**Glossary**

A/B testing - A method of comparing two different versions of a product or website to determine which one performs better.

ACID: A set of properties that guarantee reliable processing of database transactions.

Additive Facts: Facts that can be summed up across all dimensions.

Algorithm - A set of instructions that a computer program follows to solve a problem or perform a task.

Bus Matrix: A matrix that outlines the relationships between business processes and data entities.

Business Intelligence: A set of techniques and tools for analyzing and reporting data to support business decision-making.

Cloud computing - The use of remote servers to store, manage, and process data.

Conformed dimensions: Dimension tables that are shared across multiple fact tables.

Conformed Dimensions: Dimensions that are shared across multiple fact tables.

Dashboard - A visual display of key performance indicators (KPIs) and metrics.

Data exploration - The process of analyzing data to discover patterns and relationships.

Data governance - The management of data availability, usability, integrity, and security in an organization.

Data governance: The process of managing the availability, usability, integrity, and security of data used in an organization.

Data Lake: A large, centralized repository that allows for the storage and management of structured and unstructured data.

Data lineage - The complete history of a data asset, from its origin to its current state.

Data Mart: Data mart is a subset of a data warehouse. It collects data from a data warehouse.

Data mesh - A data architecture that emphasizes data ownership and domain-driven data products.

Data mining: The process of discovering patterns in large datasets.

Data modeling - The process of creating a conceptual, logical, or physical representation of data and its relationships.

Data Modelling: The process of creating a conceptual representation of data and its relationships.

Data pipeline - A set of tools and processes used to move data from its source to its destination.

Data pipeline: The sequence of steps that data goes through to be processed and analyzed.

Data quality - The accuracy, completeness, and consistency of data.

Data science - An interdisciplinary field that involves the extraction of insights from data using scientific methods and algorithms.

Data staging area: A temporary storage area used to store data before it is loaded into a data warehouse.

Data storytelling - The use of data to tell a compelling story that supports a business goal or decision.

Data Vault: A data modeling methodology designed for modeling historical data and for creating an audit trail of changes.

Data visualization: The graphical representation of data and information.

Data Warehouse: Data Warehouse is optimised to store data for analysis and reporting purposes.

Database: A Database is basically where a collection of transactional data is stored.

De-Normalisation: The process of intentionally adding redundant data to a database to increase performance.

Derived data: New datasets or tables that are created from existing data

DevOps - A set of practices that combines software development (Dev) and IT operations (Ops) to streamline software delivery.

Dimensional Modelling: A data modeling technique that organizes data into facts and dimensions for easy analysis and reporting.

ELT: Extract, Load, Transform - A process for extracting data from multiple sources, loading it into a data lake or warehouse, and then transforming it as needed.

Entity Relationship Diagram (ERD): A visual representation of entities and their relationships to each other.

ETL: Extract, Transform, Load - A process for extracting data from multiple sources, transforming it to fit a common data model, and loading it into a data warehouse.

Extract - Extract is the process of extracting data from the original data source.

Fact Tables: A table in a data warehouse that stores quantitative information, such as sales or revenue.

Factless fact tables: Fact tables that do not contain any measures or metrics, but instead serve to establish relationships between dimensions.

Git - A version control system used for software development.

Junk Dimensions: A dimension table used to store multiple low-cardinality attributes.

Key Performance Indicator (KPI): A measurable value that demonstrates how effectively a company is achieving key business objectives.

Keys: A unique identifier for a record or entity in a database table.

Load - Load is where transformed data is loaded to our data warehouse.

Master data management (MDM): The process of defining and managing the critical data of an organization to provide a single point of reference.

Normalization: The process of organizing data in a database to reduce redundancy and dependency.

NoSQL: A type of database that uses a non-relational data model.

OLAP: Online Analytical Processing - A technology for analyzing data to support decision-making.

OLTP: Online Transaction Processing - A technology for managing transaction-oriented applications.

Periodic Facts Tables: Fact tables that store information related to a specific time period.

Relational database: A type of database that organizes data into one or more tables with a predefined structure.

Role Playing Dimensions: A single dimension table that is used multiple times within a fact table.

SCD: Slowly Changing Dimensions - A technique for tracking changes to dimension data over time.

Self-service analytics: The ability for business users to access and analyze data without relying on IT or technical resources.

Semi-Additive Facts: Facts that can be summed up across some dimensions but not all.

Snowflake schema: A type of dimensional model that uses multiple levels of normalized dimension tables.

Snowflake Schema: A type of dimensional model that uses normalized dimension tables.

SQL: Structured Query Language, a programming language used to manage and manipulate relational databases.

Star schema: A type of dimensional model that uses a single, denormalized fact table with multiple dimension tables.

Star Schema: A type of dimensional model that uses denormalized dimension tables.

Streaming data - Data that is generated in real-time and can be processed and analyzed as it is generated.

Transform - Transformation is the process of changing the data by means of either enriching it or by cleaning the data.

Unstructured data - Data that does not have a predefined data model or schema, such as text or images.

Visualization - The creation of graphical representations of data to facilitate understanding and analysis.

Wide Columns: A NoSQL data model that stores data in tables with a high number of columns.

YAML - A human-readable data serialization format used for configuration files and data exchange.

SQL - Structured Query Language, a programming language used to manage and manipulate relational databases.

Table: A collection of related data organized in rows and columns.

Column: A vertical set of data in a table that represents a particular attribute of the data.

Row: A horizontal set of data in a table that represents a specific instance or record.

Primary key: A column or set of columns in a table that uniquely identifies each row.

Foreign key: A column or set of columns in a table that refers to the primary key of another table, creating a relationship between the two tables.

Index: A data structure used to improve the performance of queries by allowing for fast retrieval of data based on specific columns.

Query: A request for data from a database, written in SQL syntax.

SELECT: A SQL keyword used to retrieve data from a table or tables.

WHERE: A SQL keyword used to filter data based on a specified condition.

JOIN: A SQL keyword used to combine data from two or more tables based on a related column.

GROUP BY: A SQL keyword used to group rows together based on a specified column or columns.

ORDER BY: A SQL keyword used to sort data in ascending or descending order based on a specified column.

INSERT: A SQL keyword used to insert new rows of data into a table.

UPDATE: A SQL keyword used to modify existing data in a table.

DELETE: A SQL keyword used to delete one or more rows of data from a table.

COMMIT: A SQL command used to save changes made to a database.

ROLLBACK: A SQL command used to undo changes made to a database since the last COMMIT.

Trigger: A SQL object that executes automatically in response to certain events, such as inserting or updating data in a table.