

DataQuality

White paper

SADRŽAJ

O nama	3
Uvod	4
Cost-benefit analiza	5
DataQuality – karakteristike	6
Implementacija	11
Zaključak	12
Data Quality prednosti	13



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 rješenja 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

Donošenje poslovnih odluka temelji se na poslovnoj inteligenciji. Poslovna inteligencija je skup metodologija, koncepata i sustava koji podatke poduzeća pretvaraju u znanje poduzeća. Na temelju informacija i znanja donosi se pravovremene poslovne odluke, zato je bitno da su podaci što kvalitetniji, odnosno točniji. Pripremu podataka odrađuje ETL („Extract-Transform-Load“) proces. Sustav dohvaća podatke iz izvorišnih sustava, uvodi kvalitetne podatke, prilagođava podatke tako da se odvojeni izvori podataka mogu koristiti zajedno, i na kraju dostavlja iste u format spreman za prezentaciju tako da korisnici mogu graditi aplikaciju i donositi odluke. Potreba za kvalitetnim ETL sustavom je velika jer bez kvalitetnih podataka izrađeno skladište podataka neće imati dobre rezultate. „DataQuality“ je modul aplikacije „Agile Query“ i dio transformacijskog procesa. Aplikacija nudi mogućnost da korisnik brzo i jednostavno prepozna podatke koji su netočni, nepotpuni ili nepostojeći kako bi ih mogao ispraviti te da bi onda ti podaci mogli stići na konačno odredište potpuno spremni za upotrebu.

Cost-benefit analiza

Loši podaci mogu imati značajne poslovne posljedice za tvrtke. Primjeri ekonomske štete koju problemi s kvalitetom podataka mogu prouzročiti uključuju:

- izgubljene prodajne prilike zbog pogrešnih ili nepotpunih zapisa o kupcima
- novčane kazne za nepravilno izvješćivanje o financijskom ili regulatornom skladu itd.

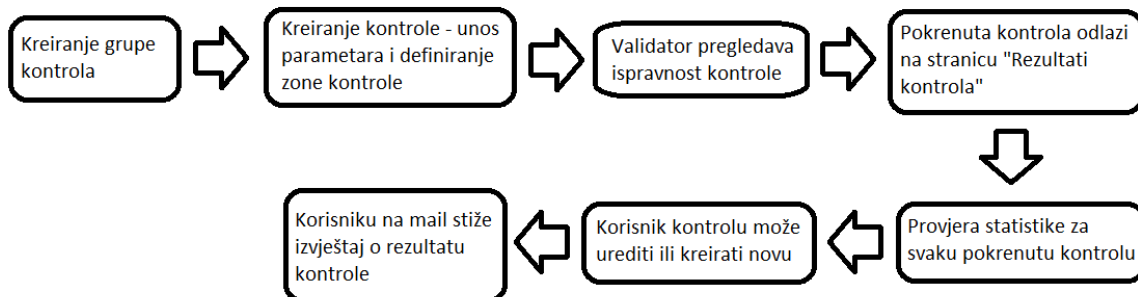
Konzultantska tvrtka Gartner izjavila je 2021. da loša kvaliteta podataka košta organizacije u prosjeku 12,9 milijuna dolara godišnje dok je npr. konzultant za kvalitetu podataka Thomas Redman procijenio da ispravljanje pogrešaka u podacima i rješavanje poslovnih problema uzrokovanih lošim podacima košta tvrtke u prosjeku 15% do 25% njihovih godišnjih prihoda.

Tu još ulazi i cijena developera koji mora podatke pratiti i prepravljati kako bi bili kvalitetni. Analiza greška je 30% popravka, što znači da korištenje ove aplikacije štedi vrijeme developeru.

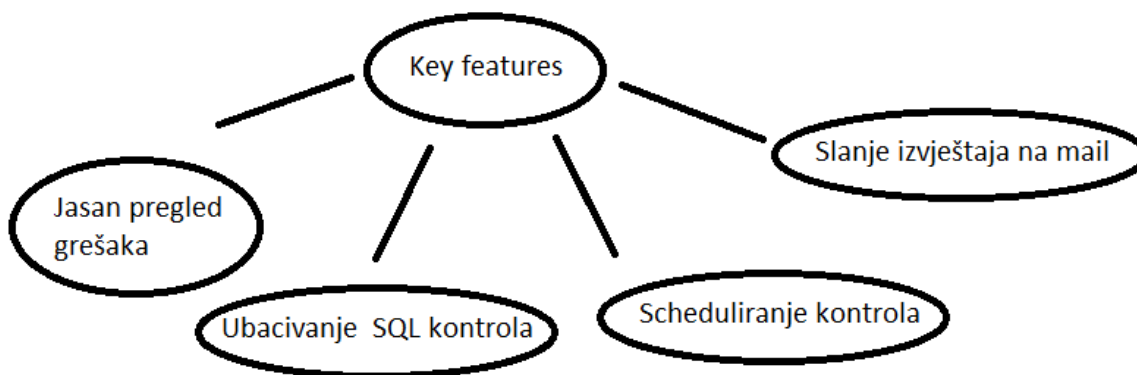
Nekvalitetni podaci mogu tvrtku koštati ne samo novčano, već joj pokvariti i reputaciju.

Zbog toga je učinkovita strategija upravljanja kvalitetom podataka neophodna.

DataQuality – karakteristike



Slika 1 – Dijagram postupka



Slika 2 – Key features DataQuality aplikacije

Prednosti korištenja ove aplikacije ujedno su i njene glavne karakteristike. Korisniku je omogućeno slanje izvještaja na e-mail adresu, zakazivanje („scheduliranje“) kontrola, ubacivanje SQL kontrola i jasna preglednost grešaka. Kako aplikacija radi ukratko je prikazano na dijagramu (slika 1).

Na slici 3 prikazano je kako početna stranica Data Quality aplikacije izgleda. Crvenom strelicom označen je navigacijski izbornik koji sadrži:

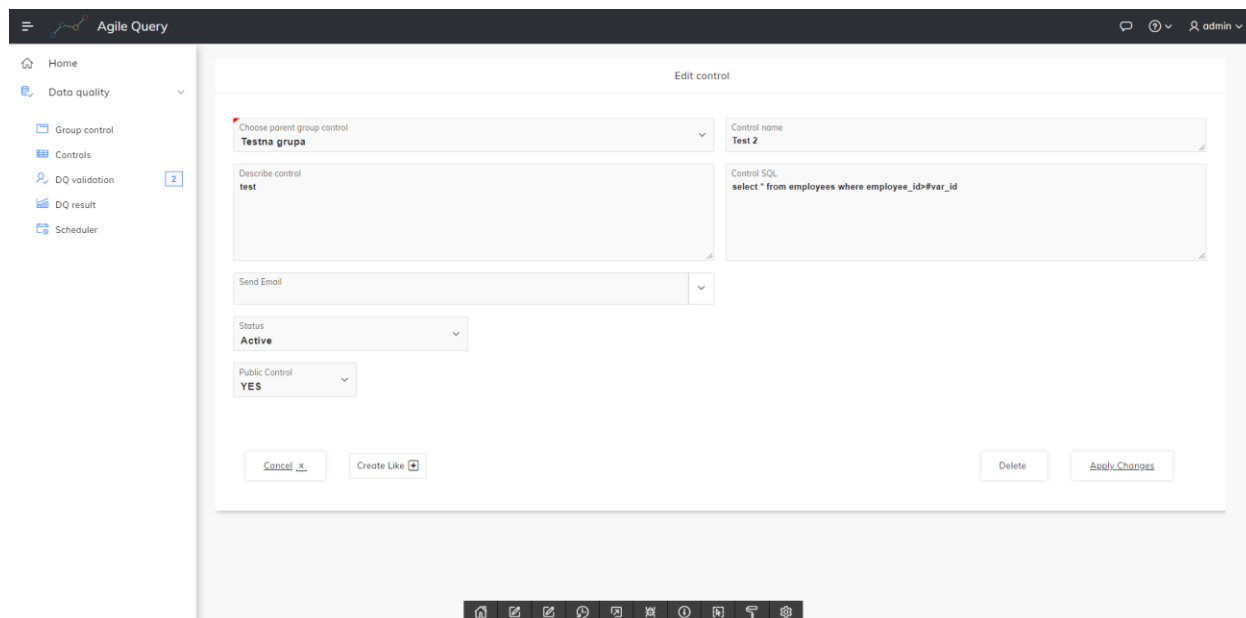
- Grupe kontrola – definirane grupe koje sadrže kontrole jednog procesa
- Kontrole – popis kreiranih kontrola – testiranje podataka
- Rezultate kontrola – rezultati izvršenih kontrola
- Validacije – validacija ispravno kreiranih upita
- Raspored izvršavanja – generiranje vremenskog perioda izvršavanja kontrola.

Narančasta i plava strelica vode redom na grupne kontrole i kontrole (do kojih se može doći i putem navigacijskog izbornika), a zelene strelice pokazuju razna izvješća (grafove) koja se odnose na posljednje izvršene kontrole.



Slika 3 – DataQuality početna stranica

Za početak u nekoliko jednostavnih koraka korisnik može kreirati grupne kontrole (prikazano na slici 4), a zatim slijedi kreiranje same kontrole – korisnik odabire grupu kojoj kontrola pripada, naziv, upisuje SQL upit, mail primatelja te želi li da bude public ili private. Korisnik može primatelje maila odabarti s drop-down liste ili može ručno unijeti informacije.



Slika 4 – Kreiranje grupne kontrole

Klikom na gumb threshold otvara se modalni prozor (prikazan na slici 5) gdje je moguće unijeti broj očekivanih grešaka, odnosno korisnik može definirati zone kontrole kako bi provjerilo je li izvršena s:

1. uspješnom
2. upozoravajućom
3. negativnom ocjenom.

Ako korisnik ne unese proizvoljne vrijednosti, postavljaju se defaultne vrijednosti koje su: *Success – 0, Warning – 5, Error – 10.*

Enter detection parameters

SUCCESS ZONE

Please enter a specific number for control detection evaluation as grade "Success"

WARNING ZONE

Please enter specific number for control detection evaluation as grade "Warning".

ERROR ZONE

Please enter a specific number for control detection evaluation as grade "Error".

Slika 5 – Definiranje zone kontrole

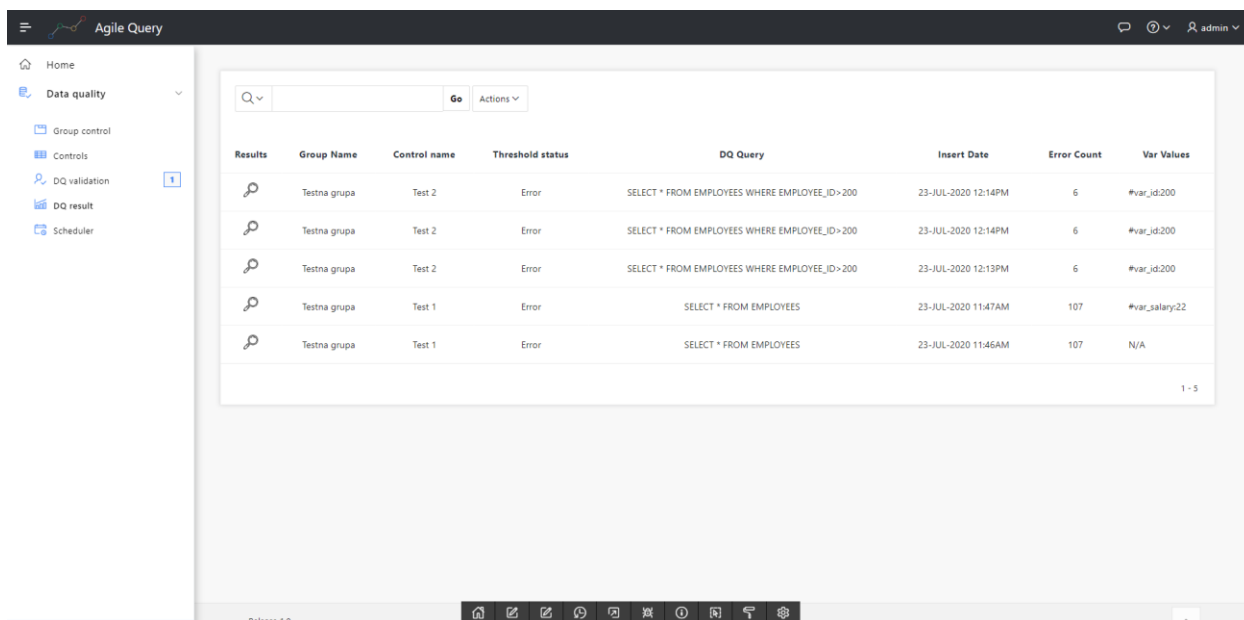
Pritiskom na „Save“ kontrola se sprema i šalje validatoru na pregled. Validator provjerava ispravnost kontrole, a može ju i pokrenuti ili schedulirati.

U trenutku kada određena kontrola bude provjerena i validirana od strane validatora, ona je vidljiva na stranici gdje se nalaze kontrole, i omogućeno je pokretanje za istu.

Nakon što je kontrola pokrenuta, odlazi se na stranicu „Rezultati kontrola“ gdje se mogu provjeriti svi podaci relevantni za kontrolu → koji je datum pokretanja, koliki je broj pogrešaka, sql upit...

Applikacija ima i opciju kreiranja kontrola s varijablom gdje korisnik treba onda definirati i sql s varijablom.

Statistiku svake pokrenute kontrole korisnik može pronaći na stranici „Kontrole-preview“ kontrole. Otvara se stranica s popisom svih okinutih kontrola te informacijama za svako pojedino okidanje. Pritiskom na gumb chart prikazuje se grafičko sučelje s informacijama za tu kontrolu.



The screenshot shows the 'Agile Query' application interface. On the left is a navigation menu with options: Home, Data quality (selected), Group control, Controls, DQ validation, DQ result, and Scheduler. The main area displays a table with the following data:

Results	Group Name	Control name	Threshold status	DQ Query	Insert Date	Error Count	Var Values
	Testna grupa	Test 2	Error	SELECT * FROM EMPLOYEES WHERE EMPLOYEE_ID>200	23-JUL-2020 12:14PM	6	#var_id:200
	Testna grupa	Test 2	Error	SELECT * FROM EMPLOYEES WHERE EMPLOYEE_ID>200	23-JUL-2020 12:14PM	6	#var_id:200
	Testna grupa	Test 2	Error	SELECT * FROM EMPLOYEES WHERE EMPLOYEE_ID>200	23-JUL-2020 12:13PM	6	#var_id:200
	Testna grupa	Test 1	Error	SELECT * FROM EMPLOYEES	23-JUL-2020 11:47AM	107	#var_salary:22
	Testna grupa	Test 1	Error	SELECT * FROM EMPLOYEES	23-JUL-2020 11:46AM	107	N/A

At the bottom of the table, there is a pagination indicator '1 - 5'. The application footer shows 'Release 1.0' and a set of standard navigation icons.

Slika 6 - Stranica rezultata, pokazuje korisniku status pokrenutih kontrola.

Nakon završetka pokretanja određene kontrole, ranije definirani primatelji izvješća, na svoje email račune će dobiti generirano izvješće.

Izvještaj

Report name	Date	Number of errors	Check errors	SQL Query
Report_491	3/3/2020	8	Check errors	Details (SQL desc)

Slika 7 – Izvješće koje korisnik dobiva na mail

- Klikom na link Check errors, dolazite na stranicu pregleda rezultata iste kontrole.
- Klikom na link Details(SQL desc), imate priliku pogledati SQL izraz iste kontrole.

Implementacija

DataQuality se može koristiti kao lokalna aplikacija. Da bi se mogla početi koristiti, trebali bi biti zadovoljeni preduvjeti instalacije Oracle infrastrukture (baza, apex). Ako su preduvjeti zadovoljeni, odmah se može krenuti na idući korak, ako nisu, instaliranje tih stavki potrajalo bi 3 dana. Nakon toga može se instalirati aplikacija i nad njom se mogu izvršiti neke početne postavke što bi potrajalo otprilike 1 radni dan. Nakon instalacije alata slijedi implementacija kontrola, čije trajanje ovisi o potrebama korisnika i željenom broju kontrola. Za provedbu edukacije bi bio dovoljan 1 dan.

U svrhu edukacije dostupni su i video materijali koji detaljno prolaze kroz postupak korištenja aplikacije.



Zaključak

Održavanje kvalitete podataka na visokoj razini omogućuje organizacijama smanjenje troškova identificiranja i popravljavanja loših podataka u njihovim sustavima. Na taj način izbjegavaju kazne ili druge dodatne troškove, štede na developerima koji moraju brinuti o kvaliteti podataka, a zadržavaju i dobru reputaciju među svojim klijentima.

DataQuality aplikacija bi u tom smislu uvelike olakšala posao korisnicima, a sa sobom nosi i financijske benefite kao što smo vidjeli na primjeru konzultantske tvrtke Gartner.

Brzo detektira probleme u podacima, a pregledni izvještaj korisnik dobiva direktno na e-mail adresu pa nakon toga analitičar podataka može razmotriti alternative za uklanjanje uzroka problema, uvođenje preventivnih tehnika i/ili poduzimanje nekih korektivnih radnji.

Osim toga, ovaj alat daje timovima za upravljanje podacima više vremena kako bi se, umjesto na čišćenje podataka, mogli usredotočiti na potencijalno produktivnije zadatke.

Data Quality prednosti



Ušteda troškova

Minimalni resursi, maksimalna produktivnost



Osigurana kvaliteta

Kvalitetni podaci



Brži razvoj

Ubrzan proces čišćenja podataka - brzo spremni za upotrebu