

1 common question.

Set 1:

1.	<p>What is the need of data warehouses?</p> <p>The need for Data Warehouse is to generate reports, feed data to Business Intelligence (BI) tools, forecast trends, and train Machine Learning models. Data Warehouse stores data from multiple sources such as APIs, Databases, Cloud Storage, etc., using the ETL (Extract Load Transform) process.</p>																						
2.	<p>Why OLAP is used?</p> <ul style="list-style-type: none">• Online Analytical Processing (OLAP): Online Analytical Processing consists of a type of software tools that are used for data analysis for business decisions. OLAP provides an environment to get insights from the database retrieved from multiple database systems at one time.• Examples – Any type of Data warehouse system is an OLAP system. The uses of OLAP are as follows:• Spotify analyzed songs by users to come up with a personalized homepage of their songs and playlist.• Netflix movie recommendation system.																						
3.	<p>How is multidimensional data model used in data warehouse?</p> <ul style="list-style-type: none">• The multi-Dimensional Data Model is a method which is used for ordering data in the database along with good arrangement and assembling of the contents in the database.• OLAP (online analytical processing) and data warehousing uses multi dimensional databases. It is used to show multiple dimensions of the data to users.• It represents data in the form of data cubes. Data cubes allow to model and view the data from many dimensions and perspectives. It is defined by dimensions and facts and is represented by a fact table. Facts are numerical measures and fact tables contain measures of the related dimensional tables or names of the facts.																						
8.	<p>How can you tell the difference between fact and dimension tables?</p> <table><tr><th>S. NO.</th><th>Fact Table</th><th>Dimension Table</th></tr><tr><td>1</td><td>Fact table contains the measuring of the attributes of a dimension table.</td><td>Dimension table contains the attributes on that truth table calculates the metric.</td></tr><tr><td>2</td><td>In fact table, There is less attributes than dimension table.</td><td>While in dimension table, There is more attributes than fact table.</td></tr><tr><td>3</td><td>In fact table, There is more records than dimension table.</td><td>While in dimension table, There is less records than fact table.</td></tr><tr><td>4</td><td>Fact table forms a vertical table.</td><td>While dimension table forms a horizontal table.</td></tr><tr><td>5</td><td>The attribute format of fact table is in numerical format and text format.</td><td>While the attribute format of dimension table is in text format.</td></tr><tr><td>6</td><td>It comes after dimension table.</td><td>While it comes before fact table.</td></tr></table>		S. NO.	Fact Table	Dimension Table	1	Fact table contains the measuring of the attributes of a dimension table.	Dimension table contains the attributes on that truth table calculates the metric.	2	In fact table, There is less attributes than dimension table.	While in dimension table, There is more attributes than fact table.	3	In fact table, There is more records than dimension table.	While in dimension table, There is less records than fact table.	4	Fact table forms a vertical table.	While dimension table forms a horizontal table.	5	The attribute format of fact table is in numerical format and text format.	While the attribute format of dimension table is in text format.	6	It comes after dimension table.	While it comes before fact table.
S. NO.	Fact Table	Dimension Table																					
1	Fact table contains the measuring of the attributes of a dimension table.	Dimension table contains the attributes on that truth table calculates the metric.																					
2	In fact table, There is less attributes than dimension table.	While in dimension table, There is more attributes than fact table.																					
3	In fact table, There is more records than dimension table.	While in dimension table, There is less records than fact table.																					
4	Fact table forms a vertical table.	While dimension table forms a horizontal table.																					
5	The attribute format of fact table is in numerical format and text format.	While the attribute format of dimension table is in text format.																					
6	It comes after dimension table.	While it comes before fact table.																					

	7	The number of fact table is less than dimension table in a schema.	While the number of dimension is more than fact table in a schema.
	8	It is used for analysis purpose and decision making.	While the main task of dimension table is to store the information about a business and its process.
2.	Compare ROLAP and MOLAP.		
	S.NO.	ROLAP	MOLAP
	1	ROLAP stands for Relational Online Analytical Processing .	While MOLAP stands for Multidimensional Online Analytical Processing .
	2	ROLAP is used for large data volumes.	While it is used for limited data volumes.
	3	The access of ROLAP is slow.	While the access of MOLAP is fast.
	4	In ROLAP, Data is stored in relation tables.	While in MOLAP, Data is stored in multidimensional array.
	5	In ROLAP, Data is fetched from data-warehouse.	While in MOLAP, Data is fetched from MDDBs database.
	6	In ROLAP, Complicated sql queries are used.	While in MOLAP, Sparse matrix is used.
	7	In ROLAP, Static multidimensional view of data is created.	While in MOLAP, Dynamic multidimensional view of data is created.
3.	How is virtual data warehouse different from distributed data warehouse?		
	<p>Distributed data warehouse: The concept of a distributed data warehouse suggests that there are two types of distributed data warehouses and their modifications for the local enterprise warehouses which are distributed throughout the enterprise and a global warehouses.</p> <p>Virtual data warehouse: Virtual data warehousing uses distributed queries on several databases, without integrating the data into one physical data warehouse.</p>		
4.	<p>Why is it that tuning is difficult in a data warehouse?</p> <p>Tuning a data warehouse is a difficult procedure due to following reasons –</p> <ul style="list-style-type: none"> • Data warehouse is dynamic; it never remains constant. • It is very difficult to predict what query the user is going to post in the future. • Business requirements change with time. • Users and their profiles keep changing. • The user can switch from one group to another. • The data load on the warehouse also changes with time. 		

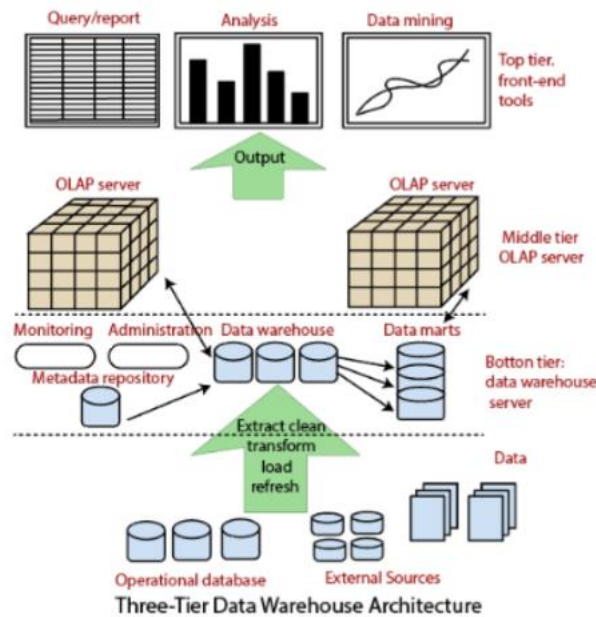
6.	<p>What are the challenges in ETL process?</p> <p>Some examples of warning signs in ETL processes include:</p> <ul style="list-style-type: none"> • Unexpected changes in data quality or structure. • Increased errors or failure rates in the pipeline. • Difficulty in maintaining or updating the pipeline. • Performance degradation in the pipeline. • Difficulty in identifying the root cause of pipeline issues.
1.	<p>What is the difference between KDD and data mining?</p> <p>KDD (Knowledge Discovery in Databases) is a field of computer science, which includes the tools and theories to help humans in extracting useful and previously unknown information (i.e., knowledge) from large collections of digitized data. KDD consists of several steps, and Data Mining is one of them. Data Mining is the application of a specific algorithm to extract patterns from data.</p> <p>KDD is the overall process of extracting knowledge from data, while Data Mining is a step inside the KDD process, which deals with identifying patterns in data.</p>
2.	<p>What do you understand by Data Mining?</p> <p>“Mining” is the process of extraction of some valuable material from the earth e.g. coal mining, diamond mining, etc. In the context of computer science, “Data Mining” can be referred to as knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. It is basically the process carried out for the extraction of useful information from a bulk of data or data warehouses. One can see that the term itself is a little confusing. In the case of coal or diamond mining, the result of the extraction process is coal or diamond. But in the case of Data Mining, the result of the extraction process is not data!! Instead, data mining results are the patterns and knowledge that we gain at the end of the extraction process. In that sense, we can think of Data Mining as a step in the process of Knowledge Discovery or Knowledge Extraction.</p>

Set 2:

4.	<p>What is the primary purpose of building the multidimensional model?</p> <ul style="list-style-type: none"> • Stage 1 : Assembling data from the client • Stage 2 : Grouping different segments of the system • Stage 3 : Noticing the different proportions • Stage 4 : Preparing the actual-time factors and their respective qualities • Stage 5 : Finding the actuality of factors which are listed previously and their qualities • Stage 6 : Building the Schema to place the data, with respect to the information collected from the steps above
7.	<p>Differentiate between a data warehouse and a data mart.</p>

	<p>A data warehouse stores data in a structured format. It is a central repository of preprocessed data for analytics and business intelligence. A data mart is a data warehouse that serves the needs of a specific business unit, like a company's finance, marketing, or sales department.</p>
11	<p>Give the differences between a database and a data warehouse.</p> <ul style="list-style-type: none"> • A database stores the current data required to power an application. • A data warehouse stores current and historical data from one or more systems in a predefined and fixed schema, which allows business analysts and data scientists to easily analyze the data.
12	<p>How role of operation makes changes on data cube?</p> <ul style="list-style-type: none"> • Rollup – decreases dimensionality by aggregating data along a certain dimension • Drill-down – increases dimensionality by splitting the data further • Slicing – decreases dimensionality by choosing a single value from a particular dimension • Dicing – picks a subset of values from each dimension • Pivoting – rotates the data cube
1.	<p>What are the major applications of data warehousing?</p> <ul style="list-style-type: none"> • Financial services • Banking services • Consumer goods • Retail sectors • Controlled manufacturing
5.	<p>What are major issues that will be faced in data warehouse development?(correct ans aa nu teriyala guys. Net pathu eduthathu..vena vera ans naalum padichikonga 😊)</p> <ul style="list-style-type: none"> • Challenge 1: Managing the data structure and optimization. ... • Challenge 2: Managing user expectations. ... • Challenge 3: The costs of data warehousing. ... • Challenge 4: Data quality. ... • Challenge 5: Data accuracy. ... • Challenge 6: Adjusting to non-technical users.
6.	<p>What are the challenges in ETL process?</p> <p>Some examples of warning signs in ETL processes include:</p> <ul style="list-style-type: none"> • Unexpected changes in data quality or structure. • Increased errors or failure rates in the pipeline. • Difficulty in maintaining or updating the pipeline. • Performance degradation in the pipeline. • Difficulty in identifying the root cause of pipeline issues.

7. What are the three tiers of data warehouse architecture?



3. Why is KDD important?

KDD is very important for organizations and businesses as it enables them to derive new insights and knowledge from their data, which can be further used to improve decision-making, enhance the customer experience, improve business processes, support strategic planning, optimize operations, and drive business growth.

4. What are the major issues in data mining?

1. Mining different kinds of knowledge in databases
2. Interactive mining of knowledge at multiple levels of abstraction
3. Incorporation of background knowledge
4. Data mining query languages and ad-hoc data mining
5. Presentation and visualization of data mining results
6. Handling noisy or incomplete data
7. Pattern evaluation
8. Efficiency and scalability of data mining algorithms
9. Parallel, distributed, and incremental mining algorithms