OPZ – vyčistenie databázy

Proces zberu dát

1. Definovanie metadát zberu (v Oracle prostredníctvom modulu)
2. Publikovanie metadát do MongoDB. Metadáta na strane MongoDB nie sú menené, jedná sa iba o repliku podmnožiny dát z Oracle potrebnú pre zber (modul „Správa zberov“)
3. Externý alebo interný používateľ vytvorí/zmení výkaz na strane MongoDB. Udalosť o prijatí výkazu sa zapíše do Kafky
4. Replikačná aplikácia postupne spracováva udalosti z Kafky a zapracováva zmeny do ODS (Operačný dátový sklad)
5. Nočný job na strane Oracle zaktualizuje pohľady podľa dát operačného dátového skladu.

Spracovaním súboru Urgencia_ZR_NOR_ZN-2015_1z2.xml boli dotknuté moduly a funkcionality ISZI minimálne v rozsahu

- Vykazovanie - Inbox SJ
- Zberový modul
- Žurnál - logovanie zmien nad databázami
- ODS – operačný dátový sklad
- Dátový sklad
- Apache Kafka - výmena dát formou publikovania udalostí medzi biznisovými modulmi

Logovanie

Každý z modulov publikuje metriky, ktorými indikuje, či spracovanie z Kafky beží, prípadne či nedochádza k chybám počas spracovania. Nad týmito metrikami sú postavené alerty, ktoré vedia rádovo do 10 minút ohlásiť problém v konkrétnom module. Podmienka detekcie je, aby modul bežal a publikoval metriky.

Všeobecný alert KafkaConsumerStuck deteguje zaseknutie ľubovoľného spracovateľa, ak má lag väčší ako 1 a zároveň za hodinu neprišlo k posunu offsetu. Pokrýva tak aj situácie keby modul nepublikoval metriky, spracovanie by sa zaseklo z technických príčin, prípadne by modul nebežal. V určitých prípadoch by mohol falošne hlásiť problém, napr. ak sa zruší funkčnosť spracovania udalostí z modulu, ale v Kafke ostal konzument evidovaný. Pri zápise nových dát do pôvodne sledovaných topicov by hlásil lag. V tomto prípade je potrebné pri nasadení novej verzie modulu so zrušeným konzumentom aj zrušiť konzumenta z evidencie Kafky, aby falošne nehlásil problémy. Nevýhodou tohto alertu je, že čas na detekciu problému je väčší ako v predošlom prípade.

MongoDB – dokumentová databáza. Tu vstupujú dáta zo súboru Urgencia_ZR_NOR_ZN-2015_1z2.xml. Tu je potrebné začať analýzu. Identifikovať duplicity, odmazať, ideálne ošetriť používané procedúry a funkcie... „Úložisko výkazov“. Tiež úložisko metadát. Tie definujú štruktúru a vzhľad zbieraných dát.

Slúži ako úložisko dát pre nasledovné biznisové moduly:

- Inbox
- Zberový modul

- Upload

MongoDB je prevádzkované formou replica set s tromi aktívnymi inštanciami v režime vysokej dostupnosti.

Oracle DB

Z MongoDB sa do Oracle replikujú dáta pre nasledovné biznisové moduly:

- Národné administratívne registre
- Výstupy do ODS

Replikácia sa deje na základe udalostí v Kafke. Systém je z tohto pohľadu eventuálne konzistentný, stav v MongoDB a Oracle bude konzistentný, keď sú spracované všetky udalosti.

Obsahuje príslušné schémy:

- Hlavnú
- Pohľady (názov je vytvorený priložením textu „pohlady“ k názvu hlavnej schémy) a službou na výpočet pohľadov, tvorenou storovanými procedúrami v hlavnej databázovej schéme začínajúcimi prefixom POHLAD_.

Zároveň slúži ako úložisko alebo zdroj dát nasledovných biznisových modulov a tiež zabezpečuje vybranú funkčnosť pre nasledovné biznisové moduly:

- ODS (Operačný dátový sklad)
- Dátový sklad
- Automatizované generovanie multidimenzionálnych kociek z metadát ISZI
- Dátové pumpy nad dátami z ISZI

Oracle REST fasáda

Poskytuje REST rozhrania pre automatizované vykonávanie zmien z Oracle pre nasledovné biznis moduly:

- Inbox
- Zberový modul

AKKA cluster

Komunikáciu medzi Frontend a Backend zabezpečuje toolkit AKKA prostredníctvom tcp protokolu. Komunikácia s MongoDB prebieha prostredníctvom MongoDB Wire protocol.

ISZI biznis moduly komunikujú prostredníctvom microservices

ODS (Operačný dátový sklad)

Vyzbierané údaje po replikácii z MongoDB do Oracle DB. Relačný obraz vstupných údajov prvotne prijatých zberovým modulom (ODS_ tabuľky + procedúry, ktoré ich napĺňajú).

Prístup cez Oracle SQL Developer. Bude potrebné skontrolovať / modifikovať na základe analýzy nad MongoDB minimálne

- dátový model ODS
- príslušné DB views
- relačné tabuľky ODS

Dátový sklad (DS)

Agregované a detailné údaje (výstupné pohľady) vypočítané z údajov výkazov

Prístup cez Oracle SQL Developer. Bude potrebné skontrolovať / modifikovať na základe analýzy nad ODS minimálne

- dátový model DS
- príslušné DB views

Po vyčistení vyššie spomenutých častí bude potrebné spustiť automatizované pregenerovanie multidimenzionálnych kociek z metadát ISZI (storované procedúry v Oracle DB). Základné informácie k problematike:

Star schémy - dimenzionálne a faktové tabuľky výstupných pohľadov (D_ pohľady a F_ pohľady)

Generovanie metadát pre vybraný typ záznamu výkazu, pre každý typ sú vygenerované metadáta základnej faktovej tabuľky a pre tabuľky za každý otvorený modul.

Metadáta zahŕňajú aj asociácie na zdieľané dimenzné tabuľky a vzorce pre transformáciu použitých biznis kódov na cudzie kľúče dimenzných tabuliek

Pre detailnú analýzu a návrh riešenia bude potrebné mať aplikačné prístupy minimálne v rozsahu:

Používateľská rola

SpravcaZberu - Správca zberu (ukončovanie zberu, povoľovanie dodatočných opráv zberu po ukončení)

AdminISZI - Administrátor ISZI (technická správa systému)

SpravcaZmien - Správca zmien

Admin

Apache Kafka

Slúži na orchestráciu systémových procesov. Integrácia modulov ISZI je realizovaná cez log udalostí v Kafke. Zmenové udalosti sú publikované do príslušných topicov v Kafke, ktoré môžu konzumovať ostatné moduly a na základe udalostí modifikovať svoje dáta.

Kafka je prevádzkovaná v clustrovom režime s tromi inštanciami. Každý topic je replikovaný na dva servery, čím je tolerovaný výpadok jedného servera. Prevádzka formou clustra umožňuje vykonávať bezodstávkové aktualizácie Kafky.

Vybrané topicy môžu mať aj viacero partícií, ktoré sú rozdistribuované cez viacero serverov. Dáta partícií sú spracovávané na báze round robin. V prípade zahltenia jednej partície nie sú ovplyvnené ostatné partície a dáta z partícií sa spracovávajú rovnomerne.

Zabezpečuje výmenu dát formou publikovania udalostí medzi nasledovnými biznisovými modulmi:

- Inbox
- Zberový modul
- Upload
- Správa zberov
- Výstupy do ODS (Operačný dátový sklad)

Zookeeper - Uchováva konfiguráciu Kafky

Kafka REST - Umožňuje Oracle publikovať udalosti do Kafky pre iné moduly

MS Power BI

Zabezpečuje prezentáciu údajov pre nasledovné biznisový modul:

- Reporting