Data Management & tools course Cheat Sheet

Prepared By: Naren Peri

Published Date: Sep 11, 2011

|  |  |  |
| --- | --- | --- |
| S. No | Topic | Reference/response/what it means |
|  | Operational Data Store | ODS, sometimes data can pass through ODS before it reaches DW, used for operational reporting |
|  | Data Warehouse | Integrated, non-volatile, time non-variant (does not change with time unlike in operational systems) subject-oriented, historical store is a data warehouse. Typically end users/analytical users are not exposed to DW tables: DW is exposed through read only views |
|  | Data Mart | Subset of Data warehouse, subject oriented, analytical/aggregated data sets. OLAP/BI tools usually connect to data mart to source aggregated data for reporting/cubes. Analytical power users/sophisticated users are given access to data mart to perform ad-hoc analyses |
|  | Data Architecture | Architecting conceptual, logical and physical data models |
|  | Star Schema | Center fact(s) surrounded by dimensions. Each dimension contributes its business key to the fact table. Fact can be detailed fact or aggregated fact |
|  | Snow Flake Schema | Dimension is flaked. E.g.: Countries and states. Country table is connected to fact, and contributes country id to fact table. State table is connected to country table but not to the ‘fact’ table. If someone needs to aggregate sales by country and state, then one has to navigate to ‘state’ through ‘country’. When data size is huge, this can reduce performance( additional join condition), however if such requests are low, then one would not denormalize the schema |
|  | Normalization/De Normalizaton | Process of architecting/organizing data to balance redundancy & retrieval speed is called normalization/denormalization |
|  | E T L C | E-extract, T-transform, L-Load, C- cleanse. In practice, ETL is the most popular word, some theorists/practitioners also use ELT, ECLT, ECTL etc |
|  | ETL Architecture | Architect/design the data and process flow to move data from source tables to target tables |
|  | RDMBS | Relational data base management software: Oracle, SQL server DB, My SQL DB, Teradata, DB2, Access .mdb are examples of RDBMS |
|  | Appliances | DW appliances provide solutions for the mid-to-large volume data warehouse market, offering low-cost performance most commonly on data volumes in the [terabyte](http://en.wikipedia.org/wiki/Terabyte) to [petabyte](http://en.wikipedia.org/wiki/Petabyte) range. E.g: Teradata, Netezza, Greenplum. They use massively parallel processing architectures |
|  | ETL Tools | Tools that do ETL ☺ e.g: Informatica, AbInitio, SSIS |
|  | BI Tools | Business Intelligence /OLAP tools: reporting, dashboarding, that allow for slicing and dicing of data: Microstrategy, Cognos, Business Objects, SSAS, SSRS |