**Data Warehousing**

* Data warehousing(BI, Business Intelligence) solution is a collection of objects that allows data to be turned into useful information.

**BI Life Cycle**

1. Interview and Identify Data
   1. Determine the solution requirements and isolating the data you will be working with
2. Plan the BI solution
   1. Document the requirements, procedures
3. Create a data warehouse
   1. Design the RDBMS Database
4. Create the ETL process
   1. Extract-Transform-Load
      1. Create
      2. test ETL scripts
      3. Connections
      4. Load data into data warehouse
5. Cube creation
   1. Create cubes and determine storage solutions
      1. ROLAP
         1. Relational Online Analytical Processing
      2. HOLAP
         1. Hybrid OLAP
      3. OLAP
         1. Online Analytical processing
6. Create Reports
   1. Report against cubes and/or other sources
7. Test and fine tune the solution
8. Go Live, Approve, release and plan for updates

A close up of text on a white background

Description automatically generated

**Main Components of a Data Warehouse**

1. Operational Source Systems
   1. There are OS captures the transactions of the business(ie: OLTP). The source systems should be thought of as outside the Data Warehouse.
2. Data Staging Area
   1. Storage area and a set of processes commonly referred to as extract-transformation-load (ETL). The data staging area is everything between the operational source systems and data presentation area.
3. Data Presentation
   1. Data is organized, and made available for querying by users and analytical applications.
4. Data Access Tools

**Why do we need a Data warehouse?**

1. To perform querying and reporting on servers/disks not used by transaction processing systems. Don’t want to interface with the functioning and performance of the transactional database.
2. To use data models and/or server tech that speed up querying and reporting that are not appropriate for transaction processing
3. To provide a repo of “cleaned up” transaction processing systems data that can be reported against.
4. To make it easier to query and report data from multiple transaction processing systems and/or from external data sources and/or from data that must be stored for query/report purposes only.
5. Security? To prevent persons who only need to query and report transaction processing system data from having any access to transaction processing system db and logic used to maintain those db.

**Who uses the Data warehouse?**

* Designed to executives, senior managers, business analysts in making complex business decisions.

**Data Warehousing Terminology**

* Dimensional Modeling
  + Relational DB(OLTP) normalize. but not in Data warehouse
  + Highly normalized designs don’t benefit Data warehouse.

**Schemas**

1. **Star Schema**
   1. Fact Table = single central table
      1. Surrounded by multiple tables(dimensions)
   2. Each **Star Schema** covers one business area
      1. E.g., ‘FactInternatSales’
   3. Fact Table is connected to all dimensions with foreign keys
   4. Usually, FK taken are in the Fact Table
   5. Fact Table is ALWAYS on many side of the relationship with the dimensions.
   6. Fact Table represents a multi dimensional hypercube.
2. **Shared Dimension / Conformed Dimension**
   1. Dimensions have FK relationships with multiple fact tables. Dimensions with connections to multiple fact tables.
3. **Snowflake Schema**
   1. Star Schema dimensions are de-normalized.
   2. But **Snowflake Schema** isn’t
   3. If you normalize Star Schema, you end up with a **Snowflake Schema**
   4. **Fact Table must always still be on many side of relationship.**
4. **Hybrid Schema**
   1. When you normalize only part of the design to accommodate shared dimensions

**Then which Schema is Appropriate?**

1. **Star Schema =** best option
2. **Hybrid Schema** = if you need to share a dimension with multiple dimensions.
3. **Snowflake Schema** = only use for quick proof of concept project
   1. This design is closest to the source DB that is already in 3NF.

**Granularity**