reporty a analýzy (kap. 3.3.2)
13. Implementovať OUT objekty a transformácie dát do požadovaného formátu pre koncového odberateľa (kap. 3.3.3)
14. Dodať nastavenia BI nástroja
15. Dodať nasledovné frameworky a nastavanie metadát týchto frameworkov:
    * pre podporu správu, riadenie a monitoring aktualizačných úloh
    * pre správu číselníkov, mostíkov, referenčných dát,  hierarchií jednotlivých dimenzií a manuálnych vstupov
    * pre nahrávanie veľkého počtu a rôznych súborov
    * pre správu, automatické spúšťanie a vyhodnocovanie dátovej kvality
    * pre správu prístupov (riadenie bezpečnosti) k dátam
    * pre správu dokumentácie k DWH a iných metadát (o modeli, o reportoch, mapovaní zdroj-cieľ) pre správu riadenia dátovej retencie
16. Poskytnúť projektové riadenie zastrešujúce dodávku dátového skladu
17. Vytvoriť a odovzdať technickú dokumentáciu implementovaného riešenia (kap. 3.4.6)
18. Zaškoliť obsluhu a používateľov DWH (kap. 3.4.6)
19. Poskytnúť súčinnosť pri inštalácii a prevádzke neprodukčných prostredí DEV a TEST dodávaného informačného systému
20. Poskytnúť súčinnosť pre podporu produkčného prostredia PROD informačného systému dátový sklad (súčasťou dodávky uchádzača je aj prevádzkový manuál) - (kap. 4.2 a 4.3)
21. Poskytnúť súčinnosť pri inštalácii riešenia na PROD prostredia (relačná časť, DataLake časť, aplikačné servery). Pod pojmom poskytnúť súčinnosť sa pre účely tejto zákazky myslí aj dodávka inštalačných materiálov pre NBS prostredie.
22. Dodať licencie pre SW špecifikáciu pracovnej stanice vývojára (rozšírenie oproti poskytnutým licenciám NBS)
23. Ďalšie podľa detailného popisu v kapitole 3.2 (technické požiadavky)

Súčasťou dodávky uchádzača NIE sú nasledovné časti:

1. Dodávka HW serverov – zabezpečí NBS v konfigurácií po dohode s vybratým uchádzačom.
2. Dodávka licencií produktov Oracle (DB, ODI) a Microsoft (DB, PoweBI). Zabezpečí NBS, alebo už existujú v NBS.
3. Inštaláciu databázy pre relačnú časť DWH zabezpečí NBS (ako DB, tak aj PowerBI)
4. Inštaláciu a prevádzku virtuálnych strojov s Linux OS zabezpečí NBS (pre DataLake časť DWH).
5. Inštaláciu a prevádzku virtuálneho stroja s Tomcat aplikačnom serverom zabezpečí NBS.
6. Dodávka licencií aplikácie pre Dátového modelovanie.
7. Inštalácia riešenia na PROD prostredia je kompetencii NBS na základe súčinnosti uchádzača.

## Realizácia riešenia NBS DWH

Verejný obstarávateľ požaduje realizáciu riešenia NBS DWH do 36 mesiacov odo dňa nadobudnutia účinnosti zmluvy o dielo. Realizácia bude mať definované fázy dodávky s jednotlivými iteráciami (biznis výstup) s postupným pribúdaním funkcionality. Fázovanie dodávky je navrhnuté v kapitole 3.4, uchádzač je povinný toto fázovanie odborne pripomienkovať.

# Opis predmetu zákazky

Táto kapitola opisuje celkový predmet zákazky a delí informácie do týchto častí:

* Východiská situácia – opis stavu prípravy a zberu dát pre štatistické účely v NBS
* Požiadavky na riešenie NBS DWH:
  + Technické požiadavky
    - Požiadavky na Relačnú časť a DataLake DWH
    - Požiadavky na podporné frameworky (podpora prevádzky a správy dát)
    - Bezpečnostné požiadavky
  + Biznisové požiadavky
    - Integrácie zdrojov
    - Interný reporting NBS
    - Výstupné oblasti NBS DWH pre tretie strany
  + Projektové požiadavky, t. j. požiadavky na organizáciu a výstupy projektu
    - Projektový plán
    - Požadovaná dokumentácia a výstupy projektového manažmentu

## Východisková situácia

Východiskový stav je ďalej popísaný z pohľadu procesov spracovania dát, hlavných systémov využívaných v NBS v súvislosti so spracovaním dát. Zvláštna pozornosť je venovaná popisu aktuálnej architektúry systémov a nástrojov používaných pre prácu s integrovanými dátami, ich reportingom a analýzou. Dôležitým faktorom pre budovanie cieľového riešenia je potom existujúce pilotné riešenie dátového skladu, ktorého niektoré komponenty sú k dispozícii pre ďalšie využitie.

### Procesy a systémy pre spracovanie dát

NBS je pre výkon svojej činnosti v prevažujúcej miere závislá na množstve externých dát zbieraných od subjektov na finančnom a nefinančnom trhu a externých poskytovateľov ekonomických informácií, a to na mikroekonomickej úrovni (dáta o konkrétnych subjektoch a finančných inštrumentoch), ako aj makroekonomickej úrovni (makro dáta – agregované ukazovatele o ekonomickej situácii a trendoch v jednotlivých sektoroch trhu). Tieto dáta sú zbierané, analyzované, čiastočne sa odvodzujú nové dáta ako produkt makroekonomických prognóz a výskumov. Niektoré dáta sú ďalej zdieľané v rámci medzinárodných inštitúcií (najmä ESAs, v rámci Eurosystému alebo Jednotného mechanizmu dohľadu) alebo aj s verejnosťou. Menšia časť dát vzniká primárne v NBS ako produkt vlastných aktivít NBS, napr. registre o subjektoch na trhu v rámci dohľadu a ďalšie dáta v rámci obchodných procesov, riadenia peňažnej hotovosti alebo finančného riadenia spoločnosti. Preto ďalej popisujeme spracovanie dát v týchto oblastiach, ktoré majú osobitný prístup k spracovaniu dát:

* Produkcia štatistických dát
* Hlavné procesy produkujúce dáta
* Analytické procesy
* Podporné procesy

#### Produkcia štatistických dát

Zber dát pre štatistické spracovanie sa deje v NBS niekoľkými kanálmi:

1. Proces zberu a produkcie štatistických dát je v NBS riešený primárne v systéme **ŠZP** (Štatistický zberový portál), pričom je za tento proces zodpovedný predovšetkým Odbor štatistiky. Systém ŠZP je zákazkové riešenie od externého zhotoviteľa na mieru pre potreby NBS. Štatistické dáta sú však zbierané aj formou e-mailových príloh, systémom RBÚZ a RS, ktoré sú vyčlenené z pôsobnosti odboru štatistiky – viď ďalej.
2. ŠZP umožňuje definovanie používateľských, štandardných výstupov s exportom do CSV a XLSX pre ďalšie spracovanie (typicky pre zabezpečenie lokálnych požiadaviek) a automatizovanú výmenu niektorých dát s ESMA (XML štandard) a EBA/EIOPA (v štandarde **XBRL**).
3. Pre účely dohľadu nad finančným sprostredkovaním a finančným poradenstvom sa pre zber dát využíva aj systém **RegFAP**, ktorého vlastníkom je odbor ODK a je vyvíjaný na odbore OFI.
4. Zber dát pre niektoré agendy je však stále riešený aj manuálne cez e-mail a dáta v **MS Excel**.
5. Zber detailných dát pre AnaCredit a ich výmena s ECB sú riešené samostatným systémom **RBUZ/RS** (Register bankových úverov a záruk / Register subjektov). Tento systém bol vytvorený pre účely dohľadu, nie štatistiky, ale z povahy pravidelného spracovania dát od vykazujúcich subjektov sleduje rovnaký proces. Z pohľadu ECB je AnaCredit tiež štatistický systém, no v NBS nebol takto implementovaný.



Obrázok 1: Oblasť spracovania dát – Produkcia štatistických dát

Uvedené riešenia pre zber dát od vykazujúcich subjektov a ich ďalšie spracovanie však nezahŕňajú nástroje pre reporting, analýzy a systematickú distribúciu výstupov. Za týmto účelom vyvinul odbor štatistiky pre svoje vlastné potreby riešenie Apoštol s analytickým nástrojom Qlik Sense, ktoré malo za cieľ suplovať neexistujúcu analytickú platformu (resp. DWH/BI).

Existujú používatelia, ktorí pristupujú k dátam priamo cez SQL, R resp. Python.

V NBS prebiehajú alebo sa plánujú tieto súvisiace aktivity:

* Moderná štatistika – konsolidácia agendy zberu dát na OST do systému ŠZP
* Upgrade ŠZP (vyššia flexibilita definovania štatistických výkazov a výstupov, zlepšenie výstupného rozhrania pre dátovú integráciu a ďalšie...)
* Zber detailných dát o úveroch fyzických osôb a vytvorenie riešenia pre register fyzických osôb
* ŠZP ESAs - presun časti funkcionality štatistického procesu zameraného na výkazy vo formáte XBRL zo ŠZP do nového dedikovaného systému ŠZP ESAs

#### Ďalšie procesy produkujúce dáta

Niektoré dáta vznikajú vnútri NBS pri výkone niektorých hlavných činností, ktoré nespadajú do oblasti produkcie štatistických dát, ako sú:

* Dohľad (zodpovednosť za vedenie rôznych registrov a údaje o individuálnych procesoch/konaniach so subjektmi)
* Bankové operácie (údaje o vykonaných obchodoch – systém IBFO/WSS)
* Riadenie rizika (analytické systémy Riskhouse, Credit Manager a Risk Manager)
* Riadenie hotovosti (údaje o stave a obehu hotovosti – systém EZO)
* Platobné systémy (údaje o transakciách – systémy SIPS, Target2)

Pre každú agendu existuje špecializovaný informačný systém, ktorý poskytuje aj operatívne výstupy. V prípade využívania spoločných informačných systémov na úrovni Európskej únie (IMAS pre procesy dohľadu nad významnými bankami, Target2 – zúčtovanie veľkých platieb) sú k dispozícii súčasne aj analytické reporty a výstupy pre uspokojenie dátových potrieb relevantných odborov.



Obrázok 2: Oblasť spracovania dát – Hlavné procesy produkujúce dáta

Dáta jednotlivých odborov (hlavných procesných oblastí) sú spracúvané a využívané viac menej pre vlastné potreby každého príslušného odboru a sú zdieľané iba obmedzene. Výnimkou sú spolupracujúce odbory Bankových operácií a Riadenia rizika, ktoré majú vytvorený vlastný analytický systém Riskhouse, ktorý napĺňa požiadavky týchto odborov bez potreby ďalšej konsolidácie s inými systémami.

Napriek malému prieniku týchto dátových oblastí prináša neexistencia jednotnej integrovanej (resp. analytickej) platformy komplikácie užívateľom niektorých odborov v oblastiach:

* Manuálna a netransparentná príprava dát pre reporting (mimo existujúcich analytických platforiem)
* Chýbajúci nástroj pre tvorbu vlastných reportov a analýz
* Duplicitné spracovanie externých dát
* Neznalosť a nemožnosť analyzovať dáta v kontexte dát z iných odborov
* Chýbajúce prepojenie registrov a IDAS/ASDR s vykazovanými údajmi dohliadaných subjektov z oblasti štatistiky

Aktuálne strategické aktivity NBS v tejto oblasti sú:

* ASDR – Implementácia jednotného agendového systému pre dohľad a reguláciu (proces manažment dohľadu), ktorý má nahradiť IDAS
* DMS / ECM – Implementácia integrovaného systému pre správu dokumentov (DMS) a Enterprise Content Managementu

#### Analytické procesy

Osobitná skupina procesov z pohľadu spracovania dát sú:

* Finančná stabilita
* Menová politika
* Výskum

Tieto procesy sú konzumentom dát publikovaných z procesu štatistiky a ďalej množstva externých ekonomických dát a zdrojov, ktoré sú predmetom rôznych analýz, modelovaní a predikcií. V každom prípade sú tieto procesy životne závislé na integrovaných dátach a súčasne poskytujú nové dáta, respektíve obohacujú agregované štatistické dáta o odvodené ukazovatele a predikcie.



Obrázok 3: Oblasť spracovania dát – Analytické proces

Dnes sú činnosti príslušných útvarov závislé na manuálnej príprave dát cez MS Excel a štatistických nástrojoch.

#### Podporné činnosti

Medzi podporné činnosti sú v NBS radené procesy:

* účtovníctvo a finančné riadenie (SAP FINU/HRO + BW)
* hospodársko-prevádzkové činnosti a služby (SAP FINU/HRO + BW)
* interný audit
* legislatívno-právna agenda
* IT
* ochrana a bezpečnosť
* medzinárodné vzťahy
* komunikácia
* riadenie ľudských zdrojov
* agenda riadiacich orgánov

V rámci týchto procesov vzniká iba obmedzené množstvo údajov, ktoré prakticky nemajú využitie v hlavných procesoch NBS.



Obrázok 4: Oblasť spracovania dát – Podporné procesy

Aktuálne strategické aktivity NBS v tejto oblasti sú:

* Migrácia SAP FINU/HRO na technológiu S4HANA

### Súčasná architektúra aplikačnej vrstvy práce s dátami

Nasledujúci obrázok znázorňuje relevantnú časť súčasnej architektúry aplikácií a systémov v NBS, pričom tieto sú zoskupené do logických skupín podľa účelu.



Obrázok 5: Aplikačná vrstva – súčasný stav

* **Externé zdroje** – zdroje dát a informácií využívané pri analýzach a reportingu. Vo väčšine prípadov sú dáta z týchto zdrojov získavané manuálne, prípadne dochádza k výmene dát cez zdieľané úložisko.
* **Primárne systémy** – aplikačné systémy a registre prevádzkované NBS pre podporu primárnych a podporných procesov a úloh. Ich dáta sú čiastočne integrované v rôznych lokálnych databázach (*viď Integračná vrstva*), resp. sú exportované do Excelov pre ďalší reporting a analýzy. Medzi zdrojové systémy z tohto pohľadu zahŕňame aj Makroekonomickú databázu, ktorá je iba zdieľaným úložiskom časových radov v MS Excel.
* **Integračná vrstva** – riešenia pre integráciu dát NBS, ktoré vznikli v priebehu histórie pre podporu individuálnych potrieb jednotlivých odborov NBS; zahŕňajú ETL procesy pre plnenie integrovanej dátovej základne a prípadné data marty. V NBS sa využívajú databázové systémy Oracle (Pilot DWH, SAP BW) a Microsoft SQL Server (Apoštol databáza, Riskhouse, AŠI) + ETL systém MS Integration Services.
* **Prezentačná / Analytická vrstva** – obsahuje nástroje a služby Business Intelligence, ktoré sprostredkovávajú prístup k dátam pre koncových užívateľov (Qlik Sense v prípade Apoštol, MS Analysis Services Tabular Model pre Pilot DWH, resp. MS Reporting Services pre Riskhouse)
* **Aplikačná vrstva BI** – predstavuje koncové klientske nástroje pre prístup k integrovaným dátam

#### Externé zdroje

Rôzne odbory využívajú nasledujúce externé zdroje:

Štátna správa

| Označenie | Názov | Komentár k používaným dátam |
| --- | --- | --- |
| ŠÚ SR | Štatistický úrad SR | vybrané datasety podľa požiadaviek odborov https://data.statistics.sk/api/html/help-sk.html |
| KaPor | Katastrálny portál |  |
| FR SR | Finančné riaditeľstvo SR | súbory údajov z daňových výkazov a z registrov daňových subjektov |
| Sociálna poisťovňa | Sociálna poisťovňa | v súčasnosti veľmi obmedzený prístup k dátam, vo vyjednávaní je dohodnutie prístupu k širším anonymizovaným dátam  (počty zamestnancov vo firmách, mikrodáta o poistencoch a poberateľoch dávok) |
| Finančná správa | Dáta z Finančnej správy na základe zmluvy | mesačné agregáty, daňové priznania PO a FO k dani z príjmov – mikrodáta - ročné, daňové priznania DPH mikrodáta – mesačné a kvartálne |
| Register trestov | oversi.gov.sk |  |
| Štátna pokladnica |  | dáta o plnení verejných rozpočtov, ručne sťahované a integrované v AŠI databáze |
| Ardal | Agentúra pre riadenie dlhu a likvidity | Výmena údajov na základe zmluvy cez SFTP úložisko |
| Eurofondy | Systém ITMS 2014+ | OEM využíva 1 dataset cez API prevolávané z Matlab a odkladané do lokálnej databázy |
| SILC | Survey of Income and Living Conditions | Dáta prieskumu o vývoji príjmov a životných podmienok domácností od ŠÚ SR – manuálny vstup |
| Vzorka podnikov | Zjednodušené finančné výkazy pre vzorku cca 5000 podnikov | CSV súbor posielaný zo ŠÚ SR cez SFTP |
| eKasa | Dáta o tržbách podnikov z Finančnej správy | Atomické anonymizované dáta; dnes prenos e-mailom |
| Počty sporiteľov | Štatistika MPSVaR o počtoch klientov a objeme príspevkov | za každú DSS polročne - manuálny vstup |

Tabuľka 5: Prehľad externých zdrojov – štátna správa

Verejné registre

|  |  |  |
| --- | --- | --- |
| Označenie | Názov | Komentár k používaným dátam |
| RÚZ | Register účtovných závierok | 1x/mesiac automatický download do RS |
| RPO | Register právnických osôb | primárny zdroj Registra subjektov (denná synchronizácia); využívané aj na jednotlivých odboroch |
| RO | Register organizácií | zdroj Registra subjektov |
| Gleif | GLEIF - Global Legal Entity Identifier Foundation | zdroj Registra subjektov |
| RPVS | Register partnerov verejného sektora | aktuálne využívané pri AML validáciách protistrán <https://rpvs.gov.sk/OpenData/swagger/ui/index> |
| ŽR SR | Živnostenský register |  |

Tabuľka 6: Prehľad externých zdrojov – verejné registre

EÚ

|  |  |  |
| --- | --- | --- |
| Označenie | Názov | Komentár k používaným dátam |
| ECB SDW | ECB Statistical Data Warehouse | všetky datasety kde figuruje krajina "SK" |
| KL | Kurzový lístok ECB |  |
| ESMA | European Securities and Market Authority | údaje zo systémov TREM, FIRDS, FITRS, EMIR a SFTR (XML ISO 200022) + shortselling, Registre ESMA (XML dáta) – veľké dáta, v súčasnosti prebieha problematické ručné spracovanie |
| CSDB | Centralised Securities Database - databáza cenných papierov | 4x CSV export manuálne stiahnutý 1x/mesiac do Apoštol |
| Eurostat | Štatistický úrad EÚ | štatistiky na európskej úrovni, ktoré umožňujú porovnávanie medzi krajinami a regiónmi |
| SHSDB | Securities Holdings Statistics Database - databáza držby cenných papierov | dáta zo SHSDB 1x za mesiac stiahnuté ručne cez Lotus Notes a nahraté do Apoštolu |
| ECB/FIS |  |  |

Tabuľka 7: Prehľad externých zdrojov – EÚ

Zdroje ekonomických informácií

|  |  |  |
| --- | --- | --- |
| Označenie | Názov | Komentár k používaným dátam |
| Bloomberg | Bloomberg | prístup cez Excel plug-in  očakáva sa možnosť flexibilného rozširovania požiadaviek na preberanie údajov z Bloomberg (API) |
| Bisnode | Bisnode - poskytovateľ digitálnych obchodných informácií | závierky vo formáte CSV cez http; 1x/mesiac do platformy Apoštol; ďalej údaje o konkurzoch a iné podnikové mikrodáta |
| CRIBIS | CRIBIS - Register informácií o firmách, majetkové a personálne prepojenia |  |
| Reuters / Refinitiv | Najväčší poskytovateľ multimediálneho spravodajstva na svete | Používaný pre monitoring spravodajstva.  nie sú k dispozícii konkrétne požiadavky na integráciu |
| Tradeweb | elektronické mimoburzové trhy na obchodovanie s produktmi s pevným výnosom, ETF a derivátmi |  |
| Macrobond | Integrátor finančných a ekonomických dát | Ad-hoc dáta pre potreby ekonomických analýz pre OEM |
| CRN | Ceny rezidenčných nehnuteľností | XLS, štruktúrované dáta, cca 100 záznamov,  manuálne poskytované / kvartálne |
| United Classifieds | Spoločnosť poskytujúca online inzerciu (nehnuteľnosti) | Dáta o cenách nehnuteľností na sekundárnom trhu, štruktúrované dáta v CSV |

Tabuľka 8: Prehľad externých zdrojov – zdroje ekonomických informácií

#### Primárne systémy

Hlavné interné systémy a dátové zdroje relevantné pre dátovú integráciu sú nasledujúce:

| Označenie | Názov | Hlavná funkcia |
| --- | --- | --- |
| ŠZP | Štatistický zberový portál | Centrálny systém pre výkazy a hlásenia od subjektov finančného trhu pre štatistické účely a výkon funkcie dohľadu - core systém pre povinné výkazníctvo, zber dát od cca 7000 subjektov |
| ŠZP ESAS (Fujitsu XBRL) | Štatistický zberový portál (European Supervisory Authorities) | Nový systém pre výkazníctvo subjektov vo formáte XBRL (EIOPA, EBA) – v súčasnosti riešené samostatným modulom ŠZP |
| RBÚZ | Register bankových úverov a záruk | Lokálna implementácia AnaCredit - zber údajov o bankových úveroch a zárukách – prepojený s Registrom subjektov a systémom AnaCredit |
| RS | Register subjektov | Master systém pre subjekty, ktorých úvery sú vykazované cez RBÚZ (právnické osoby a podnikatelia) – založený na dátach ŠÚSR, RPO, RO a ďalších zdrojoch |
| Subjekty 2.0 | Subjekty 2.0 | Systém pre správu statických údajov o subjektoch finančného trhu pre účely dohľadu |
| REGFAP | Register finančných agentov a finančných poradcov | Systém pre evidenciu dohliadaných subjektov a pre zber dát a hlásení o ich činnosti. Evidenčné informácie sa plánujú integrovať do Subjekty 2.0, zber dát do ŠZP |
| CERI | CERI - Centrálna evidencia regulovaných informácií | Databáza regulovaných informácií, ktorými sú emitenti cenných papierov - neštruktúrované informácie - prospekty k emisiám cenných papierov, oznámenia, výročné správy, pre reporting sa požadujú iba metadáta o dokumentoch pre štatistiky |
| Register OFV |  | Odborná spôsobilosť finančných agentov a poradcov - register osôb (cca 70.000 FO), ktoré absolvovali OFV alebo odbornú skúšku podľa zákona o finančnom sprostredkovaní |
| Skúšky |  | Databáza otázok na skúšky, cvičné testy, ich výsledky, termíny skúšok |
| SIPS | Platobný systém SIPS | Retailový platobný systém určený na spracovanie klientskych platieb prevádzkovaný NBS |
| SAP FINU/HRO | Finančné účtovníctvo, hospodárenie a rozpočet, platobný styk | Implementácia na báze SAP s kompletnou kustomizáciou pre NBS (moduly SAP ECC v6.0 - účtovné jadro, SAP BCA - Bank Customer Accounts modul pre správu klientskych účtov, SAP FI, CO, IM, MM, SD, PO) |
| IDAS | Agendový systém dohľadu | Lotus Notes aplikácia pre podporu agendy dohľadu ODK, OPD a ODB pre menej významné banky |
| ASDR | Agendový systém dohľadu a regulácie | Tiež označovaný ako ISD (Informačný systém dohľadu) - *Systém vo vývoji* |
| IBFO/WSS | Investičné bankovníctvo a finančné obchodovanie | Kustomizovaná implementácia Wall Street Suite (WSS) pre podporu obchodných procesov NBS – t. j. realizáciu a vysporiadanie obchodných operácií realizovaných na OBO a OVO, vrátane ich účtovania na OFR. |
| Makroek. DB | Makroekonomická databáza | Konsolidované dáta pre tvorbu modelov na oddelení makroekonomických analýz OEM - spoločné úložisko časových radov. Kópia sa publikuje na webe NBS pre verejný prístup. |
| Credit Manager |  | Aplikácie pre riadenie rizík na odbore ORR |
| Risk Manager |  |
| EZO | Euro zásoby a obeh | Funkcie pre správu emisií bankoviek a mincí, evidenciu stavu hotovosti v ekonomike a na účtoch NBS, peňažné toky medzi NBS a tretími stranami.  Interný vývoj |
| HFCS | Zisťovanie o financiách a spotrebe domácností | Dáta prieskumu o vývoji životných podmienok domácností – manuálny vstup - cca 6000 domácností |
| BLS | Bank Lending Survey | Dotazník o vývoji ponuky a dopytu na trhu úverov – manuálny vstup |

Tabuľka 9: Prehľad interných zdrojov – primárne systémy

V NBS sa využívajú tiež niektoré ďalšie systémy pre podporu primárnych procesov, ktoré ale neboli identifikované ako relevantné pre dátovú integráciu (v prílohe A NBS DWH - Systemy.xlsx v stĺpce H označené ako „N“):

* **Target2-SK** – lokálny komponent systému Target2 prevádzkovaného Eurosystémom na spracovanie veľkých platieb, externý Target2 poskytuje tiež reportingovú službu CRSS postačujúcu pre potreby OPS
* **IMAS (**SSM Information Management System) – spoločná európska procesná platforma pre podporu dohľadových procesov vo významných finančných inštitúciách
* **IDRA** (IMAS Data Reporting and Analytics) - reportingová služba nad dátami IMAS

#### Integračná vrstva

Rôzne odbory využívajú rôzne platformy pre spracovanie dát z rôznych zdrojov. Dlhodobým cieľom je migrácia reportingu na zjednotenú platformu DWH a DataLake.

|  |  |
| --- | --- |
| Označenie | Názov |
| Apoštol databáza | Analytická platforma odboru štatistiky – predstavuje databázu integrovaných dát (**MS SQL Server**), kde hlavným zdrojovým systémom je **ŠZP** a k nemu sa nahrávajú najrôznejšie ďalšie registre a manuálne vstupy. Nad touto databázou je sprevádzkovaný in-memory BI nástroj Qlik Sense, v ktorom je tiež implementované množstvo integračnej a výpočtovej logiky. |
| Pilot DWH | Hlavným cieľom pilotného projektu DWH/BI bolo vybudovanie infraštruktúry DWH/BI (**Oracle + MS Analysis Services + MS Power BI**) a vytvorenie riešenia pre základné procesy spojené so spracovaním dát DWH a analytickými výstupmi v podobe BI riešenia. Projekt, ktorý začal v 03/2020, sa zameral na spracovanie vybraných dát zo systémov **RBÚZ, RS a ŠZP**.  Projekt Pilot DWH vytvoril základ pre budúce riešenie, rešpektuje štandardné prístupy pre budovanie dátových skladov a môže byť využitý ako základ pre ďalší rozvoj DWH. |
| Riskhouse | Riskhouse predstavuje vlastný DWH systém odboru ORR (**MS SQL Server)**, ktorého vývoj začal v roku 2012 a ktorý bol navrhnutý s klasickým prístupom a architektúrou pre budovanie dátových skladov na podporu procesov zverených odboru ORR. Hlavným zdrojom dát je systém **IBFO/WSS**, ktorý poskytuje iba slabú analytickú podporu. Predpokladá sa, že riešenie Riskhouse bude v strednodobom horizontu zachované, vzhľadom k tomu, že implementuje množstvo špecifickej biznisovej logiky pre účely riadenia rizík. |
| SAP BW | SAP Business Warehouse v 7.3 predstavuje komodizovaný dátový sklad (**Oracle**), ktorý je dodávaný ako samostatný modul balíka podnikových aplikácií SAP. V NBS je nasadený SAP BW najmä pre reportingové potreby OFR, pričom jeho kontrolingová časť je kustomizovaná pre vykazovanie podľa ESCB metodiky COMCO.  V súvislosti s plánovaným prechodom na S4HANA bude treba zvážiť alternatívny prechod na BW4HANA, alebo pokrytie požiadaviek jednotným DWH systémom. |
| AŠI | Databáza (MS SQL Server) o subjektoch verejnej správy, vytvorená najmä ako dátová základňa pre analytické spracovanie **údajov zo Štátnej pokladnice** vnútri odboru OEM, oddelenia fiškálnych a štrukturálnych analýz. Databáza nebola koncipovaná ako DWH. |
| Makroek. databáza | Úložisko v **MS Excel**, v ktorom sú konsolidované dáta pre tvorbu modelov na oddelení makroekonomických analýz OEM. Databáza nebola koncipovaná ako DWH, ale spoločné úložisko časových radov.  Pre prístup sa využívajú rôzne nástroje (Matlab, EViews, WinSolve, BEAR, R Studio, Python), pomocou ktorých sa aj generujú výstupy pre management a analytické správy. Databáza sa tiež kopíruje na web NBS pre prístup zo strany verejnosti. |

Tabuľka 10: Prehľad integračných platforiem v NBS

#### Prezentačná / Analytická vrstva

Pre prístup k dátam v jednotlivých integračných platformách sa využívajú rôzne nástroje:

|  |  |  |
| --- | --- | --- |
| Prezentačná / Analytická vrstva | Integračná vrstva | Názov |
| Qlik Sense | Apoštol databáza | Qlik Sense je self-service BI systém s in-memory analytikou a schopnosťami stavať analytické riešenia nad rôznymi dátovými zdrojmi. Aj v rámci riešenia Apoštol sa silne využíva skriptovanie pre transformácie dát a dodatočné integrovanie externých a manuálnych vstupov.  Prípravu analytických priestorov a dashboardov zabezpečuje odbor OST. |
| MS Analysis Services | DWH Pilot | Dáta z data martov pilotného DWH sú spracúvané do tabulárnych modelov na MS SQL Serveri (in-memory OLAP), ku ktorým sa pristupuje cez MS Power BI klienta.  Spracovanie dát zabezpečuje OIT. |
| MS Reporting Services | Riskhouse | Prístup k dátam v integrovanej databáze Riskhouse je zabezpečený cez MS Reporting Services, ktoré sa využívajú pre operatívny reporting pre bežných užívateľov s obmedzenými možnosťami ďalšej manipulácie.  Reporty pripravuje a dokumentuje odbor ORR, ktorý sa tiež stará o celé riešenie Riskhouse. |

Tabuľka 11: Analytické vrstvy v NBS

#### Aplikačná vrstva BI

Koncové nástroje pre prístup k dátam, reportom a dashboardom boli zavedené v súvislosti s podkladovými analytickými systémami:

| Prezentačná / Analytická vrstva | Integračná / Analytická  vrstva | Názov |
| --- | --- | --- |
| Browser (webový prehliadač) | Apoštol databáza / Qlik Sense | Analytické dashboardy sa otvárajú cez tenkého klienta (webový prehliadač) |
| MS Power BI  (+ Power BI Report Server) | DWH Pilot / MS Analysis Services | Dáta tabulárnych modelov sú otvárané z klienta MS Power BI, vytvárané reporty sú publikované do MS Power BI Report Servera. |
| Browser (webový prehliadač) | Riskhouse / MS Reporting Services | Prístup k reportom publikovaným na MS Reporting Services cez webový prehliadač. |
| MS Excel  (Business Explorer add-in) | SAP BW | Prístup k dátam je realizovaný cez Business Explorer (Excel plug-in), ktorý umožňuje získanie dát zo SAP BW do Excelu, ale bez možnosti ďalšieho formátovania výstupov a tvorby grafov. Tieto výstupy sa robia následne v MS Excel ručne. |
| Štatistické nástroje | AŠI | Pre prístup sa využívajú rôzne nástroje (Matlab, EViews, WinSolve, BEAR, R Studio, Python), pomocou ktorých sa aj generujú výstupy pre management a analytické správy (vo formáte pdf a MS Excel s tabuľkami a grafmi) |
| Makroek. databáza |

Tabuľka 12: Aplikačná vrstva BI v NBS

### Pilot DWH

Hlavným cieľom projektu Pilot DWH bolo vybudovanie infraštruktúry DWH/BI, zadefinovanie pojmov a základného data governence. Hlavným cieľom bolo otestovanie rôznych technológií a ich vhodnosť pre prostredie NBS. V neposlednom rade išlo o vytvorenie riešenia pre základné procesy spojené so spracovaním dát DWH a analytickými výstupmi v podobe BI riešenia.

Projekt, ktorý začal v 03/2020, sa zameral na spracovanie vybraných dát zo systémov ŠZP, RBUZ a RS. V súčasnosti zahŕňa asi 5 % rozsahu dát NBS.

#### Logická architektúra Pilot DWH

Diagram

Description automatically generated

Obrázok 6: DWH – Logická architektúra

Pre účely DWH bola zvolená 3-vrstvová architektúra, ktorá pozostáva zo Stage vrstvy (L0), Integračnej (L1) a Prezentačnej vrstvy (PL). L1 a PL sú modelované prístupom R. Kimball, pričom v L1 vrstve sú dáta perzistované a PL vrstva obsahuje dáta perzistované (oblasť ŠZP) ako aj pohľady nad L1 vrstvou.

V rámci PL vrstvy boli logicky vytvorené viaceré star schémy nad objektami L1 vrstvy. SSAS tabulárny model obsahuje objekty štrukturálne rovnaké ako sú objekty v PL vrstve a pomocou technického účtu si dáta preberá 1:1 z PL vrstvy. PL vrstva slúži aj pre zabezpečenie prístupov, či už riadkovú ako aj stĺpcovú.

#### Integrované dáta

Zdrojové systémy pre Pilot DWH sú:

* ŠZP (**Štatistický zberový portál**) - Centrálny systém pre výkazy a hlásenia od subjektov finančného trhu pre štatistické účely a výkon funkcie dohľadu - core systém pre povinné výkazníctvo, zber dát od cca 7000 subjektov. Z toho integrované časti sú:
  + Menová a finančná štatistika
  + Cenné papiere
  + Bilančná trieda
* RBÚZ (**Register bankových úverov a záruk**) – Lokálna implementácia AnaCredit – zber údajov o bankových úveroch a zárukách – prepojený s Registrom subjektov a systémom AnaCredit. Do DWH sa prenášajú:
  + Údaje nástroja
  + Údaje protistrana – nástroj
  + Údaje spoločných záväzkov
  + Celkom cca 50 % atribútov.
* RS (**Register subjektov**) - Master systém pre subjekty, ktorých úvery sú vykazované cez RBÚZ (právnické osoby a podnikatelia) – založený na dátach ŠÚSR, RPO, RO a ďalších zdrojoch. Do DWH sa preberajú všetky autoritatívne záznamy.

Tieto zdrojové systémy z pohľadu komplexnosti poskytnutých informácií celej NBS boli v projekte Pilot DWH integrované na cca 20 %.

#### Dátové oblasti L1 vrstvy

L1 vrstva obsahuje tieto star schémy

|  |  |  |
| --- | --- | --- |
| Zdroj | Star schéma | Obsah |
| RBUZ | Inštrument a finančné dáta | Statické a finančné údaje o inštrumentoch |
| RBUZ/RS | Subjekty | Agregované finančné údaje z RBUZ na úrovni subjektu vystupujúceho v roli debtor (dlžník), informácie o subjektoch. |
| ŠZP | Trieda MFŠ | Menová a finančná štatistika dohľadaných subjektov. Star schéma je organizovaná vo forme key-value, kde key atribút určuje významovosť hodnoty |
| ŠZP | Trieda bilancie | Bilančné výkazy dohľadaných subjektov. Star schéma je organizovaná vo forme key-value, kde key atribút určuje významovosť hodnoty |
| ŠZP | Trieda Cenné papiere | Výkazy cenných papierov dohľadaných subjektov. Star schéma je organizovaná vo forme key-value, kde key atribút určuje významovosť hodnoty |
| DWH (MFŠ, RBUZ/RS) | Konsolidovaný model MFŠ-RBÚZ | Konsolidácia údajov zo štatistickej triedy MFŠ a RBUZ/RS. Porovnanie nominálnej hodnoty vykázanej ako agregát v ŠZP a napočítaného údaju z granulárnych dát RBUZ/RS podľa BSI metodiky (ECB). |
| RBUZ/RS | Prepojenie subjektu a inštrumentu | Prepojenie inštrumentu na subjekt vystupujúci v konkretnej roli (debtor, creditor...), údaje spoločných záväzkov a nominálnej hodnoty napočítanej podľa BSI štatistiky (ECB). |

Tabuľka 13: Star-schémy DWH – vrstvy L1

#### Technická architektúra Pilot DWH“

Graphical user interface

Description automatically generated

Obrázok 7: DWH – Technická architektúra

Ako základ pre DWH bola zvolená databáza od firmy Oracle. Pre BI riešenie bolo zvolené Microsoft SQL Server Analysis Services s reporting platformou Power BI.

V rámci DWH existujú 2 prostredia:

1. DEV používané ako na vývoj tak aj testovanie
2. Samotné produkčné prostredie.

Každé prostredie má samostatnú databázu a samostatný HW.

Spracovanie zdrojových dát prebieha nasledujúcim spôsobom:

1. Dáta sa extrahujú cez databázovú linku Oracle->Oracle cez extrakčné pohľady, pričom z nich si berie full množinu dát.
2. Dáta sú prenesené do temporárnych tabuliek v DWH.
3. V L0 schéme sú však za daný deň uložené full snímky ako aj inkrementálne snímky, ktoré sú tvorené porovnaním 2 po sebe idúcich full snímkov. Tabuľky v L0 schéme sú organizované podľa zdrojového systému a samotného extrakčného pohľadu. Výnimku tvorí hlavná tabuľka výkazov ŠZP, ktorá obsahuje cez 1 mld záznamov. Z tejto tabuľky sú dáta ukladané do osobitných, štrukturálne rovnakých tabuliek podľa samotných tried. Nedochádza k transformácii dát zo zdrojových systémov.
4. Ako posledný krok beží vloženie všetkých chýbajúcich záznamov tak, aby inkrementálne snímky zachovali referenčnú integritu.
5. Integrácia dát do L1 vrstvy je zabezpečená pomocou PL/SQL procedúr, ktorých úlohou je:

* Vygenerovanie umelých kľúčov pre príslušnú tabuľku
* Transformovanie dát do perzistentných LD tabuliek, ktoré sú obrazom tabuliek L1.
* Historizácia dát z LD tabuliek do L1 v závislosti na typu tabuľky

Na spracovanie dát v rámci DWH (microscheduling) sa používa custom workflow nástroj vyvinutý zhotoviteľom pilotu.

Na spracovanie dát do tabulárnych modelov SSAS sa používa SQL Server Integration Services.

Dostupná dokumentácia k pilotnému DWH/BI

* Základné princípy DWH/BI - Teoretický materiál
* Funkčný návrh DWH/BI - Logický a funkčný návrh riešenia PL vrstvy, popisuje zvolenú metodiku modelovania podľa R.Kimballa, popis star schém (datamartov), fyzický návrh modelu PL vrstvy
* Architektúra DWH/BI - Návrh integrácie okolitých systémov, ETL architektúra a workflow manažment, BI a reporting
* Technické požiadavky na zabezpečenie DWH/BI - Popis HW a SW požiadaviek na zabezpečenie serverových komponentov DWH a BI, špecifikácia prostredí
* Menné konvencie
* Technické riešenie DWH - Detailný technický popis jednotlivých vrstiev DWH/BI, ETL objektov, logovaní a zabezpečenie DWH/BI

#### Reporting

Projekt Pilot DWH zaviedol BI infraštruktúru pre reporting postavenú na nástroji MS Power BI a MS Power BI Report Server.

#### HW špecifikácia Pilot DWH

| HW komponent | Typ | Parametre |
| --- | --- | --- |
| DB server | Výrobca/ verzia | Oracle 19c EE + Partitioning option |
| Operačný systém | Oracle Linux 8 (RedHat) |
| RAM (Typ, počet) | 512GB, DB má dedikované SGA/PGA na úrovni 50GB/40GB |
| CPU (Typ, počet) | 1xCPU Xeon Gold 6244 (8 cores), 3,6 GHz; 1 slot ešte voľný |
| Súčasná utilizácia CPU | v maxime 30% |
| Veľkosť DB | 2TB (dáta a indexy) |
| Diskové pole (Typ HDD, počet, veľkosť) | 14x 3TB SSD lokálne disky, zapojené do RAID6, cez PCI RAID radič, je možné zriadiť prístup aj na centrálne diskové pole 3PAR, lebo server má 2x 16 Gbit FC karty |
| LAN na diskové pole (rýchlosť, počet) | - |
| LAN do siete (rýchlosť, počet) | 4x 10GBit |
| Reporting / BI server | Výrobca/ verzia | MS SQL Server 2019 Enterprise |
| Operačný systém | Windows Server 2022 Data Center edícia. |
| RAM | 32 GB |
| CPU | 4 vCPU |
| Diskové pole | 3PAR |
| Aplikačný server | Výrobca/ verzia | Apache Tomcat/9.0.56 |
| Operačný systém | Red Hat Enterprise Linux 7.9 |
| RAM | 10 GB |
| CPU | 6 vCPU, Intel(R) Xeon(R) Gold 6134 CPU @ 3.20GHz |
| HDD | 25 GB |

Tabuľka 14: Pilot DWH – HW špecifikácia

**Poznámka:** na databázovom serveri nebeží len Pilot DWH databáza, ale ďalšie produkčné databázy NBS. Súčasné vyťaženie CPU je na úrovni 30 %. Pilot DWH obsahuje cca 2TB dát.

## Technické požiadavky

### Úvod

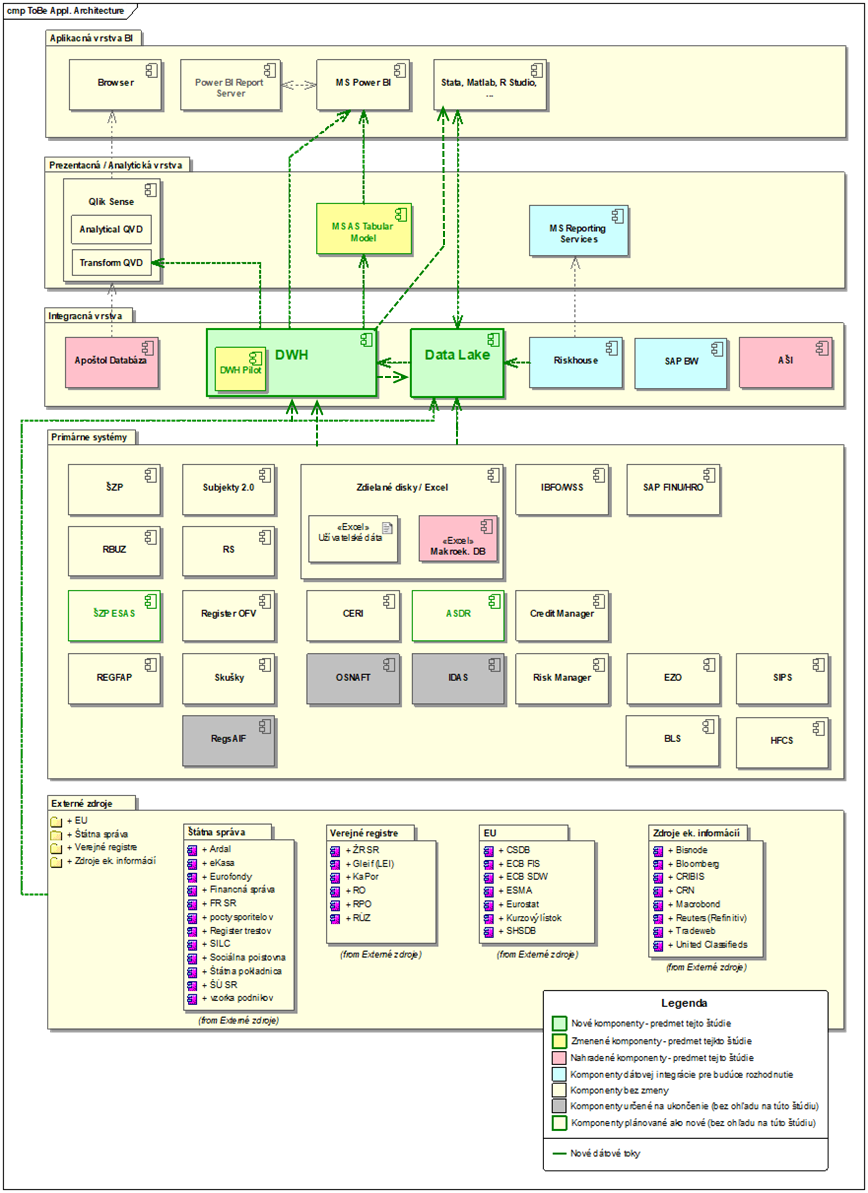
Táto kapitola obsahuje technické požiadavky na nový systém NBS DWH. Požiadavky sú rozdelené do jednotlivých oblastí ako:

1. Architektúra (logická, fyzická a aplikačná)
2. Popis vlastností jednotlivých komponentov na úrovni relačnej databázy a DataLake
3. Popis aplikačných vlastností podporných aplikácií DWH.

Rovnako sa táto kapitola zaoberá bezpečnostnými požiadavkami a požiadavkami na spustenie systému (iniciálny load).

### Požadovaná architektúra aplikačnej vrstvy práce s dátami

Plánované zmeny zachytáva nasledujúci diagram, ktorý vychádza zo schválenej dátovej stratégie bankovou radou NBS v apríli 2022:



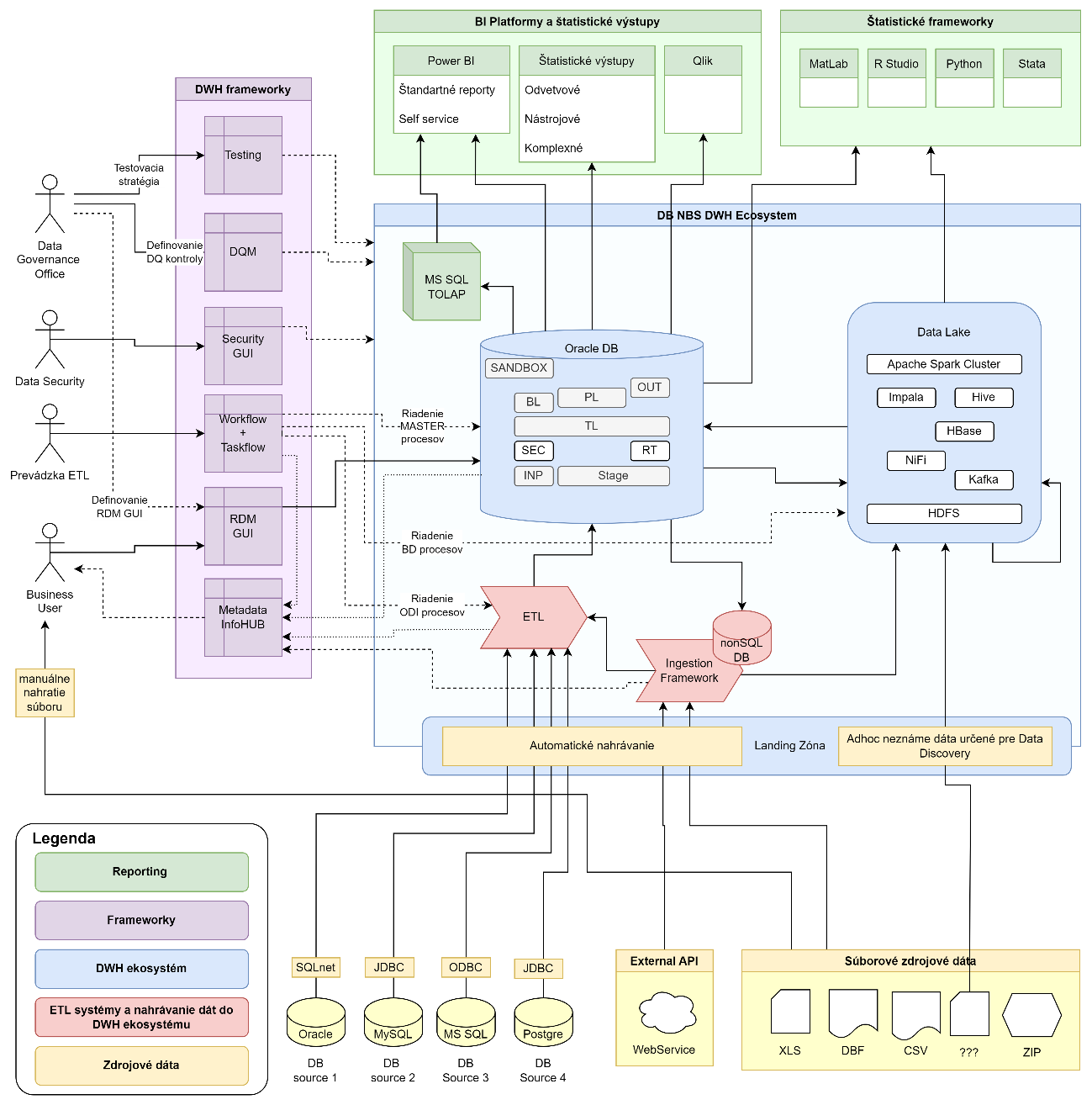
Obrázok 8: Aplikačná vrstva – budúci stav

### Logická architektúra

Riešenie novej dátovej integrácie je navrhnuté ako jeden dátový ekosystém v kombinácii dvoch platforiem:

* DataLake
* Relačná časť, ktorý sa skladá z:
  + Oracle DB, ktorá rieši problematiku dát ako je atomicita, konzistencia, izolovanosť, trvanlivosť, či unikátnosť
  + MS SQL DB s in-memory T-OLAP podporou pre PowerBI reporting.

Každá z týchto platforiem nie je osobitnou aplikáciou, ale súborom technológií, nástrojov a na mieru naprogramovaných procesov a integrácií, ktoré vo svojom celku plnia úlohu dátovej integrácie a spracovania dát pre účely reportingu a pokročilých analýz. Z tohto dôvodu sú v ďalšom texte obe platformy ďalej dekomponované na jednotlivé dátové vrstvy, dátové toky, funkčné bloky a použité technológie.



Obrázok 9: Logická architektúra nového riešenia

Predmetom riešenia sú tieto komponenty architektúry:

* **Relačná časť** – Dátový sklad, jeho dátové vrstvy, dátové toky, metadátové nastavenia a podporné frameworky
* **DL** – DataLake, jeho dátové vrstvy, dátové toky, metadátové nastavenia a podporné frameworky
* **Spoločné frameworky** – aplikácie a nástroje, ktoré nie sú súčasťou samotných platforiem dátovej integrácie, ale využívajú sa pre celkové riadenie spracovania dát a podporu odborových používateľov.

Nasledujúce časti aplikačnej architektúry sú dotknuté týmto riešením:

* **BI platformy** – existujúce BI aplikácie, ktoré budú v rámci projektu pripojené na DWH a DataLake pre čítanie dát.
* **Štatistický framework** – existujúce portfólio štatistických nástrojov, pre ktoré budú zadefinované mechanizmy prístupu k DWH a DataLake
* **Zdrojové systémy s priamym prístupom** – reprezentuje všetky systémy, ktoré budú poskytovať priamy prístup ku svojím dátam, resp. cez vyhradené databázové rozhranie
* **Externé systémy poskytujúce extrakty** - reprezentuje všetky systémy, ktoré budú poskytovať extrakt svojich dát, ktorý bude vymieňaný s Data Landing zónou cez SFTP server alebo extranet Sharepoint.
* **Zdrojové systémy poskytujúce služby** - reprezentuje všetky systémy (interné aj externé), ktoré budú poskytovať svoje dáta prostredníctvom vystavenej služby (REST API, SOAP).

### Fyzická architektúra

Verejný obstarávateľ požaduje aby, navrhnutý datawarehouse pozostával z rozširovateľných a otvorených komponentov. Všetky logické komponenty budú prevádzkované formou „On-Premise“ v dátových centrách NBS ako fyzické či virtuálne servery.

Fyzicky bude infraštruktúra rozvrhnutá nasledovne na 4 skupiny:

* Relačná časť
* DataLake
* Aplikačný server pre podporné aplikácie
* BI oblasť

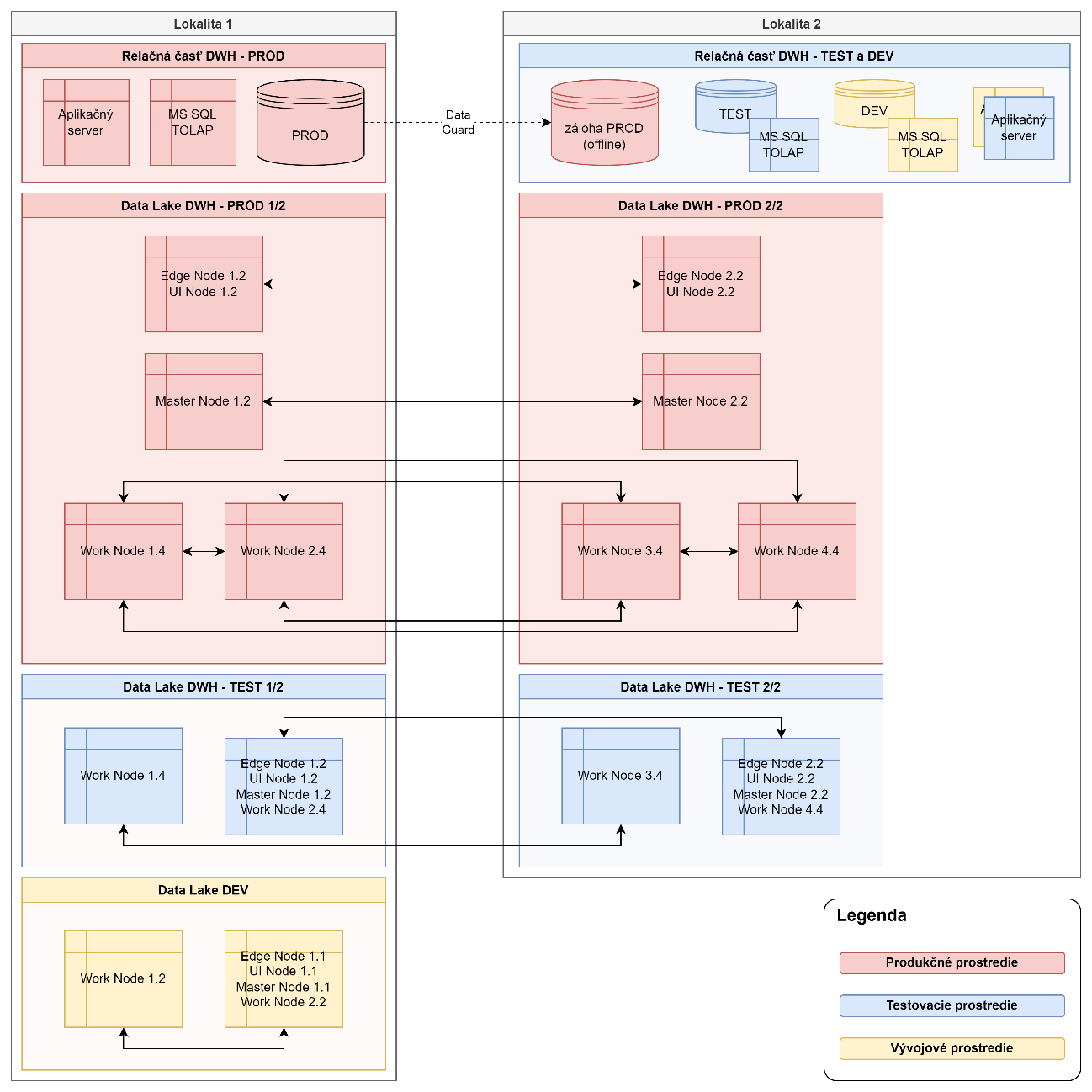
Nasledujúca tabuľka obsahuje súpis HW požiadaviek.

| Platforma | Uzly | Navrhované riešenie |
| --- | --- | --- |
| Relačná časť | PROD-DB  TEST-DB  DEV-DB | NBS DWH bude nasadený na existujúcej infraštruktúre prevádzkovaný interným oddelením OIT na 2 samostatné servery v oddelených lokáciách s rovnakým výkonom v konfigurácii:   * Existujúce fyzické servery:   + CPU: 1x X86 CPU 8 cores s potencionálnym rozšírením o ďalší CPU   + RAM: 512 GB RAM s potencionálnym rozšírením na 1 TB   + SSD: 4x3,2TB v RAID s potencionálnym rozšírením * Na serveroch je inštalovaná Oracle 19c EE + Partitioning option * Pre zabezpečenie vysokej dostupnosti je vytvorená standby produkčná inštancia cez Oracle Data Guard na vývojový server. * Na serveri budú ďalej bežať RunTime ETL nástroja Oracle Data Integrator, dodané a inštalované uchádzačom. |
| DataLake | Worker Node  Master Node  UI Node  Edge Node | Pre DataLake bude potrebné vytvoriť novú celkovú infraštruktúru cez 3 prostredia. V každom prostredí bude vytvorená sieť výpočtových uzlov, na ktorých pobežia služby Big Data.  Vzhľadom na predpokladanú veľkosť dát v NBS s funkciou vysokej dostupnosti sa očakáva vybudovanie infraštruktúry **vo virtuálnom prostredí NBS** s možnosťou prechodu na fyzické servery. Virtuálne servery budú škálované podľa licenčnej politiky vybratej Enterprise softwarovej podpory. Predpoklad je v krokoch 1vCPUx8GB RAM. Budovanie jednotlivých uzlov bude postupné.  Architektúra DataLake má charakter vysokej dostupnosti a nebude potrebné budovať standby inštanciu, ale použije sa roztiahnutie produkčných uzlov cez obe lokality.  Produkčné uzly:   * 4x Worker Node (2+2 v lokalitách)   + CPU: X86 8 cores   + RAM: 64GB   + Disk: 20 TB * 3x Master Node (1+2 v lokalitách)   + CPU: X86 8 cores   + RAM: 64GB   + Disk: 2 TB * 2x Edge Node+UI Node (1+1 v lokalitách)   + CPU: X86 4 cores   + RAM: 32GB   + Disk: 3 TB   Test uzly:   * 2x Worker Node (1+1 v lokalitách)   + CPU: X86 4 cores   + RAM: 32 GB   + Disk: 10 TB * 2x Worker Node + Master Node+ Edge Node+UI Node (1+1 v lokalitách)   + CPU: X86 4 cores   + RAM: 32GB   + Disk: 3 TB   Vývojové uzly:   * 3x Worker Node + Master Node+ Edge Node+UI Node (1 v lokalite)   + CPU: X86 4 cores   + RAM: 32GB   + Disk: 3 TB |
| Aplikačný server |  | Existujúci aplikačný server na beh aplikácií (frameworkov DWH). Tento komponent je podobne ako BI server tvorená virtualizačnou platformou NBS .   * Apache Tomcat/9.0.56 aplikačný server na Red Hat Enterprise Linux prostredí * CPU: 4 vCPU * RAM: 10 GB * Disk: 25 GB   Vysoká dostupnosť (failover) je riešená na úrovni virtualizačnej platformy NBS. |
| Reporting / BI server | PROD-BI  DEV-BI  TEST-BI | Exsitujúce BI servery sú virtualizované na platforme VMware ESX, virtualizačná platforma je v správe IT NBS. Výkon je možné prideliť podľa potreby.   * MS SQL Server 2019 Enterprise * Windows Server 2022 Data Center edícia. * CPU: 4 vCPU (s možnosťou navýšiť na 8 vCPU) * RAM: 32 GB   Vysoká dostupnosť (failover) je riešená na úrovni virtualizačnej platformy. Budú dve platformy:   1. PROD-DB (produkčný BI server) 2. DEV-DB (vývojový BI server) + TEST-BI (testovací BI server) |
| PC | PC | Inštalované komponenty, ktoré zabezpečí NBS (do 15 pracovných staníc):  VMware Horizon, MS Windows 10, 4vCPU, 16GB RAM   * SQL Server Data Tools - pre vývojára ETL pre spracovanie tabulárnych modelov v MS SQL Server Analysis Services * Oracle client pre vývojára ETL, resp. reportov * Oracle SQL developer pre vývojára ETL * Oracle Data Integrátor pre vývojára ETL   Ostatné potrebné SW komponenty musia byť dodané v rámci ponuky uchádzača , ako napríklad:   * MS Power BI Pro * MS Visual Studio |

Tabuľka 15: Prehľad HW komponent riešenia

NBS plánuje využitie svojej existujúcej HW a SW infraštruktúry pomocou svojho IT oddelenia pre prevádzku. Táto infraštruktúra musí byť uchádzačom cieľového riešenia posúdená z pohľadu kapacity a dostupnosti s prípadným dopadom na SW licencie.

Z hľadiska vytvorenia vysokej dostupnosti bude fyzická architektúra nastavená nasledovne:

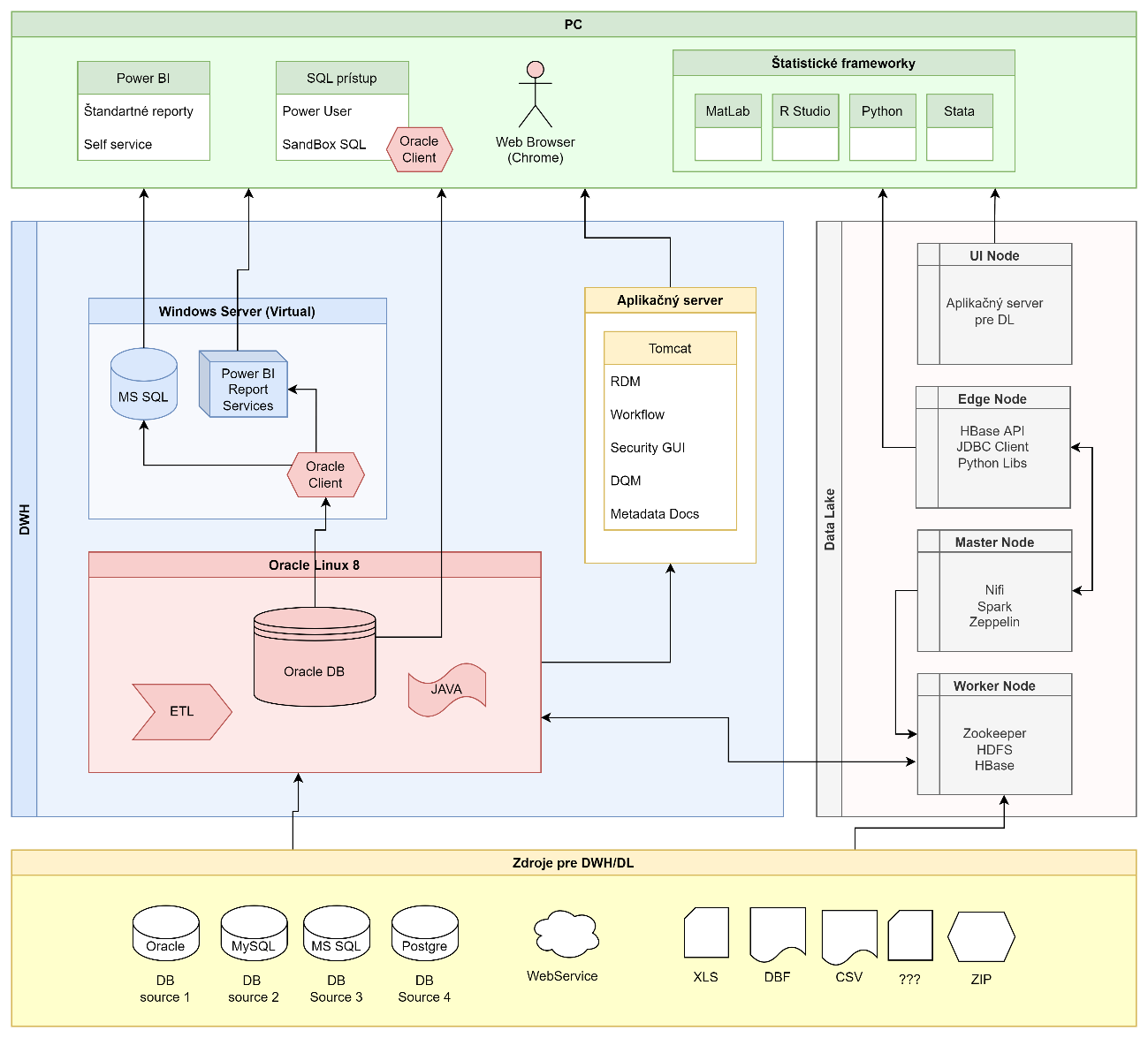


Obrázok 10: Vysoká dostupnosť nového riešenia

### Aplikačná architektúra

#### Úvod

Nasledujúci infraštruktúrny diagram popisuje požadovanú novú aplikačnú architektúru NBS DWH z  hľadiska inštalácie jednotlivých produktov.



Obrázok 11: Aplikačná architektúra nového riešenia

Navrhované riešenie tvorené:

1. relačným DWH a BI platformou bude prevádzkované s využitím existujúcej aplikačnej infraštruktúry projektu Pilot DWH.
2. DataLake - potreba vybudovania novej aplikačnej infraštruktúry.

V prípade DWH je základným software komponentom databázový systém a systém pre ETL. NBS preferuje využiť existujúcu investíciu do projektu Pilot DWH, ktorý je postavený na:

* Databáza – Oracle DB v edícii Enterprise
* Reporting – MS SQL Server Analysis Services a MS Power BI

Navyše je v riešení NBS DWH požadované vytvorenie DataLake prostredia ako oblasť pre uloženie dát pre data discovery.

Jednotlivé komponenty riešenia budú vyžadovať nasledujúci software, a to minimálne v 3 oddelených prostrediach (vývoj, test, produkcia) v súlade s detailnými požiadavkami:

| Platforma | Software | Navrhované riešenie |
| --- | --- | --- |
| DWH | Operačný systém | Výber operačného systému je podriadený voľbe databázy, prípadne podkladovej HW infraštruktúry.  V prípade existujúceho pilotného DWH sa využíva **Oracle Linux 8** (RedHat distribúcia). |
| Databáza | Existujúci Pilot DWH je postavený na relačnej databáze **Oracle 19c Enterprise Edition + Partitioning Option**. |
| Doplnkové časti DB | Súčasné nastavenie DB a vybrané opcie (**Partitioning Option**) sú dostatočné a neplánuje sa úprava. |
| DataLake | Operačný systém | Výber operačného systému pre DataLake bude podriadený voľbe Big Data platformy, ktorá musí mať pre daný operačný systém podporu.  NBS preferuje **Red Hat Linux**. |
| Big Data platforma | Výber Big Data platformy je predmetom výberového konania. |
| Platform Management | Výber Platform Management systému je predmetom výberového konania. |
| ETL systém | SQL ETL | Ako ETL framework je preferovaný nástroj **Oracle Data Integrator** |
| Big Data ETL | Výber Big Data platformy je predmetom výberového konania. Podmienka NBS je použitie Enterprise riešenia pred čistým OpenSource. |
| Ingestion framework | Projekt NBS DWH nebude o veľkosti spracovávaných dát, ale o počte rôznorodých dátových vstupov a vzorov. Denne/mesačne budú prichádzať stovky rôzne štruktúrovaných súborov, ktoré sa počas životnosti môžu a budú meniť. Preto požadujeme dodať v projekte systém na metadátové riadené spracovanie súborov. Bližšie v kapitolách 3.2.13. |
| BI | Operačný systém | Výber operačného systému je podriadený voľbe BI servera, prípadne podkladovej HW infraštruktúry.  V prípade existujúceho pilotného BI servera, ktorý je postavený na technológiách Microsoft a virtualizačnej platforme VMware ESX, sa využíva **Windows Server 2022 Data Center edícia**.  Aktuálne je zalicencovaná celá virtualizačná platforma, prípadné rozšírenie kapacity BI je preto bez dopadu na potrebu ďalších licencií OS. |
| BI server | NBS využíva **MS SQL Server 2019 Enterprise** pre BI riešenie, konkrétne **Analysis Services** (tabulárne modely) a **MS Power BI Report Server** pre publikáciu a zdieľanie reportov a analýz.  Na odbore štatistiky je to ďalej **Qlik Sense** – NBS však nepredpokladá ďalšie rozširovanie tohto systému mimo odbor štatistiky.  Tabulárne modely v pilotnom režime boli vyvíjané na odboroch OFI a OST.  Aktuálne je zalicencované a prevádzkované 1 produkčné a 1 vývojové prostredie (2x 4 jadra). |
| BI klient | NBS využíva **MS Power BI** (On-premise) pre tvorbu reportov a plánuje tento nástroj naďalej využívať ako jednotnú aplikáciu pre reporting a analýzy.  Aktuálne je k dispozícii 13 licencií MS Power BI Pro. |
| Štatistický framework | Štatistické nástroje používané analytickými oddeleniami sú rôzne a používajú sa vždy podľa konkrétneho účelu. Tento štatistický framework zahŕňa nástroje Matlab, R studio, Eviews, Stata, jazyky SQL, Python, Scala a pod.  Predmetom projektu nie je dodávka ďalších nástrojov alebo licencií. |
| AI/ML | Nástroje pre umelú inteligenciu a Machine Learning sa zatiaľ nepoužívajú. Platforma Data Lake by však mala vytvoriť predpoklad pre ich budúce nasadenie.  Predmetom projektu nie je dodávka licencií, resp. infraštruktúry pre AI/ML. |
| Aplikačné frameworky | Workflow Manager | Pre orchestráciu spracovania cez rôzne platformy bude zavedený jednotný nástroj.  Výber konkrétneho nástroja je súčasťou výberového konania. |
| Dátové modelovanie | Nástroj pre dátové modelovanie je potrebný pre zabezpečenie celého procesu vývoja dátovej integračnej platformy.  Výber konkrétneho nástroja pre dátové modelovanie je súčasťou výberového konania. |
| RDM  DQM  Metadata Management | Nástroje pre správu referenčných dát (RDM), Data Quality Management (DQM), a Metadata Management ako sú popísané v kapitolách 3.2.13  Výber konkrétnych nástrojov je súčasťou výberového konania. |

Tabuľka 16: Software komponenty riešenia

#### Relačná databáza

Existujúci Pilot DWH je postavený na relačnej databáze Oracle 19c Enterprise Edition + Partitioning Option. Riešenie NBS DWH počíta s prepoužitím už existujúcich komponentov.

**Oracle DB** prináša robustnosť a širokú škálu funkcionality pre DWH riešenia. DB je optimalizovaná pre spracovanie veľkého objemu dát pomocou hromadných operácií, s využitím partícií a kompresie je možné efektívne využívať diskový priestor a znížiť nároky na diskové pole, a to ako z pohľadu kapacity, tak aj z pohľadu výkonnosti. Medzi ďalšie výhody Oracle DB patrí možnosť využívať bitmapové a funkčné indexy, ktoré spolu s DB paralelizmom umožňujú skracovať nielen ELT čas, ale aj zbiehanie užívateľských reportov. Vo svete DWH sa veľmi často využívajú analytické funkcie, ktoré umožňujú jednoducho vytvárať aj veľmi komplexné reporty bez potreby ukladania medzivýsledkov v DB. Z pohľadu security je veľmi užitočnou vlastnosťou Virtual Private Database, ktorá umožňuje riadiť prístup nielen na úrovni objektov, ale aj na úrovni riadkov a atribútov. Oracle Database Resource Manager umožňuje administrátorom riadiť prideľovanie zdrojov (CPU a IO operácií) pre jednotlivých užívateľov a procesy. Pre komplexné monitorovanie výkonu a optimálne využívanie DB slúži Oracle Enteprise Manager spolu s Automatic Workload Repository, ktoré poskytujú kompletný prehľad o využívaní zdrojov jednotlivými procesmi. Oracle DB v prípade potreby umožňuje jednoducho rozširovať funkcionalitu o ďalšie možnosti ako je napríklad: In-memory, OLAP DML, datamining.

**Aplikačné frameworky**, ktoré majú pomôcť zrýchliť vývoj DWH, zjednodušiť správu a spríjemniť technické riešenie pre netechnických používateľov. Z aplikačného hľadiska sú to (detailný popis je nasledujúcich kapitolách):

* Nástroje pre podporu prevádzky a odborových používateľov:
  + Podporu správy, riadenia a monitoring aktualizačných úloh (Workflow Management – Data Orchestration).
  + Správa číselníkov, mostíkov, referenčných dát, hierarchií jednotlivých dimenzií a manuálnych vstupov.
  + Bezpečné nahrávanie veľkého počtu a rôznych súborov, t. j. všetky a práve raz (Ingestion framework)
  + Správa dátovej kvality - automatické spúšťanie a vyhodnocovanie testov
  + Správa prístupov (riadenie bezpečnosti) k dátam
  + Správa dokumentácie k DWH a iných metadát (o modeli, o reportoch, mapovaní zdroj-cieľ)
  + Riadenie dátovej retencie

**ETL framework**

ETL (Extraction – Transformation – Load) framework je základným subsystémom pre integráciu dát v DWH a jeho voľba do značnej miery určuje budúcnosť ďalšieho rozvoja a správy integračnej platformy.

ETL framework by mal najmä spĺňať tieto požiadavky:

* Poskytovať štandardizované prostredie pre vývoj, prevádzku a ďalší rozvoj DWH a integrácie dátových zdrojov a spracovanie dát
* Štandardizovať prístup k vývoju ETL kódu a minimalizovať technické chyby v implementácii
* Podporovať kolaboratívny prístup k vývoju, umožňovať spoluprácu vývojárov a vývojových tímov
* Poskytovať systematickú podporu pre verziovanie kódu a release management
* Poskytovať metadáta pre technickú data lineage
* Poskytnúť prostredie na aplikovanie templatov a prednastavených ETL častí. Vyžaduje sa podpora a prístup low-code, všetky parametre mapovania ETL musia byť uložené v metadátach. Samotné vytvorenie ETL kódu do SQL nastane kompiláciou contextového nastavenia topológie infraštruktúry, zvolenej technológie a znalostných templatov.
* ETL framework by mal mať API na hromadné nastavenie metadát.
* ETL framework by mal mať popísané metedáta pre hromadný export do metadata management aplikácie.
* Poskytnúť možnosť vymeniť template ETL kódu bez nutnosti zmeny nastavenia samotného mapovanie, napr. pri ladení ETL.
* Ako nástroj vývoja spracovania dát by mal byť jednoducho použiteľný a na slovenskom trhu by mal byť dostatok kvalifikovaných IT špecialistov

Na základe hore uvedených požiadaviek je odporúčaný nástroj Oracle Data Integrátor.

Oracle Data Integrátor (ODI) umožňuje jednoducho a rýchlo integrovať údaje z rôznych technológií. Knowledge Module (KM), ktoré sa využívajú pri vytváraní rozhraní prinášajú na rozdiel od tradičných ETL nástrojov, možnosť využiť natívne vlastnosti danej technológie pri spracovaní údajov v jednotlivých krokoch extrakcie, transformácie a nahrávaní. Široké možnosti definovania fyzickej a logickej topológie a ich prepojení pomocou kontextu dávajú veľkú voľnosť pri zmene infraštruktúry bez potreby zásadných zásahov do ELT procesu. Napriek tomu, že KM sú dodávané spolu s produktom ODI je ich možné upravovať podľa potreby konkrétnej riešenej úlohy. Prostredníctvom ODI API je možné vytvárať mapovania ako aj aktualizovať ODI metadáta (topológiu, security, ODI model, atď.). Táto možnosť sa využíva predovšetkým pri hromadných zmenách, prostredníctvom API je možná hromadná zmena KM spolu s vygenerovaním ODI scenárov a ich nasadením do produkčného repozitára. Pri správnom používaní ODI je veľkou výhodou možnosť robiť impact analýzu pri zmenách na zdrojových systémoch a rovnako pri dohľadaní zdroja, ktorý sa používa na aktualizáciu DWH objektov. Okrem toho, že ODI priamo podporuje vytváranie impact analýz, zdokumentované ODI metadáta slúži na generovanie online ETL dokumentácie v rámci Data Governance stratégie.

#### DataLake

**DataLake** je súbor aplikačných komponentov, ktoré majú používateľovi priniesť :

* Distribuovaný súborový systém umožňujúci spoľahlivo uchovávať dáta v rôznych formátoch
* Nástroje umožňujúce vytváranie databáz, tabuliek, partícií a iných komponentov
* Distribuovaný výpočtový framework umožňujúci spracovanie nad dátami
* Grafický interface pre potreby správy platformy
* Granulárne riadenie prístupov do platformy auditnými funkciami
* Správa zdrojov platformy
* Nástroje umožňujúce prístup k dátam použitím tradičného SQL jazyka ako aj programovacích jazykov (R, Python, Scala, ...)

**Správa DataLake**

DataLake platforma je zložená zo súboru technológií, ktoré medzi sebou komunikujú a zabezpečujú bezproblémový chod platformy. Každá technológia, ktorá je súčasťou platformy zastrešuje inú funkcionalitu, ktorá je od DataLake očakávaná. Tieto technológie sú zväčša inštalované na viaceré lokácie z dôvodu zabezpečenia vysokej dostupnosti a robustnosti platformy.

Pre efektívne spravovanie platformy je vyžadované, aby existoval nástroj s grafickým rozhraním, ktorý ponúka možnosti na spravovanie a monitorovanie platformy. V prípade, že by nebol využitý nástroj na správu platformy, je nutné všetky zmeny vykonávať manuálne alebo polo-manuálne prostredníctvom zmien v konfiguračných súboroch, ktoré sa môžu veľa krát nachádzať na rôznych serveroch. Práve z tohto dôvodu vznikli nástroje, ktoré umožňujú jednoduchú správu DataLake platformy, či už platforma pozostáva z troch serverov alebo z desiatok/stoviek.

Pomocou takéhoto nástroja bude možné jednoduchým spôsobom konfigurovať a sledovať stav platformy. Požiadavky na nástroj :

* Grafické rozhranie
* Podpora pridávania a modifikovania DataLake služieb
* Podpora reštartovania služieb, prípadne celej platformy
* Podpora real-time monitorovania stavu platformy
* Podpora analýzy historických dáta z monitorovania behu v podobe grafov a metrík
* Podpora konfigurácie custom alertov na základe aktuálneho stavu platformy
* Možnosť upgradovať platformu priamo z nástroja
* Podpora správy užívateľov

**Storage (Distribuovaný súborový systém)**

Hlavným pilierom DataLake platformy bude distribuovaný súborový systém, ktorý zabezpečuje robustné a spoľahlivé riešenie na ukladanie a následné procesovanie veľkého množstva dát. Distribuované súborové systémy DataLake platformy budú optimalizované na čítanie a zápis veľkého množstva dát v akomkoľvek formáte.

Tieto technológie zväčša využívajú master/slave architektúru. Master server uchováva jednotlivé metadáta o súboroch a na slave serveroch sú rozdistribuované jednotlivé dátové bloky. Vysoká dostupnosť dát bude zabezpečená replikáciou metadát ako aj jednotlivých dátových blokov naprieč servermi DataLake platformy. Replikácia dát zabezpečuje bezproblémový chod platformy aj v prípade výpadku viacerých diskov, poprípade celých serverov.

**ETL framework (DataLake)**

ETL (Extraction – Transformation – Load) framework je základným subsystémom pre integráciu dát v DataLake platforme obdobne ako pre DWH. ETL frameworky umožňujú jednoduchým spôsobom:

* integrovať externé dáta do platformy - štruktúrované, pološtruktúrované aj neštruktúrované dáta (csv, xml, json, logy, pdf, obrázky …)
* efektívne spracovanie dát do finálnych štruktúr
* export dát do ďalších systémov.

Moderné DataLake platformy ponúkajú rôzne prístupy riešenia ETL požiadaviek. V rámci budovania DataLake je potrebné zabezpečiť nasledujúce potreby:

* možnosť čítania dát z externých databáz pomocou štandardných JDBC/ODBC konektorov, SFTP serverov, distribuovaných file-systémov (HDFS, S3, ..)
* distribuované spracovanie dát pomocou štandardných SQL príkazov a syntaxe, ako aj možnosť použitia programovacích jazykov (Python, Java, Scala, R). Technológia by mala umožniť aj kombináciu oboch prístupov (SQL + programovací jazyk) v jednom ETL.
* možnosť exportovať napočítané dáta do štandardných relačných databáz ako aj NoSQL databáz. Taktiež možnosť exportovať dáta priamo do DataLake platformy či už vo forme súborov alebo vo forme dát v tabuľkách.

**Orchestrácia ETL úloh**

Každá úloha musí spustená práve v správnom čase a práve v presnej závislosti na iné úlohy. O tieto úlohy sa stará systém na orchestráciu. Je vyžadované:

* Každá úloha ETL frameworku DataLake-u musí logovať svoj stav do dohodnutej DB tabuľky
* Úlohy ETL DL bude možné zavolať cez WebService z Oracle prostredia aplikačného frameworku pre zabezpečenie globálneho riadenia. T. j. je vyžadované aby makro scheduller bežal na relačnej časti DWH. ETL DL bude riadený vzdialene ako ucelené bloky. Ich podčasti je možné kontrolovať mimo DWH makroscheduller, ale všetky časti bude možné skontrolovať z jedného miesta.
* ETL DL nebude plánované samostatne, ale len ako jeden celok DWH (relačná + DL časť)

**JDBC prístup**

Jednou z najdôležitejších požiadaviek na DataLake platformu je možnosť napojenia štatistických a BI nástrojov na dáta, ktoré sú súčasťou platformy. Najštandardnejším spôsobom, ktorý je využívaný v BI a štatistických nástrojoch je práve spôsob dotazovania pomocou JDBC a ODBC technológií.

DataLake platforma musí poskytovať spôsob napojenia pomocou týchto technológií.

Požiadavky na JDBC/ODBC:

* Možnosť dotazovania pomocou JDBC/ODBC spojenia
* Možnosť granulárnej konfigurácie oprávnení pre JDBC/ODBC spojenia
  + Pre technických aj biznis používateľov
  + obmedzenia na čítanie/zápis len pre vybrané objekty
* Dotaz definovaný v JDBC/ODBC spojení musí byť distribuovane spracovaný DataLake platformou.

**Import súborových dát priamo do DataLake**

DataLake platforma by mala poskytovať rozhranie na ukladanie a čítanie súborov z distribuovaného súborového systému prostredníctvom grafického rozhrania. Týmto spôsobom je možné pridávať, poprípade kontrolovať súbory ktoré sa nachádzajú na distribuovanom súborovom systéme.

Požiadavky:

* Možnosť čítania súborov na distribuovanom súborovom systéme
* Možnosť zápisu dát na distribuovaný súborový systém
* Možnosť granulárnej konfigurácie oprávnení pre manipuláciu so súbormi na úrovni súborov a adresárov.

Pomocou priameho prístupu na distribuovaný súborový systém je možné nahrávať do platformy dátové zdroje, pre ktoré neexistuje automatizovaná integrácia (tento prístup je možné použiť aj v prípade manuálnych korekcií dát).

**Resource Management**

DataLake platforma by mala poskytovať Resource Management službu, ktorá by zastrešovala prerozdeľovanie dostupného výpočtového výkonu DataLake platformy medzi technické aplikácie a biznis užívateľov. Je nutné si uvedomiť že aplikácie, ktoré budú využívať výpočtový výkon DataLake platformy sú súčasťou ETL spracovania alebo sú súčasť biznisových procesov (machine learning, ad-hoc štatistika, ...). Resource Management služba by mala byť schopná zabezpečiť možnosť definovať prioritu a dostupné zdroje (CPU + RAM) pre rozdielne ETL aplikácie a rozdielnych biznis užívateľov (poprípade skupinu užívateľov). Takýmito pravidlami je možné zabezpečiť vyváženosť prerozdeľovania výpočtového výkonu DataLake platformy medzi rozdielne aplikácie.

V prípade, že by služba Resource Managementu nebola dostupná pre DataLake platformu mohli by nastávať situácie využívania celkového výkonu platformy v prospech jednej (alebo viac) aplikácie, pričom ostatné aplikácie by neboli schopné alokovať potrebné zdroje.

Resource manager DataLake platformy by mal byť schopný zabezpečiť:

* Možnosť konfigurácie resource group na úroveň užívateľov, skupiny užívateľov a skupiny technických aplikácií
  + Minimálne a maximálne zdroje.
* Efektívne prerozdeľovanie dostupných zdrojov
* Monitoring dostupných a využitých zdrojov

**Bezpečnosť (security)**

DataLake platforma musí spĺňať bezpečnostné štandardy, ktoré zabezpečia že dáta uchovávané a spracúvané v DataLake platforme budú dostupné len používateľom a aplikáciám s prislúchajúcimi právami.

Nároky na bezpečnosť prostredia DataLake platformy:

* Centrálne riadené
* Možnosť granulárnej konfigurácie oprávnení
  + databázy, tabuľky, stĺpce
  + adresáre a súbory prislúchajúceho distribuovaného súborového systému
* Grafické rozhranie na správu prístupových oprávnení
* Auditovanie údajov (všetky aktivity na DataLake musia byť auditované a dostupné v reálnom čase)
  + Prístupy a modifikácie tabuliek a databáz
  + Prístupy a modifikácie na súborovom systéme
  + Spúšťanie distribuovaného spracovania

### Metodika vývoja

Pri vývoji DWH musia byť dodržané metodické postupy, aby návrh DWH bol:

* **Nezávislý na platforme** - Navrhnutý dátový model by mal byť v čo najväčšej miere nezávislý od technológie, na ktorej bude DWH prevádzkovaný. Dátový model by mal byť prenositeľný medzi jednotlivými platformami (MS SQL, Oracle, PostgreSQL, atď.) a nemal by sa spoliehať na špecifiká vybraného databázového systému. Rozdiely plynúce zo syntaxe resp. samotného jazyka DDL jednotlivých platforiem nie sú považované za systémové špecifikum. Súčasť dodávky však budú tvoriť DDL definície iba pre dodanú platformu. V rámci projektu NBS DWH sa prepoužije technológia z Pilotu DWH.
* **Nezávislý od zdrojových systémov** - Navrhnutý dátový model by mal byť čo najmenej závislý od zdrojových systémov. Predpokladá sa, že v dlhodobom horizonte, v čase životnosti nového dátového skladu dôjde k výmene niektorých biznis systémov.
* **Flexibilný** - Dátový sklad by mal byť z dlhodobého hľadiska rozšíriteľný na nové okruhy dát podľa budúcich požiadaviek zákazníka ako napr. nové typy produktov, zmena procesov atď.
* **Časová nezávislosť dát** - Navrhnutý dátový model musí zabezpečiť ako aktuálny, tak aj historický pohľad na dáta v DWH na základe výberu časového intervalu. Dátová retencia bude dohodnutá počas projektu. Predpokladaná dĺžka histórie je 10 rokov, granularita dát minimálne na úrovni kalendárneho dňa.
* **Aktuálnosť** - Dáta v dátovom sklade by mali byť aktualizované minimálne raz denne. V DWH by sa mal nachádzať stav dát s platnosťou minimálne D-1. Avšak môžu existovať definované typy dát, kde bude potrebné navrhnúť a použiť aj nižšiu frekvenciu aktualizácie (cyklické near-online spracovanie).
* **Interaktívny a inkremetálny vývoj** - Ktorákoľvek úloha v rámci projektu, alebo po ňom môže byť predmetom iterácie. Či iterovať alebo nie, rovnako ako počet iterácií, toto rozhodnutie je rôzne od prípadu k prípadu. Úlohy môžu byť iterované za účelom zvýšenia kvality výstupov na požadovanú úroveň, za účelom získania dostatočnej miery detailu alebo vylepšenia a rozšírenia výstupov na základe spätnej väzby používateľa.

Uchádzač bude preukázateľne disponovať skúsenosťami v oblasti dátového modelovania a návrhu dátových modelov pre dátové sklady. V rámci projektu zhotoviteľ navrhne a zdokumentuje:

* metodiku tvorby a ďalšej správy dátových modelov relevantných pre projekt
* súvisiace dizajn štandardy (štandardy a postupy návrhu dátových modelov v jednotlivých vrstvách DWH a DataLake (typy modelov, ich účel, previazanosť a forma zápisu)
* menné konvencie pre jednotlivé typy modelov a dátové vrstvy
* prevádzkový model, t. j. popíše procesy a role súvisiace so správou modelov (reverse engineering zdrojov, design nových modelov, kolaborácia používateľov a riadenie prístupu, exporty resp. publikácia metadát pre zdieľanie, release management)

Tieto štandardy budú odsúhlasené s Data Governance kanceláriou a následne aplikované počas realizácie projektu.

V rámci projektu budú zavedené minimálne tieto dátové modely a súvisiace výstupy:

* Konceptuálny dátový model – dokumentujúci biznisové koncepty a pojmy podľa definície NBS
* Fyzické dátové modely:
  + DataLake
    - Primárna vrstva
    - Discovery vrstva
  + DWH
    - Transakčné repozitory (L1 – Integračná vrstva)
    - Prezentačná vrstva (L2 - Analytická vrstva)

Úroveň detailu zachytenia fyzických parametrov databázových objektov bude dohodnutá tak, aby efektívne podporovala dizajn, vývoj, nasadenie príslušných objektov a súčasne udržateľnosť modelu a jeho dokumentačnú funkciu.

Podobne požadujeme dodanie popisu dátových modelov a procesov jednotlivých funkčných frameworkov (napr. schémy s referenčnými dátami, lookup tabuľkami, chybovými tabuľkami, metadátami a výsledkami kontrol kvality a automatických testov, manuálne korekcie, žurnál spracovania apod.)

Dátové toky a transformácie dát budú systematicky a metadátovo popísané tak, aby bolo možné tieto metadáta využiť pre navigáciu v dátových tokoch, zisťovanie dátových zdrojov pre cieľové atribúty a analýzu dopadov pri zmenách v dátových zdrojoch na cieľové štruktúry.

Nasledujúce princípy návrhu dátového modelu musia byť počas projektu dodržané a považujeme ich za kritické pre konzistentnosť, integritu, otvorenosť a výkonnosť DWH riešenia ako celku.

* Model transakčnej (L1) a prezentačnej vrstvy (L2/PL) sú integrované a spoločné. Jedna entita je mapovaná do jednej tabuľky bez ohľadu na počet inštancií danej entity v zdrojových systémoch.
* Dimenzie v prezentačnej vrstve sú denormalizované a zdieľané a využívané cez všetky biznis oblasti. Existuje iba jedna dimenzia zákazníkov, iba jedna dimenzia umiestnení, atď. Zdieľané dimenzie podporujú krížové analýzy cez jednotlivé biznis oddelenia v spoločnosti.
* Prezentačná vrstva pozostáva zo star schém, ak je to možné bez snowflake. Inými slovami faktová tabuľka je spojená priamo s dimenziou nie je nutný viacnásobný join. Je možné, že budú existovať výnimky, ktoré je potrebné pred implementáciou explicitne schváliť NBS.
* Všetky dimenzie a referenčné tabuľky sú prepojené výlučne pomocou umelých kľúčov. Umelé kľúče chránia dátový sklad pred zmenami v zdrojových systémoch a umožňujú integráciu viacerých zdrojov do jednej tabuľky v DWH.
* Všetky dimenzie, prepojovacie lookup tabuľky (v prípade potreby aj faktové tabuľky) obsahujú aj prirodzené kľúče a identifikátor zdrojového systému. To znamená, že záznam v DWH môže byť dohľadaný v zdrojovom systéme.
* Multitenant – globalizácia a efektívne využívanie zdrojov vedie k situáciám, kedy v jednom zdrojovom systéme sa prevádzkujú údaje viacerých spoločností. V tomto prípade nestačí v DWH rozlišovať z ktorého zdrojového systému prišiel do DWH daný záznam, ale je potrebné riešiť kto „vlastní“ daný záznam v zdrojovom systéme. V DWH je potrebné sledovať vlastníka („tenanta“) daného záznamu. Táto informácia sa používa v určitých typoch reportov, kde je potrebné rozdeliť údaje o jednom zákazníkovi z pohľadu daného vlastníka napriek tomu, že ide o jednu a tú istú právnickú alebo fyzickú osobu.
* V prípade potreby budú údaje v DWH unifikované. Bude dohodnuté, ktoré dáta a ako budú unifikované. Prístup k unifikácií využíva dve rozdielne techniky:
  + Riadková unifikácia, napríklad v prípade zákazníkov. V tomto prípade je jedna informácia v DWH reprezentovaná jedným záznamom bez ohľadu na počet záznamov v zdrojových systémoch. Údaje sú unifikované pomocou rôznych čistiacich a unifikačných procedúr.
  + Druhý prístup je unifikácia pomocou hierarchií. Tento prístup sa využíva v prípade dimenzií (napr. produkty). V tomto prípade má každý produkt v DWH jeden záznam rovnako ako v zdrojovom systéme. Nad takouto dimenziou je vytvorená hierarchia, ktorá v reportoch zgrupuje identické produkty do jedného záznamu.
* Všetky riadky v tabuľke majú zhodnú granularitu a význam. To znamená, že nie je akceptovateľné miešať detailné a agregované údaje alebo údaje s rôznou granularitou.
* Integritná konzistencia (primárne, unikátne a umelé kľúče) bude definovaná pre všetky tabuľky v transakčnej a prezentačnej vrstve. Avšak pre veľké tabuľky bude integrita dát kontrolovaná zabezpečená designom ETL, nebude kontrolovaná na úrovni DB, z dôvodu výkonnostných problémov ELT.
* Pre veľké tabuľky je nutné využiť Oracle DB partície, aby sa uľahčil ELT proces, zvládol nárast historických údajov a zlepšila výkonnosť ako používateľských dotazov, tak aj ELT nahrávania dát.
* Jedným zo základných princípov pri budovaní DWH je, že budovanie histórie v transakčnej vrstve DWH je s prihliadnutím, že údaje sa v DWH nikdy fyzicky nemažú. Počas projektu bude stanovená retenčná doba jednotlivých dát.
* V rámci dátového modelovania budú udržiavané prípadné ďalšie metadáta k jednotlivým štruktúram na podporu data governance procesov, ako sú napr.:
  + klasifikácia dát z pohľadu citlivosti a GDPR
  + väzby na referenčné dáta / katalógy / domény
  + účelu použitia
  + prípadne iné podľa požiadaviek Data Governence kancelárie

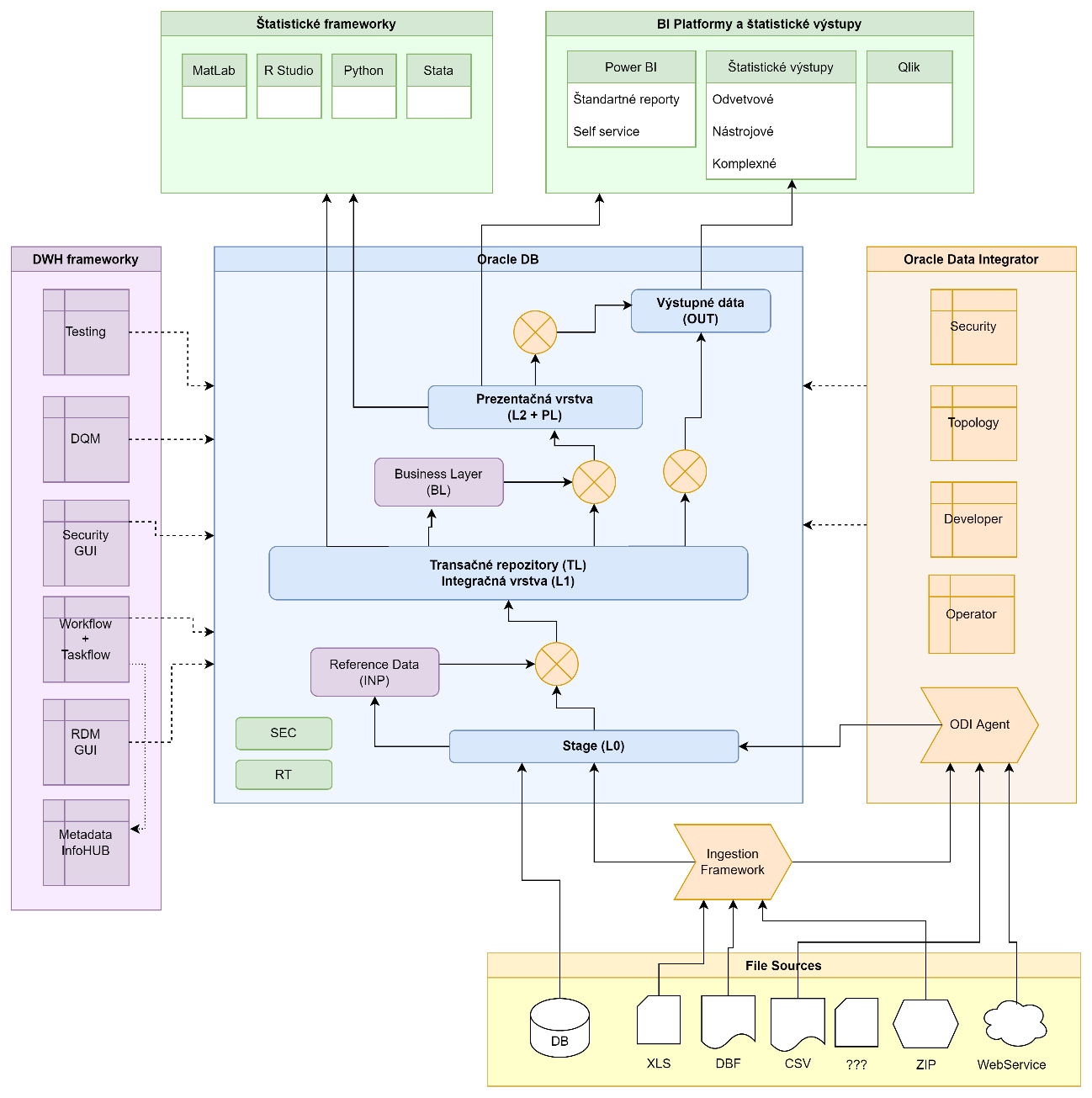
Dátový sklad na prezentačnej úrovni bude budovaný ako tzv. dimenzionálny dátový sklad, ktorý obsahuje dva základné typy objektov:

* Fakty, t. j. numerické hodnoty, ktoré sú sledované. Fakty sú spravidla aditívne, resp. semiaditívne hodnoty, ktoré je možné sčítať resp. spriemerovať podľa požadovaných kritérií. Fakty sú zoskupené do tzv. faktových tabuliek.
* Dimenzie, t. j. kritériá, na základe ktorých sú fakty členené. Príkladom je čas, nákladové stredisko, organizačná štruktúra atď. Dimenzie väčšinou obsahujú veľké množstvo atribútov a úrovní, ktoré prvok na dimenzii charakterizujú. Z úrovní je možné tvoriť hierarchie a prirodzeným spôsobom tak prechádzať z agregovanej hodnoty (vyššia úroveň) na detailnejšie dáta (nižšia úroveň). Dimenzie v dimenzionálnom dátovom sklade musia byť unifikované pre zdieľanie medzi rôznymi faktovými tabuľkami. Jedine tak je možné prechádzať („Drill Accross“) medzi faktami z rôznych faktových tabuliek - t. j. na základe rovnakých hodnôt spoločných dimenzií.

### Relačná časť NBS DWH

#### Úvod

NBS DWH bude logicky členený na niekoľko dátových vrstiev, pričom každá bude plniť špecifickú úlohu. Vrstvy sú navrhnuté v súlade so zavedenými prístupmi pre vnútorný návrh DWH.



Obrázok 12: DWH Relačná časť - vrstvy DB a jej aplikačné časti

Preto je požadované v  návrhu DWH dodržať zásady spísané v nasledujúcich podkapitolách.

#### Stage oblasť/L0

Vstupná vrstva dátového skladu, ktorá obsahuje dáta v štruktúre zdrojových systémov, resp. ich výstupných interfejsov. Účel tejto vrstvy je:

* Poskytnúť obraz vzájomne konzistentných zdrojových dát pre spracovanie v DWH na spoločnej technologickej platforme
* Štandardizovať dátové typy (zmena časového pásma dátumu/času, štandardizácia číselných údajov, ignorovanie tzv. white spaces v textových poliach, úprava kódovej stránky a podobne)
* Poskytnúť krátkodobú cache (historizáciu vstupných dát) pre prípadné dohľadávanie chýb a podozrivých stavov v dátach

Dáta môžu byť prenášané do STAGE vrstvy vo forme kompletných kópií zdrojových dát (napr. kmeňové dáta a referenčné dáta) alebo zmenových prírastkov (nové transakcie, resp. záznamy, kde došlo k zmene od posledného spracovania). Pre každý dátový interface (zdrojový systém, je jedno či ide o databázu, súbor, alebo záložku napr. v XLS) je vyžadované aby vznikol dokument tzv. „Dohoda o extrakcii“ (Interface agreement), template dodá NBS. Kde bude popísané:

* Forma interface (DB, súbor, WebService)
* Veľkosť zdrojového systému a predpokladaný denný/mesačný inkrement
* Spôsob prenosu dát (dblink, FTP, JDBC, ODBC, http, ...)
* Umiestnenie systému (IP adresa, Linka, ...)
* Predpis extrakcie a štandardizácie dátových typov
* Základný ER diagram zdroja
* Spôsob prípravy vzorky dát pre TEST a DEV prostredie
* Identifikácia citlivých a osobných údajov
* Nastavenie a spôsob (inkrement, celá snímka) historizácie vstupných dát – ak je požadovaná, alebo potrebná

Stage vrstva musí mať vlastnosť samočistenia, t. j. bude sama odmazávať a defragmentovať priestor na základe metadát z nastavenia historizácie vstupných dát.

Názvy tabuliek a stĺpcov by mali reflektovať názvy v zdrojových systémoch. Počas projektu bude dohodnutá názvová konvencia, ktorá stanoví štandardy pre názvy tabuliek (a ich prefixy/sufixy), spôsob pomenovania biznis stĺpcov a názvy technických stĺpcov.

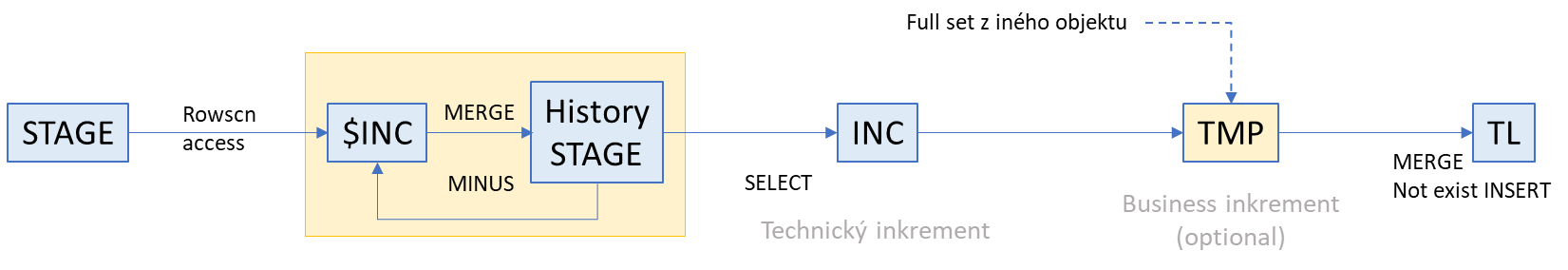
#### Princípy zachytávania zmien (Change Data Capture)

Prístup k zachytávaniu zmien údajov bude implementovaný spôsobom, ktorý minimalizuje množstvo údajov, ktoré majú byť spracovávané a taktiež zmeny riadkov, ktoré sú potrebné pre SCD. Budú použité niektoré z nasledujúcich techník, ktoré sú plne podporované ODI.

* Priamo systém poskytuje interface s potrebnými dátami na extrakciu.
  + Dáta budú zo systému priamo extrahované a uložené do STAGE. Toto uloženie bude označené metadátovou značkou, kedy dáta boli prenesené pre spätnú kontrolu kvality prípravy extraktu na zdroji – identifikácia prípadných duplicít, alebo opravných setov dát.
* Systém neposkytuje priamo dáta.
  + Pre malé a stredne veľké tabuľky budú sa robiť full extrakty zo zdrojových systémov. Tieto full extrakty sa porovnajú s údajmi, ktoré sa už nachádzajú v DWH buď množinovou operáciou MINUS alebo pomocou FULL OUTER JOIN na identifikáciu zmien.
  + Ak zdrojové tabuľky obsahujú spoľahlivú časovú známku, budú extrahované zo zdrojového systému riadky, ktoré majú časovú známku novšiu ako je predchádzajúca extrahovaná časová známka. Prenesené dáta musia byť označené metadátovou značkou.
  + V prípade, že extrakcia založená na časových známkach nie je možná, pripadá do úvahy extrakcia s využitím DB trigrov. Vyžadujú si úpravy na strane zdrojových systémov (vytvorenie DB trigrov – ako je MVIEW a podobne) a preto nepredpokladáme využitie tejto metódy v NBS DWH.

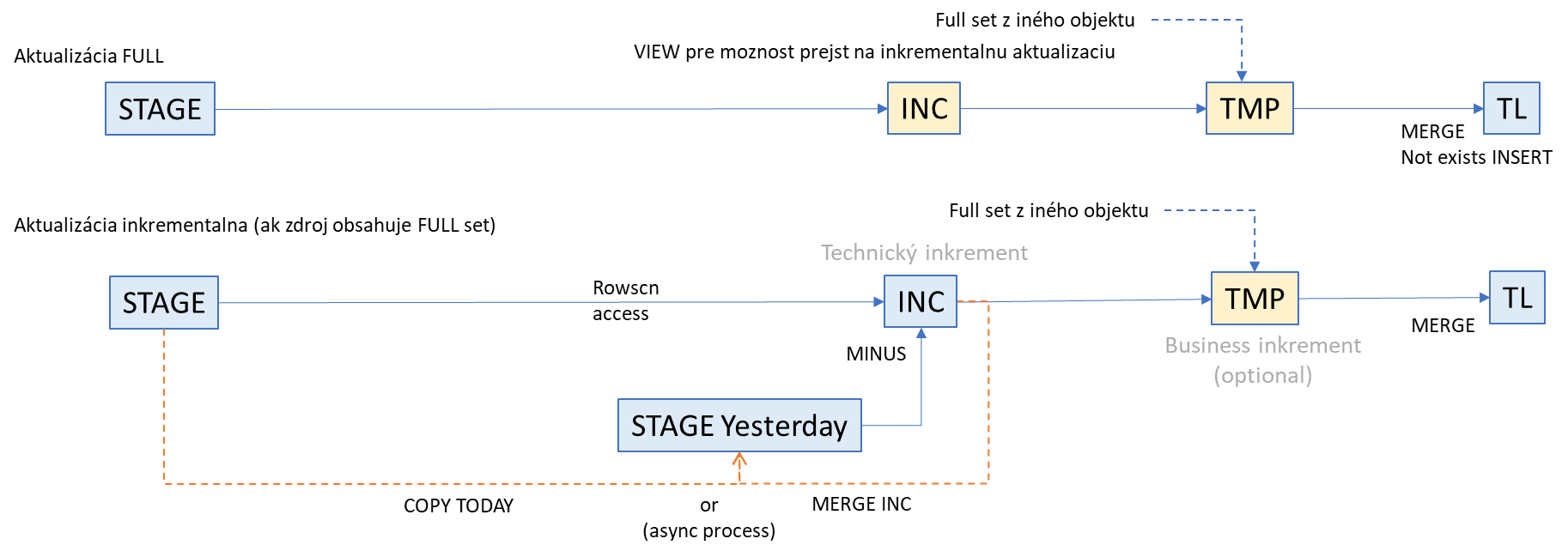
Pre účely riadenia zápisu zistených zmien na zdrojovej strane sa vyžaduje použitie nasledujúcej schémy ETL procesy:

**1. Pri použití historizovaného STAGE/L0**



Obrázok 13: ETL schéma historizovanej vrstvy STAGE

**2. Pri použití klasického nehistorizovaného STAGE/L0**



Obrázok 14: ETL schéme nehistorizovanej vrstvy STAGE

#### Reference Data Oblasť (INP)

Účel tejto vrstvy je:

* Udržať konzistenciu dát na úrovni číselníkov
* Vstupná brána pre používateľov cez aplikačné GUI pre doplnenie dát do DWH
* Poskytnúť cez REST API tieto číselníky ostatných odberateľom v NBS

Táto oblasť preto obsahuje číselníky a čistiace, či mostíkové tabuľky pre:

* mapovanie zdrojových hodnôt na DW kľúče,
* doplnenie hierarchií a údržbu master dimenzií,
* pridávať ďalšie informácie importom súborov cez aplikačné GUI

Návrh dátového modelu tejto časti je silne v tretej normálnej forme s plným nastavením primárnych a unikátnych indexov pre zabezpečenie kvality dát. Rovnako vzťahy medzi objektami typu master-slave musia byť podporené cez FK (vzdialený alebo cudzí kľúč) a iné obmedzenia.

#### Transakčné repozitory (L1 – Integračná vrstva)

Táto oblasť je základ dátového skladu, kde sa uchovávajú údaje na najnižšom detaile ako je možné. V transakčnej vrstve sa tvorí jedna pravda na jednom mieste, ktorá generuje históriu a sleduje zmeny v čase. Pri aktualizácii sú povolené výlučne technické transformácie a biznisové transformácie, ktoré nie sú ovplyvňované rozhodnutiami biznisu. Táto vrstva je striktne navrhovaná ako tretia normálna forma (metodika Billa Inmona) a vytvára deliacu čiaru medzi ďalšími komponentmi dátové skladu. Ak je potrebné zmeniť alebo doplniť ďalší zdrojový systém, upraví sa len prislúchajúca časť ETL v tejto vrstve. Všetky ostatné časti DWH sa nemenia. Navyše táto vrstva vytvára predpoklady na integráciu a normalizáciu údajov. Keďže transakčná vrstva zohľadňuje všetky relácie medzi objektmi, je možné vytvárať komplexné analýzy a dotazy, ktoré sú nad rámec dimenzionálneho sveta.

Pri návrhu transakčnej časti DWH je nutné v súčasnosti prihliadať aj na nové poznatky a metodiky, ktoré dopĺňajú a rozširujú skôr spomenuté prístupy modelovania. Pomáhajú zrýchliť, zjednodušiť ETL procesy napríklad použiť aj princípy modelovania metódy Data Vault 2.0[[1]](#footnote-2) .

Ako nadstavba nad týmto transakčným repozitárom je vytvorená prezentačná vrstva v zmysle dimenzionálneho modelu ako ju navrhuje Ralph Kimball pre vytváranie reportov a analýz.

Dátový model zachytáva konkrétne koncepty, entity, ich vzťahy a klasifikácie nezávisle na tom, ako sú zodpovedajúce dáta organizované v zdrojových systémoch. Vlastníkom dátového modelu je NBS.

T. j. cieľom tejto vrstvy dát je:

* Poskytnúť aktuálny aj historický pohľad na dáta, resp. ich historické verzie k akémukoľvek bodu v čase (použitie umelých kľúčov, vymedzenie platnosti záznamov – tzv. SCD2 mechanizmus u „pomaly sa meniacich“ dimenzií a väzobných tabuliek, resp. transakčné a snímkovacie tabuľky pre „rýchlo sa meniace údaje“)
* Organizovať dáta zrozumiteľne podľa biznisových entít (entitno-relačný model), t. j. dáta z rôznych systémov zodpovedajúce jednej entite sú uložené spoločne, t. j. tabuľky odpovedajú reálnym konceptom/entitám, ktorým rozumie biznis užívateľ (napr. subjekt, transakcia, inštrument, cenný papier, účet, výkaz, položka výkazu, apod.)
* Štandardizovať, resp. unifikovať dáta (vynútenie jednotného spôsobu zápisu údajov s rovnakým významom)
* Nahradiť neznáme, prípadne nekvalitné dáta dohodnutým spôsobom (default hodnoty a ošetrenie tzv. „late arriving“ údajov)
* Konsolidovať klasifikácie dát z rôznych zdrojov pre spoločný reporting (okrem uloženia pôvodných referenčných hodnôt to prebieha ako premapovanie cez konsolidované číselníky pomocou dát uložených v „Reference Data Oblasti“).
* Konsolidovať dátové entity podľa konceptuálneho návrhu dátového skladu. Ako sú napr.:
  + údaje o subjektoch (právnických aj fyzických osobách), ktoré môžu byť reprezentované v rôznych zdrojových systémoch pomocou rôznych identifikátorov a kódov – ich prepojenie bude implementované cez dohodnutý algoritmus tak, aby bolo možné prepojiť údaje súvisiace s konkrétnym výskytom subjektu pochádzajúcich z rôznych zdrojov.
  + údaje o entitách typu cenné papier/finančný inštrument.

Pri návrhu bude potrebné počítať minimálne s týmito požiadavkami:

* Dáta budú spracúvané a ukladané čo najoptimálnejším spôsobom podľa best-practice budovania dátových skladov s využitím technológií databáze (partitioning, paralelné spracovanie, správa indexov, optimálne rozloženie dát podľa ich povahy a volatility apod.)
* Bude vytvorený mechanismus pre archiváciu (resp. presun do záložných štruktúr) a výmaz nepotrebných historických dát z DWH podľa zadaných parametrov na úrovni jednotlivých tabuliek.

Pre každú tabuľku bude zabezpečený iniciálny load, ktorý zabezpečí prvotné naplnenie všetkými dátami, ktoré sú potrebné pre ďalšie inkrementálne spracovanie so zachovaním referenčnej integrity. Historický load bude zabezpečený pre zvolené dátové oblasti podľa definície jednotlivých dátových zdrojov, t. j. pre každý dátový zdroj je individuálne definovaná hĺbka histórie, spracované budú relevantné cieľové tabuľky DWH. Cieľom historického loadu je rekonštruovať historické stavy dát, t. j.:

* historické transakcie / udalosti
* historické položky dimenzionálnych tabuliek a číselníkov (kmeňových dát), na ktoré sa odkazujú historické fakty pre zachovanie referenčnej integrity
* historické stavy hodnôt atribútov v dimenzionálnych tabuľkách
* historické snímky stavových ukazovateľov a relevantných popisných údajov
* historické platnosti vzťahov (relácií) medzi entitami

#### Biznisová oblasť

Podobne ako oblasť DWH nazývaná „referenčné dáta“, ktorá obsahuje dimenzie a číselníky pre doplnenie ETL procesu počas normalizácie a konsolidácie zo stage do transakčnej oblasti, tak oblasť nazývaná „biznisová oblasť“ obsahuje mostíky, šablóny a prioritizačné tabuľky pre riadenie (ak je to potrebné, nie každé ETL to potrebuje) spracovanie z transakčnej oblasti do prezentačnej vrstvy.

Na prvý pohľad číselník a mostík vyzerajú rovnako. Do číselníka môže biznisový používateľ cez RDM systém zadávať hodnoty, pridávať hierarchie, alebo definovať vzťahy medzi dátami. Ale väčšinou ide o jednu informáciu v jednej dimenzii. V mostíku sa môže definovať vzťah medzi viacerými dimenziami a ako sa má tento vzťah konvertovať na novú dimenziu. T. j. z jednej a viac vzniká nová odvodená (premostená) hodnota dimenzie. Kde toto spojenie nie je triviálne čo do počtu kombinácií (nakoľko ide o kartézsky súčin všetkých hodnôt v dimenziách) požadujeme vytvoriť systém využívajúce šablóny, kde biznisový používateľ bude vedieť definovať pravidlá cez „stanovené postačujúce podmienky“ (presne definované pravidlá) a zvyšné časti opakujúcich sa spojení systém dogeneruje počas (alebo pred) ETL spracovania.

#### Prezentačná vrstva (PL + L2 - Analytická vrstva)

Prezentačná vrstva je replikovaná časť transakčného repozitára, vytvorená na základe špeciálnej „biznisovej“ požiadavky. Zvyčajne sa navrhuje metódami Ralpha Kimballa a má dimenzionálny charakter, t. j. táto vrstva DWH poskytuje pohľad na dáta v štruktúre, ktorá je vhodná pre konkrétne analytické resp. reportingové účely. Organizácia dát je podriadená konkrétnym požiadavkám užívateľov a BI aplikácií.

Na rozdiel od technických transformácii sa v priebehu ETL procesu z transakčnej do prezentačnej vrstvy využívajú predovšetkým biznisové transformácie, ktoré sa v čase môžu zmeniť. V prípade zmeny biznisovej definície je možné podľa požiadavky prepočítať spätne časť alebo celú históriu na základe údajov uchovaných v transakčnej vrstve. V ETL proces sa využívajú špeciálne transformačné techniky (na základe definícií uložených v biznisovej vrstve DWH), ktoré umožňujú údaje transformovať na základe používateľských mostíkov, šablón a priorít.

Účelom tejto vrstvy je teda:

* Organizovať dáta optimálne pre konkrétne analytické a reportingové požiadavky do star-schém s potrebnou denormalizáciou a prípadnou agregáciou
* Poskytovať dodatočné vypočítané/odvodené dáta, ktoré sú špecifické pre danú analytickú potrebu.

Pri návrhu bude potrebné počítať minimálne s týmito požiadavkami (okrem tých čo sú uvedené pre L1 vrstvu a v celkovej metodike vývoja):

* Dátový model bude navrhnutý ako dimenzionálny s potrebnou úrovňou denormalizácie dátového modelu. Bude obsahovať detailné, prípadne aj agregované dáta. Nepredpokladá sa však ukladanie dát v štruktúre výstupných reportov.

#### Výstupná vrstva (OUT)

Výstupná vrstva je obdobou „Prezentačnej vrstvy“, dáta sú však organizované do dátových štruktúr, ktoré sú požadované na výstupe odoberajúcim systémom, resp. externým užívateľom (open data, ECB, iné inštitúcie, Off loading do DataLake a pod.). Táto štruktúra teda nemusí spĺňať princípy pre modelovanie analytických dát pre reporting, je podriadená špecifickým požiadavkám následného použitia, resp. zverejnenia.

Z tejto vrstvy sa generujú výstupné extrakty pre distribúciu (disemináciu).

Podobne ako stage vrstva je definovaná dokumentom „Interface Agreement“ (Dohoda o extrakcii), tak OUT vrstva musí byť definovaná a dohodnutá cez dokument „Dohoda o poskytnutí dát z DWH“. Podobne ako dokument „Dohoda o extrakcii“ musí byť dokument popisujúci OUT vrstvu podpísaný oboma stranami dátovej komunikácie.

Dáta v OUT vrstve sú tvorené ETL procesmi z transakčnej vrstvy, ale aj z dát prezentačných datamartov.

#### DWH SandBox

DWH SandBox je dohodnuté miesto tvorené špeciálnym používateľom, ktorý má právo:

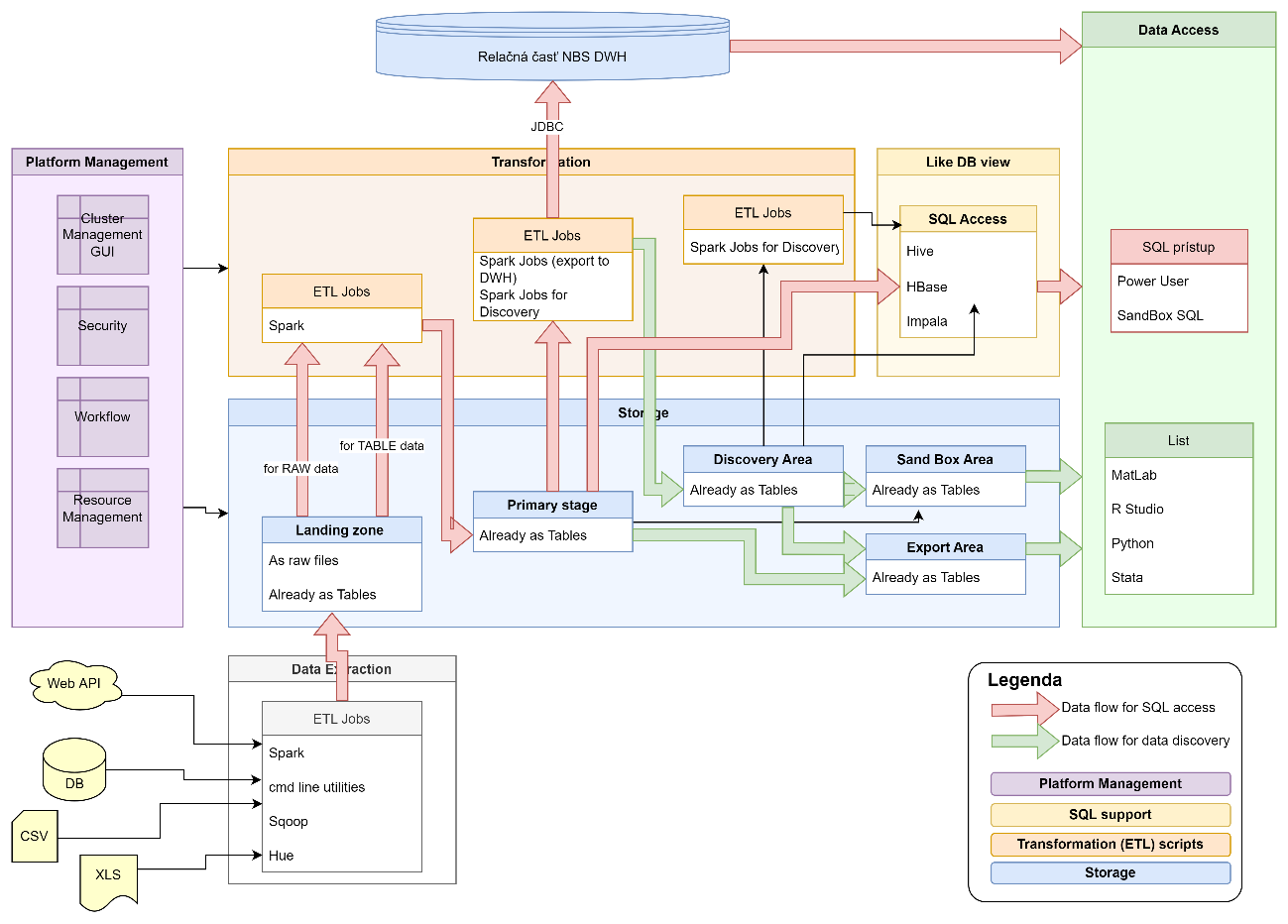
1. čítať dohodnuté dáta z DWH/DL
2. vytvárať nové tabuľky a pohľady
3. napĺňať tieto tabuľky
4. vymazávať dáta a celé objekty
5. pre pokročilých používateľov aj vytvárať procedúry a iné programy, ktoré môžu slúžiť na spustenie časti ETL a napr. pre overenie určitej hypotézy.

Sandbox nemá slúžiť na pravidelné prevádzkovanie časti spracovania DWH/DL. Pokiaľ je používateľ spokojný z pokusom je veľmi dôležité zapojiť IT do finálnej optimalizácie a zaradenia do pravidelného (denného) behu vytvorenej novej časti spracovania.

### DataLake časť NBS DWH

#### Úvod

Podobne ako pri návrhu relačnej časti DWH musia tieto predpoklady platiť pri vývoji DataLake. Nasledujúce podkapitoly popisujú úlohu DataLake platformy v navrhovanej architektúre.



Obrázok 15: DWH DataLake vrstvy a jej toky dát

Dátové vrstvy DataLake budú logicky rozdelené na niekoľko vrstiev tak ako v prípade logického členenia dátových vrstiev DWH. Dátové vrstvy DataLake platformy môžu kopírovať aj DWH model.

#### Landing zóna

Dáta nachádzajúce sa v tejto oblasti obsahujú 1:1 informácie zo zdrojových systémov. Dáta sa môžu nachádzať buď vo forme tabuliek (štruktúrované dáta exportované zo zdrojových DB) alebo vo forme RAW súborov rôznych formátov pripravených na ďalšie spracovanie. Dáta môžu prichádzať buď inkrementálne alebo vo forme celkových snímok.

DataLake by mal poskytovať mechanizmy na import dát z nasledujúcich typov zdrojových systémov:

* Databázové (za použitia JDBC/ODBC).
* Súborové (možnosť stiahnuť dáta z externého SFTP, interného SFTP , alebo interného Sharepoint SSL úložiska (časť dát z externých inštitúcii, kde sa momentálne využíva protokol WebDAV sa presunie na zabezpečené Sharepoint SSL úložisko).
* API + WEB (možnosť importovať dáta dostupné cez API rozhranie (SOAP/REST), alebo dát dostupných za použitia parsovania public webu).

Podobne musí DataLake umožňovať priamo zapisovať dáta do Landing zóny iným ETL nástrojom.

#### Primárna stage

Dáta nachádzajúce sa v tejto oblasti obsahujú informácie, ktoré boli spracované z Landing vrstvy. Dáta v tejto vrstve majú vopred určenú štruktúru na základe biznisových a technických požiadaviek (vo forme tabuliek/partícií alebo dátových súborov). Dáta v tejto vrstve už nemusia obsahovať všetky informácie, ktoré boli exportované zo zdrojových systémov. Typickým príkladom v prostredí NBS môžu byť rozsiahle XML súbory, ktoré sú exportované zo zdrojových systémov, pričom len malé množstvo atribútov je následne procesovaných vo vyšších vrstvách. Pred vstupom dát do Primárnej stage vrstvy sú dáta kontrolované na dátovú kvalitu a transformované do finálnej podoby. Čiže hlavné transformácie dát spracúvaného obdobia zo zdrojových štruktúr do štruktúr Landing integračnej vrstvy sú:

* Štandardizácie/unifikácie dát
* Technických kontrol dátovej kvality
* Ošetrenia default hodnôt
* Výpočet spoločných odvodených údajov
* Historizácie dát v Primárnej stage

Dáta z primárnej stage vrstvy môžu byť exportované do iných systémov pre potreby ďalšieho spracovania.

#### Discovery

Dáta nachádzajúce sa v Discovery vrstve obsahujú spracované dáta pripravené na použitie pre analytické a štatistické potreby. Tieto dáta sú obohacované a transformované na základe biznisových a technických požiadaviek ETL procesmi, ktoré zabezpečujú potrebnú kvalitu dát.

Transformácia dát spracúvaného obdobia zo zdrojových štruktúr Primárnej stage do Discovery analytickej vrstvy prebieha najmä z dôvodu:

* Denormalizácie dát
* Dodatočných výpočtov
* Optimalizácie pre reporting (napr. materializácie výpočtov, ich joinov a agregácií)
* Historizácie dát v Discovery vrstve podľa konkrétnej analytickej potreby

Štatistické a analytické nástroje využívajú primárne dáta z tejto vrstvy na vytváranie analýz a reportov.

#### DataLake Sandbox

DataLake platforma by mala umožňovať dátovým vedcom vytváranie “sandboxov“ pre potreby importovania a manipulácie aj väčšieho množstva dát. Výhodou takýchto “sandboxov“ je možnosť spájania externých dát s dátami, ktoré sú dostupné v DataLake platforme bez nutnosti zapojenia IT. Platforma by mala poskytovať jednoduché rozhranie, ktoré by dátovým vedcom umožňovalo import externých dát v podobe súborov a možnosť nad takto importovanými súbormi vytvárať databázové tabuľky, poprípade používať tieto súbory pre účely analýz za použitia programovacích/štatistických jazykov. DataLake platforma by mala umožňovať pri analyzovaní takto importovaných súborov využívať celkový výpočtový potenciál platformy.

Dáta v sandboxoch by mali byť logicky oddelené a nemali by byť súčasťou dát v Discovery vrstve.

### Informačná bezpečnosť

#### Autentifikácia používateľov

Systém musí zabezpečiť jednoznačné overenie používateľa pred akýmkoľvek prístupom. Požiadavky na autentifikáciu:

* Každý používateľ musí mať jedinečný prístupový identifikátor/užívateľské meno (UID)
* Autentifikácia sa bude riadiť autentifikačnými štandardami NBS, je nutné predpokladať, že bude použitá dvojfaktorová autentifikácia
* Každý užívateľ sa musí autentifikovať minimálne pomocou UID a hesla.
* Heslá musia byť v systéme uložené tak aby boli nedešifrovateľné
* Systém by mal ponúkať možnosť napojenia na SSO riešenie

#### Prístup k dátam

Systém musí poskytnúť možnosť na riadenie prístupov na všetkých úrovniach technickej architektúry. Aplikované postupy a princípy riadenia oprávnení používateľov budú zdokumentované v relevantnej dokumentácii riešenia.

Prístup k dátam musí byť definovateľný:

* Na úrovni rolí jednotlivých používateľov minimálne pomocou riadenia prístupu k jednotlivým entitám dátového modelu pre analytické role.
* V systéme sú identifikované nezlučiteľné role a oprávnenia. T. z. sú definované také role alebo oprávnenia, ktoré nemá mať pridelený jeden používateľ alebo správca.
* Riadenie prístupu k objektom. V rámci technického riešenia očakávame dve riešenia:
  + Security Framework pre relačnú časť - riadenie prístupu na úrovni DB objektov
    - Riadenie prístupu k záznamom (riadková bezpečnosť)
    - Riadenie prístupu k informačným typom (stĺpcová bezpečnosť) najmä z dôvodu zavedenia chránenia citlivých dát
  + DataLake nástroj pre možnosť granulárnej konfigurácie oprávnení pre adresáre a súbory uložené v HDFS z grafického rozhrania.
    - Z hľadiska riadenia prístupu k citlivým dátam uložených v DataLake bude vypracovaný v rámci bezpečnostného dokumentu postup ako modelovať „dátový model“ DataLake, aby spĺňal požiadavku ochrany citlivých dát.
* Osobitné oddelenie manažmentu rolí a prideľovania oprávnení užívateľom

Pri komunikácii PowerBI reporting nástroja s Relačnou časťou DWH požadujeme, aby nástroj preberal nastavenia práv viditeľnosti dát z tejto časti DWH. Vytvorenie špeciálnej vrstvy v DWH len za účelom poskytnutia zabezpečenia dát sa nám javí ako nebezpečné, t. j. neúplné a netransparentné z pohľadu riadenia prístupu a viditeľnosti dát.

#### Logovanie používateľských aktivít

Systém musí zabezpečiť záznam všetkých používateľských aktivít ohľadne prístupu k dátam (databáza, DataLake). Zo záznamov by malo byť spätne identifikovateľné kto, kedy a k akým dátam pristupoval. Záznam by mal obsahovať minimálne:

* Dátum a čas operácie
* Identifikátor používateľa
* Názov a adresa pracovnej stanice užívateľa
* Vykonaná operácia a/alebo dátová požiadavka (napr. kópia SQL príkazu)
* Úspech/neúspech vykonanej operácie

Používateľské biznis frameworky musia zachytávať minimálne nasledovné udalosti:

* Prihlásenie/odhlásenie užívateľa
* Používateľský prístup k dátam
* Aktivita

Prístup k správe logov musí byť nasledovný:

* V logoch nie je možné ukladať citlivé informácie.
* Riadený prístup k logom systému (t. j. na základe prístupový práv)
* Minimálne udalosti popísané hore musia byť propagovateľné cez API do bezpečnostného systému NBS, ktorým je SIEM.

#### Návrh aplikačnej vrstvy

Návrh aplikačnej vrstvy DWH musí spĺňať tieto požiadavky:

* Komponenty systému sú umiestnené on-premise, t. j. komponenty nesmú byť umiestnené v cloude.
* Komponenty systému sú konfigurované v súlade s vhodným bezpečnostným štandardom (hardening), napr. CIS benchmark. Odchýlky od zvoleného štandardu musia byť zdokumentované a zdôvodnené.
* V prípade realizácie bezpečnostného testovania poskytne uchádzať súčinnosť a odstráni na vlastné náklady identifikované zraniteľnosti.
* Včasné nasadzovanie bezpečnostných opráv vydaných výrobcami jednotlivých komponentov systému, ktoré dodáva v rámci projektu uchádzač.
* Systém umožní zálohovať konfiguráciu komponentov a údajov:
  + Všetky metadáta a konfigurácie aplikačných frameworkov musia byť uložené v DWH databáze pre jednoduché (synchrónne) zálohovanie.
  + Ak niektoré konfigurácie sú uložené priamo v súborovom systéme serverov (napr. DataLake), toto uloženie musí byť jasne a presne zdokumentované.
* Komunikácia používateľov aj správcov s komponentami systému musí byť realizovaná bezpečným komunikačným kanálom, t. j. komunikácia cez webové rozhranie, komunikácia "tučného" klienta, vzdialený prístup správcov do OS a pod.
* Komunikácia samostatných komponentov systému musí byť realizovaná bezpečným komunikačným kanálom.
* Komunikácia systému s inými systémami (externými alebo internými, s databázou, mailovým serverom, doménovými radičmi a pod.) musí byť realizovaná bezpečným komunikačným kanálom.
* Použitie štandardov pre bezpečný vývoj.

### Iniciálny load

Pre každú tabuľku DWH v transakčnej vrstve bude zabezpečený tzv. „iniciálny load“ - prvotné nahratie kompletného obrazu dát zo zdrojového systému. To môže obsahovať aj historické položky číselníkov a kmeňových dát, ak existujú, prípadne historické transakcie a snímky, ak ich daný systém udržiava v produkčných tabuľkách (veľmi často teda ide napr. o historické transakcie - k nim musia byť dohraté aj relevantné referenčné a kmeňové dáta, aj keď dnes už nemusia byť aktívne - dlhopis, platobná karta, subjekt).

Zvyčajne sa nerekonštruujú historické stavy statických, resp. pomaly sa meniacich dát, ak tieto zmeny neobsahuje aj samotný zdrojový systém vo svojich základných zdrojových štruktúrach.

### Migrácia dát z projektu Pilot DWH

Nie je vyžadovaná migrácie dát z riešenia Pilot DWH do nového riešenia.

### Metodika fyzického vývoja a inštalácie na produkciu

Z dôvodu zabezpečenia bezpečnej prevádzky a ochrany údajov v DWH budú požité na vývoj, testovanie a samotnú prevádzku 3 prostredia:

* DEV – development prostredie, jeho charakteristiky:
  + málo dát,
  + anonymizované, alebo vybratá reprezentatívna vzorka, ktorá ale neporušuje predpisy ochrany údajov
  + vykonanie systém testov
* TEST – testovacie prostredie, jeho charakteristiky:
  + väčšie množstvo dát pre vykonanie určitých performance testov
  + vykonanie integračných testov
  + úvodné UAT testy biznis používateľov
* PROD – produkčné prostredie
  + záverečné UAT testy biznis používateľov
  + bežná prevádzka

Vývoj bude prebiehať v nasledujúcom predpise:

* DEV prostredie je plne R/W. Vývoj bude prebiehať pre ETL/Workflow len nastavením metadát. Nie je dovolené používať PL/SQL procedúry pre ETL úlohy. Všetky aplikačné PL/SQL procedúry musia byť uložené ako metadáta (anonymné bloky) vo workflow frameworku, kde majú svoju historickú verziu a metadáta dôvodu vzniku.
* TEST prostredie je plne R/W ale v koordinácii s DEV. Prostredia musia byť synchronizované pravidelne na metadátovej úrovni. ETL/Workflow, ktorý je vo vývoji bude mať metaznačku.
* PROD prostredie je čisté R/O prostredie pre metadáta. Prenos z TEST do PROD sa vykoná cez inštalačný balíček. Zápis do metadát ETL/Workflow je možný len počas riešenia produkčného incidentu 1. stupňa (hot fix). Pri iných riešeniach incidentov bude použitý štandardný postup DEV->TEST->PROD.

Pri fyzickej inštalácii jednotlivých častí diela požadujeme dodržať:

* Všetky dodané UI budú podporovať prehliadač Chrome, ktorý je štandardom v NBS prostredí. Okrem výnimiek popísaných nižšie.
* Všetky metadáta frameworkov budú uložené v DWH relačnej časti. Okrem presne dohodnutých a NBS vopred odsúhlasených oblastí (napr. pre DataLake, ale aj tu požadujeme dohodu a preferujeme použiť DWH relačnú časť pred iným riešeniami). Názvoslovie technických používateľov dodá NBS.
* Fyzická implementácia bude na HW a SW ako je popísaný vyššie (popis je vo Fyzickej infraštruktúre).
* Pre inštaláciu riešenia nebude potrebný Admin alebo Root alebo iný super používateľ. Okrem presne striktne vopred dohodnutých postupov, ktoré musia byť vopred (čiže na začiatku projektu) oznámené, spísané v dokumentácii a NBS schválené.

### Nástroje na podporu prevádzky

Súčasťou dodávky budú aj aplikácie a nástroje na podporu prevádzky DWH, novo vznikajúcich procesov a zjednodušenie/sprehľadnenie práce odborných používateľov. Najmä požadujeme adresovať tieto oblasti či procesy:

* Podporu správy, riadenia a monitoring aktualizačných úloh (Workflow Management – Data Orchestration).
* Správa číselníkov, mostíkov, referenčných dát, hierarchií jednotlivých dimenzií a manuálnych vstupov.
* Bezpečné nahrávanie veľkého počtu a rôznych súborov, t. j. všetky a práve raz (Ingestion framework).
* Správa dátovej kvality - automatické spúšťanie a vyhodnocovanie testov.
* Správa prístupov (riadenie bezpečnosti) k dátam.
* Správa dokumentácie k DWH a iných metadát (o modeli, o reportoch, mapovaní zdroj-cieľ).
* Riadenie dátovej retencie.

Tieto nástroje budú dodané a nasadené v kontextu implementácie platforiem DataLake a DWH, ale rozsah ich využitia by nemal byť v budúcnosti obmedzený iba na tieto platformy. Tieto nástroje by mali byť schopné aplikovať potrebné koncepty a procesy aj v rámci ďalších systémov NBS.

#### Workflow Management – Data Orchestration

Kvalitné riadenie a správa aktualizácie dátového skladu je jedným z kritických ukazovateľov úspešnosti implementácie dátového skladu. Len aktualizácia, ktorá dáva presný obraz o svojom behu dáva dôveru biznis používateľovi, že dáta uložené v dátovom sklade sú korektné a pripravené spracovanie v reportoch. Preto takýto framework musí umožňovať:

* Nastavenie kalendára a času spustenia jednotlivých úloh
* Riadenie závislostí (dynamické zapínanie a vypínanie)
* Error handling – znovu spustenie, nastavenie chybovej vetvy behu
* Manuálne spustenie úlohy
* Nastavenie cyklického spracovania dát
* Riadenie heterogénneho prostredia, t. j. ako relačnej časti DWH, tak aj DataLake časti
* Logovanie behu úloh
* Analytiku nad logmi behov jednotlivých úloh
* Nastavenie používateľských práv a prístupov
* Riadenie prístupu do aplikácie cez login modul napojeného na NBS LDAP/AD.
* Webové rozhranie použiteľne v prehliadači Chrome, ktorý je štandardom v NBS prostredí. Editor metadát môže byť riešený ako inštalovaná aplikácia vo MS Windows 10 prostredí.
* Všetky metadáta uložené v databáze s prehľadným API

Framework je možné rozdeliť do viacerých logických celkov:

1. Požiadavky na editovanie metadát schedullera/workflow manažéra:
   1. Vytváranie/Vymazávanie úloh
   2. Prioritizácia úloh
   3. Rozdelenie úloh do vertikálnych a horizontálnych skupín, presun existujúcej úlohy medzi skupinami, možnosť riadenia skupín samostatne
   4. Možnosť definovania parametru pre maximálny paralelný beh jednotlivých úloh v rámci skupiny
   5. Rozdelenie úloh na automaticky spúšťané podľa kalendára, manuálne spúšťané, alebo automaticky v definovaných cykloch
   6. Definovanie závislostí medzi úlohami v heterogénnom prostredí (PL/SQL, ODI, JAVA, WebService, ShellScript a iné)
   7. Možnosť definovania závislosti aj medzi úlohami v rôznych skupinách
   8. Možnosť zapínať a vypínať jednotlivé skupiny (zaradenie a vyradenie z denného behu).
   9. Možnosť definovania anonymného bloku SQL namiesto zavolania externej úlohy
   10. Definovanie kalendára
   11. Vyhľadávanie v metadátach (Fulltext, podľa definovaných polí) a vložených anonymných blokov SQL
   12. Verzionovanie zmien, vyhľadávanie v zmenách, porovnanie verzií a detailné zobrazenie rozdielov
   13. Logovanie používateľských aktivít
   14. Nastavenie prístupových práv používateľov pre právo tvorby metadát podľa vertikálnych a horizontálnych skupín, t. j. môže existovať používateľ, ktorý má len právo čítať všetky metadáta
   15. Podpora spolupráce medzi používateľmi:
       1. Check-in/check-out metóda (nutnosť riešiť problematiku hraničného prípadu nečakaného pádu aplikácie a automatického check-out po stanovenej časovej hranici – napr. 4 hodiny nečinnosti).
       2. iný používatelia uvidia v GUI, že daný uzol ETL je v režime check-in (uzamknutý)
       3. vkladanie poznámok
   16. Vkladanie metadát pomocou predpripravených šablón. T. j. pri vkladaní uzla úlohy do stromu závislostí úloh sa napríklad:
       1. automaticky vytvoria podúlohy na riadenie chybových stavov pre reštart a samoopravu preddefinovaných chybových stavov
       2. pre rôzne typy úloh (ODI, SQL, Ozee, Nifi, ...) sú povinné iné nastavenia, ktoré definuje šablóna
       3. možnosť doplnenia šablóny (pridať ďalšiu funkcionalitu vkladania uzla do stromu závislostí) prostredníctvom plugin, t. j. bez nutnosti novej verzie aplikácie (GUI)
   17. Základný monitoring behu úlohy:
       1. Prehľad logu behu, vyhľadávanie v správach, filtrovanie len chybových správ
       2. Stav predchodcov – analýza zisťovania, prečo sa úloha ešte nespustila
       3. Reporting konzumácie pamäťových častí servera – TEMP, PGA
   18. Ak je úloha volaná do ODI agenta, tak editor by mal mať možnosť zobraziť metadáta ODI (aj cez verzie) a tiež generovaný SQL skript systémom ODI na kontrolu kvality skriptov
   19. Základné vlastnosti uzla úlohy:
       1. Id, meno
       2. Typ úlohy (SQL, PL/SQL, UNIX shell script, JAVA volanie, ODI volanie, WebService volanie)
       3. Čas spustenia – ak nie je zadaný, spustí hneď ako sú splnené závislosti
       4. Čas pozdržania pri chybe pre vyskúšanie samoobnovy. Ak je definovaná chybová vetva spustenia, tá sa spustí okamžite. Tento parameter sa ignoruje
       5. Parametre uzla – context spustenia (vstupy pre ODI, bankový deň, prostredie, používateľ)
       6. Závislosti na predchodcov a nasledovníkov
       7. 32kB freetext pre skript
   20. GUI pre editovanie metadát by malo jednotlivé uzly ETL procesu vedieť zobraziť ako stromovú štruktúru (podobne ako súborový prehliadač v MS Windows):
       1. podľa vertikálnych a horizontálnych skupín
       2. závislostí a vedieť sa v závislostiach pohybovať cez prelinky (predchodcovia, nasledovníci)
       3. možnosť zobrazenia viacerých editačných okien pre viacero uzlov naraz
       4. možnosť hromadného zavretia nepotrebných okien
2. Požiadavky pre RunTime časť:
   1. Automatická príprava default kalendára (ak neexistuje konkrétny pre daný deň) a plánu spustenia úloh, t. j. RunTime vytvorí dennú kópiu metadát, ktoré bude postupne spúšťať
   2. RunTime časť by mala používať DBMS\_SCHEDULER balíček DWH relačnej časti a to z dôvodu riešenia hraničných stavov počas pádu celej databázy a transparentného nábehu.
   3. Spúšťanie úloh podľa kalendára, závislostí a stavov (zaradenie a vyradenie) ich skupín. Úloha sa môže pustiť len práve raz a včas.
   4. Automatická reštartovateľnosť úlohy pri definovanej chybe
   5. Automatická prioritizácia úloh podľa výsledku behu
   6. Úlohy, ktoré nemajú definovaný presný čas spustenia, sa spustia okamžite po dobehnutí všetkých jej predchodcov.
   7. Cyklická úloha ak náhodou bežala dlhšie ako bol definovaný čas dĺžky cyklického samospustenia, sa spustí okamžite ako skončí. Za predpokladu, že nemá iné závislosti.
   8. Logovanie behu
   9. Manažment Runtimu nesmie zaťažovať server nadmiernou aktivitou. Toto bude overené cez AWR reporting.
   10. Manažment Runtimu nesmie vykazovať v relačnej časti DWH žiadne LATCH, alebo DEAD LOCK. Zamykanie metadát musí byť maximálne na úrovni 1% behu všetkých úloh. Bude merané cez AWR reporting.
3. Požiadavky pre sledovanie, kontrolu a riadenie behu (monitoring):
   1. GUI podporujúci admin konzolu pre bežný dohľad - Celkový pohľad na end-to-end spracovanie z jedného miesta
   2. Sledovanie stavu behu ETL:
      1. na najvyššej úrovni v near-online režime (+/- 5 minút)
         1. stav kontrolných bodov (počet úloh do dosiahnutie cieľa; percento spracovania; linka na úlohu, ktorá blokuje beh);
         2. celkový pohľad na stav skupiny úloh
         3. počet aktuálne chybných úloh
         4. zoznam posledných 10 úloh, ktoré boli v chybe
         5. kontrola behu podľa tradičného (vážená snímka minulosti) behu
         6. Grafy a reporty s podporou drill-down na detailné zobrazenia.
      2. Detailný pohľad na stav úlohy v reálnom režime s funkcionalitou:
         1. Pozastaviť úlohu
         2. Znovu spustiť
         3. Úplne zastaviť
         4. Zrušiť beh – t. j. strom nasledujúcich úloh sa rozbehne ako keby úloha dobehla korektne
         5. Dátum čas spustenia, ukončenia úlohy
      3. Detail ETL úloh:
         1. Sledovanie logu a kódu, ktorý zbehne/zbiehal (SQL, ODI, OOZEE, ...)
         2. Nastavenia úlohy – tradičný čas behu, v akej skupine sa úloha nachádza a podobne.
         3. Prechádzať na nasledujúcu/predchádzajúcu úlohu
      4. Zobrazenie kritickej cesty pre zvolenú úlohu a zvolený deň spracovania so zobrazením časov behu aktualizácie, dĺžky jednotlivých úloh a meškaní (predbiehaní) oproti tradičnému behu.
      5. GUI pre manuálny štart neautomatizovaných úloh s možnosťou zadania parametrov
      6. Analýza štatistík o spúšťaní jednotlivých úloh v histórii až 180 dní v dvoch grafoch:
         1. Časy spustenia a dokončenia úlohy s porovnaním na tradičný beh. Ak úloha sa reštartovala z dôvodu chybového stavu je zobrazený minimálny čas spustenia a maximálny čas dobehnutia.
         2. Dĺžka behu úlohy s porovnaním na tradičný beh. Ak úloha sa reštartovala z dôvodu chybového stavu je zobrazený len posledný úspešný beh.
      7. Analýza dohodnutej SLA o dobehnutí ETL úloh
   3. GUI podporujúci zjednodušenú konzolu pre kontrolný dohľad, napr. vybraným biznisovým používateľom
   4. Manuálna reštartovateľnosť úloh v prípade poruchy – schopnosť nadviazať na už vykonané úlohy
   5. GUI pre spustenie výpočtu tradičného (vážená snímka minulosti) behu
   6. Notifikácie (e-mail) v prípade dobehnutia alebo chybových stavov spracovania
   7. Definovanie kontrolných bodov a reporting na sledovanie dosiahnuteľnosti týchto kontrolných bodov:
      1. Počet úloh, ktoré musia zbehnúť v tomto behu na dosiahnutie kontrolného bodu
      2. Predpokladaný čas na dosiahnutie kontrolného bodu
      3. Tradičný čas na dosiahnutie kontrolného bodu
      4. Indikátor (smerník) na úlohu, ktorá práve blokuje dosiahnutie kontrolného bodu
      5. Graf času dosiahnutia kontrolného bodu v histórii až 180 dní
4. procedúry pre logovanie a samotné logovacie tabuľky:
   1. Popísané API na jednotné logovanie
   2. Integrované API (VIEW) pre čítanie ODI logov
   3. Existencia API pre zasielanie informácií do centrálneho monitoringu IT prostredia (položky ako v bode 3.g. bez grafu)

Riadenie spracovania dát v prostredí NBS bude riešené jedným spoločným nástrojom, ktorý bude dostatočne technologicky zdatný dizajnovať, spúšťať a administrovať tok spracovania cez rôzne technické platformy. Táto požiadavka zabezpečí, že jednotlivé kroky dátového spracovania bude možné vykonávať na rozdielnych platformách, čo bude mať za dôsledok efektívnejšiu správu zdrojov nad platformami.

#### Správu číselníkov, mostíkov, referenčných dát a hierarchií

Pre klasifikáciu údajov v DWH/BI je možné vychádzať z číselníkov a klasifikácií, ktoré sú poskytované zdrojovými systémami ako súčasť jeho natívneho dátového modelu.

Pre reporting však vznikajú aj požiadavky na triedenie dát podľa odvodených charakteristík, prípadne unifikovaných klasifikácií, ktorých kódovníky nie sú v zdrojových systémoch k dispozícii, resp. zdrojové systémy neposkytujú potrebné údaje (napr. užívateľské hierarchie, súčtové prvky, príznaky, deskriptívne názvy a pod.).

Na správu týchto rôznych číselníkov a sú základom pre kvalitný reporting, je nevyhnutné mať viac ako menej rôznych formulárov, ktoré podporia:

* Správa referenčných dát a hierarchií údajov.
  + číselníky / kódovníky
  + klasifikácie a iné malé dimenzie
  + definície hierarchie
  + lookup tabuľky pre konsolidáciu číselníkov a ich obohatenie o dodatočné informácie (popisy, príznaky)
* Správa konverzných mostíkov (napr. aj pre potreby ETL – transformácia technických transakcií na biznisové definície pre niektoré hviezdicové schémy):
  + parametre (zoznamy, podmnožiny hodnôt) používané v rozhodovacej logike spracovania dát
  + rozhodovacie tabuľky pre odvodenie dát a aplikáciu obchodných pravidiel pri spracovaní dát v DWH
* Udržovanie nastavení DWH metadát – napr. spôsob distribúcie niektorých automatických reportov cez e-mail a podobne.
* Manuálne vstupy statických dát malého rozsahu, ktoré nie sú primárne uložené a obhospodarované v inom systéme

Tento systém by mal poskytovať minimálne tieto funkcie a nástroje pre správu referenčných dát:

* umožniť používateľom (odborným aj technickým) modifikovať referenčné dáta cez jednoduchý GUI formou selfservice:
  + pridanie nového záznamu (ak je používateľ oprávnený)
  + odstránenie záznamu (ak je používateľ oprávnený)
  + zneplatnenie záznamu (ak je používateľ oprávnený)
  + editácia záznamu (ak je používateľ oprávnený)
  + podpora triedenia, filtrovania, pop-up nápovedy
  + podpora master-slave dát, t. j. zobrazenie dát v drop-down/pickliste z master objektu
  + možnosť nastaviť niektoré (definovateľné) stĺpce ako R/O (len na čítanie)
  + možnosť exportu dát do dohodnutého formátu - CSV, XSL, JSON.
  + v závislosti od návrhu číselníkov mať možnosť nastaviť platnosť dát
* všetky dáta a metadáta budú uložené v technickom používateľovi v Relačnej časti DWH. ETL DWH bude tieto dáta aktualizovať pravidelna počas ETL procesov v dennom režime a micro-batch režime v priebehu dňa - napr. pri zmene v dátach, rovnako ako pre Relačnú časť tak aj pre DataLake.
* riadiť prístup k jednotlivým referenčným dátam s autentifikáciou a autorizáciou užívateľov k jednotlivým referenčným dátam
* auditovať všetky akcie a verzovať zmeny v záznamoch s časovými značkami
* umožňovať definíciu nových referenčných dát s validáciami vstupných dát (minimálne na doménovú a referenčnú integritu)
* sprístupniť dáta pre ETL proces cez definované rozhranie
* poskytne mechanismus pre notifikáciu resp. spustenie micro-batch spracovania v prípade zmeny v dátach
* poskytovať online služby pre ostatné aplikačné systémy, ktoré chcú svoje referenčné dáta delegovať na RDM, t. j. systém bude vedieť vystaviť REST API pre definovaný číselník s funkciou READ vo formáte JSON.
* vytváranie číselníkov (pridávanie nových referenčných štruktúr) bude s podporou a prístupom low-code/no-code (bez softvérového vývoja), všetky parametre číselníkov musia byť uložené v metadátach.
* automatická notifikácia cez e-mail definovanému odbornému pracovníkovi, ak nastane nejaká udalosť v číselníku (napr. vloženie nového záznamu s default hodnotou)
* Webové rozhranie použiteľne v prehliadači Chrome, ktorý je štandardom v NBS prostredí.
* Riadenie prístupu do aplikácie cez login modul napojeného na NBS LDAP/AD.

Riešenie RDM bude koncipované ako „s úložiskom priamo v DWH“ s možnosťou batchovej či okamžitej synchronizácie do DWH/DataLake. Rovnako musí prostredie RDM poskytnúť REST API pre prístup k dátam iným aplikačným systémom.

#### Manuálne vstupy – selfservice pre používateľov

Táto požiadavka rozširuje požiadavku pre GUI RDM pre “Manuálne vstupy statických dát malého rozsahu” a funkcionalitu vytvorenia importného medzikroku na kontrolu dátovej kvality importovaných dát na základe zadefinovaných pravidiel.

Proces manuálneho importu používateľom bude následne vyzerať (navyše okrem štandardnej RDM funkcionality):

* Importované dáta zo známeho a predom dohodnutého formátu sú transformované (rozšírené o mapované dohodnuté dimenzie) do medzipamäte
* GUI zobrazí stav importu s tým, že chybné dáta sa zobrazia červenou farbou. Chybné dáta budú rozoznateľné cez nastavené pravidlá:
  + neexistuje JOIN na číselník
  + hodnota je mimo definovaný povolený rozsah
* GUI nemusí mať podporu na opravu týchto dát
* Po kontrole a správnom nastavení všetkých riadkov sa importný balík označí metaznačkou v STAGE vrstve DWH
* Importovacie GUI musí obsahovať sumárny report (súbor, kedy, stav, počet chýb, atď.) pre sledovanie importnej činnosti priamo používateľom.
* Webové rozhranie použiteľne v prehliadači Chrome, ktorý je štandardom v NBS prostredí.
* Riadenie prístupu do aplikácie cez login modul napojeného na NBS LDAP/AD.

#### Ingestion framework - nahrávania veľkého počtu rôznych súborov

Súčasťou DWH/DL platformy by mal byť Ingestion framework, ktorý by primárne zabezpečoval importovanie externých zdrojov do Stage Zóny platformy. Externé zdroje sú dostupné prostredníctvom API rozhraní, parsovania public  webových stránok alebo sťahovania dát zo vzdialených SFTP serverov mimo alebo v NBS infraštruktúre. Spracovanie Ingestion frameworku musí byť riadené prostredníctvom Workflow managera.

Požiadavky na Ingestion Framework :

* Schopnosť pridávania nových externých zdrojov do frameworku:
  + SOAP/REST API, parsovanie Web stránok, SFTP, Kafka, JDBC, súborový systém, MS Sharepoint, HDFS
* Možnosť nastavenia odkladania spracovaných dát na „archívne“ miesto
* Možnosť nastavenia vymazania zdroj dát
* Schopnosť metadátovo definovať parser čítača pre známe formáty ako je CSV, AVRO, JSON , XML, DBF a podobne.
* Schopnosť priamo čítať ZIP na vstupe
* Schopnosť metadátovo definovať mapovania zdroj-cieľ (low-code/no-code programovanie)
* Schopnosť priamo počas spracovania dát tieto dáta obohacovať o iné dáta definovanou reláciou na mapovacie dimenzie.
  + toto obohacovanie a transformácia dát má prebiehať nezávisle od zdroja a formátu dát aj s možnosťou zmeny zdroja bez nutnosti úpravy transformácií
  + schopnosť spracovať transformácie in-memory
  + relácie môžu byť definované aj ako best-match (hľadaj najbližší najlepší), t. j. nie je nutné mať definované tvrdé 1:1 prepojenie
  + mapovania by mali byť definovateľné deklaratívne spôsobom low-code
* Spúšťaný z centralizovaného Workflow managera.
* Schopnosť importovať len inkrementy dát pričom Ingestion framework by mal byť zodpovedný za udržiavanie kompletného prehľadu o už spracovaných dávkach. T. j. musí mať mechanizmy ako spracovať „práve raz a všetky“ vstupné zdroje. napr. pri zdroji so súborového systému je možné definovať spracovanie cez masku súboru. Framework vyberie a spracuje len tie súbory, ktoré vyhovujú nastaveniu. Pri iných typoch zdrojov sú to iné nastavenia, ktoré podporia požiadavku udržania kompletnosti spracovania.
* Spracovanie musí byť kontrolované, auditované a reportovateľné – interface, beh, čas, počet spracovaných vstupných riadkov, počet výstupných riadkov, počet chybových riadkov, počet zahodených riadkov z dôvodu duplicita a atd.
* Udržiavať si metadáta o externých zdrojoch (API kľúče, mená, heslá, tokeny, logy, aktuálny stav spracovania + v prípade inkrementálneho spracovania všetky potrebné informácie potrebné pre stiahnutie ďalšieho inkrementu, počty exportovaných záznamov pre každé spustenie)
* Schopnosť spracované dáta uložiť cez JDBC do STAGE DWH, alebo cez HDFS konektor priamo do zóny DL.
* Schopnosť robiť tzv. multi-insert, t. z. keď vstupný tok dát obsahuje viac rôznych dátových formátov, tak dodaný framework tieto formáty bude vedieť rozoznať a uložiť do rôznych cieľových objektov v rôznych technológiách, t. j. či DataLake, alebo Relačná časť DWH.
* Schopnosť v cykle vyčítať jednotlivé záznamy z WebService, kedy nie je možné vyčítať všetky záznamy, ale je nutné volať WebAPI cez parameter. Zoznam všetkých hodnôt parametrov bude si framework vedieť prečítať z databázovej tabuľky, ktorá bude metadátovo nastaviteľná pre každé vlákno zvlášť.
* Schopnosť odložiť chybné záznamy počas spracovania do iného dohodnutého objektu na ďalšiu analýzu. Túto schopnosť bude možné nastaviť metadátovo.
* Schopnosť identifikovať duplicity v dohodnutom spracovávanom okne a v rámci jedného streamu. Kontrolu duplicít bude možné nastaviť metadátovo.
* Podpora tvorby dokumentácie z kódu spracovania vo formáte HTML, PDF, Excel, API na vkladanie do databázy
* Rýchlosť spracovania (bez transformácie dát) na úrovni 1 milióna riadkov z XML súboru za 15 minút pri jednovláknovom spracovaní do Oracle DB na konfigurácii CPU podľa kapitoly 3.2.4.
* Detailné logovanie behu a spracovania:
  + Online monitoring a management procesov spracovanie cez JMX
  + Rozdelenie logov podľa typov a príslušnosti procesu. Detailné logy môžu byť vo file systéme servera
  + Jeden centrálny log (najlepšie v databáze) s integráciou na workflow manažér, ktorý hovorí o aktuálnom dni spracovania, ktorý súbor (alebo skupina súborov) je už spracovaný, či čaká na spracovanie a s akým výsledkom (počet vstupných, výstupných, odložených, chybných riadkov). Dátum čas začatia a skončenia procesu.

**Prvý príklad** použitia vyššie uvedených požiadaviek je čítanie dát z ESMA zdroja (<https://www.esma.europa.eu/>). Tento zdroj kategorizujú tieto atribúty:

* Veľký objem dát – rádovo stovky GB až desiatky TB
* Dáta sú vo formáte XML
* Súbory sú ZIPnuté
* Jeden súbor XML obsahuje viac podstromov

t. z., že je vyžadované pri tomto zdroji čítať priamo ZIP (nebude miesto na unzip), v pamäti parsovať XML, vyberať definované podstromy a robiť MULTI-INSERT do rôznych objektov v DWH bez rozdielu cieľa technológie, t. j. či DataLake, alebo Relačná časť.

Ako **druhý príklad** je ITMS2014 (<https://opendata.itms2014.sk/swagger/?url=/v2/swagger.json>) kde je nutné cez REST API vyčítavať informácie v cykle, kde nie je známe o aké a koľko dát pôjde. Tieto dáta budú ukladané podľa definície do rôznych (napr. /v2/ciselniky) alebo rovnakých tabuliek (napr. /v2/dodavatelia/{dodavatelId}). Niektoré vstupné vlákna obsahujú rozvetvené podstromy.

**Tretí príklad** je import dát z Katastrálneho úradu Slovenskej republiky. Tieto dáta po podpísaní zmluvy sa dajú stiahnuť v stanovený deň a čas zo zabezpečeného úložiska ako ZIP súbor, ktorý obsahuje niekoľko DBF (formát dBase) súborov. T. j. systém musí vedieť všetky a práve raz tieto súbory načítať do určených objektov. Pokiaľ sa nepodarí nahrať všetky súbory do cieľového stavu, musí framework vedieť generovať chybový kód na zastavenie behu vo Workflow managerovi.

#### Data Quality Management

Systém pre DQM (Prostredie pre správu, spúšťanie a vyhodnocovanie dátovej kvality) bude riešiť nasledujúcu funkcionalitu:

* Konfigurovať pravidlá pre meranie dátovej kvality (DQ indikátory)
* Organizovať pravidlá DQ do skupín a plánov ich spúšťania
* Spúšťať a plánovať spúšťanie vybraných skupín indikátorov
* Monitorovať výsledky meraní cez reporty a grafy
* Možnosť zasielať výsledky kontrol na e-mail - generovať podklady pre Data Quality Incident
* odkladanie chybných záznamov do špeciálnych tabuliek pre následnú kontrolu a riešenie incidentov
* zastavenie ETL procesu pri prekročení hraničného počtu chýb
* zakomponovanie spustenia DQ kontroly do procesu ETL – ako samostatný krok spracovania

Tento systém bude pripojený na DWH, nad ktorým bude prebiehať meranie indikátorov kvality cez rôzne dátové vrstvy. Riešenie by malo obsahovať komponenty pre podporu riadenia dátovej kvality na úrovni:

* Technických kontrol – štandardné/generické kontroly vyplývajúce z dátového modelu
* Biznisový kontrol – užívateľsky definované parametre kvality na úrovni atribútov (data quality indikátory)

**Technické kontroly** by mali zahŕňať:

* Dátové typy
* Obmedzenia (Constraints) – povinné hodnoty, unikátnosť, referenčná integrita
* Doménové kontroly – validácia na prípustné hodnoty
* Kvalita zbehnutia SCD

Spúšťanie technických kontrol by malo byť integrálnou súčasťou ETL procesu.

**Biznisové kontroly** môžu byť voľne definované zo strany užívateľov (resp. Data Stewardov) vo forme tzv. *indikátorov dátovej kvality (DQI)*, napríklad:

* hraničné (podozrivé) hodnoty atribútov
* neočakávané kombinácie hodnôt súvisiacich atribútov
* neočakávané vzory v dátach a pod.

Tieto indikátory môžu mať komplexnú definíciu, a teda môžu byť definované vo forme SQL dotazu.

#### Testovací framework

Testovací framework (softwarový komponent, ktorý je možné konfigurovať matadátami) predstavuje funkcie pre definíciu a automatizované spúšťanie a vyhodnocovanie testovacích scenárov cez rôzne prostredia. Typicky slúži na automatizované (alebo opakované) vykonávanie rôznych testovacích prípadov v rámci SDLC procesu, napr.:

* kontroly stavu prostredia
* funkčné testy (správne plnenie cieľových štruktúr)
* regresné testy

Framework by mal poskytovať minimálne tieto funkcie:

* repozitár testovacích skriptov (SQL)
* mechanizmus pre automatizované spúšťanie zvolených testovacích skriptov
* logovanie výsledkov testov a ich sprístupnenie pre vyhodnotenie (SQL, resp. v reportoch cez BI nástroj)
* Webové rozhranie použiteľne v prehliadači Chrome, ktorý je štandardom v NBS prostredí.
  + Jednoduchý reporting GUI pre sledovanie zbehnutia a výsledku zbehnutia
  + Riadenie prístupu do aplikácie cez login modul napojeného na NBS LDAP/AD.
* Všetky metadáta uložené v databáze s prehľadným API

Je možné tento framework spojiť s DQM podporou, ale potom sa vyžaduje aby nastavenie metadát DQM a jeho exekúcia, prípadne reporting boli jasne, zrozumiteľne a ľahko oddeliteľné od zadaných Testovacích scenárov.

#### Security GUI (relačná časť DWH)

Framework pre správu prístupov je kritický komponent, ktorý umožňuje riadiť viditeľnosť dát a celkové práva používateľov DWH systému.

Framework má mať dve časti:

* GUI
  + Webové rozhranie použiteľne v prehliadači Chrome, ktorý je štandardom v NBS prostredí.
  + Riadenie prístupu do aplikácie cez login modul napojeného na NBS LDAP/AD.
* Server DB časť

Framework GUI musí umožňovať:

* Definícia skupiny používateľov a priradenie reálneho používateľa z DB
* Definícia objektovej skupiny (prístupovej role) a priradenie reálnej role z DB
* Definícia riadkovej bezpečnosti objektu
* Definícia stĺpcovej bezpečnosti objektu
* Priradenie riadkovej a stĺpcovej bezpečnosti objektu k objektovej skupine
* Priradenie role objektová skupina na skupinu používateľov
* Možnosť simulovaného behu nastavenia rolí a bezpečnostných pravidiel

Framework nesmie umožňovať z bezpečnostných dôvodov:

* Definícia DB používateľa,
* Definícia DB role,

tieto informácie musí vedieť v DWH identifikovať a ponúknuť v zozname (pick list) pri nastavení práv.

Framework Server DB časť musí vedieť na požiadanie:

* Vygenerovať požadované pridelenie rolí, nastaviť VPD politiky a vytvoriť požadovanú stĺpcovú bezpečnosť pre jednotlivých používateľov,
* Skontrolovať nastavenia a prípadné nezrovnalosti v DB nastavení so schváleným nastavením v metadátach odobrať. T. j. ide o odobratie prístupových práv ak by niekto urobil zmeny ručne mimo systém.

#### Metadata Management / Wikipédia o DWH

V procesoch DWH vzniká množstvo metadát, ktoré popisujú vecnú aj technickú stránku spracúvaných a publikovaných dát, ktoré je potrebné efektívne sprístupniť širšej základni používateľov a analytikov, a uľahčiť tak ich navigáciu v dostupných dátach.

Budúci systém integrácie dát by mal byť doplnený o riešenie pre správu metadát, ktoré bude poskytovať NBS nástroje pre správu biznis katalógov, data lineage, data katalog a navigáciu v nich.

Medzi tieto metadáta patria:

* Biznis slovník pojmov
* Dátové katalógy – popisujúce dostupné dáta, zdroje, DataLake, DWH
* Ďalšie katalógy, napr. KPI katalóg, DQI katalóg, katalóg reportov
* Relácie medzi pojmami a dátovými modelmi resp. dátovými katalógmi
* Údaje o dostupnosti a kvalite dát

Nástroj pre manažment metadát bude umožňovať tieto funkcie:

* Správa a audit Biznisového slovníka:
  + Nástroje pre správu biznisového slovníka, t. j. centrálnu definíciu dôležitých konceptov a pojmov, ich významu a klasifikácie.
* Správa a audit Dátového katalógu s popisom dát v dátových zdrojoch, DWH a DataLake
* Automatické (na spustenie) naplnenie metadát z dostupných metadát DB, ODI, Workflow Managera a podobne. Vyžaduje sa minimalizácia (na úrovni jednotiek hodín do mesiaca) manuálnej správy informácií.
* Funkcionalita data lineage a impact analýzy nad objektami zaznamenanými v katalógoch (biznisové pojmy, KPI, DQI) a importovanými metadátami o objektoch.
* Jednoduchá navigácia medzi objektami dátového modelu na základe zachytených väzieb (lineage informácie)
* Funkcia fulltextového vyhľadávania v rámci všetkých atribútov katalógov a importovaných metadát
* Portál s webovým frontend-om pre navigáciu biznisových používateľov v metadátach:
  + Stránky sa musia generovať dynamicky (t. j. môže ich byť ľubovoľne veľa a v čase sa rozširovať) na základe nastavených metadát a definovaného template. Požadované template:
    - Jednoduchý report
      * Názov
      * Popis
      * Výsledok SQL SELECT z Relačnej časti DWH databázy (vrátené hodnoty riadkov môžu obsahovať dohodnuté TAGY na prelinkovanie na ďalšie objekty/stránky)
    - Informačná stránka, alebo skupinová stránka
      * Názov
      * Popis
      * Obrázok (napr. diagram dátového modelu pre vizualizáciu entít a vzťahov medzi nimi)
      * Linky na ďalšie podstránky
      * Linky na dokumenty uložené v Sharepointe NBS (nebude sa zobrazovať ak k danej stránky nebudú žiadne priradené)
    - Objekt (automatický import z DB)
      * Názov
      * Popis
      * Stĺpec
      * Popis (okrem popisu objektu, bude toto pole obsahovať dohodnuté TAGY na prelinkovanie na ďalšie objekty/stránky)
    - Metadata z  ETL mapovania (obsah podľa metadát ODI), t. j. minimálne:
      * Názov
      * Popis
      * Zdroje (s prelinkovaním na ne)
      * Cieľe (s prelinkovaním na ne)
      * Použitý Knowledge modul s jeho nastaveniami v danom mapovaej úlohe
      * Parametre SQL, ako sú typy a nastavenia JOIN, WHERE a atď.
      * Detainé mapovanie stĺpcov
    - Metadata z Workflow managera (obsah podľa metadát), t. j. minimálne:
      * Názov uzla ETL úlohy
      * Popis
      * Predchodca a nasledovník (s prelinkovaním na ne)
      * Typ uzla (SQL, REST API, volanie DL úlohy, volanie Ingestion úlohy, atď.)
      * Nastavenie
      * Plné znenie volania (text SQL, alebo volanie REST API atď) uzla ETL úlohy
    - Metadata z RDM
      * Zoznam číselníkov a mostíkov s popisom
      * Zoznam hlavných gestorov číselníkov
      * Zoznam používateľov číselníkov s pridelenými právami
  + Fulltextové prehľadávanie a navigácia na vyhľadané topicy, či TAGY
  + Portál bude nastavený v design manuáli NBS
* Nastavenie používateľských práv a prístupov
* Podpora LDAP/AD.
* Webové rozhranie použiteľne v prehliadači Chrome, ktorý je štandardom v NBS prostredí.

#### Riadenie dátovej retencie – management partícií

Dátový sklad umožňuje uložiť dáta v histórii podľa požiadaviek používateľov, zvyčajne v nekonečnej. Avšak musí existovať retenčná politika napr. z dôvodu ochrany investícií do IT infraštruktúry, alebo z prípadných legislatívnych dôvodov.

Je požadované aby v rámci projektu NBS DWH bol dodaný framework kde:

* Dátová retencia bude definovaná metadátami pre každý objekt (tabuľka) zvlášť.
* Spustenie frameworku bude riadiť workflow manažér
* Framework musí vedieť zalogovať svoj stav do riadiacej konzoly
* Framework musí podporovať tvorbu názvoslovia databázových tablespace a partícií
* Detailné požiadavky na automatizáciu riadenia časových partícií:
  + pridávanie,
  + rozdelenie poslednej časovej partície
  + odpájanie (retencia) „starých “ dátových particií
  + presúvanie a defragmentovanie s nastaviteľnou možnosťou a výberom kompresie a tablespace.
* Riadenie zámeny dát cez partície
  + Výmena (zámena) partícií v objekte – zámena dát
  + podpora automatického znovu vytvorenia databázových indexov po procese zámeny

#### Dátové modelovanie

Nástroj pre dátové modelovanie by mal podporovať vývojárov pri analýze, dizajne, implementácii, testovaní a nasadzovaní softwarových riešení. Okrem toho plní dokumentačnú funkciu, je zdrojom metadát dokumentujúcich dátové štruktúry a závislosti pre business data lineage (závislosti dát DWH na ultimatívnych zdrojových štruktúrach). V jednotlivých fázach SDLC sa využíva nasledujúcim spôsobom:

* Analýza – reverse-engineering a dokumentácia dátových zdrojov, dopadová analýza zmien v zdrojových štruktúrach na štruktúry v DL/DWH, komunikačný prostriedok so zadávateľom
* Dizajn – návrh a dokumentácia nových dátových štruktúr, definícia logiky mapovania na dátové zdroje, generovanie časti dokumentácie
* Implementácia – generovanie kódu pre tvorbu a modifikáciu databázových štruktúr, exportovanie metadát pre generovanie niektorých transformácií a technických kontrol pri spracovaní dát v DL/DWH
* Testovanie – zdroj informácií pre testerov k dátovému modelu a očakávanej transformačnej logike
* Nasadzovanie - generovanie kódu a metadát pre nasadenie, sledovanie stavu vývoja a nasadzovania štruktúr cez prostredia

Voľba konkrétneho nástroja pre dátové modelovanie tak môže výrazne ovplyvniť spôsob a kvalitu implementácie. Prosíme odporučiť nástroj na podporu hore popísaného procesu, ktorý bude zakúpený separátne. Preferovaným nástrojom v NBS je Sparx Enterprise Architect (existujúce licencie) a/alebo Oracle Data Modeler (voľná licencia).

### Nástroje pre Data Governance a riešenia Master Data Managamentu

Projekt celobankového NBS DWH nebude pokrývať komplexitu MDM/DG procesu. Túto tému bude analyzovať paralelná štúdia pre MDM a zavedenia Data Governance. Táto veta ale neznamená, že vyššie popísané technické požiadavky kapitoly 3.2 nie je nutné naplniť, práve naopak.

Až táto štúdia a ďalšie diskusie udajú implementačný zámer v celej šírke problematiky DG a MDM.

## Biznisové požiadavky

V rámci dodávky projektu požadujeme implementáciu end to end (E2E)reportingového nástroja, ktorý bude podporovať analytické procesy.

Na základe pred-analýzy boli identifikované kľúčové oblasti pre:

1. Dátovú integráciu zdrojových systémov
2. Reporting a analýzy
3. Výstupné dátové oblasti

### Zdrojové (vstupné) systémy

Tabuľka nižšie obsahuje zoznam zdrojových systémov, ktoré je potrebné integrovať do nového NBS DWH. Tabuľka obsahuje systémy označené ako:

1. Interné – systém, ktorý nesie dáta požadované pre integráciu do DWH a je v internej sieti (infraštruktúre) NBS. Dáta sú uložené buď v databázach, alebo v extraktoch vo forme súborov.
2. Externé – systém, ktorý obsahuje dáta požadované pre integráciu a je mimo infraštruktúry NBS. K dátam sa môže pristupovať rôzne cez API (Web Service) až po súborové rozhrania (HTTPS/SFTP a iné.)
3. Manuálne dátové vstupy - ide o dáta získavané z neštruktúrovaných dátových zdrojov ako je napr. e-mail, web, ročenka, výročné správy, ktoré je možné ukladať v štruktúrovanej podobe do strojovo spracovateľných formátov, avšak ide o výsledok manuálnej práce spracovateľov týchto dát. V prehľade nižšie je rámcový výpis takýchto zdrojov rozdelených podľa účelu použitia v aktuálne dostupnom formáte. Predpokladom pre automatizáciu bude aj vytvorenie rozhrania, cez ktoré sa budú rôznorodé manuálne vstupy do procesu integrovať. Momentálne takéto rozhranie neexistuje a každé spracovanie používa prispôsobené logiky pre prácu s manuálnymi vstupmi a potrebné dáta extrahuje cez vlastný dekóder. Pozri kapitolu 3.2.13

Význam ďalších atribútov tabuľky je nasledovný:

* Fáza- predpokladaná fáza dodávky dátovej integrácie do DWH (pozri kapitolu 3.4.5)
* Dôvod integrácie:
  + BI - dáta z označeného systému pôjdu primárne použiť v BI reportingu
  + OUT - dáta z takto označeného systému sú potrebné pre výstupy z DWH na odosielanie mimo NBS v presne definovanej štruktúre
* Databázový systém
  + DB - dáta uložené v NBS databáze (Oracle, MS SQL, MySQL)
  + Súbor - lokálne uložené súbor (CSV, MS Excel) na FTP úložisku alebo SharePoint v prostredí NBS
  + WebAPI - Exsitencia WebAPI na externom serveri mimo NBS
  + Lotus Notes - dokumentačná databáza LN
* Objem dát (GB) - predpokladaný objem dát v Gigabajtoch (1.000.000.000 bajtov)
* Počet objektov - tabuliek/súborov/API endpointov

| Názov systému | Popis systému | Interný /externý | Fáza | Dôvod integrácie | Databázový systém | Objem dát (GB) | Počet objektov |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **KL** | Kurzový lístok ECB | externý | 1 | BI, OUT | WEB API (XML) | 1 | 2 |
| **RPO** | Register právnických osôb | externý | 1 | BI | WEB API (XML) | 2 | 10 |
| **RS** | Register subjektov | interný | 1 | BI | DB – Oracle (Tabuľky) | 60 | 18 |
| **Subjekty 2.0** | Subjekty 2.0 | interný | 1 | BI | DB – MySQL (Tabuľky/JSON) | 0,4 | 4 |
| **ESMA** | European Securities and Market Authority | externý | 2 | BI | Súbor (XML) | 500 | 5 |
| **LEI kódy / GLEIF** |  | externý | 2 | BI | WEB API (XML) | 0,05 | 1 |
| **RBUZ** | Register bankových úverov a záruk | interný | 2 | BI | DB – Oracle (Tabuľky) | 30,6 | 24 |
| **REGFAP** | Register finančných agentov a finančných poradcov | interný | 2 | BI | DB – MySQL (Tabuľky) | 1 | 30 |
| **RO** | Register organizácií (pre spracovanie MFI BSI, FA) | externý | 2 | OUT | Súbor (CSV) | 0,3 | 2 |
| **RUZ** | Register účtovných závierok | externý | 2 | BI | WEB API (JSON) | 1 | 6 |
| **SU SR** | Číselníky a registre poskytované štatistickým úradom SR (pre spracovanie GFS, IC, BoP, FA a iné) | externý | 2 | BI, OUT | Súbor (JSON, XML (štandard SDMX-ML), XLSX, CSV, ACCDB) | 1 | 30 |
| **ŠZP ESAS** | ŠZP ESAS (XBRL Fujitsu) | interný | 2 | BI | Súbor (XBRL) | 80 | 100 |
| **ŠZP ESAS** | ŠZP ESAS (XBRL Fujitsu) | interný | 2 | BI, OUT | DB – Oracle (Tabuľky/BLOB obsahujú XML v ZIPe) | 1 | 100 |
| **ŠZP** | Štatistický zberový portál | interný | 2 | BI | DB – Oracle (Tabuľky) | 250 | 100 |
| **ARDAL** | Agentúra pre riadenie dlhu alikvidity - údaje o úveroch, vkladoch, CP, derivátoch, repo obchodoch (stavy a transakcie); (pre spracovanie GFS, BoP, FA) | externý | 3 | OUT | Súbor (XLSX) | 1 | 2 |
| **BISNODE** | Databáza firiem - kvantitatívne akvalitatívne informácie o subjektoch | externý | 3 | BI, OUT | WEB API (CSV) | 1,3 | 14 |
| **CERI** | Centrálna evidencia regulovaných informácií | interný | 3 | BI | MySQL DB (Tabuľky/JSON) | 1 | 3 |
| **CSDB** | Centralised Securities Database | interný | 3 | BI, OUT | WEB API (XML) | 170 | 4 |
| **ESKN** | Katastrálny úrad (nehnuteľnosti, listy vlastníctva, vlastníci, prepojenia medzi dátami) | externý | 3 | BI | Súbor (DBF) | 2 | 15 |
| **Eurofondy** | Exporty dát z čerpania eurofondov | externý | 3 | BI | WEB API (JSON) | 50 | 39 |
| **Finančná správy SR** | Tabuľky Finančnej správy SR pre spracovanie BoP | externý | 3 | OUT | Súbor (XLSX) | 0,5 | 2 |
| **Makroek. DB** | Makroekonomická databáza | interný | 3 | BI | Súbor (XLS/CSV) | 5 | 8 |
| **manuálne vstupy** | Konfiguračné metadáta pre všetky spracovania | interný | 3 | OUT | Súbor (XLS/CSV) | 1 | 60 |
| **manuálne vstupy** | Týždenné štatistické ukazovatele | interný | 3 | OUT | Súbor (XLS/CSV) | 8 | 100 |
| **manuálne vstupy** | Ročné štatistické ukazovatele | interný | 3 | OUT | Súbor (XLS/CSV) | 2 | 20 |
| **manuálne vstupy** | Manuálne výkazy a kompilácie údajov NBS | interný | 3 | OUT | Súbor (XLS/CSV) | 3 | 40 |
| **Ministerstvo financií SR** | Tabuľky MF SR pre spracovanie BoP, FA | externý | 3 | OUT | Súbor (XLSX) | 1 | 3 |
| **SHSDB** | Securities Holdings Statistics Database | interný | 3 | BI, OUT | WEB API (XML) | 170 | 4 |
| **Skúšky** | Databáza otázok na skúšky, cvičné testy, ich výsledky, termíny skúšok | interný | 3 | BI | DB – MySQL (Tabuľky) | 0,1 | 3 |
| **Štátna pokladnica** | Tabuľky Štátnej pokladnice pre spracovanie GFS | externý | 3 | OUT | Súbor (XLSX) | 5 | 5 |
| **BLMBRG** | Bloomberg - obchodná platform | externý | 4 | BI | WEB API (JSON/XML) | 1 | 50 |
| **CRIBIS** | CRIBIS - databáza firiem, majetkové a personálne prepojenia | externý | 4 | BI | REST API | 1 | 1 |
| **ECB SDW** | European Central Bank Statistical Data Warehouse - všetky datasety kde figuruje krajina "SK" | externý | 4 | BI | Súbor (CSV/XML) | 1 | 50 |
| **IDAS** | Agendový systém dohľadu | interný | 4 | BI | Súbor (CSV) | 55 | 10 |
| **Register OFV** |  | interný | 4 | BI | DB – Oracle (Tabuľky) | 0,1 | 2 |
| **Sociálna poisťovňa** | Tabuľky Sociálnej poisťovne pre spracovanie BoP | externý | 4 | BI, OUT | Súbor (XLSX) | 0,2 | 4 |
| **ASDR** | Agendový systém dohľadu a regulácie - ISD (Informačný systém dohľadu) | interný | 5 | BI | Oracle DB | 9 | 100 |
| **oversi.gov.sk** | Register trestov | externý | 5 | BI | WEB API (-) | 1 | 1 |

Tabuľka 17: Zoznam hlavných zdrojových systémov pre NBS DWH, ktoré budú integrované do relačnej časti DWH

V rámci dátovej architektúry NBS existujú aj ďalšie databázy a importy z externých zdrojov, ktoré plánujeme počas projektu integrovať. Detaily sú v prílohe A - Zdrojové systémy NBS DWH. Stĺpec H označuje systémy relevantné pre tento projekt. Celkovo systémov na integráciu je 65. V hornej tabuľke č. 17 je zobrazených 38 systémov a priradených do jednotlivých fáz projektu. Tých, ktoré nie sú uvedené v hornej tabuľke ostáva 27. Tieto systémy majú rôzny rozsahu náročnosti integrácie. Niektoré z týchto systémov plánujeme integrovať priamo do relačnej časti DWH (registre, číselníky) a niektoré dátové zdroje plánujeme požadovať uložiť len do DataLake Stage vrstvy pre Data discovery účely. Táto integrácia prebehne v 3. až 5. etape projektu a je požadované tieto práce vykonať uchádzačom. Pre nacenenie projektu môžeme počítať pravidlo, že spôsob integrácie bude nasledovný:

* 38 systémov primárne integrované do relačnej časti DWH (tabuľka č. 17) v TL a PL vrstvách podľa potrieb reportingu, viď. nasledovná kapitola. Označené v stĺpci B ako „BI“, „OUT“, „BI, OUT“.
* 12 systémov z prílohy pôjde tiež do primárne do relačnej časti DWH (registre, číselníky), integrácia dát bude požadovaná na úrovni TL vrstvy. Označené v stĺpci B ako „REGISTER“.
* 15 systémov z oblasti transakčných dát, t. j. tie ktoré nie sú registrami, či číselníkmi budú cez Ingestion framework načítané do DataLake v tabulárnom formáte bez zložitých transformácií, max 5 mapovaní na číselníky, či registre. Označené v stĺpci B ako „DISCOVERY“.

### Reportingová oblasť

Integrácia dátových zdrojov do konsolidovanej dátovej základne bude hlavným predpokladom pre vybudovanie analytickej dátovej vrstvy – datamarty. Sú požadované vytvoriť aj ukážkové reporty, ale predpokladáme že ďalšie reporty a analýzy si NBS vytvorí na základe naučených vedomostí z projektu.

Tieto datamarty budú slúžiť za účelom:

* „self-service“ vytvárania analytických výstupov (produktov):
  + pre podporu manažérskeho rozhodovania,
  + pre podporu strategických činností NBS – dohľad, regulácia, štatistika, ochrana spotrebiteľa,
  + pre podporu operatívneho rozhodovania.
* zdieľania analytických údajov a prehľadov v interaktívnej grafickej podobe naprieč spoločnosťou,
* ukladanie a historizáciu analytických údajov,
* prepojenie individuálnych a agregovaných údajov (a ich rekonciliácií),
* vytvorenie konsolidovanej informácie o subjektoch finančného trhu.

Je požadované mať možnosť nastaviť v metadátach datamartoch, ktoré dáta za aké obdobie majú byť prepočítavané (a v akej periodicite) a ktoré obdobia majú byť už uzamknuté.

V rámci požadovaného projektu je zadefinovaných týchto 11 analytických oblastí:

| ID | Označenie datamartu | Fáza | Popis datamartu | Rámcový popis požadovaných prvkov |
| --- | --- | --- | --- | --- |
| 1 | REPORTS | 2 | * Datamarty pre štatistické údaje vyzbierané cez výkazy portálu ŠZP. | * časové rady ekonomických položiek, medzimesačné zmeny objemov súvahových položiek podľa rôznych ekonomických členení, výpočet vážených priemerov, etc. |
| 2 | MFI\_ESAS\_ANACREDIT | 2 | * Datamarty pre údaje o bankových úveroch a ich protistranách. * Datamart pre konsolidáciu agregovaných údajov ŠZP (menová a finančná štatistika) a individuálnych údajov RBÚZ. * Datamart na rekonciliáciu údajov ESAS (COREP a FINREP) na RBÚZ, resp. kosolidované porovnanie týchto dát | * časové rady ekonomických veličín (nominálna hodnota, priemerná úroková sadzba, etc.), prehľady podľa subjektu, nástroja, prijatého zabezpečenia, rizika protistrany, etc. konsolidované porovnanie nominálnych hodnôt agregátov menovej štatistiky a individuálnych nástrojov s možnosťou drill-down analýzy v zmysle BSI metodiky časové rady položiek, zmeny medzi referenčnými dátumami položiek (objemy a percentá), poradia subjektov |
| 5 | AML | 2 | * Datamart pre spracovanie dát AML pre risk assessment. | * výpočet AML rizikových indexov z výkazov ŠZP |
| 6 | EXR | 2 | * Datamart pre kurzový lístok | * prehľad kurzového lístka, časové rady výmenných kurzov |
| 4 | SECURITIES | 3 | * Datamart pre konsolidáciu údajov o cenných papieroch CSDB a vykazovaných údajoch o cenných papieroch ŠZP | * konsolidované porovnanie atribútov cenných papierov z výkazov ŠZP voči databáze CSDB a SHSDB |
| 11 | ESMA | 3 | * Datamart pre cenné papiere Datamart pre ESMA dáta | * Obsah reportov približne obdobne ako v publikáciách ESMA:  https://www.esma.europa.eu/file/122189/download?token=kTi4gzPC https://www.esma.europa.eu/file/122190/download?token=Cd0YfTB2 |
| 16 | MACROP\_\_\_1 | 3 | * Datamart pre monitoring fin. inštitúcií + CBD * Datamart pre proticyklický smerovník | * Reporty zamerané na základný monitoring a analýzu rizík predovšetkým v bankovom sektore v členení podľa jednotlivých oblastí - ziskovosť, kapitálová primeranosť, kvalita aktív, štruktúra bilancie, likvidita a pod. Možnosť uskutočňovať medzinárodné porovnania indikátorov pomocou údajov z databázy ECB SDW. * Report na výpočet konsolidovaných bankových údajov (CBD), ktoré sa zasielajú na pravidelnej báze do ECB. * Reporty zamerané na základný monitoring a analýzu rizík v nebankových sektoroch v členení podľa jednotlivých sektorov - DSS, DDS, KI, OCP. Report ma kvantifikáciu trhových rizík a výpočty stresového testovania pre fin, inštitúcie. |
| 19 | MACROP\_\_\_3 | 3 | * Datamart pre podnikový sektor | * Na komplexné pokrytie monitoringu podnikového sektora (primárne na účely finančnej stability) |
| 9 | ESAS | 4 | * Datamart pre dohľad nad kapitálovým trhom * Datamart pre dohľad nad platobnými službami * Datamart pre risk-based supervison * Datamart pre dohľad nad poisťovníctvom * Datamart pre údaje bankového dohľadu v portáli ESAs | * súvaha, výsledovka, analýza 1 subjektu, trhová analýza, zloženie portfólia aktív, výkonnosť, časové rady, medziročné zmeny, vážené priemery, definovanie vhodného metapopisu na integrované dáta, risk dashboard, SREP skóre, kapitálová primeranosť, pákový efekt, systémové riziko. * Sektory: centrálny depozitár, burza, obchodníci s cennými papiermi, správcovské spoločnosti a ich fondy, limity pre fondy časový pohľad na vývoj vybraných atribútov (napr. fin. ukazovateľov) 1 subjektu vrátane agregovaného výstupu za sektor vo forme tabuľky príp. grafu vrátane medzimesačného a medziročného porovnania; * analýza 1 subjektu v podobe tabuľky (zahŕňajúca údaje z viacerých zdrojov od finančných ukazovateľov, rôzne kvalitatívne zdroje napr. údaje o osobách, konaniach...) * risk dashboard zahrňujúci analýzu časových radov, prehľad osôb, etc. časový pohľad na portfólio aktív v tabuľkových a grafických prezentáciách (statické tabuľky, kontingenčné tabuľky, koláčové grafy, etc.), medzimesačné, medziročné vývojové indexy časové rady položiek, zmeny medzi referenčnými dátumami položiek (objemy a percentá), poradia subjektov |
| 17 | MACROP\_\_\_2 | 4 | * Datamart pre databázu fin. aktív držaných fin. inštitúciami | * Reporty zamerané na základný monitoring a analýzu rizík v nebankových sektoroch v členení podľa jednotlivých sektorov - DSS, DDS, KI, OCP. * Report ma kvantifikáciu trhových rizík a výpočty stresového testovania pre fin, inštitúcie. |
| 14 | ASDR | 5 | * Datamart pre účely ASDR * Datamart pre dohľad nad kapitálovým trhom * Datamart pre dohľad nad kapitálovým trhom | * Spájanie údajov (kvalitatívne, kvantitatívne a procesné údaje) pre účely dohľadu, ktoré má NBS k dispozícií o danom subjekte podľa jednotlivých sektorov. * Early warnings, vykazovanie načas, KPIs, benchmarkové porovnania. * Prehľady subjektov podľa sektorov, pôsobenie osôb vo funkciách, konflikty záujmov, nedôverihodnosť, odborná spôsobilosť... |

Tabuľka 18: Zoznam DataMartov pre projekt NBS DWH

V nasledujúcej tabuľke je kvantifikácia komplexnosti dátových oblastí pre BI reporting. Každá oblasť je definovaná požadovanou dátovou integráciou, predpokladanou náročnosťou výpočtu jednotlivých KPI a predpokladaným rozsahom faktových tabuliek. Počet a rozdelenie dátových oblastí sa môže meniť (ale nie zásadne) pri návrhu realizovaného riešenia.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ID | Označenie datamartu | Počet dátových zdrojov na integráciu | Odhad počtu DIM tabuliek | Odhad počtu Faktových tabuliek | Odhad počtu atribútov | Odhad počtu metrík | Zložitosť výpočtu metrík | Počet vzorových reportov |
| 1 | REPORTS | 3 | 100 | 50 | 400 | 50 | Stredná | 10 |
| 2 | MFI\_ESAS\_ANACREDIT | 4 | 67 | 3 | 270 | 30 | Nízka | 5 |
| 5 | AML | 1 | 14 | 1 | 80 | 50 | Vysoká | 2 |
| 6 | EXR | 1 | 3 | 2 | 10 | 0 | Nízka | 1 |
| 4 | SECURITIES | 4 | 18 | 2 | 100 | 10 | Nízka | 2 |
| 11 | ESMA | 8 | N/A | 25 | 1000 | 100 | Vysoká | 10 |
| 16 | MACROP\_\_\_1 | 4 | N/A | N/A | N/A | 100 | Vysoká | 5 |
| 19 | MACROP\_\_\_3 | 6 | N/A | 10 | 100 | 50 | Nízka | 4 |
| 9 | ESAS | 13 | 75 | 25 | 500 | 100 | Stredná | 10 |
| 17 | MACROP\_\_\_2 | 5 | 75 | 25 | 500 | 10 | Vysoká | 5 |
| 14 | ASDR | 12 | N/A | N/A | N/A | N/A | N/A | N/A |
|  | 11 DataMartov |  | 352 | 143 | 2960 | 500 |  | 54 |

Tabuľka 19: Predpokladaná maximálna náročnosť jednotlivých DataMartov a prezentačnej vrstvy (PL) relačnej časti DWH

Poznámka 1: Odhady komplexnosti jednotlivých DataMartov vychádzajú z výsledkov Pilotu DWH a metodiky vývoja podľa Ralpha Kimbala

Poznámka 2: Detailná tabuľka je v prílohe „DataMarty NBS DWH“ .

### Výstupné dátové oblasti

Druhá veľká časť výstupov NBS DWH sú výstupy dát v presne definovanej štruktúre pre potreby odosielania týchto dát mimo NBS.

Cieľom tejto časti je rámcový popis požiadavky na automatizáciu procesov diseminácie v štatistickom spracovaní dát z pohľadu:

* integrácie požadovaných dát z identifikovaných dátových zdrojov (referencia na tabuľku) do databáz DWH/DL,
* vytvorenia dátových modelov pre účely štatistického spracovania dát v prostredí databáz DWH/DL a nástrojov BI,
* návrhu transformačných a výpočtových procedúr vo vrstvách DWH/DL a nástrojov BI
* prípravy a archivácie výstupov v exportných formátoch (GESMES/TS / SDMX/ML),
* automatizácie kontrol dátovej kvality výstupov.

Výstupy zo štatistického spracovania je možné kategorizovať do nasledovných oblastí:

* odvetvové
* nástrojové
* komplexné

V rámci odvetvových a nástrojových oblastí spracovania je požadované zaviesť automatizované a semi-automatizované procesy, a generovanie výstupných formátov (GESMES/TS/SDMX/ML/TSV/CSV) podľa špecifikácie konkrétneho spracovania. V súčasnej dobe sú využívané dostupné self-service nástroje, pričom logiky spracovania sú decentralizované v rámci týchto nástrojov (VBA, self-service BI, SQL, manuálna práca MS EXCEL). Rámcové ohraničenie rozsahu a štruktúry dátových zdrojov, nad ktorými sa budú vykonávať výpočtové operácie sú uvedené v úvode tejto kapitoly.

V prípade výstupov označených ako komplexné, nie je vzhľadom na povahu a rozsah možné zabezpečenie plne automatizovaného „end-to-end“ spracovania a to najmä kvôli častej potrebe manuálnych korekcií a ad-hoc intervenciám v procese kompilácie výstupov zo strany zostavovateľov štatistík. V tomto prípade bude požadované minimálne zabezpečenie aktuálnych a dostupných dát zo všetkých dátových zdrojov pre efektívne spracovanie výstupov prostredníctvom na to určených „self-service“ nástrojov.

Oblasti štatistického spracovania sa pri tvorbe výstupných súborových formátov vyznačujú rôznorodosťou potrebných dátových zdrojov v kombinácií s ich vysokou dostupnosťou. Tieto dátové zdroje možno rozčleniť na:

* manuálne dátové vstupy (dáta, ktoré nemožno získať z interných alebo externých systémov a sú výsledkami práce expertov danej domény),
* dáta z interných systémov NBS (interné databázy, portály, súbory),
* dáta z externých inštitúcii (štatistické údaje v rôznych prenosových formátoch, ktorých rozsah a spôsob výmeny je dohodnutý v bilaterálnych zmluvách).

Zoznam **štatistických domén**, v ktorých je požadované zvýšenie úrovne automatizácie, sa nachádza v tabuľke nižšie:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Výstup | Doména | Typ | | | Spôsob spracovania dát v DWH | | | | Počet dátových zdrojov |
| Komplexné | Sektorové / odvetvové | Nástrojové | príprava dátových zdrojov | Transformácia dát | Tvorba OUT | Nastavenie DQ |
| DM.001 | Government Finance Statistics | X |  |  | X |  |  |  | 24 |
| DM.002 | MFI Interest Rate Statistics |  | X |  | X | X | X | X | 2 |
| DM.003 | MFI Balance Sheet Items |  | X |  | X | X | X | X | 4 |
| DM.004 | Securities Issues |  |  | X | X |  |  |  | 9 |
| DM.005 | Securities Holdings SHSE |  |  | X | X | X | X | X | 2 |
| DM.006 | Securities Holdings SHSS |  |  | X | X |  |  |  | 4 |
| DM.007 | Investment Funds |  | X |  | X | X | X | X | 3 |
| DM.008 | Factoring, Leasing and Consumer Credit Companies |  | X |  | X | X | X | X | 1 |
| DM.009 | Financial Accounts | X |  |  | X |  |  |  | 8 |
| DM.010 | Insurance Corporations, (sektorovo / odvetvové) |  | X |  | X | X | X | X | 8 |
| DM.011 | Pensions Funds |  | X |  | X | X | X | X | 6 |
| DM.012 | Balance of Payments | X |  |  | X |  |  |  | 13 |
| DM.013 | Payment statistics |  |  | X | X | X | X | X | 3 |
| DM.014 | Payment institutions and institutions of electronic money |  | X |  | X | X | X | X | 9 |
| DM.015 | Bank for the Accounts of Companies Harmonised |  | X |  | X |  |  |  | 5 |

Tabuľka 20: Zoznam štatistických domén

V DWH databáze sú požadované **vytvorenie v OUT vrstve** príslušné tabuľky v štruktúre požadovanej pre jednotlivé výstupy. Výstupy budú generované ako 1:1 tieto tabuľky so zmenou do požadovaného formátu:

|  | Doména | Cieľ | Výmenný formát |
| --- | --- | --- | --- |
| OP.001 | MFI Interest Rate Statistics | ECB EXDI | XML (SDMX-ML 3.0) |
| OP.002 | MFI Interest Rate Statistics | WEB NBS | CSV |
| OP.003 | MFI Balance Sheet Items | ECB EXDI | XML (SDMX-ML 3.0) |
| OP.004 | MFI Balance Sheet Items | WEB NBS | CSV |
| OP.005 | Securities Holdings SHSE | ECB EXDI | CSV |
| OP.006 | Investment Funds | ECB EXDI | XML (SDMX-ML 3.0) |
| OP.007 | Investment Funds | WEB NBS | CSV |
| OP.008 | Factoring, Leasing and Consumer Credit Companies | WEB NBS | CSV |
| OP.009 | Factoring, Leasing and Consumer Credit Companies | ECB EXDI | XML (SDMX-ML 3.0) |
| OP.010 | Inssurance Corporations | ECB EXDI | XML (SDMX-ML 3.0) |
| OP.011 | Inssurance Corporations | EIOPA | XML (XBRL) |
| OP.012 | Pension Funds | ECB EXDI | XML (SDMX-ML 3.0) |
| OP.013 | Pension Funds | EIOPA | XML (XBRL) |
| OP.014 | Payment Statistics | ECB EXDI | XML (SDMX-ML 3.0) |
| OP.015 | Payment Statistics | EBA | XML (SDMX-ML 3.0) |

Tabuľka 21Zoznam výstupov

Zmena na požadovaný formát a odosielanie (cez SFTP) na distribučný server je súčasťou projektu. Vytvorenie distribučného servera nie je predmetom tohto projektu.

Spracovanie **vyžaduje historizáciu zmien** vstupných a aj výstupných dát vrátane archivácie výmenných formátov vo frekvencii, v akej sú zo zdrojových systémov zbierané a v cieľovom systéme spracovávané.

**Transformačné logiky jednotlivých spracovaní** sú v súčasnosti decentralizované a umiestnené v rôznych komponentoch v závislosti od použitých nástrojov. Jedným zo základných predpokladov automatizácie je zmapovanie a centralizácia týchto logík do prostredia, v ktorom sa bude spracovanie vykonávať. Tieto operácie sa opakujú a možno zhrnúť do nasledovných oblastí:

|  | Operácia | Popis |
| --- | --- | --- |
| a. | Preklad a mapovanie | Aplikovanie pravidiel pre preklad prvkov číselníkov, atribútov riadkov, stĺpcov zo vstupného na požadovaný výstupný formát.  Napr.: Preklad adresy dátového bodu národného dátového modelu na adresu dátového bodu dátového modelu ECB (typ 1:N, N:1). |
| b. | Obohacovanie metadátami | Denormalizácia, pripájanie a kombinovanie s inými dátovými zdrojmi na základe mapovacích pravidiel.  Napr.: Pripojenie atribútov cenného papiera z databázy cenných papierov na základe jednoznačného ID cenného papiera k vykázaným údajom z databázy IS ŠZP. |
| c. | Filtrovanie | Výber relevantných objektov (záznamy, ich atribúty) na základe filtrovacích pravidiel. |
| d. | Agregácia a sumarizácia | Výpočet (aj viackrokový) dátových bodov v požadovanej granularite aplikovaním sumarizačných výpočtov.  Napr.: Výpočet váženého priemeru na základe predpočítanej váhy a hodnoty, súčet za podmienky, počet, etc. |
| e. | Formátovanie | Príprava predpočítaných dát do finálnej štruktúry výstupného formátu – súboru.  Napr.: Formátovanie vypočítaných dátových bodov do formátu XML v štandarde SDMX-ML 3.0. |
| f. | Matematické operácie | Aplikovanie matematických výpočtov na dáta podľa stanovených pravidiel a metodík.  Napr.: Výpočet odhadov, prenásobenie koeficientom. |

Tabuľka 22: Transformačné logiky jednotlivých spracovaní pre OUT vrstvu

### Funkčné požiadavky

Z pohľadu biznis používateľského komfortu ale aj z pohľadu rýchlosti na reagovanie zmien, poprípade testovania nových častí výpočtov požadujeme:

* ETL procesy budú programované ako fixné, t. j. mapovanie, transformácie, spájanie a agregácie bude definované cez nástroj Oracle Data Integrátor.
* Parametre mapovaní, t. j. napr. ktoré číselníkové dáta vstupujú, prípadne s akými váhami je striktné požadované vytvoriť ako parametrické. Tieto parametre bude používateľ mať možnosť cez mostíkový číselník meniť bez zadania zmenového listu na odbor IT a zmeny ETL úloh .
* Rovnako sa počíta, že transformácia do výstupných formátoch bude definovaná v ODI.
* Pokiaľ výstupné formáty budú príliš komplikované na ODI (je zakázané, až na dohodnuté výnimky programovať – požadujeme low-code metódy implementácie) je možné použiť na transformáciu Ingestion framework, ktorý by mal poznať formáty na spracovanie. Preto predpokladáme, že keď bude vedieť čítať, mal vedieť aj ich generovať.
* Súčasťou projektu je aj analýza primárneho zdroja dát a prepojenie číselníkov rôznych zdrojových systémov cez Referenčné hodnoty. Biznisová požiadavka je mať nástroj na udržiavanie týchto hodnôt bez priamej podpory IT ako je popísané v kapitole 3.2.13 Správa číselníkov.
* Dáta budú primárne integrované do relačnej časti DWH. Niektoré vybrané zdroje budú spracovávané/ukladané cez DataLake. Toto bude dohodnuté v mapovacom dokumente počas Návrhu riešenia. V rámci spracovania dát sa požaduje:
  + automatické sťahovanie a ukladanie dát
  + transformácia do relačnej podoby pre analýzy
  + sprístupnenie relačnej štruktúry pre dátovú analýzu
  + asistencia pri definovaní relevantnej množiny záznamov a atribútov pre ďalšie spracovanie (analýza a rozhodnutie o dátach na strane NBS)
  + integrácia relevantnej podmnožiny dát do DWH
  + sprístupnenie relevantných dát pre reporting nad DWH
  + automatizácia spracovania v danom rozsahu
  + vzhľadom ku charakteru rozsahu spracúvaných dát bude navrhnutý efektívny mechanizmus ukladanie v rámci DataLake
  + dáta z DWH bude možné off-loadovať do DataLake, napr. z výstupnej vrstvy DWH (OUT) alebo výpočty z Prezentačnej vrstvy DWH (L2).
* V rámci projektu požadujeme dodať metodiku a implementáciu konsolidačných mechanizmov pre oblasti (identifikácia primárnej informácie, vytvorenie kmeňového záznamu, nastavenie kvalitatívnych ukazovateľov (DQI)):
  + Subjekty
  + Cenné papiere
  + Adresy Slovenskej republiky
* V rámci dodávky bude sprevádzkovaný prístup koncových nástrojov MS Power BI, QlikSense k relevantným vrstvám DWH a DataLake, tj. bude demonštrovaný aj SQL prístup k štruktúram DWH a DataLake s rešpektovaním uvedených prevádzkových a bezpečnostných požiadaviek.
* Vytvorenie OUT vrstvy pre generovanie štandardných a dohodnutých exportov do rôznych systémov NBS a aj mimo NBS (OpenData, ECB, EBA a podobne).

## Požiadavky na organizáciu a výstupy projektu

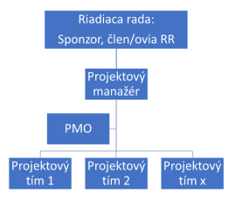
### Projektové riadenie

Požiadavky na zabezpečenie projektového riadenia zo strany verejného obstarávateľa zahŕňajú:

* zabezpečenie projektového riadenia v súlade s metodikou projektového riadenia v NBS, ktorú určuje pracovný predpis NBS č. 20/2020 o projektovom riadení. Metodika je založená na princípoch metodiky PRINCE2 viď príloha C.
* vytvorenie plánu kvality v súlade s metodikou riadenia kvality projektového riadenia v NBS
* prácu s MS Project Professional v minimálnom rozsahu tvorby a aktualizácie plánu projektu

### Prístup k projektu a RACI matica

Organizácia projektu je ukotvená schémou v organizačnej štruktúre projektu viď obrázok:



Obrázok 16: Organizačná štruktúra projektu

Projekt bude dodávaný v piatich (5) implementačných fázach, z ktorých každá musí obsahovať nasledovné aktivity:

1. Analýza
2. Dizajn
3. Implementácia
4. Nábeh do produkcie
5. Produkcia

Na strane NBS sa predpokladá zapojenie nasledovných rolí:

| Rola NBS | Popis | Predpokladané fázy účasti na projekte |
| --- | --- | --- |
| Projektový manažér | Riadenie projektu – interné | Analýza a Návrh, Implementácia a Nasadenie |
| Solution architekt | Špecialista pre validáciu a odsúhlasenie návrhu architektúry | Analýza a Návrh, a Nasadenie |
| Databázový špecialista | Špecialista pre validáciu a odsúhlasenie návrhu úložiska a databáz | Analýza a Návrh, Implementácia a Nasadenie |
| Testeri pre akceptačné testovanie | Zástupca odborného útvaru, ktorý bude používateľom | Testovanie a nasadenie |
| Biznis používatelia | Špecialisti nominovaní pre formuláciu funkčných požiadaviek a odsúhlasovanie špecifikácie | Analýza a Návrh, Akceptačné testovanie |
| Biznis vlastníci | Vedúci pracovníci, ktorí sú nominovaný na finálne odsúhlasovanie výstupov za odbory. | Analýza a Návrh, Akceptačné testovanie |
| IT aplikačný špecialista | Špecialista pre validáciu a odsúhlasenie návrhu aplikácií | Analýza a Návrh, Akceptačné testovanie a Nasadenie |

Tabuľka 23: Projektové role na strane NBS

Na strane uchádzača požadujeme zapojenie minimálne kľúčových expertov, a to: Kľúčový expert č. 1 – Projektový manažér, Kľúčový expert č. 2 – Architekt riešenia (Relačná časť DWH), Kľúčový expert č. 3 – Architekt riešenia (BigData/DataLake časť DWH), Kľúčový expert č. 4 – Senior Dátový Analytik, Kľúčový expert č. 5 – Data Ingestion Špecialista, Kľúčový expert č. 6 – ETL Špecialista, Kľúčový expert č. 7 – BI Špecialista ktorí sú uvedení v zmluve o dielo, v podmienkach účasti, ktoré sú súčasťou týchto súťažných podkladov verejného obstarávania.

#### Rozhodovanie sporov

V prípade rozporov medzi PM NBS a PM zhotoviteľa prerokuje spory Riadiaca rada projektu a rozhodnutie vydáva výlučne Sponzor projektu, ktorý nesie hlavnú zodpovednosť za celý projekt.

V prípade, že toto rozhodnutie je v zásadnom rozpore s platnou zmluvou, iniciuje Sponzor projektu negociácie a popr. konanie k zmenovým požiadavkách.

RACI matica základných projektových aktivít je uvedená prílohe D.

### Pravidlá pre riadenie komunikácie

Táto kapitola obsahuje popis spôsobu a frekvencie komunikácie jednotlivých zainteresovaných strán (ako interných tak aj externých). Takýto spôsob zabezpečuje kontrolovanú a správne usmerňovanú komunikáciu všetkých zainteresovaných strán v realizačných projektoch.

#### Zainteresované strany

* riadiaca rada (RR),
* hlavný zhotoviteľ projektu,
* projektový manažér objednávateľa ,
* projektový manažér zhotoviteľa,
* projektový tím objednávateľa
* projektový tím zhotoviteľa

#### Zaobchádzanie s informáciami

Berúc do úvahy význam dôležitosti niektorých informácií získaných v projekte a potreby udržiavať tieto informácie dôvernými majú všetky role v projekte povinnosť zaobchádzať so získanými informáciami, ktoré sú verejne neprístupné, a to technické, obchodné, finančné, komerčné alebo prevádzkové informácie, špecifikácie, plány, dáta, počítačové programy, softvér alebo dokumentácia v akejkoľvek podobe, či už zachytené hmotne alebo ústne poskytnuté ako s chránenými.

Každá rola v projekte musí dodržiavať a prijať zodpovedajúce technické, organizačné a iné opatrenia potrebné na ochranu informácií, ktoré mu boli alebo mu budú poskytnuté, alebo sprístupnené, pred neoprávnenou manipuláciou s nimi, minimálne však také opatrenia, ktoré sú porovnateľné s opatreniami aké dodržiava pri ochrane vlastných údajov podobnej povahy a dôležitosti.

#### Riadenie komunikácie

Pre úspešné naplnenie cieľov realizačného projektu je potrebné vytvoriť vhodné podmienky pre efektívnu komunikáciu. Pre všetkých účastníkov projektu je kľúčové, aby pochopili svoje úlohy a zodpovednosti, ako aj úlohy a zodpovednosti ostatných členov projektu.

Komunikácia a jej toky sú primárne definované a odvíjajú sa od nastavenia riadenia projektu, t. j. od štruktúry riadenia projektu.

Z hľadiska formálnosti sa komunikácia v rámci projektu rozdeľuje na formálnu a neformálnu.

Komunikácia na projekte bude prebiehať v zmysle čiar autority, a to na nasledovných úrovniach:

* riadená komunikácia (využívajúca všetky dostupné spôsoby komunikácie),
* stretnutia,
* eskalácia.

##### Riadená komunikácia

Najdôležitejším aspektom komunikácie na projekte je komunikácia projektového manažéra. Predmetom tejto komunikácie je najmä:

* operatívne riadenie projektu pozostávajúce najmä z/zo:
  + pravidelných a ad hoc stretnutí, ktorých výsledkom je:
    - aktualizácia projektových registrov,
    - (pre)rozdelenie a (pre)plánovanie projektových úloh,
    - eskalovanie problematických záležitostí,
    - riešenie incidentov,
  + vedenia registra rizík,
  + správy verzií výstupov,
  + zabezpečenia kvality výstupov,
* preberanie a odovzdávanie výstupov projektu.

Všetky formálne náležitosti sú komunikované emailovou komunikáciou, resp. úradnou listinnou komunikáciou.

Využitie elektronickej pošty sa predpokladá v rámci všetkých komunikačných procesov projektu. Prostredníctvom elektronickej pošty sa vymieňajú jednak informácie obsiahnuté v správach a jednak elektronické súbory v prílohe správ. Každá elektronická správa obsahujúca informácie alebo súbory týkajúce sa projektu musí mať identifikáciu projektu v predmete správy.

V projekte sa nepredpokladá masívne využívanie listovej korešpondencie a pokiaľ nie je stanovené inak, primárnym nástrojom komunikácie je elektronická pošta. Listovú korešpondenciu je možné odoslať prostredníctvom podateľne, doporučeným listom, prostredníctvom kuriérskej služby alebo faxom.

V projekte sa predpokladá aj využitie telekonferenčných a videokonferenčných zariadení. Pri organizácii telekonferencií a videokonferencií platia rovnaké pravidlá ako pre organizáciu bežných stretnutí. Navyše však zvolávateľ stretnutia zabezpečuje včasnú distribúciu údajov potrebných pre pripojenie vzdialených účastníkov.

V projekte sa predpokladá časté využitie telefonických zariadení, najmä pri riešení operatívnych a organizačných záležitostí. Využitie telefónu nie je vhodné pre zadávanie úloh a rozhodovanie o dôležitých záležitostiach. V prípade, že je výsledkom telefonického rozhovoru rozhodnutie alebo zadanie úlohy je potrebné potvrdiť túto skutočnosť aj prostredníctvom elektronickej pošty.

Neformálnou komunikáciou sa rozumie doplnkový spôsob komunikácie pre efektívne riešenie každodenných náležitostí. Pre tento spôsob komunikácie môžu byť využívané adekvátne nástroje a technológie, ktoré zabezpečia efektívny a nenáročný spôsob komunikácie (videohovory, chaty a pod.).

#### Projektové stretnutia

Okrem riadenej komunikácie a predkladania požadovaných výstupov budú v rámci realizácie projektu organizované rôzne typy pravidelných aj nepravidelných stretnutí.

Tieto stretnutia budú realizované za účelom:

* úvodného oboznámenia sa s projektom a jeho nastavením realizácie,
* plánu realizácie úloh,
* revízie plnenia úloh,
* revízie stavu projektu,
* odovzdávania/preberania projektových výstupov a poskytovania vyžiadanej súčinnosti,
* výmeny ponaučení a odporúčaní,
* sumarizácie a zhodnotenia realizácie projektu.

Pre potreby zdieľania dokumentov je zriadené spoločné zdieľané projektové úložisko dokumentov na SharePointe, kde budú ukladané pracovné verzie dokumentov ako aj ich finálne verzie, ktoré budú vytvorené po schválení výstupov/produktov na RR. Komunikácia bude podporovaná projektovými dokumentmi, predovšetkým zápismi zo stretnutí, výstupmi jednotlivých úloh a nástrojmi projektového riadenia ako sú projektový plán, register úloh, register rizík, a pod.

**Základnými typmi projektových stretnutí sú:**

| Názov | Agenda | Periodicita | Účastníci | Čas a trvanie |
| --- | --- | --- | --- | --- |
| Zasadnutie RR | schvaľovanie kľúčových dokumentov a výstupov, odpočet a stanovovanie strategických úloh, strategické riadenie projektu zodpovednosť za organizáciu projektu, schvaľovanie zmenových požiadaviek projektu | Minimálne 1x za 2 mesiace | členovia RR a prizvaní účastníci | podľa dohody, cca 1 hodina |
| Stretnutia s projektovým vlastníkom | operatívne prijímať rozhodnutia k predchádzaniu problémov, rizikám, riešenie otázok, konzultácia zmien | ad hoc podľa potreby | Vlastníci projektu a prizvaní účastníci | podľa dohody, cca1 hodina |
| Projektový tím (PTO) míting | odpočet a stanovovanie operatívnych úloh, operatívne riadenie projektu, návrhy a rozhodovanie o zmenách v projekte, aktualizácia rizík, otvorených otázok a pod. | 1x týždenne | projektový manažér, PMO, vedúci tímu, členovia projektového tímu podľa potreby a kľúčoví experti podľa potreby | určený deň v týždni, cca 1-2 hodiny |
| Pracovný tím | stretnutie pracovných tímov ku konkrétnym témam projektu (analýzy, procesy, funkčná špecifikácia, finančný model, zmluva, legislatíva, prevádzka a pod.) | podľa potreby | Expert, zástupcovia pracovných tímov a prizvaní účastníci | podľa potreby |

Tabuľka 24: Projektové stretnutia

Zápis vo všeobecnosti vyhotovuje organizátor stretnutia, prípadne ním poverená osoba.

Zápis zo stretnutia musí obsahovať:

* identifikáciu projektu,
* číslo zápisu,
* dátum a čas,
* miesto konania,
* typ stretnutia,
* zoznam prítomných účastníkov,
* tému stretnutia,
* agendu stretnutia,
* stručný záznam z rokovania,
* závery stretnutia rozdelené v zmysle nasledovnej kategorizácie:
* informácie,
* prijaté rozhodnutia,
* otvorené otázky,
* predpoklady,
* riziko,
* odsúhlasené úlohy vrátane určenia zodpovedných osôb za úlohy a termínov úloh,
* prílohy (podľa potreby).

Za správnosť návrhu zápisu zo stretnutia, jeho prípadných medziverzií a jeho finálneho znenia podľa vyššie uvedených požiadaviek zodpovedá organizátor stretnutia, ak sa nedohodne inak; je ďalej zodpovedný aj za správne uloženie návrhu zápisu zo stretnutia a jeho prípadných medziverzií na spoločné zdieľané projektové úložisko dokumentov.

#### Postupy eskalácie

Počas životného cyklu projektu môže dôjsť k neočakávaným situáciám, ktoré nie je možné vyriešiť bez zapojenia vyšších zložiek riadenia alebo ak je potrebné prijať rozhodnutie, vyžadujúce si schválenie vyšším manažmentom. Pre dosiahnutia riešenia, resp. rozhodnutia pre takéto situácie bude v rámci projektu zavedený eskalačný mechanizmus, ktorý bude využívať nasledujúce eskalačné úrovne:

* prvý stupeň eskalácie – projektový manažér,
* druhý stupeň eskalácie – vyšší manažment (riadiaci pracovníci/manažment – vlastníci projektu ako nositelia úlohy z Bankovej rady),
* tretí stupeň eskalácie – Riadiaca rada projektu (zloženie v zmysle Štatútu).

Pri eskalácii je vždy potrebné písomné oboznámenie so vzniknutou situáciou, ktoré okrem popisu vzniknutej situácie popisuje aj dopad na chod projektu, identifikovanie dotknutých osôb, návrh možného riešenia, a pod.

Problematická situácia bude eskalovaná do doby, kým nebude vyriešená, pričom o rozhodnutí o riešení eskalovanej situácie sú informované všetky dotknuté osoby. Eskaláciu je potrebné používať ako nástroj nemožnosti vyriešenia úlohy na danej úrovni. Všetky eskalačné požiadavky musia byť zaznamenané písomnou formou.

### Projektový plán

Požadovaný termín na finalizáciu dodávky riešenia je do 36 mesiacov odo dňa účinnosti zmluvy o dielo. Navrhovaný projektový plán je súčasťou prílohy E.



Obrázok 17: Časový plán projektu

Projekt je rozdelený do piatich fáz (dropov):

| Fáza | Popis | Platobný rozsah (z investičnej ceny implementácie celého diela) |
| --- | --- | --- |
| 1 | * Nastavenie relačnej časti DWH, príprava infraštruktúry, technických používateľov * Integrácia zdrojových systémov pre 1. fázu pre relačnú časť DWH * Návrh a implementácia modelu DWH transakčnej vrstvy (L1) * Inštalácia frameworkov (Ingestion, Workflow manager, Security, Dátová retencia), ich úvodné nastavenie a odsúhlasenie ich funkcionalít | 16 % |
| 2 | * Integrácia zdrojových systémov pre 2. fázu * Návrh a implementácia modelu DWH transakčnej vrstvy (L1) * Návrh a implementácia prezentačnej vrstvy relačnej časti (L2) * Návrh a implementácia OUT vrstvy * Konsolidačná procedúra pre Subjekty * Nastavenie DataLake časti DWH, príprava infraštruktúry, technických používateľov, inštalácia BigData komponentov * Inštalácia frameworkov (Manuálne vstupy, RDM, Test framework, Data Quality framework) | 24 % |
| 3 | * Integrácia zdrojových systémov pre 3. fázu * Návrh a implementácia modelu DWH transakčnej vrstvy (L1) * Návrh a implementácia prezentačnej vrstvy relačnej časti (L2) a príprava Datamartov a Reportov * Návrh a implementácia OUT vrstvy * Konsolidačná procedúra pre Adresy Slovenskej republiky * Implementácia DataLake časti (nábeh do produkcie) * Inštalácia frameworku – Metadata Management (InfoHUB) | 29 % |
| 4 | * Integrácia zdrojových systémov pre 4. fázu * Návrh a implementácia modelu DWH transakčnej vrstvy (L1) * Návrh a implementácia prezentačnej vrstva relačnej časti (L2) a príprava Datamartov a Reportov * Konsolidačná procedúra pre Cenné papiere * Implementácia DataDiscovery infraštruktúry na DataLake * Vytvorenie Sand boxov pre power-userov | 19 % |
| 5 | * Integrácia zdrojových systémov pre 5. fázu * Tvorba prezentačnej vrstva relačnej časti (L2) a príprava Datamartov a Reportov | 12 % |

Tabuľka 25 Projektové fázy a finančné rozdelenie

Toto rozdelenie je za týchto podmienok:

* Každá fáza obsahuje svoje časti Analýza, Design, Implementácia, Nábeh do produkcie, Produkcia, t. j. každá fáza bude implementovaná inkrementálne a postupne. Bude mať produkčný nábeh a biznisový výstup pre NBS používateľov. Očakávame, že jednotlivé systémy sa budú zapájať postupne a postupne testovať.
* Druhá fáza bude môcť začať až keď skončí prvá fáza, ostatné môžu mať (čiastočne) paralelný súbeh a môžu začať aj pred ukončením akceptácie predchádzajúcich fáz.

V tomto bode je požadované aby vybratý uchádzač riešil:

* Riadenie celkovej dodávky riešenia v rozsahu, ktorý je popísaný v tomto dokumente.
* Vypracovanie detailného projektového plánu na začiatku projektu a jeho ďalšia aktualizácia po celú dĺžku trvania projektu.
* Pravidelný reporting o stave projektu pre riadiace orgány projektu s frekvenciou podľa požiadaviek objednávateľa.
* Zabezpečenie a organizácia zdrojov projektu na strane zhotoviteľa a poskytovanie súčinnosti vedeniu projektu na strane banky.
* Manažovanie projektových závislostí a rizík a organizovanie aktivít na ich predchádzanie a riešenie na strane zhotoviteľa.
* Príprava a predkladanie špecifických výstupov riadenia projektu, ako sú status prezentácie, prezentácie pre riadiacu radu projektu, projektové plány, žurnál projektových rizík a incidentov (risk log/issue log)

### Požadované výstupy projektu

Nasledujúce kapitoly obsahujú minimálny zoznam dodaných projektových výstupov v jednotlivých častiach projektu.

#### Analýza

Najvyšším cieľom tejto fázy je získať jednotný názor medzi všetkými účastníkmi projektu na ciele životného cyklu projektu. Preto je táto fáza kritická pre každý projekt, pretože rozsah následných prác, high-level požiadavky a dôležité riziká musia byť pochopené, kým projekt môže pokračovať.

V tejto fáze projektu sú systematicky zbierané požiadavky biznisu na cieľové riešenie a je vykonaná ich prioritizácia. Vo fáze analýzy sú typicky vedené analytické interview s užívateľmi a workshopy, na ktorých sú zisťované, triedené (klasifikované) a overované jednotlivé požiadavky používateľov na dátový sklad, posudzovaná ich realizovateľnosť v IT prostredí NBS a vzťah k iným existujúcim systémom, pričom sú jednotlivé požiadavky popísané vo forme používateľských scenárov (use cases), tak aby jednoznačne reprezentovali a interpretovali ich požiadavky.

| Výstup | Formát | Popis |
| --- | --- | --- |
| Špecifikácia požiadaviek  (Requirements Traceability Matrix) | Dokument  (XLSX) | Tabuľka priorít projektu, ktorá sumárne popisuje dohodnutý rozsah spôsobom „Must“, „Should“, „Could“, „Won‘t“.  Tabuľka definuje priority a hlavné ciele. |
| Opis Architektúry  (Project Blueprint) | Dokument  (DOCX) | Dokument popisujúci architektúru a príslušné rozhrania v rámci implementácie riešenia, rovnako obsahujúci analýzu vplyvu implementácie projektu na NBS. Tento dokument môže byť predmetom aktualizácie v ďalších fázach projektu, ale iba v kontexte pridávania úrovne detailu. Dokument pokrýva detailné rozpracovanie nasledujúcich oblastí:   * Návrh prostredí - vývojového, testovacieho a produkčného prostredia vrátane konfiguračného manažmentu:   + Logická architektúra riešenia (RDBMS, integračná platforma, API, toky dát, riadenie dátovej kvality, metadáta).   + Fyzická architektúra riešenia,   + HW a SW architektúra,   + Topológia riešenia (ETL architektúra a dátové toky pre DataLake/DWH/BI) * Workflow management, monitoring a notifikácie * Požiadavky na dostupnosť (service level expectations). * High availability stratégia. * Back-up a recovery procesy. * Stratégia historizácie a archivácie dát. * Skupiny používateľov, prístup k bezpečnosti dát (matica a riadenie prístupových práv) |
| Analýza zdrojových systémov  (Interface Agreement) | Dokument  (XLSX) | Analýza zdrojových systémov definuje (popisuje) hlavný komunikačné cesty zo zdrojových systémov do DWH. Dokument taktiež popisuje základný subjektový model zdrojového systému.  Template pre IA dodá NBS OIT. |
| Dokumentácia k testovaniu | Dokument  (DOCX, XLSX) | Dokumentácia musí obsahovať nasledovné časti:  Prístup k testovaniu  Testovací plán  Testovacie scenáre  Akceptačné kritériá pre testovanie |

Tabuľka 26: Tabuľka výstupov fázy Analýza

#### Dizajn

Táto fáza do detailu rozpracováva výstupy predchádzajúcej projektovej fáze, t. j. do detailu popisuje funkčné a nefunkčné požiadavky na riešenie. Jej obsahom sú tieto oblasti (výstupy):

| Výstup | Formát | Popis |
| --- | --- | --- |
| *Návrh akvizície a konverzie dát*  (Interface Agreement) | Dokument  (XLSX) | Doplnenie dokumentov z fázy Analýzy o:   * Rozhrania zdrojových a cieľových systémov, pre každý systém špecifikácia rozhrania, typu (text file, db link, API, service access atď.), typu aktualizácie (full vs. prírastkový), frekvencie aktualizácie (on-line, batch, frekvencia) a obsahu. * Architektúra dátových tokov, vrátane zoznamu a detailného popisu dátových tokov a frekvencie, závislostí, detailný návrh workflow a job scheduling managementu, vrátane notifikácie užívateľov o prebehnutej aktualizácii dát. * Inicializačný load, vrátane popisu zdrojových systémov, rozhraní, požadovaných a dostupných entít, požadovanej a dostupnej histórie, očakávaných objemov a prístupov automatizácie. * Data refresh strategy – detailný popis postupov aktualizácie dát vo všetkých vrstvách DWH a pre všetky typy dát (číselníky, dimenzie, fakty, odvodené atribúty atd.). * Procesy dátovej kvality pre data acquisition – error handling, error notification, error recovery. |
| Funkčný detailný design – Logický model a mapovanie | Dokument  (DOCX;XLSX a EAP) | * Definuje logický dátový model poskytujúci detailnú definíciu biznis dátového modelu. * Obsahuje entity, atribúty, vzájomné vzťahy, domény, unikátne a alternatívne identifikátory. Atribúty a domény sú opísané formátmi, povolenými hodnotami alebo rozsahmi hodnôt. * Obsahuje identifikáciu citlivých a osobných údajov, taktiž definuje ochranu týchto údajov * Definuje hlavné predpoklady pre akvizíciu dát, mapovanie medzi zdrojovými dátami a cieľovými objektmi logickej databázy, mapovacie pravidlá a logiku, ktoré sú potrebné na vytvorenie konverzií a rozhraní v rámci riešenia DWH. Účelom dokumentu je poskytnúť potrebné informácie pre vývojový tím na vytváranie správnej extrakčnej (load) a transformačnej logiky |
| Security dokument | Dokument  (DOCX) | * Definuje na ktorom komponente sú aké osobné údaje a spôsob ich ochrany (pseudonymizácia, šifrovanie a pod). * Definuje metodiku tvorby prístupových práv, tvorby rolí a prístupov * Definuje metodiku celého životného cyklu dát:   + Administrácia osobných údajov   + Nastavenie dátovej retencie |
| Stratégia prechodu do produkcie | Dokument  (DOCX) | High-level opis krokov uvedenia predmetného riešenia do prevádzky. |
| ETL vývojové štandardy | Dokument  (DOCX) | Detaily o využívaní ETL nástroja pre projekt. Je kritickým elementom pre samotný ETL vývoj nakoľko definuje nevyhnutné štandardy. Dokument adresuje nižšie uvedené oblasti pre ODI (Oracle Data Integrator) v kontexte štandardov pre vývoj a prístup do ETL managera:   * Vývojové cykly * Používanie verzionovania * Definícia vykonateľného „jobu“ * Definícia štandardov a pravidiel * Menné konvencie * Používanie Workflow frameworku pre definíciu závislostí * Preukázateľné používanie revízie kódu s cieľom nájsť a odstrániť programové chyby. |
| Prístup k testovaniu riešenia | Dokument  (DOCX) | Dokument bude obsahovať:   * požiadavky na otestovanie riešenia, akceptačné kritériá * zároveň adresuje spôsob akým úspešný uchádzač interne otestuje predmetné riešenie a akú formu podpory dodá pri akceptačných testoch |
| Prístup k metadátam | Dokument  (DOCX) | * Katalóg metadát * Prístup k metadátam * Prístup k bezpečnosti |

Tabuľka 27: Tabuľka výstupov fázy Design

#### Implementácia

Cieľom implementačnej fázy je prevedenie riešenia od detailných modelov požiadaviek, cez konfiguráciu štandardných balíkov softvérovej funkcionality, vývoj a testovanie frameworkových komponentov a integrácia do formy systému, ktorý je pripravený na prvý „release“ na uvedenie do prevádzky. V skratke ide o kompletizáciu vývoja aplikácie, validáciu či všetky komponenty spolu „sedia“, a prípravu systému na akceptačné testy a nasadenie.

| Výstup | Formát | Popis |
| --- | --- | --- |
| Fyzický model, ETL komponenty, frameworky | SW  (export z EAP, ODI Metadata, súbory pre inštaláciu na aplikačné servere) | Implementácia systému podľa špecifikácií získaných v predchádzajúcich fázach. Budú konfigurované všetky reusable komponenty a vyvinuté a otestované komponenty definované pre Data Acquisition a Data Access. Logický dátový model bude základom pre vývoj a implementáciu fyzického dátového modelu. Na základe metadát bude vygenerovaný ETL kód, nastavené závislosti úloh a kalendár automatického spúšťania. |
| Komentované zdrojové kódy dodávane v rámci riešenia zhotoviteľom | Zdrojový kód podľa použitého programovacieho jazyka | Zdrojový kód obohatený o komentáre s vysvetlením obsahu programovej funkčnej časti (procedúry, funkcie) |
| Image pracovnej stanice vývojára | Súbor vo formáte príslušnej virtualizácie | Uložená HW a SW konfigurácia pracovnej stanice vývojára podľa špecifikácie dodanej zhotoviteľom |
| BI oblasť | SW | * Implementácia dátových modelov v analytickej vrstve MS Analysis Services (Tabulárne modely) podľa dohodnutých design princípov a konvencií. * Prototyping a implementácia analytických výstupov (reporty a grafy) v prostredí MS Power BI s publikáciu na MS Power BI Report Server. |
| Dokumentácia k testovaniu | Dokument  (DOCX, XLSX) | Dokumentácia musí obsahovať nasledovné časti:   * Detailný testovací plán * Testovacie scenáre * Žurnál z testovania * Akceptačné protokoly |
| Návrh iniciálneho nahratia dát | Dokument  (DOCX) | Popis ako sa projekt zachová k dátami v zdrojových systémoch, t. j.:   * Ktoré dáta budú nahrané do DWH ako celok * Ktoré dáta vyžadujú iný spôsob ETL pre denný beh a celkový nábeh * Návrh podmienok iniciálnej migrácie tak, aby nebola ohrozené prevádzka a výkon zdrojových systémov |
| Návrh infraštruktúry podpory produkcie | Dokument  (DOCX) | * definuje požiadavky na prevádzkovanie riešenia a adresuje ich z procesného hľadiska. Tento dokument je predmetom aktualizácie v ďalších fázach projektu a na jeho konci kompletne definuje proces podpory a prevádzky riešenia * návrh obnovy komponentov systému a údajov zo zálohy. |
| Školenie | Školenie a dokument  (PPTX, video vo formáte MP4) | Školenie vybraných používateľov NBS. Do 30 používateľov:   * Vykonanie jedného jednodňového školenia pre bežných užívateľov. * Vykonanie jedného jednodňového školenia pre analytikov a kľúčových užívateľov. * Vykonanie jedného jednodňového školenia pre operátorov.   Súčasťou školenia bude dodávka školiace materiálu vo forme PPT. |
| Používateľská a technická dokumentácia | Dokument  (DOCX) | Rozšírenie školiaceho materiálu o ďalšie informácie dôležité počas prevádzky a rozvoja systému. |

Tabuľka 28: Tabuľka výstupov fázy Implementácia

#### Nábeh do produkcie

Cieľom fázy nábehu do produkcie je uviesť ucelené riešenie od inštalácie na produkčný systém cez akceptačné testy po naštartovanie živej aplikácie, otvorenej a pripravenej na biznis. Validuje že riešenie je testované systematicky a je sprístupnené pre koncových používateľov.

Úlohy v tejto fáze:

* Uzatvorenie vývojovej verzie aplikácií v definovanom stave (code freeze).
* Príprava skriptov pre vytváranie / modifikáciu systémových, databázových a programových súčastí riešenia – tzv. inštalačný balíček
* Validácia technickej dokumentácie (predovšetkým dátového modelu).
* Migrácia testovacieho prostredia na produkčné prostredie a nasadenie aplikovateľných extraktorov na produkčné zdrojové systémy.
* Spustenie (exekúcia) skriptov pre vytváranie / modifikáciu systémových, databázových a programových súčastí riešenia na produkčnom prostredí.
* Súčasťou implementácie a akceptačných testov musí byť overenie obnovy komponentov systému a údajov zo zálohy.
* Akceptácia dokumentácie
* Skúšobná prevádzka

| Výstup | Formát | Popis |
| --- | --- | --- |
| Inštalácia na produkcii | SW | Spustenie inštalačných balíčkov |
| Vykonanie iniciálneho loadu | SW / FTE | Prvotná inicializácia DWH na základe dohodnutých pravidiel.  Inicializačný load bude musieť byť s najväčšou pravdepodobnosťou opakovaný niekoľkonásobne, z dôvodov jeho technického, funkčného a používateľského testovania. Požadujeme podporu celkovo pre 2 úspešné iniciálne nahratia dát.  Pred spustením produkčného iniciálneho loadu musí prebehnúť performance test, ktorý preukáže, že iniciálny load nebude trvať dlhšie ako 24 hodín. Pokiaľ sa preukáže, že iniciálny load nevie prebehnúť počas 24 hodinového okna, dodá zhotoviteľ plán rozdelenia iniciálneho nahrávania tak, aby sa využil už normálny inkrementálny beh. |
| Finálna akceptácia | UAT | Finálne testovanie na testovacom prostredí, na ktorom sa overí finálna kvalita a odsúhlasenie riešenia biznisovými používateľmi |
| Performance test | FTE | Vykonanie testu výkonnosti systému. V tejto úlohe požadujeme dodať protokol, kde bude zhodnotený beh ETL úloh z pohľadu komlexnosti ETL a počtu spracovaných dát a kvality samotného ETL scriptu. Požiadavka je aby beh ETL v NBS DWH bola ukončená do 08:00 |

Tabuľka 29: Tabuľka výstupov fázy Nábeh do produkcie

#### Produkcia

Cieľom fázy „produkcia“ je prevádzkovanie novo vyvinutého systému, vyhodnotenie úspešnosti nového systému, monitorovanie systémových problémov a ich adresovanie. Toto zahŕňa monitorovanie systému a vhodné zásahy aby bola udržaná prevádzková schopnosť, meranie výkonu systému, operačné a prevádzkové systémy, odpovedania na požiadavky o pomoc a asistenciu, na chybové hlásenia a funkčné požiadavky užívateľov; a riadenie aplikovateľného procesu kontroly zmien tak aby defekty a nové funkcionality boli prioritizované a pridelené do budúcich „releasov“ a bol zostavený plán budúcich rozšírení aplikácie rovnako ako určenie, vývoj a implementácia požadovaných aktualizácií.

|  |  |  |
| --- | --- | --- |
| Výstup | Formát | Popis |
| Onsite podpora | FTE | Požadujeme počas záverečnej fázy projektu aby úspešný uchádzač dodal podporu systému vo forme 1 onsite konzultanta po dobu 10 dní. Po uplynutí tejto periódy je kompetencia za riešenie presunutá na interný prevádzkový tím. |
| Bezpečnostné testovanie | FTE | Realizácia po akceptačnom testovaní (zabezpečí NBS). Potrebná súčinnosť pri realizácii bezpečnostného testovania. Oprava identifikovaných zraniteľností. |
| Overenie vykonateľnosti havarijného plánu | FTE | Realizácia po akceptačnom testovaní (zabezpečí NBS). Potrebná súčinnosť pri realizácii testu postupu obnoviteľnosti systému. |

Tabuľka 30: Tabuľka výstupov fázy Produkcia

#### Projektová dokumentácia

| Výstup | Formát | Popis |
| --- | --- | --- |
| Zadanie projektu | DOCX/PDF/XLSX/MPP | Projektová inicializačná dokumentácia |
| Plán projektu, revízie plánu projektu a etapové plány projektu, | DOCX/PDF/XLSX/MPP | Dokument popisujúci projektový plán v zmysle dohodnutej projektovej metodiky |
| Plán komunikácie projektu (organizácia, plán meetingov, spoločný projektový priestor, zápisy) | DOCX/PDF/XLSX/MPP | Dokument popisujúci organizáciu projektu a komunikáciu na projekte v zmysle dohodnutej projektovej metodiky |
| Štandardy pre zaistenie kvality realizácie projektu a zaistenie kvality dodávaného systému. | DOCX/PDF/XLSX/MPP | Dokument popisujúci prístup ku kvalite, spôsob jej merania a vyhodnocovania. |
| Plán školení používateľov, | DOCX/PDF/XLSX/MPP | Dokument popisujúci prístup ku školeniam, formu, obsah a časové rámce školení jednotlivých rolí používateľov. |
| Plán skúšobnej prevádzky | DOCX/PDF/XLSX/MPP |  |
| Správy o stave realizácie projektu | DOCX/PDF/XLSX/MPP | Dokument popisujúci aktuálny stav projektu dodávaný zhotoviteľom |
| Dokumentácia k ukončeniu projektu | DOCX/PDF/XLSX/MPP | Záverečná správa projektu v zmysle dohodnutej projektovej metodiky |
| Register rizík | DOCX/PDF/XLSX/MPP | Evidenčný zoznam projektových rizík |
| Register problémov | DOCX/PDF/XLSX/MPP | Evidenčný zoznam riešených problémov |
| Register zmien projektu | DOCX/PDF/XLSX/MPP | Evidenčný zoznam zmien na projekte |
| Register výstupov projektu | DOCX/PDF/XLSX/MPP | Zoznam odovzdávanej výstupnej dokumentácie pri akceptačných míľnikoch |

Tabuľka 31: Tabuľka projektovej dokumentácie

### Akceptačné kritériá

Chybovosť dodávky pred nasadením do PROD:

* 0 závažných nedostatkov z UAT
* 0 zásadných nedostatkov z UAT
* Max 10 nepodstatných nedostatkov z UAT

Exekúcia 100% odsúhlasených testovacích scenárov.

Priebežná príprava a predkladanie výstupnej dokumentácie

Akceptované všetky požadované výstupy projektu pre danú fázu podľa kapitoly 3.4.5.

Zhotoviteľ dodá mesačné výkazy prác kľúčových expertov definovaných v súťažných podkladoch.

Zhotoviteľ zabezpečí prítomnosť minimálne dvoch pracovníkov v priestoroch banky počas celého trvania projektu. (Maximálna kapacita priestorov objednávateľom poskytnutých pre takýto typ práce je 12 pracovníkov).

### Riziká projektu

V rámci verejného obstarania požadujeme vytvoriť register rizík projektu, ktoré budú mať vplyv na projekt v podobe nasledujúcej tabuľky:

|  |  |  |  |
| --- | --- | --- | --- |
| Názov a popis rizika | Kategória dopadu na projekt (H/M,/L) | Kategória možnosti výskytu rizika (H/M,/L) | Popis návrhu na zmiernenie dopadu |
|  |  |  |  |
|  |  |  |  |

Tabuľka 32: Riziká projektu

# Opis predmetu zákazky pre služby podpory prevádzky a rozvoja

## Úvod

Zabezpečenie prevádzky vychádza z predpokladu trvalej udržateľnosti celého riešenia NBS DWH po dobu 3 rokov od ukončenia projektu (akceptácie diela) a nasadenia do produkcie.

**Úroveň L1 podpory** prevádzky NBS DWH bude zabezpečená vlastnými personálnymi kapacitami NBS, ktorá zabezpečí nasledovné služby prevádzky:

1. technická podpora pracovných staníc používateľov,
2. riadenie oprávnení,
3. správu centrálneho systému riadenia identít
4. riešenie prípadných problémov komunikačnej infraštruktúry (databázy, sieť, aplikačné servere) na strane NBS.
5. riadenie nasadzovania aplikácií,
6. prevádzku centrálneho SeriveDesk-u
7. poskytovanie základného poradenstva pre koncových používateľov informačného systému.

**Úroveň L2 podpory** prevádzky NBS DWH bude zabezpečená interným DWH tímom OIT NBS.

**Úroveň L3 podpory** prevádzky NBS DWH zabezpečí vybraný externý poskytovať na základe Zmluvy o poskytovaní služieb podpory prevádzky a rozvoja NBS DWH, ktorý bude garantovať:

* dostupnosť, funkčnosť a technickú podporu aplikačných komponentov riešenia
* riešenie aplikačných incidentov
* realizáciu vyžiadaných úprav a zmenových konaní v informačnom systéme

a to podľa požiadaviek uvedených v nasledujúcich častiach opisu.

## Požiadavky na služby aplikačnej podpory prevádzky NBS DWH

Služby podpory prevádzky NBS DWH zahŕňajú predovšetkým:

1. Zabezpečenie bežnej servisnej podpory NBS DWH odbornými zamestnancami poskytovateľa pre riešenie hlásených incidentov, a to denne, v čase od 8:00 do 16:00 hodiny počas pracovných dní SR.
2. Odstraňovanie hlásených incidentov a odstraňovanie chýb v jednotlivých komponentoch NBS DWH, brániacich bezproblémovému používaniu informačného systému, s dodržaním požadovaných reakčných časov, resp. poskytnutie náhradného riešenia pri riešení incidentov.
3. Raz za štvrťrok vytvorenie reportu ako zhodnotenie stavu a priebehu aktualizácie dát v DWH

### Zoznam činností vykonávaných v rámci Služieb podpory prevádzky

Služby podpory prevádzky NBS DWH zahŕňajú nasledovné činnosti:

* **Podpora (Hotline)**
  + Poskytovanie poradenstva súvisiaceho s prevádzkou
  + Vykonávanie bežných operatívnych zásahov (bez zmeny funkcionality)
  + pravidelná údržba riešenia – monitorovanie prostredia, pravidelné aplikovanie publikovaných SW záplat (okrem databáz).
  + denný monitoring ETL procesov (vyhodnocovanie systémových logov priebehu aktualizácie DWH) a opravovanie prípadných chýb implementácie.
  + proaktívna kontrola (profylaktika) ETL procesu – jej výsledkom je odporúčanie aktivít pre krátky a stredný časový horizont na zabezpečenie úspešného a optimálneho chodu aktualizácie
  + reportovanie zamerané na spracovávanie požadovaných reportov o incidentoch a operatívnych informácií o ich riešení,
  + poskytnúť súčinnosť požadovanú NBS pri prenesení uzavretého Problému do Produkčného prostredia.
  + Proaktívne informovať NBS o chybách v DWH systémoch, ktoré zistil sám
* **Údržba (riadenie incidentov)**
  + vedenie evidencie nahlásených Incidentov v systéme na evidenciu problémov.
  + identifikácia problému, jeho analýza a samotné riešenie
  + poskytovanie konzultácii o incidentoch
  + projektové riadenie servisných činností
  + riešenie eskalácií
* **Školenia**
  + Zabezpečovanie odborných školení podľa požiadaviek NBS (opcia na ročný rozsah vzdelávania a hodinové sadzby)
* **Konzultačné služby**
  + poskytovanie rád (odborných telefonických, mailových a osobných konzultácií) administrátorom a používateľom DWH ekosystému (opcia na ročný rozsah vzdelávania a hodinové sadzby)
* **Implementácia (riadenie zmien)**
  + analýza a realizácia požiadaviek na zmeny zo stany NBS – popísané v kap. 1.3
* **Podpora SW tretích strán**
  + Sledovanie nových verzií SW tretích strán
  + Podpora pri nasadzovaní nových verzií SW tretích strán
  + Sledovanie záplat pre SW tretích strán
  + Podpora pri nasadzovaní záplat SW tretích strán
* **Doplnkové služby**
  + Exit služba – podpora pri prechode na nového poskytovateľa
  + Podpora nového poskytovateľa 6 mesiacov po zmene poskytovateľa

Služby podpory prevádzky NBS DWH nezahŕňajú nasledovné činnosti:

* Riešenie problémov súvisiacich s kvalitou dát, resp. spracovaním dát v zdrojových systémoch
* Práce súvisiace s administráciou prostredí (UNIX, databázy, aplikačný server, BI nástroje)
* Realizácia požiadaviek na zmenu – tie sú riešené podľa kapitoly 4.3

### Parametre kvality poskytovanej služby podpory prevádzky

Kvalita služieb podpory prevádzky DWH je definovaná nasledujúcimi ukazovateľmi:

* Dostupnosť podpory systému v rozsahu 8x5, t. j. denne, v čase od 8:00 do 16:00 hodiny počas pracovných dní SR.
* Reakčný čas odozvy na hlásený incident podľa zaradenia incidentu do kategórie. Zaradenie incidentu do príslušnej kategórie je na prvotnom posúdení verejného obstarávateľa. V prípade rozporu o zaradení incidentu budú obe strany rokovať o zaradení do príslušnej kategórie.

Detailné podmienky sú v návrhu zmluvy o prevádzke a údržbe systému (viď príloha).

## Požiadavky na služby rozvoja systému

### Požiadavky na realizáciu zmien na základe požiadaviek verejného obstarávateľa

Služby, ktoré sú požadované v rámci podpory riešenia:

* Realizácia zmien v riešení (formou požiadaviek na zmenu):
  + analýza požiadavky
  + vytvorenie technickej špecifikácie na riešenie požiadavky
  + implementácia nových ETL častí DWH riešenia
    - Vytvorenie programového kódu dohodnutého rozsahu
    - Nastavenie metadát príslušných frameworkov
    - Testovanie
    - Súčinnosť pri nasadení na produkciu
    - Údržba dokumentácie
  + implementácia dátových integrácií
  + úprava a dopĺňanie nových výstupných zostáv a reportov
* reinštalácie a rekonfigurácie modulov a softvérových komponentov DWH v prípade technologických zmien v prostredí NBS,
* verejným obstarávateľom vyžiadané školenia pre nové verzie frameworkov, ETL nastavení, číselníkov referenčných dát a iné
* verejným obstarávateľom vyžiadané administrátorské práce.

### Parametre kvality poskytovanej služby rozvoja systému

Detailné podmienky sú uvedené v Zmluve o prevádzke, údržbe a rozvoji systému v prílohe F.

### Servisné služby čiastkového plnenia

Verejný obstarávateľ požaduje poskytovanie týchto služieb od nasadenia jednotlivých fáz riešenia do produkcie. Služby budú poskytované do začiatku účinnosti Zmluvy o prevádzke, údržbe a rozvoji systému. Špecifikácia týchto služieb ja popísaná v prílohe G.

# Prílohy

## A – Zdrojové systémy NBS DWH



## B – DataMarty NBS DWH



## C – Metodika k predpisu NBS č.20/2020 o projektovom riadení



## D – RACI matica



## E – Projektový plán



## F - Zmluva o prevádzke, údržbe a rozvoji systému

– osobitný dokument

## G -Špecifikácia Servisných služieb čiastočného plnenia



1. Daniel Linstedt, Michael Olschimke (2015). Building a Scalable Data Warehouse with Data Vault 2.0. Morgan Kaufmann. ISBN: 978-0-12-802510-9 [↑](#footnote-ref-2)