

# Big Data con Hadoop y Spark

Módulo 01 – Frameworks

# Frameworks

# Descripción de Frameworks

<b>Apache Hive</b>	Consultas SQL sobre Hadoop
<b>Apache Sqoop</b>	Transferencia de datos entre bases relacionales y Hadoop
<b>Apache Spark</b>	Procesamiento en memoria de ETL's, Streaming, Machine Learning y Grafos
<b>Apache Kafka</b>	Sistema de colas de mensajería que utiliza el patrón productor/consumidor
<b>Apache HBase</b>	Base de datos NoSQL de tipo columnar que se ejecuta sobre HDFS
<b>Apache Ranger</b>	Administración de políticas de seguridad sobre componentes de Hadoop
<b>Apache Atlas</b>	Herramienta que provee funcionalidades de Data Governance sobre Hadoop
<b>Apache Nifi</b>	Orquestación de flujos de datos desde y hacia Hadoop

# Vendors (Hadoop empresarial)

- Cloudera <https://www.cloudera.com/>
- Amazon EMR <https://aws.amazon.com/es/emr/>
- Azure HDInsight <https://azure.microsoft.com/es-es/services/hdinsight/>
- IBM Analytics Engine <https://cloud.ibm.com/catalog/services/analytics-engine>
- Google Dataproc <https://cloud.google.com/dataproc>
- MapR <https://mapr.com/>

# Otras Tecnologías Big Data



## Revisión

- ¿Qué es Big Data?
- ¿Qué es un Data Lake?
- ¿Qué es Hadoop?
- ¿Cuáles son los componentes core de Hadoop?
- ¿Qué significa escalabilidad horizontal?



# ¡Muchas gracias!

## ¡Sigamos trabajando!