

Question 1

- Fact table :
 - Presents itself as the core of a data warehouse containing aggregated data from sources
 - Single point storage of all key performance indicators pertaining to a business process
 - Records data in specific level of granularity such as daily, weekly, or monthly.
 - Decreases query performance by not doing any joins and its child dimension tables can be indexed in such a way that enables fast aggregation and filtering
 - Fact tables are connected to dimension tables using foreign key relationships. The foreign keys in the fact table link to the primary keys in the dimension tables, creating a star or snowflake schema
 - They can be periodically updated with transaction data with some means of ETL
- Star Schema
 - Acts as a simplified and easy to understandable data model
 - Makes navigation around tables of data warehouse easy
 - Its de-normalized design reduces the need for joins and enables fast querying. That is users can directly access fact tables and its index dimensions rather using large number of tables for joins
 - Provides lot of flexibility to add new data
 - Easy to scale and maintain , when the data grows large the fixed number of fact tables and dimension table enables scale to the need of incoming data
- Transactional database as OLAP
 - It is “not ideal” to use transactional database for housing an OLAP
 - Transactional database engines are designed to run super fast IO reads and writes as real time transactions by maintaining data consistency, but OLAP are designed to execute complex aggregation queries to answer analytical questions hence using a transactional database might cause performance issues
 - OLAP databases usually contain redundant de-normalized schema like star or snowflake to query data with ease rather transactional database will have normalised database tables aiming on minimal data redundancy

Question 2

Data Warehouse :

- Used to store historical data from various repositories and data sources from within an organisation
- Primary goal is to support Business intelligence and reporting activities
- Used ETL practices to extract and load data from multiple sources into a common format which is optimized for analytically processing and helpful in running complex queries without substantial joins
- example :
 - An e-commerce company can integrate all its data from various data sources like transaction, orders, inventories , customer data which provided valuable amount of business intelligence

Data mart :

- Contains data which is a Subset of data warehouse
- Designed for specific group of customers for their analytic needs
- These are created by extracting data from warehouse
- Provides user friendly and targeted environment
- example :
 - Within the above mentioned e-commerce warehouse we can create a separate data mart that contains data and focuses mainly on inventory items. These data can then be used to build models or gain intelligence on the stocks and inventory trends on all times of years which is a valuable source of information.

Data Lake :

- Vast super set of data warehouse
- Contains structured , unstructured and all variety and types of data
- Stored in its raw form until it is needed for processing or analytics
- example :
 - A racing car company will maintain a data lake where it stores vast amount of car data , engine points, inventory list , racing data , images , car sensor data which enables data scientists, researchers, and analysts to explore and analyze insights, trends, and patterns.

Links referred :

- <https://www.holistics.io/blog/data-lake-vs-data-warehouse-vs-data-mart/>
- <https://aws.amazon.com/compare/the-difference-between-a-data-warehouse-data-lake-and-data-mart/>

Question 3

- Designed a fact table with dimensions for “time-based” analytical processing
- The fact table has all the information about strikes like flights airports information
- The dimension of granularity this fact table is designed with is date and time
- The fact table can solve the problem of answering when, which month and year or quarter the strikes happened most, which year it was above average what part of the day was the strike relatively high etc
- Design : <https://i.imgur.com/0hMZbc1.jpg>