**Introduction**

Data Warehouse ~ Storage of Databases

DWH is built on top of DB, but its purpose is different

Mostly used for Structured Data

### **DWH Vs DB**

| **DWH** | **DB** |
| --- | --- |
| Store large volume of data | Stores small volume of data |
| Designed for read heavy operations | Designed for write heavy operations |
| High latency | Low latency |
| Denormalized  => Data Redundancy is more | Highly Normalized  => Data Redundancy is less |
| Columnar Storage  => Parquet, ORC etc | Row based |
| Parallel processing of requests | Not optimized for parallel processing |
| OLAP (Online Analytical Processing) | OLTP (Online Transactional Processing) |
| **Teradata**  **Exadata**  **Snowflake**  **Red-Shift** | **MySQL**  **Oracle**  **Postgres** |

### **Rules of DWH**

**1. Integrated** - Multiple sources are integrated to a DWH to get data

**2. Subject Oriented** - Reports can be generated by analyzing data in DWH

**3. Time Variant** - Historical Data

**4. Non-Volatile** - Data should not change at a point of time

### **DWH Vs Data Lake**

| **DWH** | **Data Lake** |
| --- | --- |
| Can share a limited storage | Can share infinitely large volume of data |
| Only supports Structured data and Semi-Structured data | Support Structured, Semi-Structured and Unstructured data |
| Required pre-defined schema | Doesn't require pre-defined schema |
| ETL | ELT |
| Higher Cost | Cost is Cheaper |
| Schema on write | Schema on read |