

## **Q1: Talk about the data life cycle.**

The data lifecycle is the journey data takes from the moment it is created or collected to when it is stored or deleted. This journey includes steps like gathering the data, cleaning it to fix errors, saving it safely, using it to find useful information, and finally getting rid of it when it's no longer needed. Each step is important to keep the data accurate, easy to use, and secure.

## **Q2: What is Database Normalization and Denormalization?**

**Normalization** organizes data into smaller tables to avoid duplication and ensure consistency. It's best for faster updates and maintaining data accuracy.

**Denormalization** combines tables or adds duplicate data to speed up searches. It's used when faster reading is more important than reducing data duplication.

## **Q3: What's the difference between ETL and ELT?**

**ETL (Extract, Transform, Load):** Data is prepared (transformed) before it goes into the warehouse.

**ELT (Extract, Load, Transform):** Data is loaded into the warehouse first, then prepared (transformed) inside the warehouse.

## Q4: What's the difference between WHERE and HAVING in SQL?

**WHERE:** Filters rows before grouping or aggregation. For example: Find students older than 18.

```
SELECT S_Name, Age
FROM Student
WHERE Age ≥ 18;
```

**HAVING:** Filters groups after aggregation. For example: Find age groups with more than 2 students.

```
SELECT Age, COUNT(*)
FROM Student
GROUP BY Age
HAVING COUNT(*) > 2;
```

## Q5: What is the difference between DELETE and TRUNCATE?

**DELETE** removes specific rows from a table and can be undone. It can be slow because it deletes one row at a time.

**TRUNCATE** quickly removes all rows from a table at once and cannot be undone. It's faster but permanent, meaning that it can't be undone.

## **Q6: What's the difference between structured and unstructured data?**

**Structured Data** is organized and easy for computers to read, like tables in a database. Examples include spreadsheets and customer records.

**Unstructured Data** is not organized in a specific way, making it harder to analyze. It includes things like text documents, images, and social media posts.