Data Engineering

Data Engineering Introduction

“Data engineering refers to the building of systems to enable the collection and usage of data. This data is usually used to enable subsequent analysis and data science; which often involves machine learning.**”**

We have different Data Engineering Technologies. They are:

1. Apache Data Engineering
2. AWS Data Engineering (Amazon)
3. Azure Data Engineering (Microsoft)
4. GCP Data Engineering (Google)
   Etc..

Apache Data Engineering

1. Apache Hadoop (Storage & Processing)
2. Apache Hadoop Frameworks or Ecosystems
   Apache Hive
   Apache HBase
   Etc..
3. Apache Kafka
4. Apache Spark (Java or Scala or Python or R)
5. Apache Hadoop Cluster or Cloudera Hadoop Cluster or MapR Hadoop Cluster

AWS Data Engineering

1. AWS Basic Services
2. AWS S3
3. AWS RDS
4. AWS Glue
5. AWS Athena
6. AWS Redshift
7. AWS Kenesis
   Other AWS Service

Azure Data Engineering

1. Azure Basic Services
2. Azure SQL
3. Azure Storage
4. Azure Data Factory
5. Azure Synapse

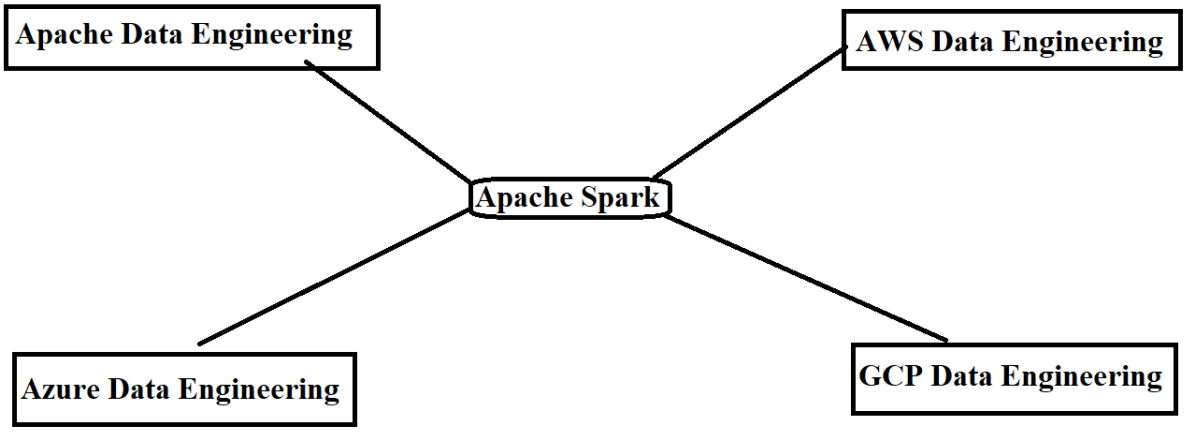
6. Azure Streams
   Etc..

GCP Data Engineering:-

1. GCP Basic Services
2. GCP Cloud Storage
3. GCP Database
4. GCP Data Processing
5. GCP Pub/Sub
   Etc..

Importance of Spark:-

In All Data Engineering Technologies, For Data Processing, Industry using "**Spark".**
Spark we can implement using Java (Spark with Java) or Scala (Spark & Scala) or Python **(PySpark)** or R (SparkR)



Pre-Requisites for Data Engineering:-

1. **Language** – Java or Scala or **Python**
2. **SQL** – SQL with Any RDBMS (MySQL or Oracle or PostgreSQL or M.S SQL Server etc..)
3. **Operating System –** Linux Essentials, Shell Scripting (Optional but Recommended)
4. **Any Cloud Basic Knowledge**

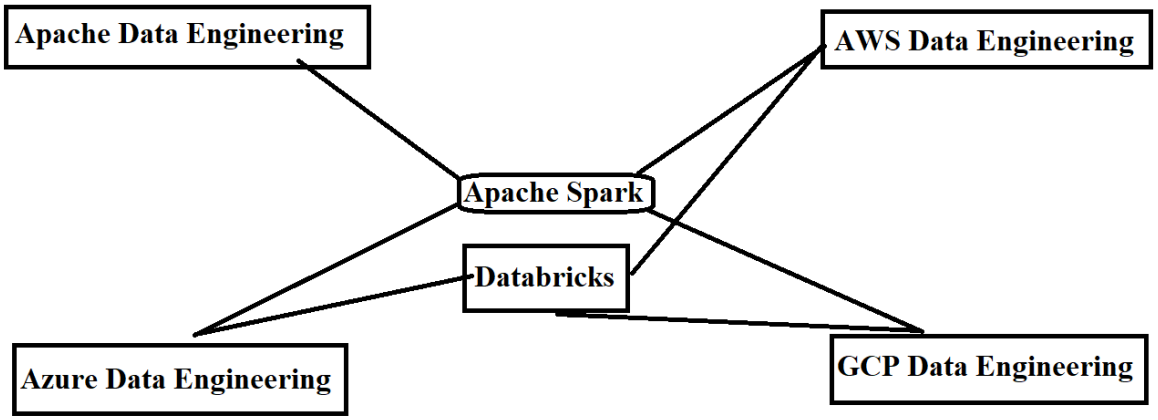
Importance of Apache Airflow:-

Shell Scripting + Crontab ➔ Used for defining workflow & Scheduling Workflows
Apache Oozie
Apache Airflow (Python based Framework)
Cloud Vendors also has their own frameworks

Data Engineering

Databricks:-

➔ Databricks currently support 3 Cloud Vendors. They are AWS, Azure & GCP.



By

Akkem Sreenivasulu – Founder of CFAMILY IT Solutions

Website: www.cfamilyit.com
eMail: info@cfamilyit.com
Contact/Msg/WhatsApp: +91-9133151144, +91-9133161144
Telegram Channel: https://t.me/cfamilyit
YouTube: www.youtube.com/@cfamilyit
Facebook: https://www.facebook.com/cfamilyit/
Instagram: https://www.instagram.com/cfamilyit/
LinkedIn: https://www.linkedin.com/in/cfamilyit
Twitter: https://twitter.com/cfamilyit
GitHub: https://github.com/cfamilyit/

CFAMILY IT Solutions

cfamily IT

Learn from Experts & Feel the difference.

Data Engineering

GCP (Google) Data Engineering:-

GCP Basic Services

- GCP Regions & Zones
- GCP Services Introduction
- IAM
- Provisioning Compute Engine
- Deploy Application on App Engine
- etc..

GCP Cloud Storage

- Google Cloud Storage
- Cloud Block Storage
- etc..

GCP Database

- Google Cloud SQL
- Cloud Filestore
- Cloud Datastore
- Cloud Memorystore
- Cloud Bigtable
- etc..

Data Engineering

GCP Data Processing

- Google Cloud Pub/Sub
- Google Bigquery
- Cloud DataFlow
- Cloud DataProc
- Cloud Data Fusion
- etc..

CFAMILY IT Solutions

cfamily IT

Learn from Experts & Feel the difference.

Data Engineering

AWS Data Engineering

1. Introduction
2. Python for Data Engineering
3. Linux Essentials for Data Engineering
4. SQL for Data Engineering
5. AWS Services for Data Engineering
6. AWS S3
7. PySpark Cluster Setup on AWS
8. PySpark Development
   PySpark Core Programming
   PySpark SQL Programming
9. AWS Kenesis
10. PySpark Streaming
11. AWS Redshift
12. PySpark Integrations
    PySpark Integration with AWS S3
    PySpark Integration with AWS Kenesis
    PySpark Integration with AWS Redshift
13. Databricks with AWS Introduction
14. AWS Glue Introduction
15. AWS Athena Introduction
    Etc..

Data Engineering

By
Akkem Sreenivasulu – Founder of CFAMILY IT Solutions
Website: www.cfamilyit.com
eMail: info@cfamilyit.com
Contact/Msg/WhatsApp: +91-9133151144, +91-9133161144
Telegram Channel: https://t.me/cfamilyit
YouTube: www.youtube.com/@cfamilyit
Facebook: https://www.facebook.com/cfamilyit/
Instagram: https://www.instagram.com/cfamilyit/
LinkedIn: https://www.linkedin.com/in/cfamilyit
Twitter: https://twitter.com/cfamilyit
GitHub: https://github.com/cfamilyit/

CFAMILY IT Solutions

cfamily IT

Learn from Experts & Feel the difference.

Data Engineering