📘 Lesson-04: What's Cloud

Summary

Cloud computing is a powerful tool for data scientists. It helps store, process, and analyze large datasets using external servers and tools. This eliminates the need for high-end local machines. Cloud platforms also allow teams around the world to collaborate in real time. IBM's Data Scientist Workbench is one such cloud-based tool that supports learning and working with big data using technologies like R, OpenRefine, and Spark.

Key Notes

- **Cloud** is a remote storage and computing environment accessible via the internet.
- It helps overcome **local system limitations** by offering powerful processing and storage externally.
- Enables **collaborative work** multiple users can access and analyze the same data simultaneously.
- Common cloud services for data science include **storage, processing, algorithms, and computing power**.
- **Data Scientist Workbench (by IBM)** provides tools like:
  - **R and RStudio** for statistical analysis.
  - **OpenRefine** for data cleaning.
  - **Apache Spark** for big data processing.
- Benefits of cloud tools:
  - No need for installation, updates, or maintenance.
  - Scalable for small and large datasets.
  - Accessible from anywhere via the internet.

What I Learned

I now understand that cloud computing is essential for modern data science. It allows me to handle massive datasets and run complex algorithms without needing expensive hardware. I also learned how cloud platforms like IBM's Data Scientist Workbench make data science tools easily accessible and user-friendly.

How I Will Use This

In my projects, I will use cloud platforms like Google Colab, AWS, or IBM's tools to manage large datasets and run machine learning models. I can also collaborate with others more easily and avoid the technical hassle of local setup and performance limits.
