

METODOLOGIJA MIGRACIJE PODATAKA VELIKIH INFORMACIONIH SISTEMA LARGE INFORMATION SYSTEMS DATA MIGRATION METHODOLOGY

Sava Jelisavčić, Ivan Jelisavčić, *Elektroprivreda Srbije, Termoelektrane "Nikola Tesla"*
Vladisav Jelisavčić

Sadržaj – *Namena metodologije za migraciju podataka je da pruži opšti okvir sa unapred definisanim i upravljivim aktivnostima koje rezultiraju uspešnim i efikasnim procesom migracije podataka iz jednog informacionog sistema u drugi bez obzira na njihovu unutrašnju organizaciju ili platformu na kojoj su postavljeni. Ovaj rad predlaže jedno takvo rešenje.*

Abstract – *The purpose for existence of methodology for data migration is to propose a general framework of predefined and at all times controllable activities that result in efficient and successful process of migration of data from one information system to another regardless of these systems' internal organization or platform. This paper proposes one such solution.*

1. UVOD

Migracija podataka je složen proces, sastavljen od više aktivnosti, kojim se iz starog informacionog sistema podaci prenose u novi sistem, pri čemu se na podacima vrše sve potrebne transformacije prema zahtevima novog sistema.

Migraciju često komplikuje činjenica da se podaci nalaze na više lokacija ili u više različitih vrsta izvora. Izvori podataka ne moraju obavezno biti organizovani u zapise jedne baze podataka, već se mogu nalaziti u više različitih baza različitih proizvođača ili u obliku tabela ili čak tekstualnih fajlova. U životnom veku starog sistema uglavnom je bilo više dogradnji i primene različitih informacionih tehnologija što za posledicu ima stanje u kome podaci nisu organizovani na jedinstveno uređen način.

Proces prenosa nije jednostavan ni u slučajevima kada se podaci prenose iz samo jedne baze podataka starog sistema u bazu novog sistema zbog toga što je dizajn baze novog informacionog sistema uvek bitno različit od dizajna baze starog sistema. Poredeći sadržaj starog i novog informacionog sistema generalno se dolazi do saznanja da su podaci često različitih tipova, smešteni u različite tabele i u različitim formatima. Novi sistem uglavnom zahteva i podatke koji ne postoje u starom sistemu. Podaci u novom sistemu su međusobno povezani na drugačiji način nego u starom zbog primene različitih poslovnih pravila i različitih aplikativnih rešenja. Čak i kod veoma uređenih izvora podataka migracija nije jednostavna, jer se može raditi o više stotina tabela i više miliona zapisa.

Zbog svega ovoga migracija podataka je mnogo širi pojam nego što je samo prenos podataka. Ona obuhvata sve aktivnosti potrebne da se iz starog informacionog sistema podaci prikupe, transformišu, prenesu u novi sistem i izvrši provera ispravnosti prenosa.

2. RAZLOZI

Razloga za zamenu postojećeg informacionog sistema novim, a samim tim i za migraciju podataka ima više, među

kojima je najčešći modernizacija. Kompanije rastu, menjaju se i usvajaju nove biznis strategije. Troškovi održavanja zastarelih informacionih sistema, kao i razvoj zastarelog aplikativnog softvera mogu da budu veoma značajna stavka u IT budžetu. Visok, skoro eksponencijalni porast obima podataka koje informacioni sistem obrađuje i čuva zahteva stalnu modernizaciju. Migracija postojećih podataka na moderan, stabilan i snažan sistem za upravljanje bazom podataka postaje neophodan uslov da bi kompanija mogla da prati brze promene poslovnog okruženja i regulative.

3. RIZICI I TEŠKOĆE

Najčešća prepreka za uspešnu migraciju predstavlja nepostojanje kvalitetne dokumentacije, kao i kadrova, dobrih poznavaoaca izvornog sistema. Izvorni sistem je uglavnom razvijan u dugom vremenskom periodu, nisu sve izmene dokumentovane, a kadrovi koji su učestvovali u razvoju nisu više raspoloživi. Nerazumevanje izvornog sistema dovodi do produženja rokova i povećanja troškova što predstavlja ozbiljnu pretnju uspehu migracije. Mnoge kompanije nisu svesne ovog rizika i podrazumevaju da dobro poznaju organizaciju svojih podataka i da će proces migracije biti trivijalan.

Problem kvaliteta podataka izvornog sistema takođe je drastična pretnja uspehu migracije. Novi informacioni sistem po pravilu zahteva veoma strogu kontrolu kvaliteta prenetih podataka. Kroz životni vek izvorni informacioni sistem je menjan u skladu sa različitim poslovnim zahtevima tako da veze između podataka nisu svuda jedinstvene, neki od podataka nedostaju, podaci su u drugačijem formatu, postoje redundantni podaci u različitim izvorima starog sistema, šema baze podataka ciljnog sistema je potpuno drugačija od izvornog sistema, entiteti i relacije su različiti ili imaju drugačije funkcionalno značenje i poslovna pravila kojima podležu podaci su sasvim različita.

Zbog neshvatanja složenosti procesa migracije kompanije ponekad ne vrše odgovarajuću pripremu i planiranje i primenjuju neadekvatnu metodologiju ili alate.

4. ZAHTEVI

Metodologija migracije podataka objedinjuje postupke koji treba da obezbede:

- integritet podataka
- detektovanje grešaka
- mehanizam popravke grešaka prenosa
- mehanizam popravke kvaliteta izvornih podataka
- transformacije podataka prema zahtevima ciljnog sistema
- automatizaciju procesa
- izveštavanje o podacima koji su preneti i o podacima koji nisu preneti
- validaciju prenetih podataka
- mogućnost vraćanja u stanje pre početka procesa
- ponavljanje na zahtev celog procesa ili delova
- da se za izvršenje koristi minimalno vreme
- da ceo proces prenosa podataka bude kontrolisan i upravljiv
- da uticaj na poslovne procese bude minimalan

Podaci moraju biti preneti kompletno i tačno. Integritet prenetih podataka osnovni je uslov za uspešnu migraciju. Svi nekompletni, netačni ili duplirani podaci moraju biti identifikovani i sređeni pre prenosa.

Migracija podataka je komplikovana, jer obezbeđuje prenos velikog broja podataka iz širokog spektra različitih izvora, pri čemu ih u velikoj meri i transformiše. Pri tome se događaju greške različite prirode. Sve greške moraju biti otkrivene pre konačnog prenosa podataka. Postoje dve tipične vrste greške od kojih prva vrsta nastaje u samom procesu prenosa i transformacije podataka i njihov uzrok su greške u softverskim alatima za migraciju ili njihovo neadekvatno podešavanje. Ova vrsta grešaka se otklanja ispravkom programskog koda softverskih alata za migraciju ili drugačijim postavljanjem parametara. Uzrok druge vrste grešaka je slab kvalitet nekih podataka izvornog sistema. U tom slučaju pribegava se popravci ulaznih podataka ako je to moguće ili izostavljanjem tzv. „nečistih“ podataka ako popravka nije moguća.

Svaka migracija podataka zahteva transformaciju podataka koja može biti veoma kompleksna. Postoji više uzroka zahtevanih transformacija i one su uvek vezane za metapodatke koje nameće šema ciljne baze podataka. Takođe, uvek se zahtevaju i transformacije koje nameće aplikativni softver sa ugrađenim poslovnim pravilima u ciljnom informacionom sistemu. Ako se kao novi informacioni sistem uvodi tzv. ERP sistem (Enterprise Resource Planning) transformacije su još složenije, jer zavise i od tekuće kastomizacije (prilagođenja) sistema.

Veliki informacioni sistemi imaju ogroman broj tabela i zapisa u bazama podataka. Uvek postoji i složena međuzavisnost podataka tako da redosled prenetih podataka ne može biti proizvoljan. Ručnim prenosom to nikako nije moguće obaviti na zadovoljavajući način. Ako se tome dodaju i složeni zahtevi za transformacijom podataka, jasno je da je automatizacija procesa kroz odgovarajuće softverske alate, procedure i skripte jedino moguće rešenje.

Prenos ogromnog broja zapisa i mogućnost različitih grešaka može se ispratiti samo kroz dobro organizovan

sistem izveštavanja koji će registrovati sve prenešene i neprenešene zapise.

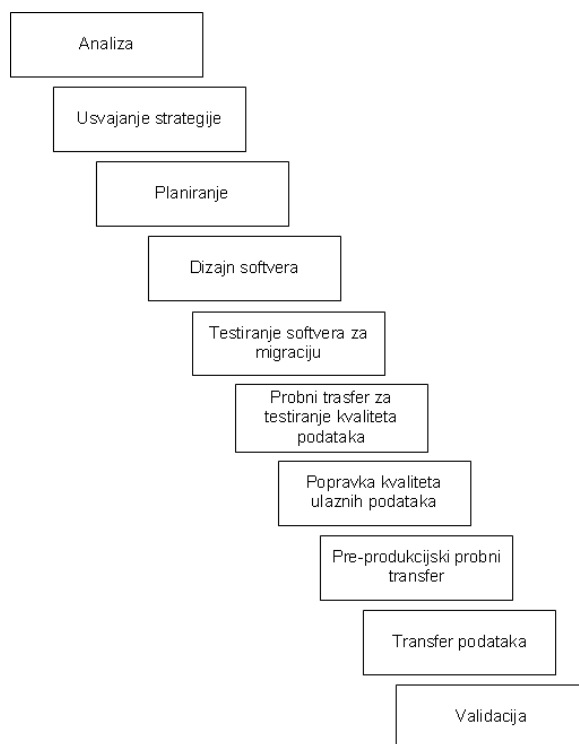
Da li je migracija podataka obavljena uspešno, biće potvrđeno kroz sistem za validaciju. Ukoliko nešto nije u redu potrebno je da se ponovi cela migracija ili samo jedan deo.

Stari sistem je po pravilu u eksploataciji do dana prelaska na novi. Najčešći je zahtev da vreme prelaska sa starog na novi sistem bude što kraće kako se ne bi ometali poslovni procesi kompanije i zbog toga je neophodno da se primeni metodologija migracije kojom se sve priprema tako da se vreme neoperativnosti sistema svodi na minimum. Metodologija migracije podataka koja se primenjuje na velikim informacionim sistemima mora da obezbedi ispunjenje svih ovih zahteva kako bi sama migracija bila uspešna.

5. FAZE MIGRACIJE

Da bi se zadovoljili svi postavljeni zahtevi metodologija mora da obuhvati više raznorodnih aktivnosti. Aktivnosti su međusobno zavisne i mogu se zaokružiti i grupisati u više faza migracije:

1. Analiza
2. Usvajanje strategije
3. Planiranje
4. Dizajn i priprema softvera
5. Testiranje softvera za migraciju
6. Probni transfer podataka za testiranje kvaliteta ulaznih podataka
7. Popravka kvaliteta ulaznih podataka
8. Pre-produkcijski probni transfer podataka
9. Transfer podataka
10. Validacija



Sl.1. Faze migracije

5.1. ANALIZA SISTEMA

Zadatak analize sistema je da katalogizuje sve elemente vezane za migraciju i definiše osnovu za planiranje sadržaja i obima migracionih aktivnosti. Analiza se uvek vrši i za izvorni i za ciljni sistem. Po pravilu, za svaki sistem se određuje tim za analizu.

5.1.1. ANALIZA IZVORNOG SISTEMA

Tim za analizu izvornog sistema se sastavlja od IT kadrova koji su dobri poznavaooci tog sistema, a mogu mu se pridodati i korisnici koji dobro poznaju sistem. Ovo je ključna faza, jer je rezultat ove analize osnova za dalji tok migracije i zato je od najveće važnosti da se izvorni sistem upozna u najranijoj fazi procesa. Nažalost, često ne postoji dobra dokumentacija koja detaljno opisuje izvorni sistem. U svakom slučaju, tim za analizu ne može završiti posao dok izvorni sistem ne bude kompletno i detaljno opisan.

Prvi zadatak je da se identifikuju izvori podataka. Bez obzira što postoje izvori podataka različite strukture, analiza mora da proizvede jedinstven katalog podataka u kome se za svaki podatak izvornog sistema identifikuje i dokumentuje:

- lokacija (server, storidž, disk ...)
- šema baze kojoj podatak pripada
- tabela
- naziv polja (kolone) gde je smešten
- format
- opseg mogućih vrednosti
- opis za lakše prepoznavanje ...

Svi drugi rezultati analize koji mogu da budu od značaja za uspeh migracije treba da budu dokumentovani, na primer: način pristupa podacima, kvalitet svakog podatka (tačnost, ažurnost itd.) ili za relacione baze podataka veza entitet-relacija.

Ako izvorni sistem ima više izvora podataka, svaki od njih mora biti predmet kompletne analize.

Definisanje preciznih perioda kada je svaki od izvora dostupan za eksport podataka neophodan je deo analize. Zavisno od količine podataka eksport može da traje vrlo dugo, pa je podatak o periodima raspoloživosti osnova za dobro planiranje celog procesa migracije.

Veličina izvorne baze podataka značajan je podatak za planiranje vremena potrebnog za migraciju.

5.1.2. ANALIZA CILJNOG SISTEMA

Analiza ciljnog sistema zasniva se na dobroj dokumentaciji. Na raspolaganju je obučen tim proizvođača aplikativnog softvera ili tim sistem-integratora, implementatora ako se uvodi ERP. U osnovi je po pravilu relaciona baza podataka. Analiza ciljnog sistema treba da dokumentuje:

- rečnik baze podataka

- šeme
- entitete
- relacije
- tabele
- tipove podataka
- formate
- opsege vrednosti
- poslovna pravila ugrađena u aplikativni softver
- kriterijume za ispravnost unosa
- opise kolona za lakše prepoznavanje

Rezultat analize je skup svih zahteva ciljnog sistema.

5.1.3. KONAČAN REZULTAT ANALIZE OBA SISTEMA

Konačan rezultat analize zaokružuje pet procesa:

- 1) dokumentovanje izvornog sistema (sadržaj, strukturu, kvalitet i integritet podataka)
- 2) dokumentovanje ciljnog sistema (sadržaj, strukturu i ugrađena pravila)
- 3) mapiranje između podataka izvornog i podataka ciljnog sistema
- 4) definisanje svih potrebnih transformacija podataka
- 5) definisanje redosleda migracije podataka da bi se obezbedio integritet ciljnog sistema

5.2. USVAJANJE STRATEGIJE

Usvajanje strategije daje odgovor na ključna pitanja vezana za migraciju i uz rezultate analize sistema postavlja osnov za pravljenje planova:

1) Prvo se rešava važno strateško pitanje da li se migracija obavlja odjednom, u više koraka ili sinhronizovano u realnom vremenu? Odgovor na ovo pitanje diktira mogućnost isključenja izvornog sistema iz redovnog rada za vreme migracije. Nekada je bila uobičajena praksa da stari sistem prestaje sa radom za korisnike u petak uveče, preko vikenda se njegovi podaci zamrzavaju i prenose, a u ponedeljak ujutru počinje sa radom novi informacioni sistem umesto starog. Danas je situacija često sasvim drugačija jer zbog podrške poslovnim procesima informacioni sistem mora da radi 24 časa / 7 dana u nedelji, a uticaj migracije na poslovne tokove mora da bude minimalan. Pošto postojanje neoperativnog perioda nije poželjno, migracija mora da se vrši na podacima koji se neprestano menjaju. Ovo predstavlja sasvim novi izazov za migraciju.

2) Pravi se lista prioriteta za migraciju prema modulima ili podsistemima sa najmanjom međusobnom zavisnošću.

3) Sledeće pitanje je koji se podaci prenose u novi sistem? Ova strateška odluka donosi se na osnovu analize sistema. Čest je slučaj da se neki istorijski podaci ne mogu na pravilan način preneti u novi sistem.

4) Određuju se kriterijumi za merenje uspešnosti migracije (procenat prenetih tabela, prenetih zapisa, zapisa sa greškom itd.)

5) Određuje se da li se razvijaju sopstveni softverski alati, skriptovi, procedure ili se koriste gotova, komercijalna rešenja.

6) Određuje se da li migraciju izvodi tim sopstvenih kadrova ili se angažuje tim specijalista

7) Način eksporta podataka iz izvornog sistema je takođe od najvećeg značaja. Uobičajeni načini su da se alatima izvornog sistema podaci eksportuju i smeštaju u fajlove u tekstualnom ili CSV formatu, a takođe je moguće izvući podatke u pomoćne ODBC ili JDBC tabele odakle se alatima ciljnog sistema čitaju, transformišu i smeštaju u bazu podataka ciljnog sistema. Moguće je da se da se alati ciljnog sistema povežu direktno na izvorni sistem, ali se to u principu ne preporučuje.

8) Određuje se tačan trenutak migracije

5.3. PLANIRANJE

Osnov za planiranje je analiza sistema i usvojena strategija. Faza planiranja definiše:

- Popis svih preduslova za uspešno obavljanje migracije
- Plan mapiranja podataka izvornog sistema na ciljni
- Plan izvršenja transformacija na svakom podatku
- Plan redosleda migriranja podataka
- Plan kreiranja softverskih alata
- Plan testiranja softvera
- Plan testiranja kvaliteta izvornih podataka
- Plan testiranja prenosa podataka
- Plan popravke kvaliteta podataka
- Plan izvršenja konačne migracije
- Plan validacije
- Plan angažovanja kadrova i sastavljanja projektnih timova

Preduslovi za uspešno obavljanje migracije obuhvataju pripremu IT okruženja. Potrebno je pripremiti listu hardvera, softvera, mrežne zahteve, potreban disk prostor, prava pristupa itd.

Plan mapiranja podataka kao i plan izvršenja transformacija direktno sledi iz analize sistema.

Plan redosleda migriranja podataka mora da objedini zahteve ciljnog sistema za referencijalnim integritetom podataka, kao i međuzavisnosti podataka izvornog sistema.

Plan kreiranja softverskih alata obuhvata zahteve za generisanjem programskog koda, procedura ili skriptova neophodnih za automatizaciju migracije. Ovim planom su obuhvaćene sve procedure za eksport, transformaciju i upis podataka kao i automatski skriptovi za testiranje.

Plan testiranja softvera predviđa scenarije i skupove kontrolnih podataka za testiranje pripremljenih softverskih alata.

Plan testiranja izvornih podataka objedinjuje kontrolna pravila za testiranje eksportovanih izvornih podataka pre sledećih koraka migracije.

Plan testiranja prenosa obuhvata probni prenos podataka po pojedinačnim podsistemima i izvorima podataka. Definiše se način razvrstavanja grešaka, način njihovog detektovanja kao i kriterijumi za ocenu uspešnosti migracije.

Plan popravke kvaliteta podataka obrađuje klasifikovanja grešaka i akcije na njihovo pojavljivanje.

Plan konačne migracije obuhvata termin izvršenja kao i sve korake od eksporta podataka, transformacije i upisa u novi sistem pri čemu precizno navodi redosled procedura, skriptova kontrolnih izveštaja do završetka konačne validacije prenetih podataka.

Plan validacije obuhvata sve postupke i automatske skriptove za validaciju migriranih podataka kao i kriterijume za uspešnost migracije.

Plan angažovanja kadrova definiše ulogu svakog pojedinca učesnika u migraciji.

5.4. DIZAJN I PRIPREMA SOFTVERA

U ovoj fazi se vrši izrada softvera za automatizaciju migracije, što se zapravo svodi na izradu programa, procedura, skriptova i njihovo povezivanje u jedinstvenu celinu. Ako je usvojena strategija da se koristi gotov komercijalni softver, vrši se njegovo podešavanje i postavljanje parametara. Softver mora da obezbedi automatizaciju sledećih koraka:

- Eksport podataka izvornog sistema u fajlove za eksport
- Mapiranje ulaznih i izlaznih podataka
- Transformacije eksportovanih podataka
- Normalizaciju
- Importovanje podataka u novi sistem
- Validaciju
- Izveštavanje

Ova faza je gotova tek kada je softver podešen i spreman da sve ove radnje obavi automatski u celini ili na zahtev korak po korak.

5.5. TESTIRANJE SOFTVERA

Svaka migracija je jedinstvena i softver mora biti podešen specifično za nju. U velikim sistemima postoji ogroman broj mapiranja i ogroman broj transformacija. Bez testiranja podešenosti softvera migracija ne bi bila sigurna. Testiranje se vrši po Planu testiranja softvera. Na setovima kontrolnih podataka testira se mapiranje kao i sve vrste transformacija.

5.6. PROBNI TRANSFER ZA TESTIRANJE KVALITETA ULAZNIH PODATAKA

Da bi migracija bila što efikasnija mora se kvalitet ulaznih podataka proveriti i po mogućstvu popraviti pre konačne migracije. Zbog toga se vrši probni transfer kroz koji će se detektovati svi podaci slabog kvaliteta. Posle pokušaja

popravke podataka probni transfer se ponavlja da bi se ustanovilo da li su svi podaci „čisti“.

5.7. POPRAVKA KVALITETA ULAZNIH PODATAKA

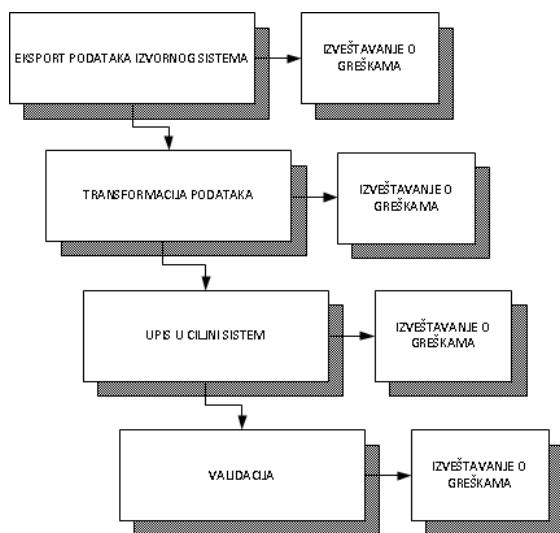
Popravka kvaliteta ulaznih podataka treba da se uradi što ranije, a svakako pre konačne migracije. Najbolje je da se uočeni slabi podaci popravljaju u samom izvornom sistemu. Zatim se ponovo vrši eksport podataka i ponavlja testiranje. Ovaj postupak se može ponavljati sve dok je potrebno. Ukoliko nije moguće vršiti popravku kvaliteta u izvornom sistemu, ona se vrši u fajlovima eksportovanih podataka. Cilj je da u dalji proces migracije ulaze samo podaci zadovoljavajućeg kvaliteta.

5.8. PRE-PRODUKCIJSKI PROBNI TRANSFER PODATAKA

Tačan trenutak migracije određen je strategijom migracije. Ukoliko je moguće, koriste se trenuci kada je to najpogodnije. Recimo na početku godine, odmah po formiranju početnog stanja za tu godinu, knjigovodstvenih podataka ima najmanje. To je idealan trenutak za migraciju. Međutim, pre nego što dođe vreme da se migracija izvrši moramo biti sigurni da će biti uspešna proveravanjem svih faza. Pre-produkcijski probni transfer podataka će nam dati potvrdu da li je sve spremno. On se uobičajeno vrši na punom setu svih podataka uključenih u migraciju.

5.9. KONAČNI TRANSFER PODATAKA

Podrazumeva se da su sve prethodne faze migracije već uspešno obavljene.



Sl.2. Tok konačnog transfera podataka

Proces konačnog transfera podataka vrši se automatski u više koraka:

- Započinje eksportom podataka izvornog sistema.

- Generiše se izveštaj o greškama eksporta, broju uspešno eksportovanih zapisa i broju neuspešno eksportovanih zapisa, kontrolne sume i totali izvornih i eksportovanih podataka.
- Vrše se transformacije podataka
- Generišu se izveštaji o greškama, kontrolne sume transformisanih podataka, odstupanja od opsega vrednosti transformisanih podataka
- Vrši se normalizacija
- Generišu se izveštaji o greškama
- Vrši se upis pripremljenih podataka u novi sistem
- Istovremeno sa upisom vrši se i validacija na nivou zapisa
- Generiše se izveštaj o greškama upisa, broju uspešno upisanih zapisa i broju neuspešno upisanih zapisa, kontrolne sume i totali upisanih podataka.

5.10. VALIDACIJA

Validacija je neophodna faza pre puštanja novog sistema u rad. Kod velikih sistema validacija mora biti automatizovana. Pri upisu podataka u novi sistem vrši se validacija na svakom slogu. Na ovaj način se proverava da li podaci zadovoljavaju referencijalni integritet novog sistema i poslovna pravila. Po završetku upisa svih zapisa neophodno je proveriti da li je migracija uspešno izvršena, to jest, da li su podaci preneti po specifikacijama. Kroz validaciju se dolazi do izveštaja koji pokazuju razlike između rezultata migracije i očekivanih rezultata. Ponekad se pristupa paralelnom radu izvornog i ciljnog sistema da bi se utvrdile moguće razlike ili gubitak podataka u prenosu. Posebno se vrši provera:

- Lokacije prenetih podataka
- Kompletnosti
- Ispravnosti
- Integriteta
- Pravila za transformaciju
- Poslovnih pravila

Validaciju obavlja tim sastavljen od IT kadrova i stručnih korisnika za pojedine module informacionog sistema.

6. IZVEŠTAVANJE

Izveštavanje je od presudnog značaja za uspeh migracije. Kod eksporta podataka iz izvornog sistema, kod transformacija, kod upisa u novi sistem izveštaji o greškama su osnov za praćenje toka migracije. Bez izveštavanja ne bi bilo moguće sprovesti popravku kvaliteta podataka. Na osnovu izveštaja o greškama delovi migracije se ponavljaju do zadovoljavajućeg rezultata.

7. ZAKLJUČAK

Primenom metodologije, kod velikih informacionih sistema migracija podataka može da se svede na niz potpuno definisanih i kontrolisanih aktivnosti koje garantuju uspešno izvršavanje i najsloženijih migracija podataka.