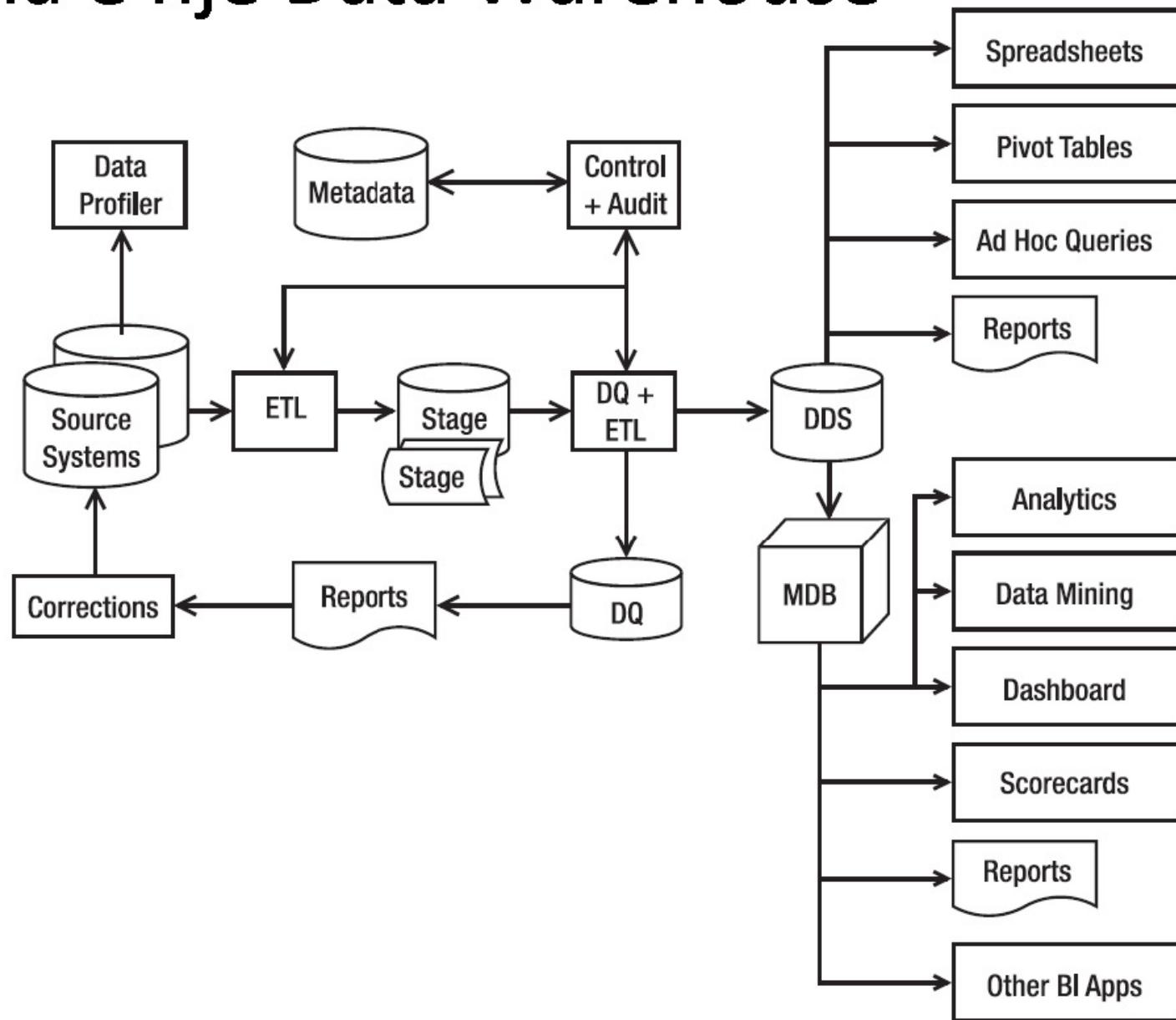


Hyrje në Data Warehouse

- Një **Data Warehouse** është një sistem që **mbledh** dhe **konsolidon** të dhëna në mënyrë **periodike** nga sistemet burim drejt një data store **dimensional** ose të **normalizuar**. Zakonisht mban **historikun** e shumë viteve dhë përdoret për të ekzekutuar **query** në funksion të **business intelligence** ose **aktivitete** të tjera analizuese.

Skema e një Data Warehouse



- **Sistemet burim** – sisteme OLTP që mbajnë të dhënat që do të merren dhe integrohen në Data Warehouse
- **Data Profiler** – ndihmon në ekzaminimin dhe njojen e të dhënavë të sistemeve burim (për shembull sa rreshta ka një tabelë, sa kolona kanë vlerën NULL, etj)

ETL (Extract, Transform, Load)

- Përbën një sistem që :
 - Mund të lidhet me sistemet burim
 - Lexon të dhënat nga këto sisteme
 - Transformon të dhënat (ndryshon të dhënat në varësi të formatit apo kërkesave të sistemit target, derivon vlera të reja nga ato ekzistuese, etj)
 - I ngarkon në një sistem target (stage)
 - Më pas integron(konsolidon) të dhënat në një data store dimensional (si ne figurën e mësipërme) ose të normalizuar

ETL (Extract, Transform, Load)

- ‘Pastron’ të dhënat

ETL kontrollon nëse të dhënat plotësojnë rregullat e cilësisë të përcaktuara në metadata. Në varësi të rezultatit, ato mund të lejohen, ndalohen apo të ruhen në databazën DQ nga ku do të gjenerohen raporte për korrigjime të mëtejshme.

- Një formë tjetër e ETL është sistemi ELT(Extract, Load, Transform) ku të dhënat ngarkohen fillimisht në data Warehouse dhe më pas i nënshtrohen transformimeve. Në këtë rast nuk ka nevojë për një ETL server

Konsolidimi (integrimi) i të dhënavë

- Konsolidimi(integrimi) i të dhënavë lind i nevojshëm duke qenë se të dhënata merren nga disa sisteme burim. Gjatë konsolidimit duhen marrë parasysh:

1. Disponueshmëria e të dhënavë

Një pjesë të dhënash mund të jetë e disponueshme në një sistem burim ndërsa në një tjetër jo. (për shembull një sistem ruan (*adrese1, adrese2, qyteti, shteti*) ndërsa një tjetër ruan (*adrese1, qyteti, shteti*))

2. Periudha kohore

Tek disa sisteme mund të kemi të njëjtat të dhëna por për periudha të ndryshme kohore. (Per shembull kostoja mesatare në një sistem llogaritet javore ndërsa tek një tjetër mujore.) Duhet pasur kujdes për periudhën kohore që aplikohet ndaj të dhënavë sepse rrezikohet që të ruajmë të dhëna të pasakta.

Konsolidimi (integrimi) i të dhënavë

3. Përcaktimet e të dhënavë

Të dhëna që emërtohen në të njëjtën mënyrë nuk është e thënë të ruajnë vlera për të njëjtën gjë. (*Për shembull trafiku favor i një website tek një sistem i referohet aksesimeve pa përsëritje ndërsa tek një tjetër atyre me përsëritje*) Duhet analizuar gjithmonë kuptimi i çdo të dhëne.

4. Konvertimet

Shpesh herë është e nevojshme të bëhen konvertime të të dhënavë kur në sisteme të ndryshme ato ruhen në njësi të ndryshme matëse. (*Për shembull kur paratë ruhen në monedha të ndryshme apo data dhe ora ndryshon për zona të ndryshme gjeografike*)

5. Përputhshmëria e të dhënavë

Duhet pasur kujdes që të dhënat që bashkohen nga sisteme të ndryshme ti referohen të njëjtit entitet në të gjitha sistemet (*Për shembull kur bashkojmë të dhënat e një personi të shpërndara në disa sisteme, duhet të sigurohem që të gjitha sistemet bëhet fjalë për të njëtin person*)

Update periodik i të dhënavë

- Marrja dhe konsolidimi i të dhënavë nga sistemet burim nuk ndodh vetëm një herë por përsëritet në mënyrë periodike. Periodiciteti varet nga:
 - Kërkesat e biznesit (për shembull sa shpesh kërkohen gjenerimi i raporteve)
 - Frekuenca e ndryshimit të të dhënavë në sistemet burim Pavarësisht se kohët e fundit gjithmonë e më shumë përdoren Data Warehouse që update-ohen në kohë reale, në të shumtën e rasteve update-imet e Data Warehouse nuk ndodhin me të njëtin shpeshtësi si sistemet burim për:
 - Të ruajtur stabilitetin e Data Warehouse-it sepse po të ndryshonin të dhënat në çdo moment, do të ishte tepër e vështirë analizimi i tyre
 - Të mos ulur performancën e sistemeve burim të cilat do të duhet të ofronin mekanizma shtesë(si psh trigera) për të reflektuar ndryshimet në kohe reale edhe tek Data Warehouse

Data Store Dimensional

- Një Dimensional Data Store(DDS) përbëhet nga një ose disa databaza që përbajnë një koleksion data marts dimensionale. Një data mart dimensional është një grup tabelash faktike dhe tabelat korresponduese të dimensioneve.
- Një DDS është e denormalizuar (pra shpesh kemi tepri të dhënash)
- Mund të implementohet në tre skema kryesore:
 - Star

Një dimension nuk ka subdimensione

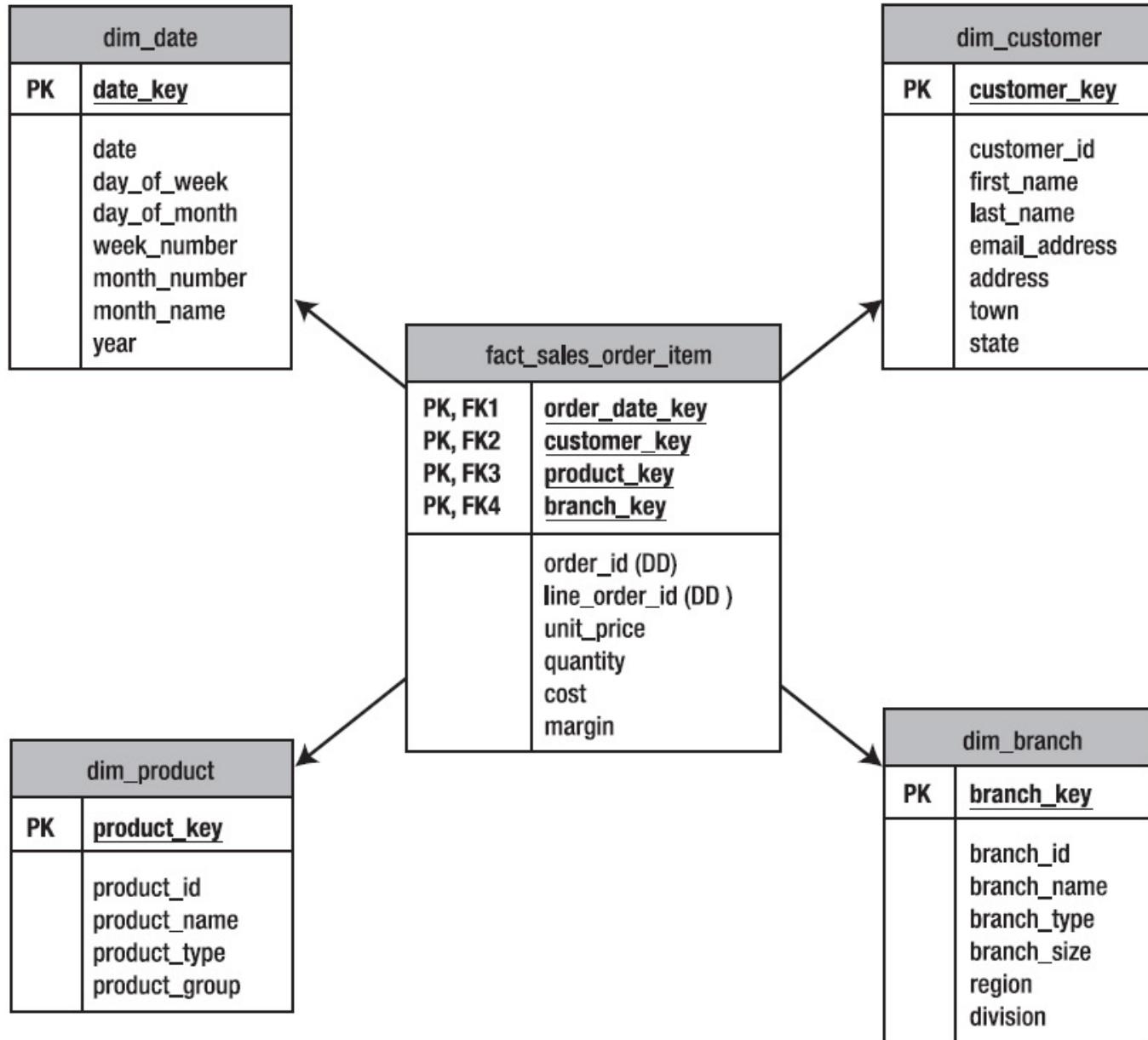
- Snoëflake

Një dimension mund të ketë subdimensione

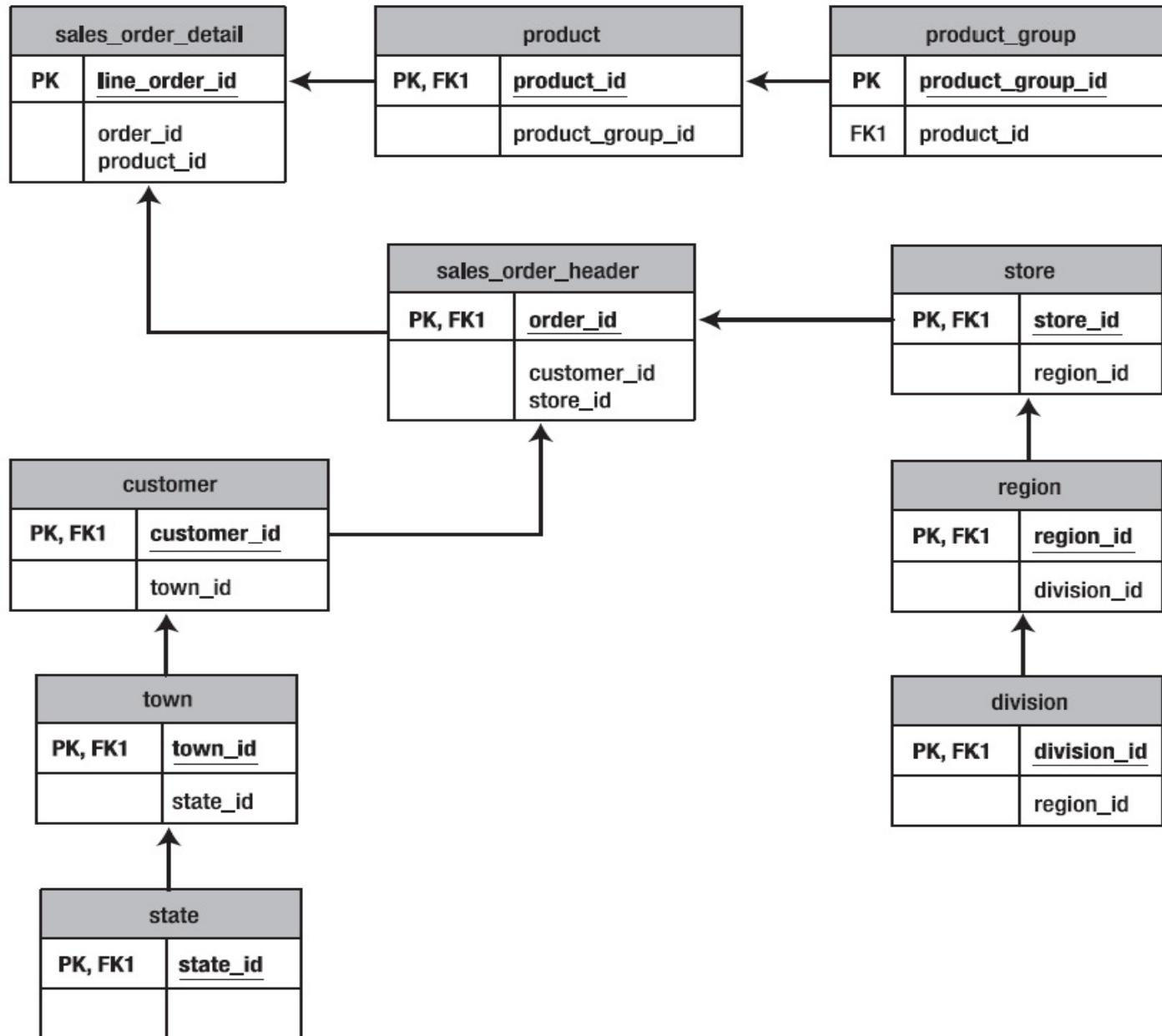
- Galaxy

Mund të kemi dy ose më shumë tabela faktike të lidhura midis tyre dhe me dimensionet përkatëse

Data Store Dimensional



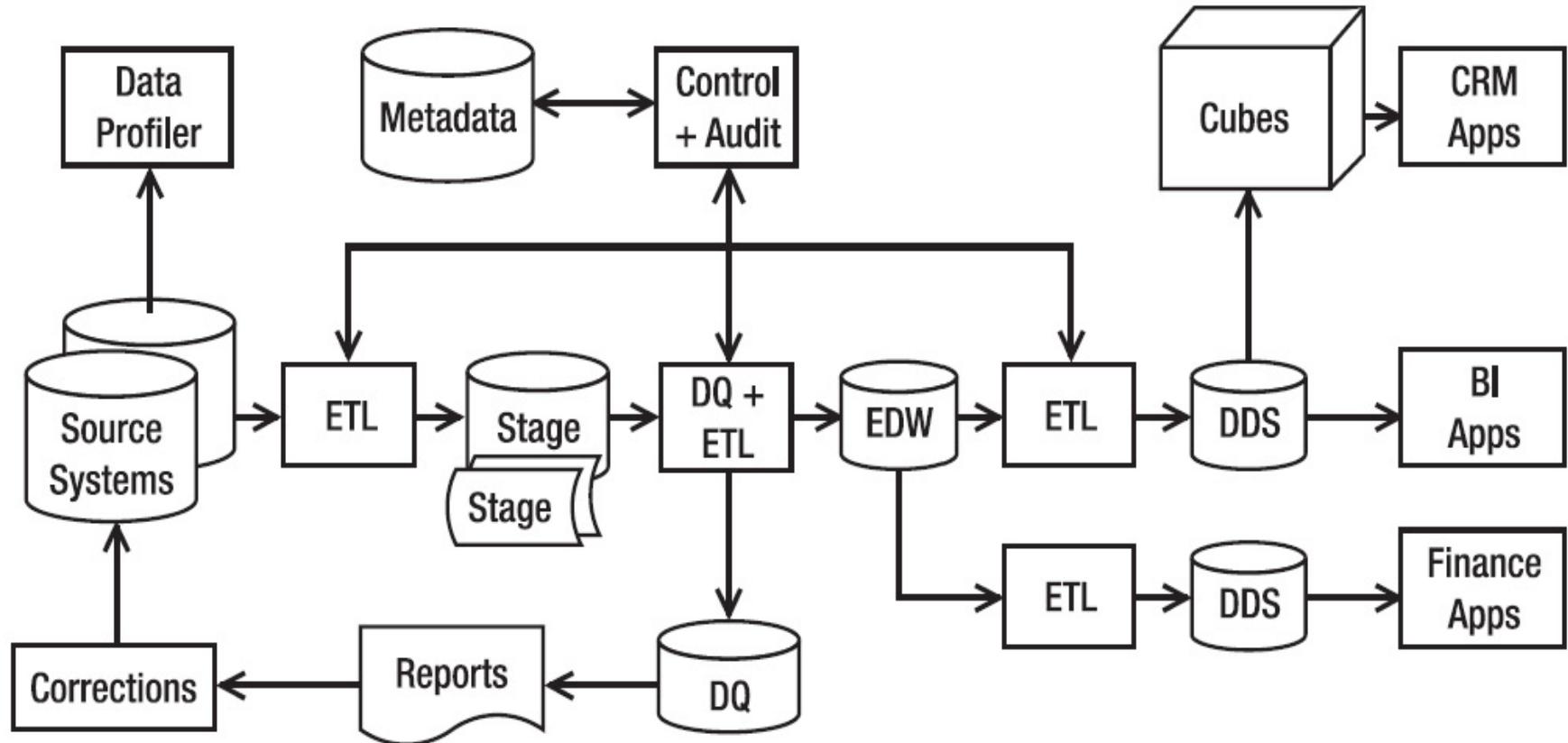
Data Store i Normalizuar



Data Store i Normalizuar

- Përbëhet nga një ose disa databaza relacionale të normalizuara të paktën në formën e tretë normale.
- Normalized Data Store (NDS) nuk ka tepri të dhënash.
- NDS është shumë e përshtatshme për integrimin e të dhënavë nga disa sisteme burim duke qenë se nuk këmi tepri të dhënat dhe shtimet e modifikimet kryehen vetëm në një vend.
- DDS është shumë më i përshtatshëm se NDS për ekzekutimin e query komplekse. Performanca e ekzekutimit te qyery ndaj NDS është e ulët duke qenë se përgjithësisht të dhënat duhet të merren nga shumë tabela me disa Join. Ky problem nuk ndeshet tek DDS duke qenë se kjo e fundit është e denormalizuar

Data Store të normalizuara përdoren zakonisht në Enterprise Data Warehouse per të marre dhe ruajtur më shpejt të dhënat nga sistemet burim. Më pas, për të rritur performancën në qyery ndaj Data Warehouse-it, ETL kujdeset në denormalizimin e databazes dhe ruajtjen e të dhënavëve në një DDS si në figurën e mëposhtme:



- Nga DDS është shumë e thjeshtë të kalohet në një MDB (multidimensional database) që njihet ndryshe edhe si kub
- Një MDB është një databazë ku të dhënrat ruhen në qeliza pozicioni i së cilave varet nga vlera e disa dimensioneve.
- MDB është shumë e përshtatshme për realizimin e analizave të thelluara

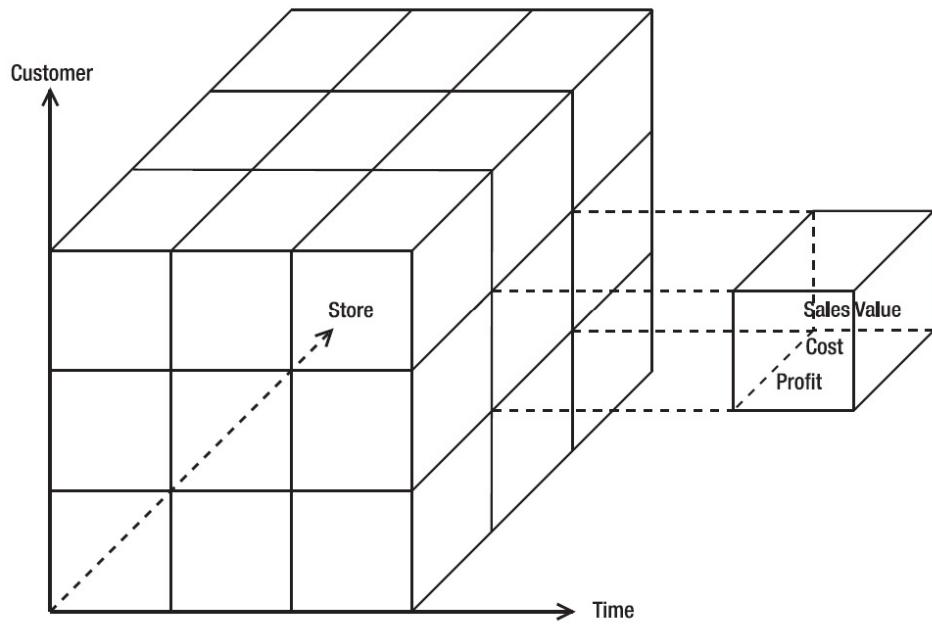


Figura tregon një kub me tre dimensione (Kohë, Dyqan dhe Blerës). çdo qelizë përfaqëson ngjarjen ku një blerës ble në një dyqan në një kohë të caktuar. Qeliza në brendësi të saj mund të ruajë një ose disa vlera si: vlera e blerjes, kosto, fitimi, etj

Historiku i të dhënavë

- Zakonisht një Data Warehouse , ndryshe nga OLTP të zakonshme, ruan historikun e shumë viteve të dhënash.
- Slowly Changing Dimensions (SCD) është teknika që përdoret në modelimin dimensional për ruajtjen e historikut të të dhënavë dimensionale
- SCD tipi 2 ruan historikun në rreshta të rinj
- SCD tipi 3 ruan historikun në kolona të reja
- SCD tipi 1 nuk ruan historik të të dhënavë
- Ruajtja e historikut mund të ndikojë në uljen e performances. Për të shmangur këtë shpesh përdoren teknika si tabelat e particionuara apo ekzekutimi i query në paralel.

Përdorimet e Data Warehouse

- Business Intelligence përpinqet të kuqtojë situatën e bizneseve duke realizuar analizime ndaj të dhënavë të kompanive për të ndihmuar në marrjen e vendimeve strategjike, taktike dhe operacionale në funksion të rritjes së performancës së biznesit.

Mund të përfshijë:

- Menaxhim të performancës së biznesit
- Analiza të përfirueshmërisë së klientëve
- Analiza statistikore
- Analiza parashikuese
- etj.

Përdorimet e Data Warehouse

- Data Warehouse gjithashtu përdoret në aktivitete të tjera analitike jashtë fushës së biznesit si: kërkime shkencore, vendimmarrje qeveritare(statistika, analiza ekonomike, parashikime, etj), situata emergjente, organizata bamirëse, analiza trafiku të rrjetit, etj.
- Përdoret në Customer Relationship Management(CRM) për analiza ndaj klientëve
- Përdoret në Web Analytics për të kuptuar sjelljen dhe karakteristikat e trafikut të një website.
- etj

- Jo çdo Data Warehouse përmban të gjithë elementët e përshkruar më lart
- Minimalisht ajo duhet të përbëhet nga një ose disa sisteme burim, një sistem ETL dhe një data store dimensional.



- Nqs hiqet qoftë edhe një prej këtyre elementëve, nuk mund të quhet më një data Warehouse