

M4 DATA ENGINEER : Esquema Teórico

1. Big Data - Hadoop y componentes Core

- Data Lake
- Hadoop
- Componentes Core: HDFS, Yarn, MapReduce

2. Frameworks Hadoop y Virtualización

- Frameworks Hadoop
- Virtualización: Maquinas Virtuales y Docker

3. Hive

- Hive: Tipos de Tablas y Datos, Formatos de Almacenamiento, Particiones
- Hue (Hadoop User Experience)
- Gobernanza del Dato

4. Formatos de Almacenamiento

- Formatos de Elección: CSV, JSON, Avro y Parquet

5. SQL - NoSQL y Teorema CAP

- SQL
- Nosql
- ACID Vs BASE
- Teorema CAP

6. Motores NO-SQL

- Tipos de Motores de Bases de Datos NoSQL: DynamoDB, Hbase, Cassandra, Neo4j y MongoDB

7. Spark [Procesamiento Streamming vs Procesamiento Batch]

- Apache Spark
- Arquitectura Spark
- Módulos Spark
- Catalyst Optimizer
- Procesamiento Batch
- Procesamiento Streamming - Herramientas

8. Orquestadores de flujos de datos

- Flujos de Trabajo
- DAG
- Notación CRON
- Oszie
- Nifi
- Airflow