

Data Lineage

White paper

SADRŽAJ

O nama	3
Sažetak	Error! Bookmark not defined.
Opis rješenja	4
Ključne funkcionalnosti	5
Način rada rješenja	6
Data lineage vizualizacija	7
Integracije	10
Cost-benefit analiza	4
Data Lineage prednosti	11



O nama

Agilos IT tim čine visoko kvalificirani i marljivi stručnjaci koji svojim iskustvom te stečenim znanjem dolaze do kvalitetnih rješenja. Naši DWH konzultanti već dugi niz godina surađuju s bankama kako bi pomogli postići izvrsnost u DWH arhitekturi, automatizaciji i optimizaciji. Naši stručnjaci usko surađuju s klijentima kako bi pružili najprikladnije rješenje koje zadovoljava njihove potrebe. Uska specijalnost nam je regulatorni reporting, visoka kvaliteta izvršenog posla te dizanje normi rada u svakom pogledu. Primarni cilj naših konzultanata je razvoj poslovnih rješenja za poslovnu inteligenciju (DWH, ETL, BI). Razumijevanjem podatkovnog modela te naprednim korištenjem SQL upita i PL/SQL-a dolazimo do razvoja DWH rješenja.



Uvod

Data Lineage alat analizira SQL strukture podataka i/ili procedure pisane SQL jezikom, parsira ih do najniže razine te od istih gradi objektni model svih metadata elemenata. Ovaj objektni model se kasnije može eksportirati u bilo koji oblik koji je korisniku potreban, prenijeti u drugi alat, ili se iz njega mogu kreirati vizualizacije povezanosti podataka, data lineage, odnosno pregled tijeka podataka ili pak skripte za ETL bez korištenja klasičnih ETL alata.

Alat za vizualizaciju data lineage-a je namijenjen analizi veza i transformacija nad podacima, kako bi se od strane poslovnih i tehničkih korisnika jasno utvrdio tijek i povezanost podataka, njihova međuovisnost i utjecaj na naknadne izmjene istih.

Cost-benefit analiza

Ovisno o primjeni i učestalosti uporabe Data Lineage alata na poslovne procese on potencijalno može znatno optimizirati razvojni i analitički workflow. Prosječno utrošeno vrijeme na analiziranje toka podataka i utjecaja zahvata nad podatkovnim procesima iznosi 50% ukupnog vremena developera ili analitičara. Optimizacijom tog procesa korištenjem Data Lineage alata vjerujemo da tu brojku možemo spustiti na 10% utrošenog radnog vremena. Ako uzmemo u obzir prosječnu cijenu radnog dana jednog developera (oko 300€) na razini mjeseca moguće je uštedjeti oko 19,000€ po svakom zaposleniku.

Također, oslobođeno vrijeme se onda može utrošiti na druge radne prioritete te time dodatno povećati isplativost i efikasnost rada. Reduciranje vremena utrošenog na repetitivne i zamorne poslove poput analize toka podataka i predikcije utjecaja izmjena u podatkovnim procesima ostavlja više prostora za efikasniju upotrebu ljudskih resursa. Korištenjem ovog alata može se ostvariti osjetna logistička, vremenska i financijska dobit.

Ključne funkcionalnosti

Ključne funkcionalnosti sustava su:

- Automatsko prikupljanje metapodataka u Data Lineage repozitorij (dnevno)
- Detaljni prikaz odnosa sistema i procesa u njemu
- Detaljni prikaz ETL-a (OWB/ODI)
- Prikaz vlasnika sistema, procedura, tablica
- Linkanje svih vezanih dokumenta i zahtjeva
- Analiza „idealnog redoslijeda“ izvršavanja procesa (optimizacija loada do 50%)
- Generiranje funkcionalne specifikacije za novi change request
 - Na osnovu odabira gdje i od kuda se puni nova kolona generira se dokument s uputom za developera koje sisteme, procese ,tablice i kolone mora mijenjati
- Generiranje story pointa za novi change request
- Sustav je prilagođen za business usere, business analyste te developere

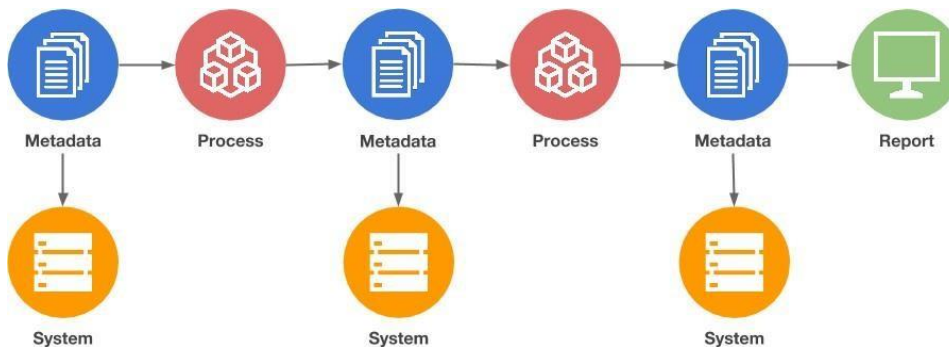
Glavna svrha sustava je:

- Imati jedno mjesto istine o svim meta podacima
- Olakšati proces kreiranja FSD-a
- Imati na jednom mjestu DG informacije

Implementacija

Opis rješenja

Data Lineage je namijenjen generiranju metadata modela za složene arhitekture SQL baza podataka. Metadata se može generirati na temelju struktura podataka i njihovih poveznica specificiranim u poveznim ključevima, no prava snaga alata dolazi do izražaja u generiranju modela temeljenim na složenim SQL upitima, gdje se parsiranjem pojedinog upita do najniže razine gradi objektni metadata model za svaki pojedini upit te njihova povezanost, što je temelj za uspostavu data lineage praćenja životnog ciklusa podataka. Praćenje data lineage-a, kao jednog od osnovnih data governance poslova, podrazumijeva uspostavljanje procesa razumijevanja tijekom povezanosti podataka, procesa koji ih povezuju i transformiraju te na kraju konzumiraju. Proces uspostave praćenja temeljen je na metadata modelima podataka i njihovom povezanošću.

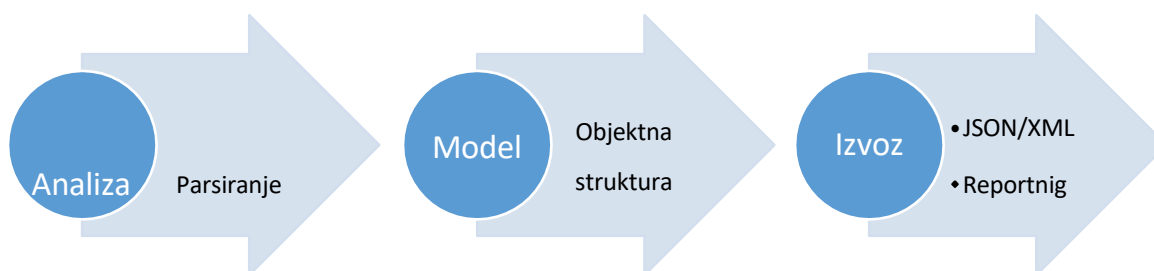


Slika 1 - Data Lineage proces

Upravljanje i razumijevanje metadata struktura je temelj izgradnje sustava za praćenje i vizualizaciju data lineage-a.

Način rada rješenja

Data Lineage omogućuje definiranje izvora strukture podataka na temelju kojeg se gradi metadata model. Iako se kao izvor može definirati postojeća struktura podatkovnih tablica u bazi (za njih je metadata model moguće kreirati i jednostavnim alatima), prava snaga alata je vidljiva kada se kao izvor postave složene procedure ili SQL skripte koje povezuju više izvora podataka s mnogostrukim kriterijima dohvata.



Prvi korak je parsiranje SQL upita koji počinje razdvajanjem na osnovne dijelove (najčešće SELECT, FROM, WHERE itd.), zatim se svaki od tih dijelova parsira do najniže razine svake pojedine tablice, polja i uvjeta.

Ovakav model izgradnje metadata modela se ujedno smatra i najnaprednijim, obzirom da zahtijeva potpuno razumijevanje i najsloženijih struktura SQL koda.

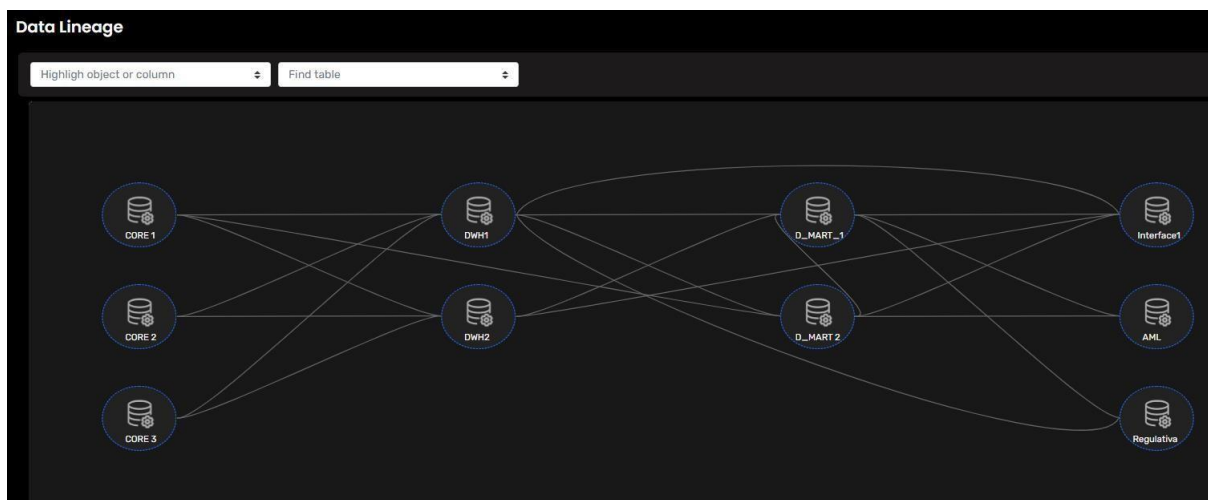
Iz parsiranih podataka aplikacija gradi hijerarhijski objektni model sa svim strukturama i njihovim metadata deskriptorima te njihovim poveznicama.

Finalni korak u implementaciji rješenja je definiranje željene strukture, formata i načina izvoza ovih metadata struktura. Postoji više mogućnosti, od jednostavnih JSON ili XML datoteka i njihovog slanja u druge aplikacije pa do kreiranja više vrsta izvještaja te data lineage vizualizacija.

Data lineage vizualizacija

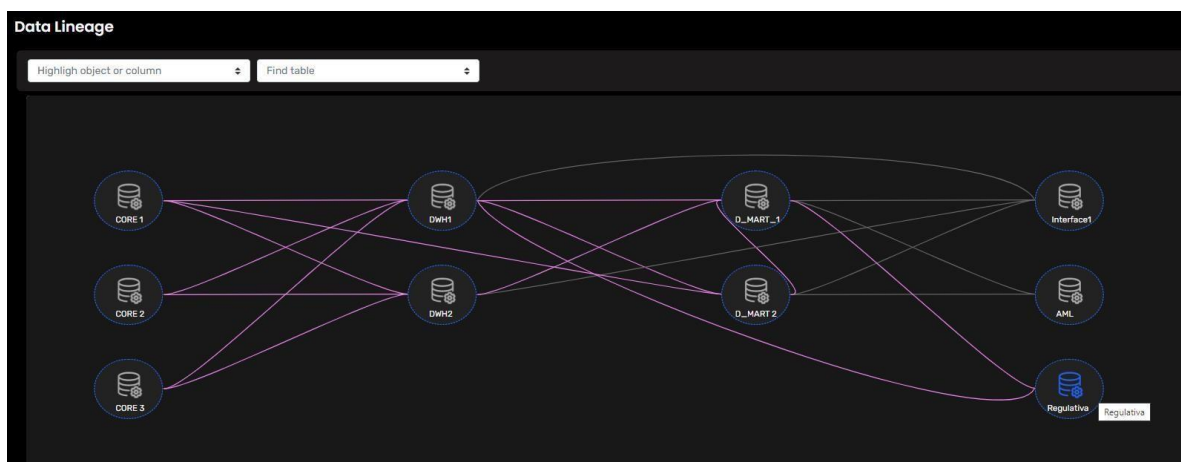
Data Lineage visualisation tool omogućuje detaljnu i preglednu vizualizaciju svih procesa i potprocesu u ETL sustavu. Kompletan sustav je sistematiziran u sisteme i procese koji odgovaraju strukturi podataka i arhitekturi podatkovnih procesa na kojima se primjenjuje Data Lineage.

Razvojna, produkcijska i sva ostala okruženja zajedno s pripadajućim međuosobnostima su prikazana u obliku povezanih sistema.



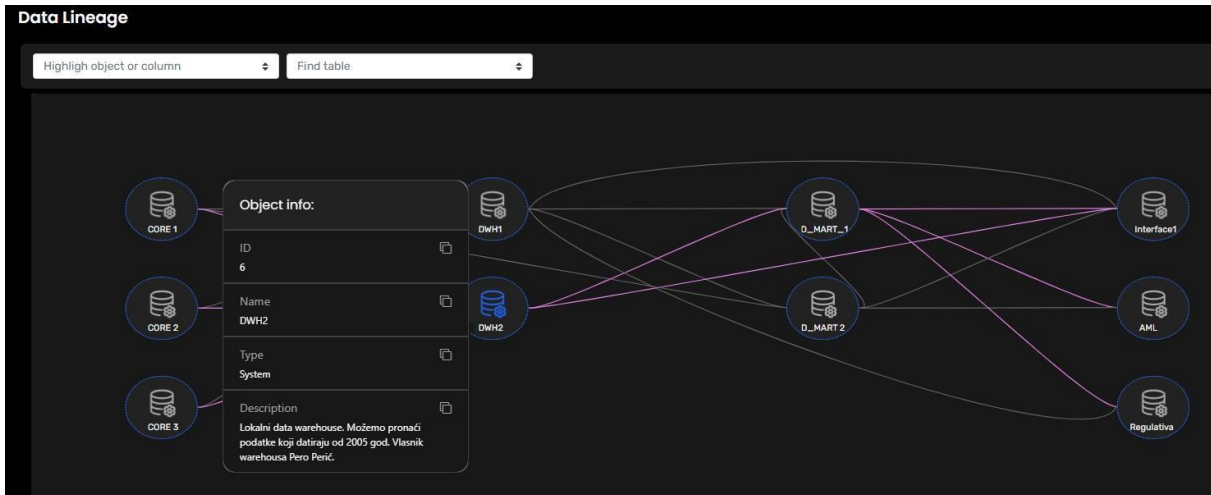
Slika 2 - Prikaz lineaga između sistema

Odabirom jednog od sistema vidimo označen tok podataka iz njemu ovisnih sistema. Jednim klikom dobijemo informaciju o izvorima relevantnim za tok podataka odabranog sistema.



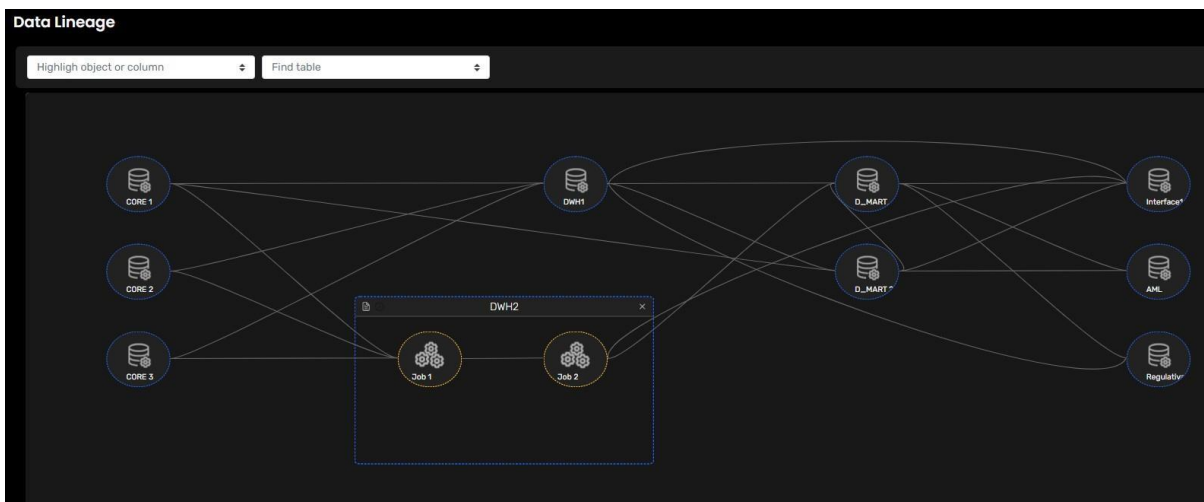
Slika 3 - Prikaz lineagea između sistema, oznčeni ovisni sistemi na sistem Reporting

Također, odabirom sistema možemo dobiti uvid u njegove detalje i specifičnosti poput vlasnika, kratkog opisa ili dokumenata vezanih na sistem.



Slika 4 - Prikaz lineaga između sistema, prikaz detaljnih informacija o sistemu.

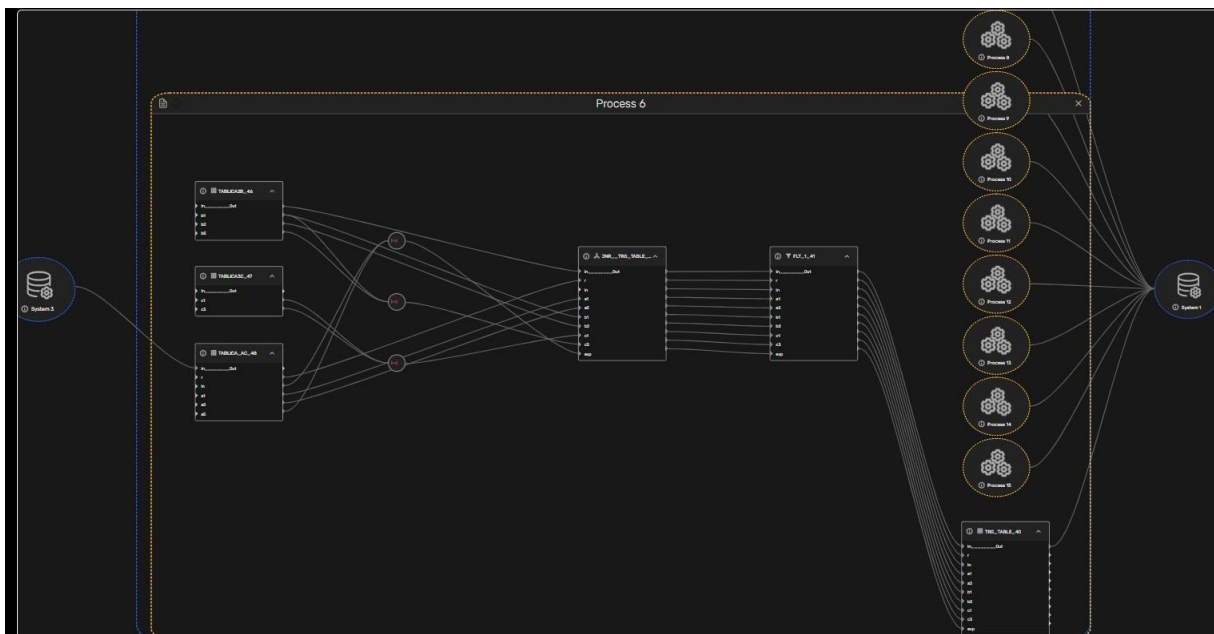
Klikom na sistem generiramo prikaz procesa koji se nalaze ispod njega. Procesi su sastavni dio izvještajnog okruženja i najčešće nam je potreban prikaz njihovog toka podataka. U praksi procesi Data Lineagea predstavljaju obrade kreirane u ETL alatima za skladišta podataka.



Slika 5 - Prikaz lineaga između sistema, drilldown iz sistema u procese sistema

Daljnijim ulaskom u strukturu procesa generiramo detaljan prikaz ETL logike od koje se proces sastoji. Sličnim prikazom kao i kod sistema možemo vidjeti tok podataka između njihovih izvorišnih i odredišnih struktura. Svaki korak toka je grafički prikazan na jednostavan i intuitivan način pomoću

kojeg u vrlo malo vremena dobijemo kompletan uvid u sve korake i operacije koje podatak prođe, poput filera, joinera, expressiona i opisa kolona i tablica.



Slika 6 i 7 - Prikaz detaljnog lineaga ETL-a (ODI/OWB).

Integracije

Alat Data Lineage je razvijen prema standardima svih suvremenih podatkovnih programskih jezika i okruženja. Nudi potpunu kompatibilnost i jednostavnu integraciju sa svim aktualnim tehnologijama koje se koriste u obradi podataka u poslovnoj primjeni. Posebno ističemo sljedeće tehnologije:

- PL/SQL
- OWB
- ODI
- SQL
- Postgre
- T-SQL

Zaključak

Ušteda vremena i resursa uz drastično olakšanje radnih zadataka vezanih uz analizu podatkovnih tokova, generiranje dokumentacije i procjenu utjecaja izmjena u sustavu su srž prednosti koje Data Lineage alat nudi korisnicima. Kroz intuitivnu implementaciju i pregledno korisničko sučelje vjerujemo da je moguće uštediti do 40% utrošenog vremena što je značajna dobit za svaku organizaciju koja se bavi analizom podataka.

Data Lineage prednosti



Ušteda troškova

Ušteda troškova, maksimalna efikasnost



Bolja distribucija resursa

Projekti mogu biti završeni s manje developera i testera



Osigurana kvaliteta

Kvalitetni izvještaji i analize



Brži razvoj

Ubrzan proces podatkovne analitike i generiranja analiza, izvještaja i dokumentacije