

ETL testing

What are the preconditions for ETL testing?

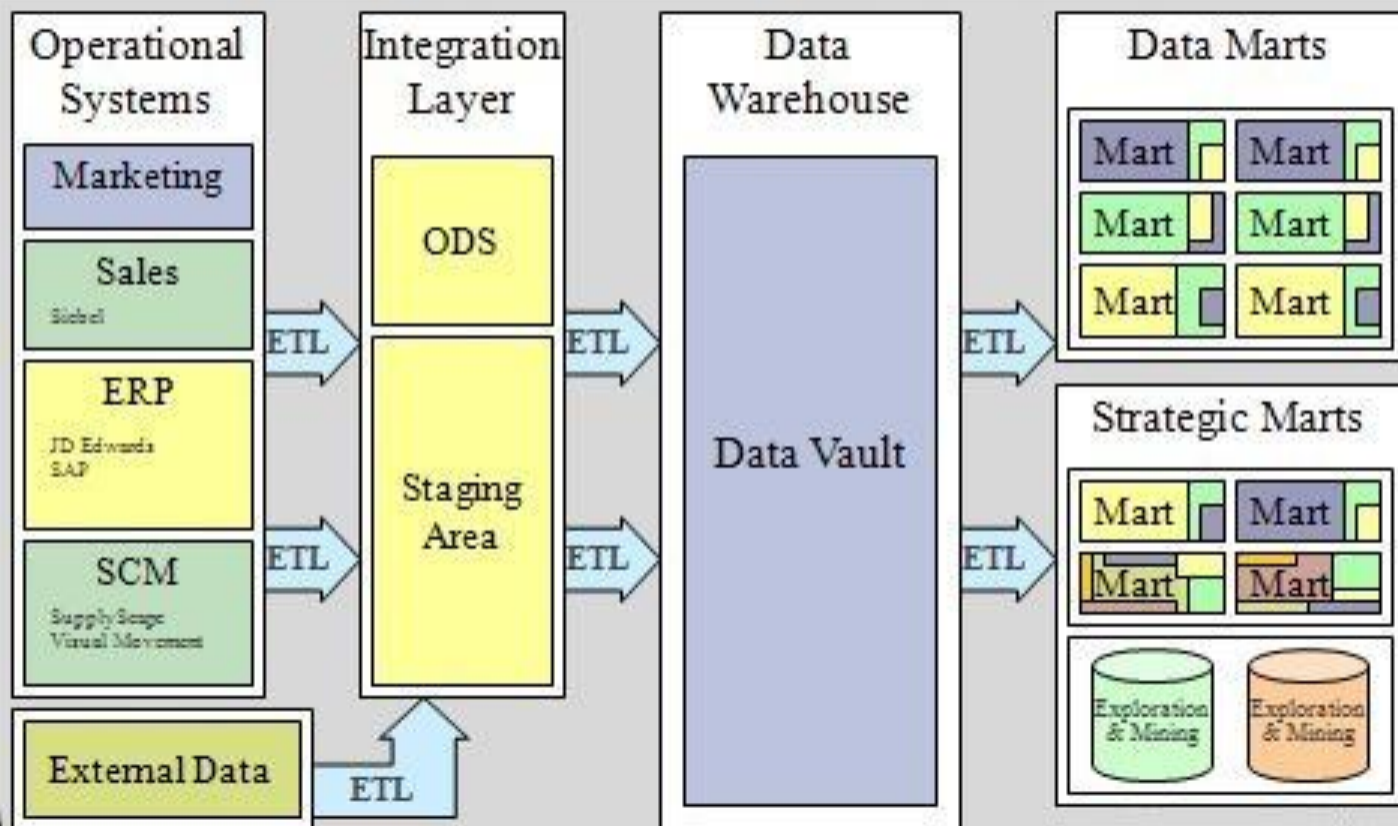


What are the preconditions for ETL testing?

1. Understanding theory
2. ETL development requirements
3. Testing types
4. Testing tools
5. Discussion time

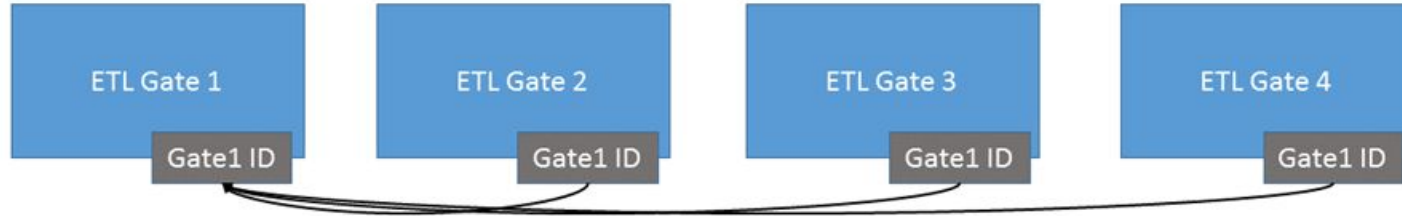
Understanding theory

Data Warehouse

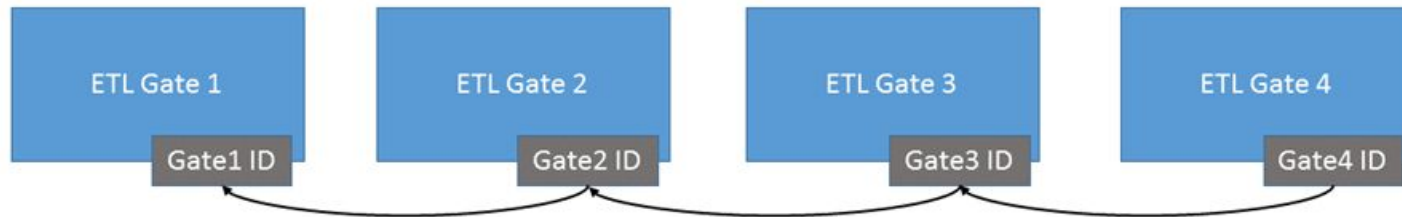


Data Lineage is important, but often forgotten

Absolute Data Lineage

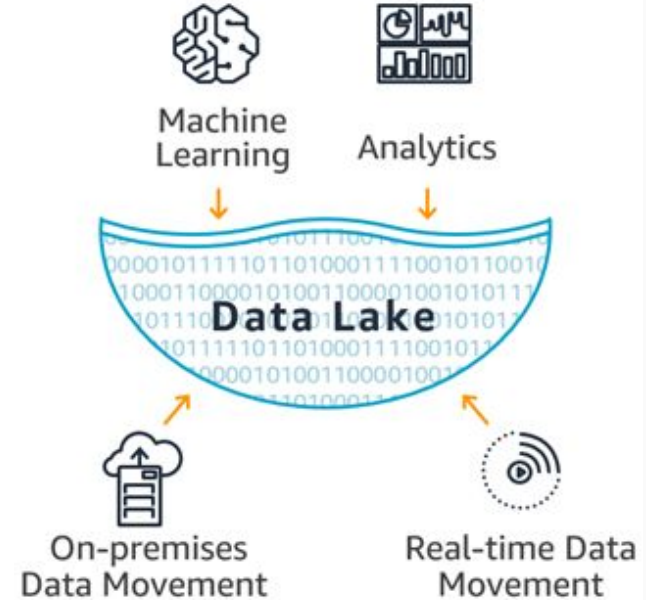
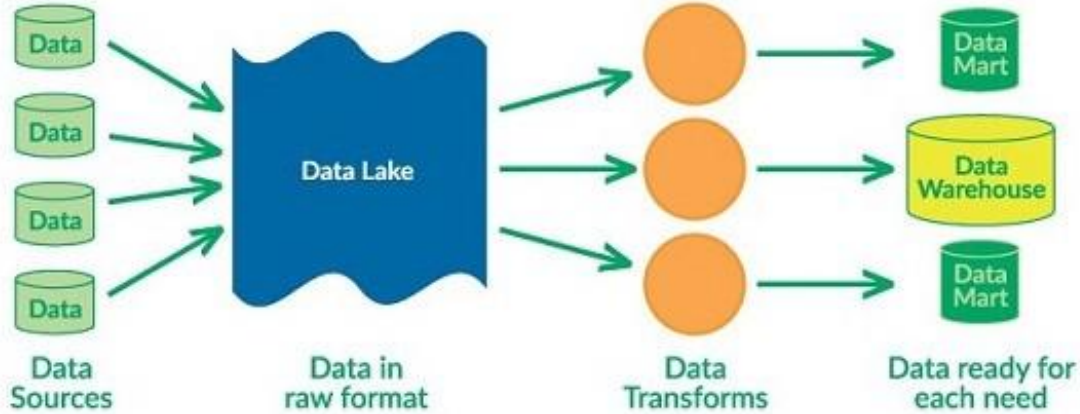


Relative Data Lineage

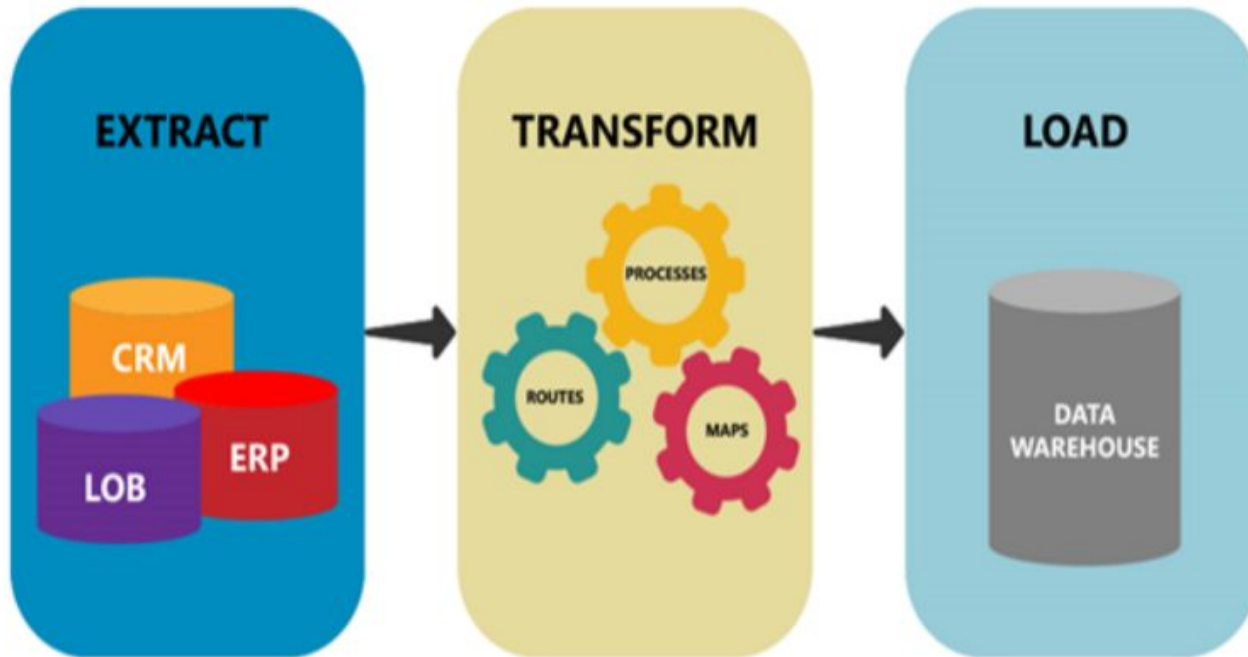


Data Lake

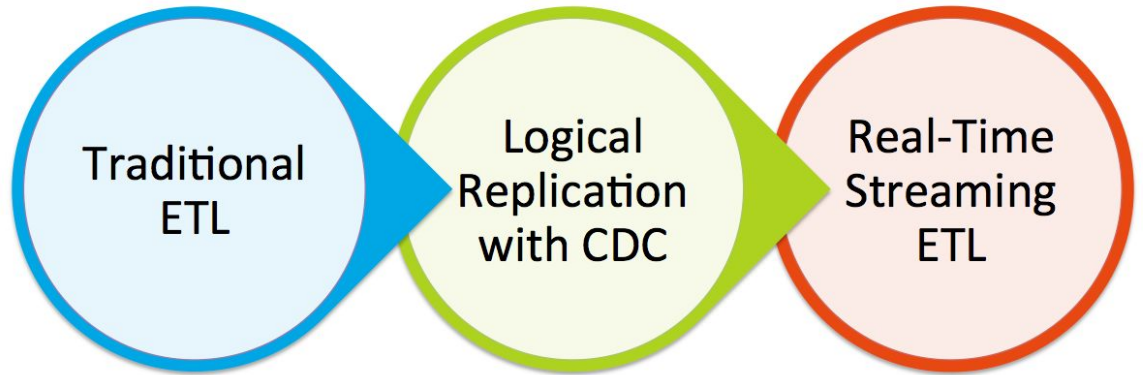
The Data Lake Pattern



ETL - Extract, Transform, Load (data integration?)



ETL evolution












- Batch-based, higher latency
- On-disk advanced transformations
- Designed mainly for databases

- Real-time or low latency
- Lacks transformation capabilities
- Designed mainly for databases

- Combines low-latency of replication with transformation capabilities of ETL
- Performs in-line transformations in-memory
- Designed to support varied sources and targets

ETL requirements

What are the ETL development requirements?

1. Business needs 
2. Compliance 
3. Data Quality (Profiling) 
4. Security 
5. Data Integration 
6. Data Latency 
7. Archiving and Lineage 
8. End User Delivery 
9. *Data transformation 







Test Types

1. Creating new DWH/job
2. Updating existing DWH/job

New DWH/job

1. Documents and requirements (BA) ?
2. Infrastructure (DevOps) ✓
3. Unit tests (RnD) ✓
4. Test plans, test suites, test scripts (QA) ?
5. Performance (DBA) ?
6. UAT ✓

Updating or Migrating DWH/job

1. Compare data with previous version 
2. Data flow is not broken 
3. Compare with test backups 
4. Deployment is not broken 
5. Unique identifiers are still unique 
6. There are no duplicate records 

Testing tools





Payed

1. QuerySurge
2. Informatica Data Validation
3. QualiDI
4. ICEDQ
5. ETL Validator

Free

1. Junit/TestNG + JDBC
2. Spark
 - a. [Local run](#)
 - b. [Spark-testing-base](#)
 - c. [Spark-fast-tests](#)
3. [MiniOozie](#)
4. [SpringJUnit4ClassRunner](#)
 - a. Unit
 - b. Steps
 - c. E2E

What are the preconditions for ETL testing?

1. Understanding theory 
2. ETL development requirements 
3. Testing types 
4. Testing tools 

DISCUSSION



TIME

That's all Folks!