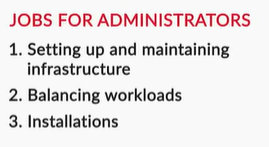
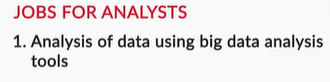
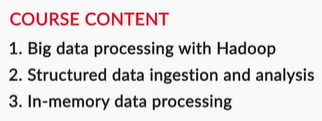
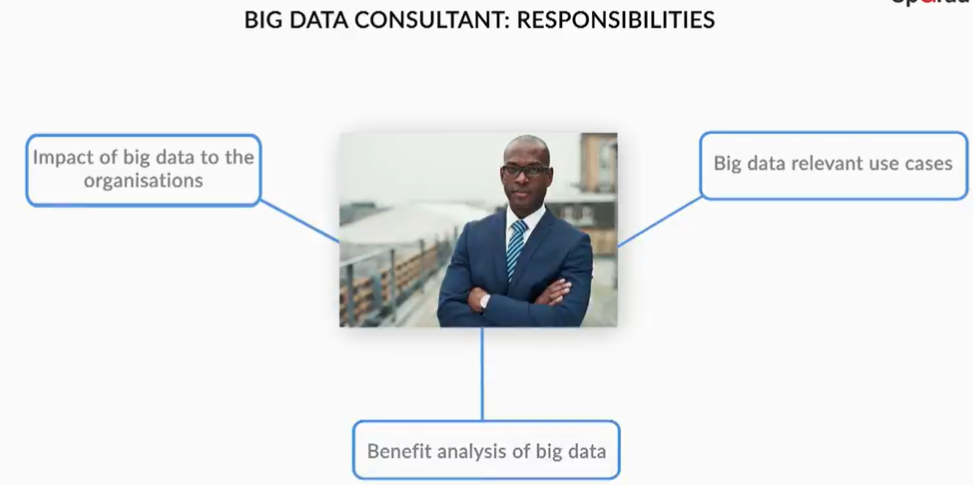
Understanding Big Data:

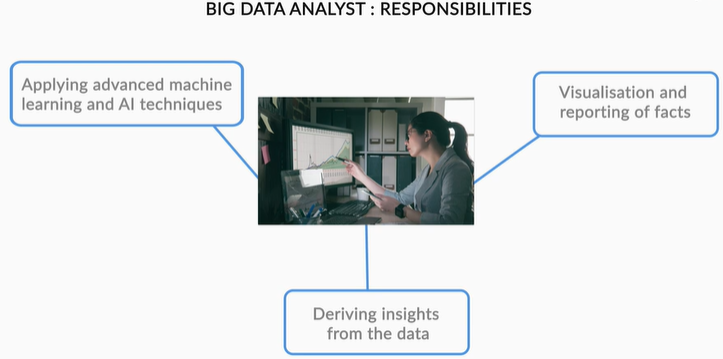
Big Data Job Roles  
  

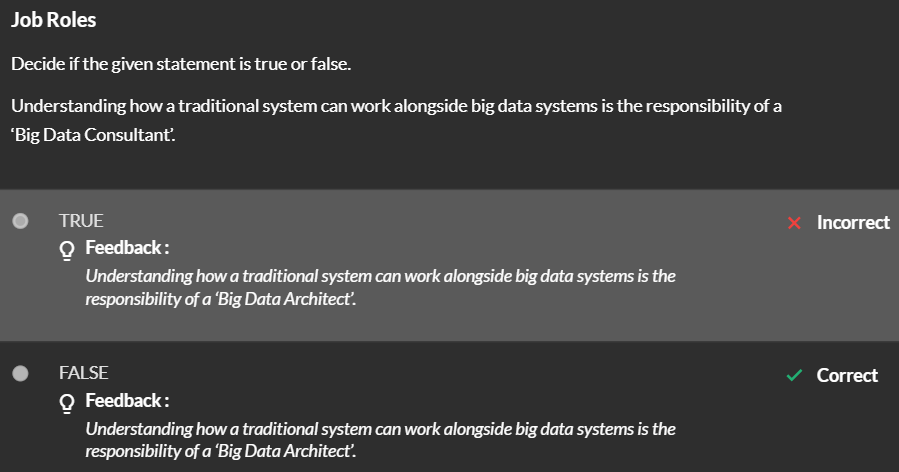




Consultant Roles  


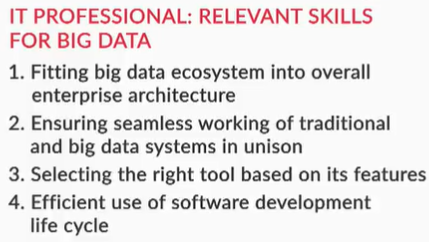
Architect Roles   

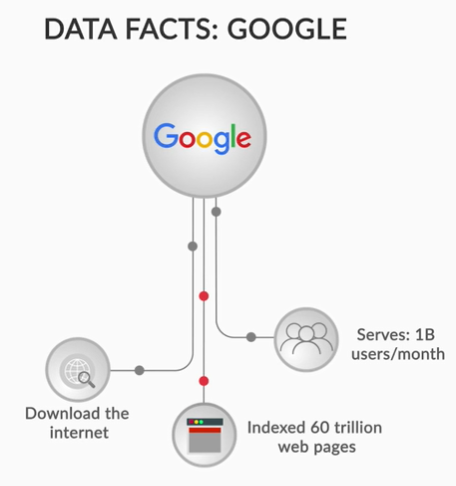
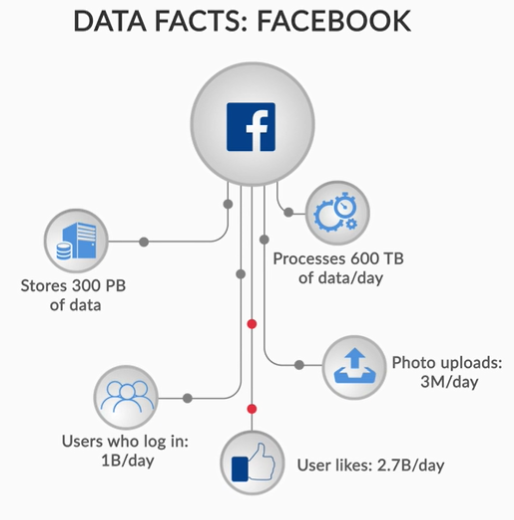

Big Data Analyst:  






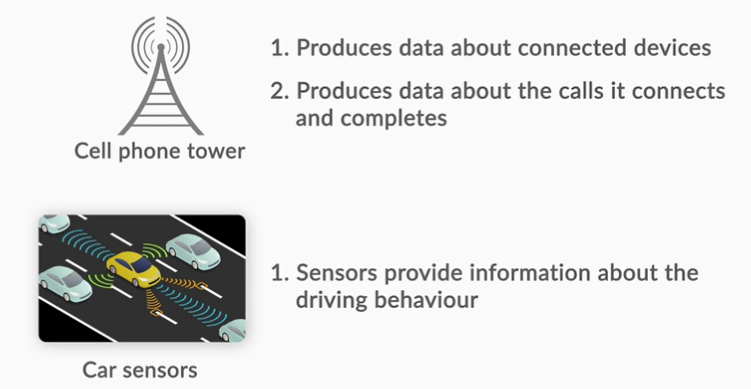






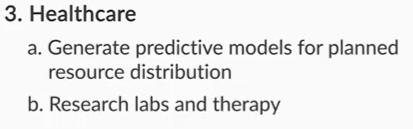
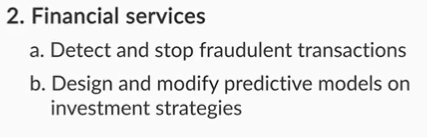
Sources of Big Data

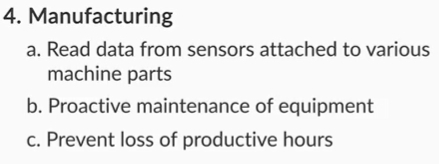


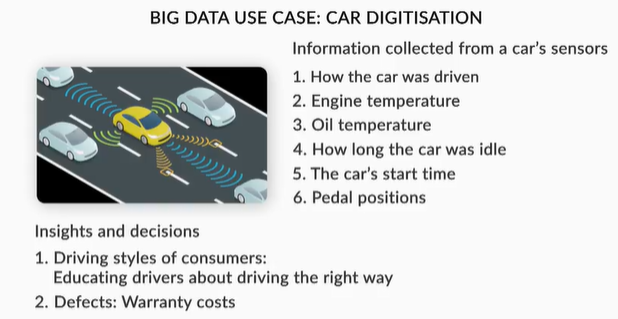


Big Data – Industry Use cases



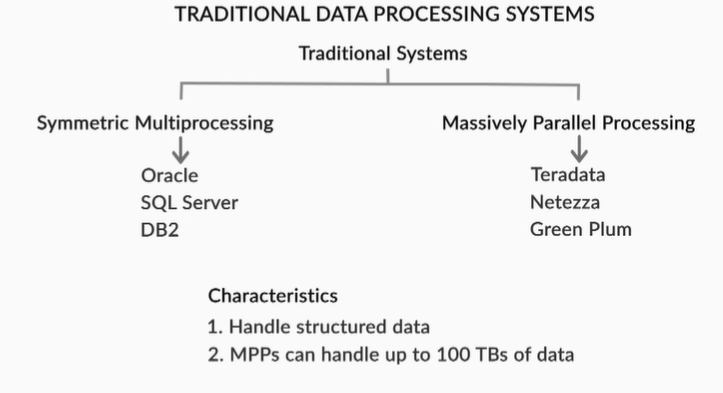


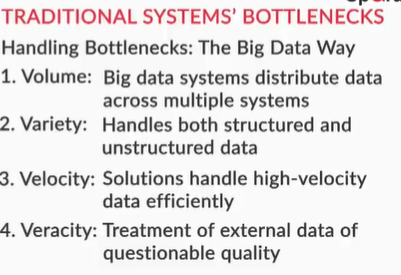




The four Vs of big data include:

* **Volume**: This represents the amount of data generated by a company/organisation. The size of big data typically ranges from petabytes (1000 TBs) to exabytes (10^6 TBs). For example, remember the amount of data generated in an internet minute, which leads to the data explosion. Search engines such as Google, Yahoo, etc. deal with enormous volumes of data.
* **Velocity**: This indicates the rate at which data is generated/consumed. Social media sites such as Twitter, Facebook, etc. generate data from every activity that a user undertakes, leading to an enormous amount of data every minute.
* **Variety**: This represents the different types of data being generated. For example, the different data forms for Gmail may be sign-up/registration data, user login data, inbox emails, sent emails, etc.
* **Veracity**: This represents the quality and accuracy of the data. Previously, veracity was not considered to be a characteristic of big data. But with increasing analyses on generated data, veracity plays an important role.





You also understood the difference between **structured, semi-structured, and unstructured data:**

* **Structured data**typically has a row-and-column type of structure. The tables in an RDBMS are examples of structured data.
* **Semi-structured**data has a defined structure, though it is often non-tabular. Popular examples of semi-structured data are XML and HTML files.
* **Unstructured data**does not have any specific structure or formatting. **Video files, log records,** and **PDF documents** are typical examples of unstructured data.