1) Dynamic Schema:- Detecting the schema of the data that is entered into the system. the schema is flexible and allows easier modification of the database.

→ NoSQL databases - MongoDB, CouchDB, CassandraDB support dynamic schema - schema-less / semi-structured data.

Usecase:→ The data generated by IOT devices
The data generated by sensors, the schema is detected after receiving data from various locations. The data is from various environmental parameters such as temperature, humidity, air quality, noise level - data based on the dynamic schema defined by the sensor capabilities.

2) The 5 V's of big data are:-
1) Volume - the amount of data
2) Velocity - the speed at which data is generated
3) Variety - the different forms and formats of data.
4) Veracity - the uncertainity and unreliability of data, important to validate the data
5) Value - insights and knowledge that can be gained from analyzing the data.

Usecase of Big data:-
Uber/Lyft The ride-sharing (riding) applications generates lots of data every day. The data includes cars information, passengers info, riders info, rides info, coupons / promo codes info street traffic data, GPS data, trips data.

3) The 3 types of cloud are ① Public ② Private ③ Hybrid
① Public cloud is a type of hosting in which cloud services are delivered over a network for public use.
② Private cloud is a type of hosting in which cloud services are solely used by one organization. → great control over security, designed for specific organization.
③ Hybrid cloud is a type of hosting where they have dual opportunities to store private info and info used by organization members on private cloud and the info for customers, marketing, and other public info on public cloud.

→ The on-prem and cloud infrastructure choosing is based upon different factors - cost, control, security, compliance

We choose On-prem environment when it is a legacy system, data security concern, have enough money and space to maintain.

We choose cloud environment when we consider cost, maintenance, space, security, accessibility - any time usage. Only concerns about accessing the resources, the other major issues are managed by cloud providers.

4) The different file formats are parquet, json, .csv, avro, ORC, spreadsheets.

Examples of structured data → schema on write - excel files, csv files, json data

Unstructured data → schema on read - text documents, images, audio & video files, social media posts

Semi-structured data → xml files, json files including nested arrays and objects, NoSQL databases, sensors collected data

5) Scalability is the ability of a system to handle the amount of load or demand by adding or reducing resources

→ Horizontal Scalability = called as Scaling Out → dividing the load over distributed machines to handle increased load for effective processing.
Aws uses load balancer to distribute the incoming load of a website over different servers equally to prevent bottleneck

→ Vertical Scalability = called as Scaling up → increasing the capacity of a single machine by adding more resources (CPU, memory, storage)
A company with own data center increases the resources if amount of data and user activity grows.