

1. Project Architecture

Contents

| | |
|--|-----------|
| 1. Project Architecture | 2 |
| 2. Software application | 2 |
| 3. Different kind online applications | 2 |
| 4. Generating data..... | 3 |
| 5. Project architecture components | 4 |
| 6. User Interface | 5 |
| 6.1. Different UI technical stack..... | 5 |
| 7. Database | 6 |
| 7.1. Different Database technical stack | 6 |
| 8. Service layer | 8 |
| 8.1. Different Service layer technical stack..... | 8 |
| 9. Data Ware House | 9 |
| 9.1. Different Service layer technical stack..... | 10 |
| 10. OLTP and Data Ware House | 11 |
| 11. ETL Tools | 12 |
| 11.1. Different ETL Tools..... | 12 |
| 12. Cluster Storage | 13 |
| 12.1. Different type of clusters | 13 |
| 13. So, Client never bothers... .. | 15 |
| 14. Visualization tools | 16 |
| 14.1. Different type of visualization tools..... | 16 |
| 15. Client questions | 17 |
| 16. Data Science Team | 18 |

1. Project Architecture

1. Project Architecture

- ✓ Any software application should follow architecture.
- ✓ Project Architecture means, different **layers** in project with **flow** diagram.
- ✓ Here layers can be called as tier too.
- ✓ Suppose our project contains 3 layer like,
 - Client layer.
 - Business layer.
 - Data layer & etc.

2. Software application

- ✓ Software application performs the specific task for the users by interacting directly.
 - **Banking** software application helps to customers quickly to finish common activities over online.
 - **Ecommerce** software application helps to customers to order the products over online.

3. Different kind online applications

- ✓ Banking applications.
- ✓ E-commerce applications.
- ✓ Telecom applications.
- ✓ Insurance applications.
- ✓ Healthcare applications & etc.

4. Generating data

- ✓ Every minute all companies are generating data.
- ✓ Data is more valuable to every company.
- ✓ So, every company supposed to store this data in place.
- ✓ Based on requirement, the stored data can be process for future analysis.

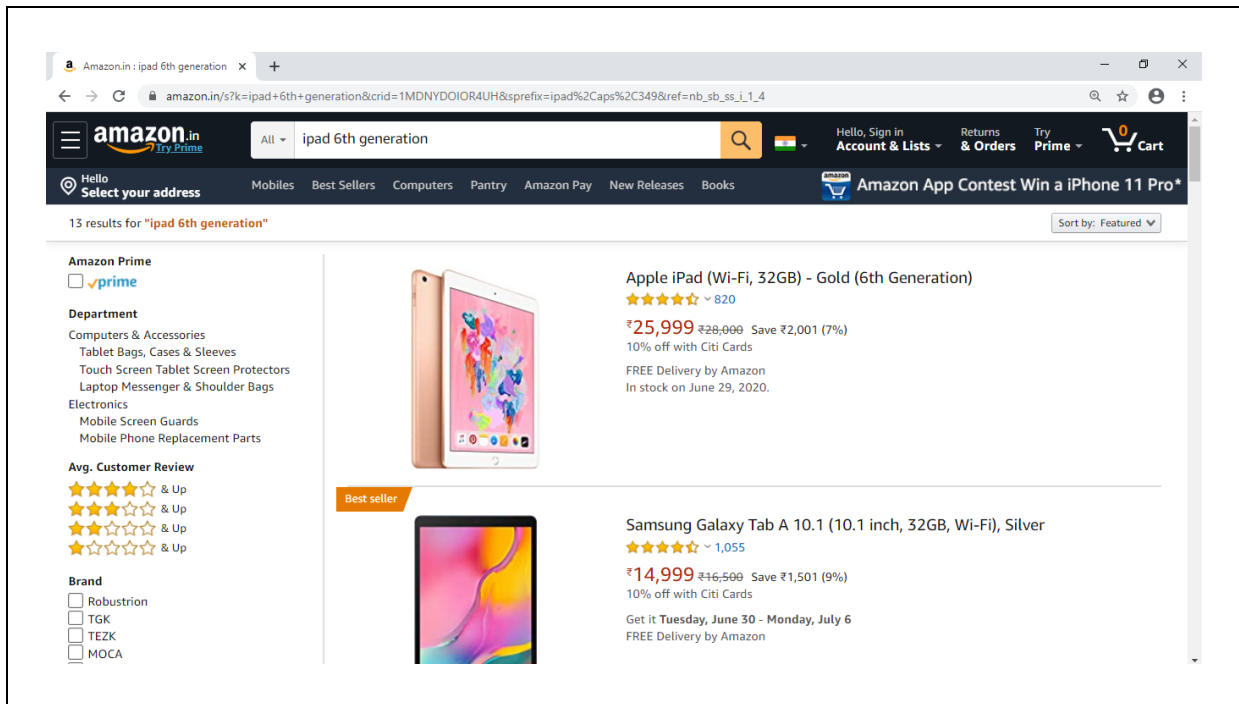


5. Project architecture components

- ✓ Any software application will follow one common architecture.
- ✓ This architecture having different components as below,
 - User Interface.
 - Database.
 - Service layer.
 - Data Ware House.
 - ETL tools.
 - Storage Cluster.
 - Admin team.
 - Visualization tools.
 - Data Science Team.

6. User Interface

- ✓ The user interface (UI) is the point of **human-computer** interaction channel.
- ✓ Communication will be done by using any device.
- ✓ So, generally customers used to interact with software applications by using User Interface.
- ✓ Example:
 - Assuming that a customer wanted to buy one product in amazon then he needs to open amazon website, search product, if interested then **proceed to buy**
 - If customer bought the product then transaction details will be stored in **database**.



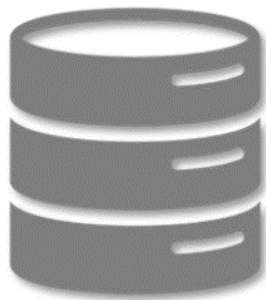
- ✓ This User Interface will be implemented with different technical stack.

6.1. Different UI technical stack

- ✓ Html
- ✓ Java script
- ✓ jQuery
- ✓ AngularJS & etc.

7. Database

- ✓ Every company needs to store the collected data for future analysis.
- ✓ Companies are using database software to store the data.
- ✓ These are also called as OLTP servers.
- ✓ The full form of OLTP is OnLine Transaction Processing.
- ✓ This OLTP uses a Relational Database Management Systems.



What work will be done by OLTP?

- ✓ OLTP is a process of storing online transactional data.
- ✓ These transactions are always related to inserting, deleting, updating data into the database.
- ✓ There are different types of databases.

7.1. Different Database technical stack

- ✓ Oracle
- ✓ MySQL
- ✓ Mongo DB
- ✓ Cassandra & etc.

Where OLTP is good?

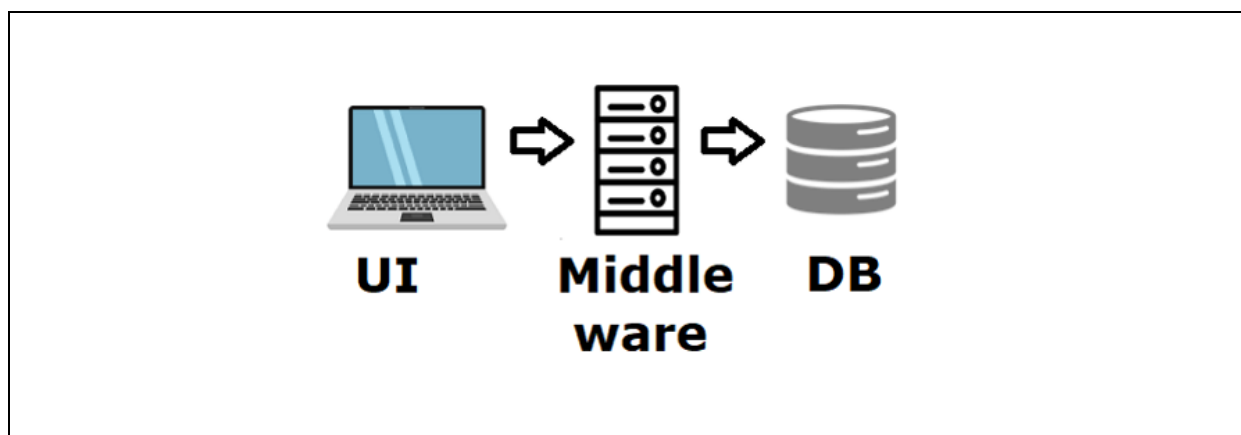
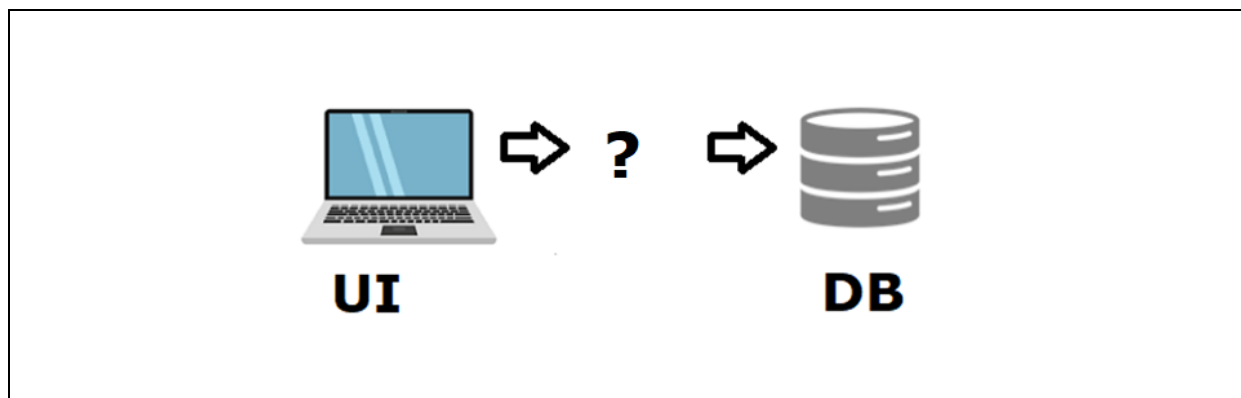
- ✓ OLTP server works perfectly for online transactions.

Limitations of OLTP servers

- ✓ OLTP servers cannot store huge data.
- ✓ Based on data storage policy, every OLTP server will be store only a couple of year's data or maximum three years of data.

8. Service layer

- ✓ Interaction between User Interface and Database will be done by using service layer
- ✓ In simple, a service layers works here to interact from UI to Database.
- ✓ Service layer main objective is ,
 - Transfers the information from UI to Database.
- ✓ This layer will be implemented with different technical stack.

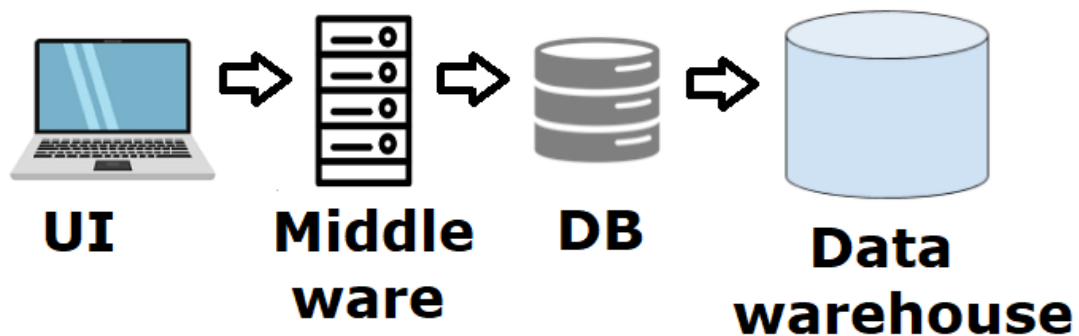
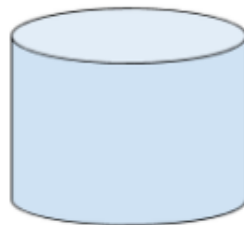


8.1. Different Service layer technical stack

- ✓ Java
- ✓ Dot net
- ✓ Python
- ✓ Php & etc.

9. Data Ware House

- ✓ As discussed, Database having storage limitations.
- ✓ To overcome this limitation Data Ware House came into the picture.
- ✓ Companies are using a separate environment to store historical data, this is called as Data Ware house
- ✓ Data Ware House means Relational Database Management Systems only but it's more powerful compare to databases.
- ✓ There are different types of databases.



9.1. Different Data ware house technical stack

- ✓ Teradata
- ✓ Netezza
- ✓ Vertica & etc.

Limitations in Data Ware House

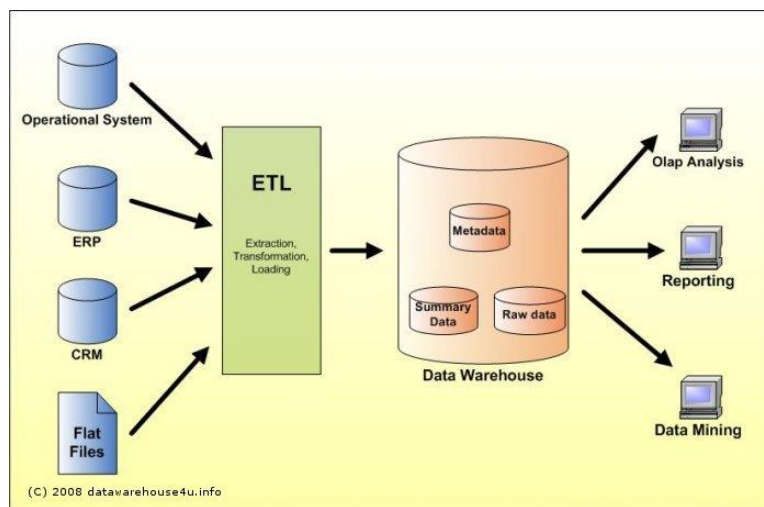
- ✓ Based on data storage policy, these Data Ware House works best if the data is less than or equal to 200 Tera Bytes.
- ✓ If data is beyond 200 TB then these Data Ware Houses are not fit to store.

10. OLTP and Data Ware House

- ✓ OLTP
 - Recent transactions will be stored in OLTP servers.
 - Money withdraw information.
 - Booking train ticket details.
- ✓ Data Ware House
 - Users last 10 years transactions will be store in Data Ware House.

11. ETL Tools

- ✓ The full form of ETL is,
 - Extract Transform Load.
- ✓ The main task of ETL team is,
 - Get data from OLTP, send this data to Data Ware House.
 - Technically speaking, connecting to the databases and fetching the data, processing the data and loading into target databases.
- ✓ Extract Transform Load will be done by ETL tools.

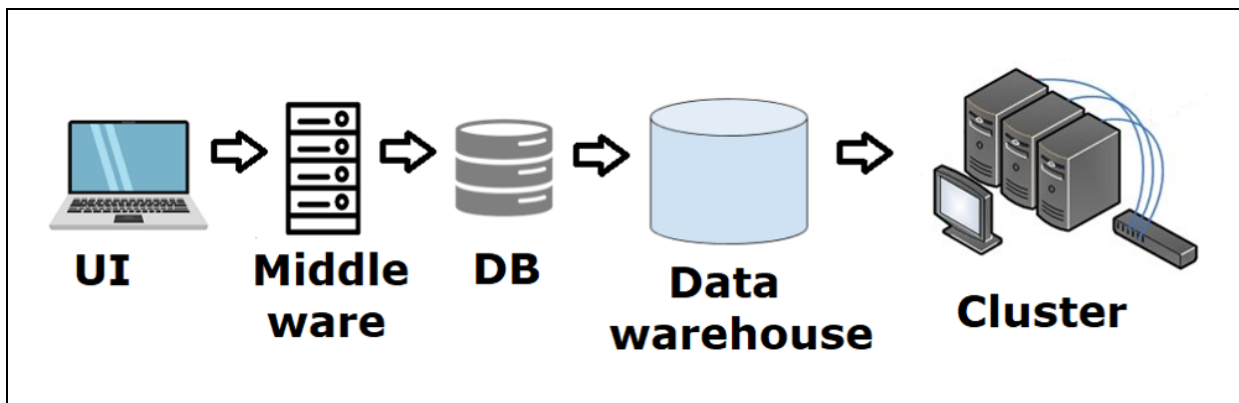
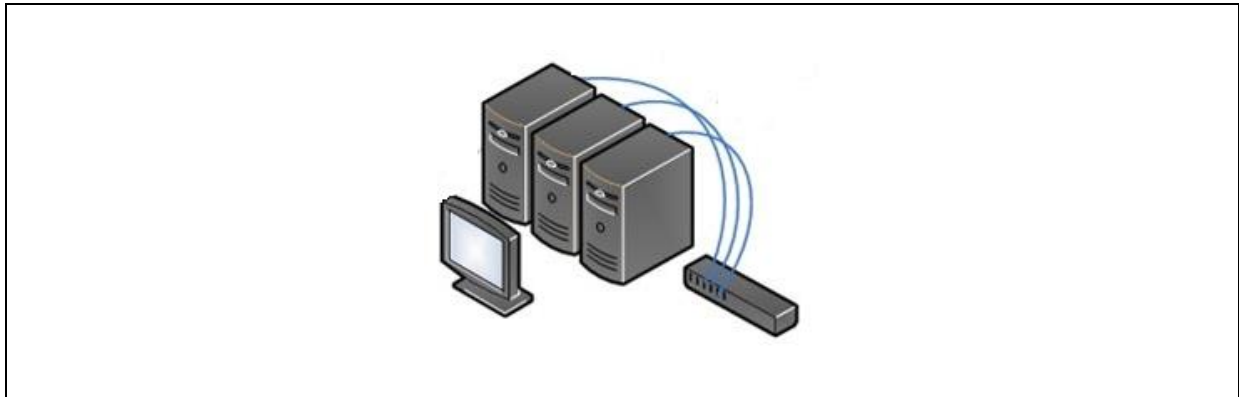


11.1. Different ETL Tools

- ✓ Extract Transform Load will be done by different kind of tools.
 - Informatica
 - DataStage
 - Abinitio
 - Pentaho
 - Talend & etc.

12. Cluster Storage

- ✓ **Node** is a computer device to store and process the data.
- ✓ **Rack** is a group of interconnected nodes.
- ✓ **Cluster** is a group of interconnected racks.
- ✓ Companies using clusters to store very HUGE data.
- ✓ Clusters are very high performance while processing the data.



12.1. Different type of clusters

- ✓ One of the best example for cluster is **HDFS**
- ✓ HDFS is related to Hadoop concept.
- ✓ The full form of HDFS is **H**adoop **D**istributed **F**ile **S**ystem.
- ✓ HDFS is used to store huge data like,
 - Terabytes
 - Petabytes
 - Exabyte & etc

Spark

- ✓ Spark is framework/engine to process huge data

BigData developer

- ✓ Access and process historical data and generate the results from cluster.
 - Hadoop & PySpark developer

13. So, Client never bothers...

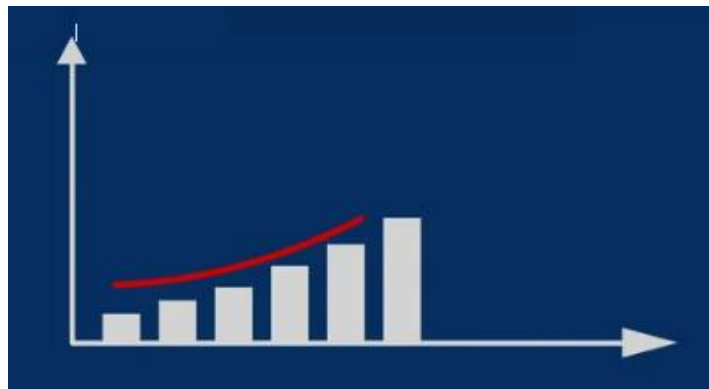
- ✓ Client never bothers about these technical things.
- ✓ Let's take an example,
 - Amazon owner needs info like, yesterday and today how much business (sales) done.
 - Occasionally he may expect visualization reports too.

Why visualization reports why not manual reports?

- ✓ Manual reports take a lot of time to read.
- ✓ Visualization reports can understand very easily

14. Visualization tools

- ✓ Visualization tools helps to visualize the data.
- ✓ Assuming that last 6 years reports,
 - First box indicates 2017 sales info.
 - Second box indicates 2018 sales info.
 - Third box indicates 2019 sales info & etc
- ✓ So, if we observe clearly then every year business is increasing.



14.1. Different type of visualization tools

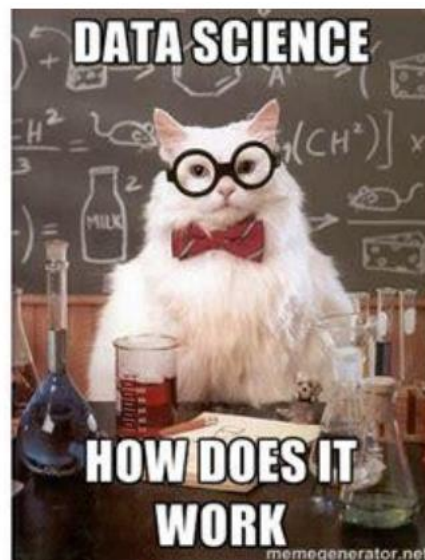
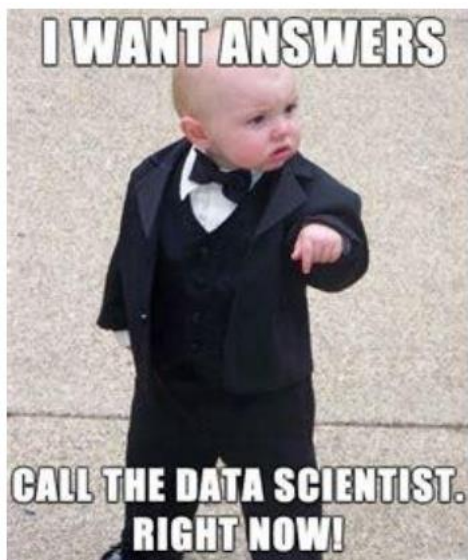
- ✓ Tableau
- ✓ PowerBI
- ✓ Quickview
- ✓ SAP BO

15. Client questions

- ✓ Assuming that client is asking few questions like,
 - Sales comparison of last 3 years.
 - If sales increase then, **why** sales got **increase**?
 - If sales decrease then, **why** sales got **reduced**?
 - How to **improve business** in upcoming year?
 - What are the **top ten reasons** which explain about business got **increased**?
 - What are the **top ten reasons** which explain about business got **reduced**?

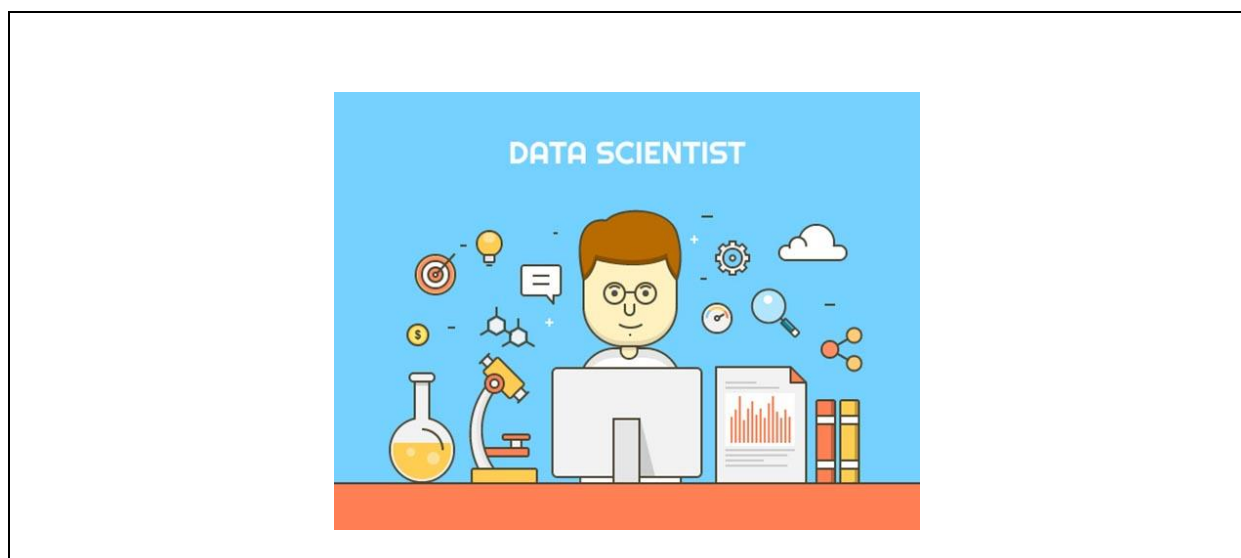
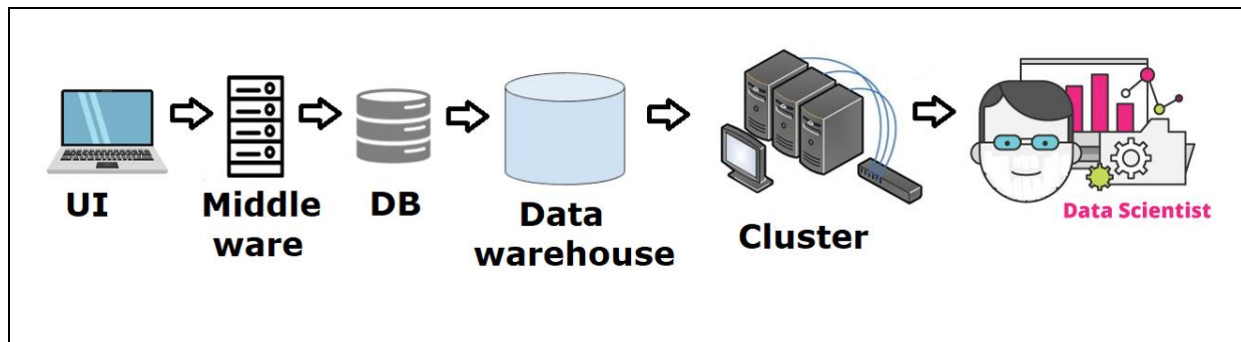
Who will answer for above questions?

- ✓ UI team?
- ✓ Middleware team?
- ✓ Database team?
- ✓ ETL team?
- ✓ Visualization team?
- ✓ BigData dev team?



16. Data Science Team

- ✓ A Data Science team will answer above question with practically.
- ✓ Data scientist prepares and trains the model
- ✓ This helps to find out root cause of everything.



Data Scientist responsibility...

- ✓ Data Scientist responsibility is,
 - **Train** the models.
 - Model **acquires** the knowledge
 - Test the **accuracy** of the trained model.
 - If satisfied with accuracy then **deploy** the model