Data Warehousing

* Data Engineering → Input → (Data) → Source

* DW → efficient way of managing data, provides support on mgmt decisions
* Warehouse → storage place, DW → Warehouse on datacenters
* Features → Subject oriented → Organize data acc to subject
  Integrated → Integrated different data source
  Time-variant
  Non-Volatile → No updates allowed.

* Data Engineering → Data Storage
                   → Data Processing
* spark → data processing
* ETL → Extracting, Transporting, Loading
* DSS → Decision Support System → support decision-making
  ↓
  Structured component → directly helps for decision
  Unstructured " → Human Interaction required

* DSS architectural style → OLTP (Online Transaction Processing)
  - used by RDBMS
  → OLAP (Online Analytical Processing
  - used by DW

* DW → database for analytics

* ~~the~~

* DSS data vs Operational Data → time span, granularity, dimensionality

* ETL Pipelin → staging, integration & access layers

* First data to arrive at DW is Operational Data store (ODS)

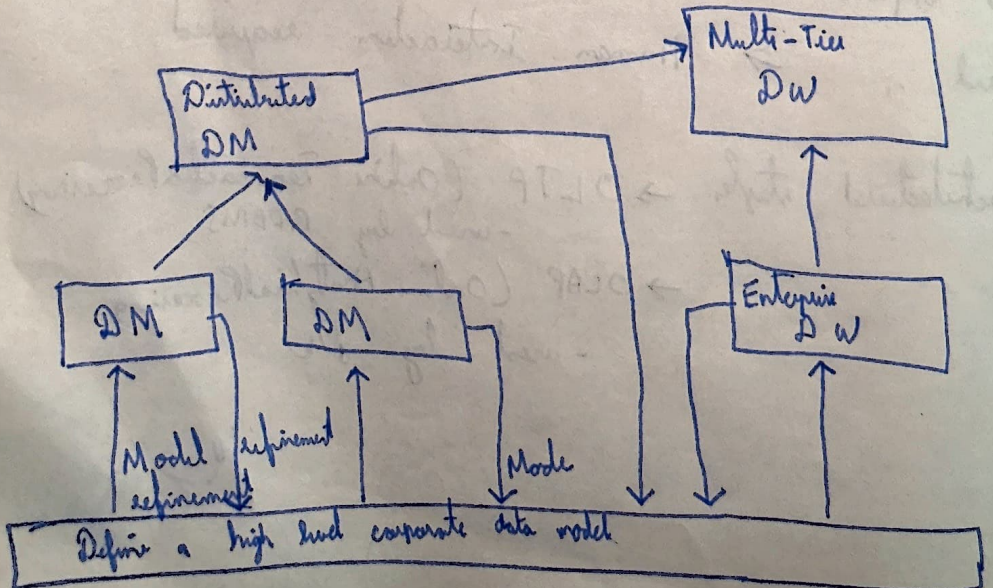
* DSS components → Data store
  " Extraction
  " Filtering
  End user query tool
  End user presentation tool

* Data in DW is stored in form of Data marts

* Data Mart is a subset of DW oriented to a specific business line

---

Data Mart Architecture :



OLAP → Approach to solve Multi-dimensional queries. relational reporting & data mining

ROLAP → multi-dim analysis of data stored in relational DB

D.W Applications:

* Information Processing → querying, basic statistical analysis
* Analytical " → multidimensional querying, drilling, pivoting
* Data Mining → Find Hidden Patterns

Big Data:

- If data runs beyond processing power & storage capacity, then we call it big data.