

Agenda

Big Data y sus Tecnologías	8
Plan	9
Introducción	9
Big Data – Una Definición	10
Big Data – Definición	11
Big Data - ¿Por qué?	12
Big Data - ¿Por Qué?	13
Disponibilidad de Datos	13
Datos Estructurados vs No Estructurados	14
Capacidad Analítica	15
Big Data - ¿Por qué?	15
Aplicaciones	15
Técnicas y Tecnologías	16
Sistemas de Archivos Distribuidos	17
Scale up vs. Scale Out	17
Sistemas de Archivos Distribuidos HDFS	17
Sistemas de Archivos Distribuidos - S3	19
Sistemas de Archivos Distribuidos CEPH	20
Técnicas - Algoritmos Distribuidos	21
Técnicas - Algoritmos Distribuidos	22
MapReduce	22
Spark	24
Flux vs. Batch	25
Apache Storm	25
Apache Samza	26
Spark Streaming	26
Transformaciones de DStream	27
Flux vs. Batch (Flujo vs Lote)	27
Tecnologías – Sistemas BD	28
Sistemas de Bases de Datos	29
NoSQL	29
Documento	30
Key-value (Clave-valor)	30
HBase	31
Cassandra	32
Grafos	33
Estructurado	34
Kudu	34
CockroachDB	34
Otros	35
Sistemas de Bases de Datos	35

Tecnologías - Sistemas de Organización	37
Sistemas de Organización	38
Dark Data (Datos Oscuros)	39
Dark Data (Datos Oscuros)	40
1. Sistemas Tradicionales de Gestión de Bases de Datos	40
2. Seguridad de Datos Tradicional	40
3. ¿Qué tan valioso es Dark Data?	41
4. Características de los Dark Data	41
Conclusión	42
Práctica 1 - Procesamiento por Lotes con Hadoop HDFS y Map Reduce	43
Objetivos del Trabajo Práctico	44
Hadoop	44
Introducción a Hadoop	46
Interfaces Web para Hadoop	48
Map Reduce	49
Wordcount	49
Launch Map Reduce on the Cluster	52
Práctica 2 - Procesamiento por Lotes y Streaming con Spark	56
Objetivos del Trabajo Práctico	57
Spark	57
Testing Spark with Spark-Shell	59
Spark API	60
Spark Batch en Java	65
Pruebas del Código Local	67
Lanzamiento del Código en el Clúster	68
Spark Streaming	70
Práctica 3 - Recopilación de Datos con el Kafka Bus	75
Objetivos del Trabajo Práctico	76
Kafka	76
Arquitectura de Kafka	78
Kafka y Zookeeper	79
Primer Uso de Kafka	80
Creación de una Aplicación Personalizada	82
Integración de Kafka con Spark	85
Práctica 4 - Almacenamiento de Datos en una Base de Datos NOSQL con HBase	89
Objetivos del Trabajo Práctico	90
Apache HBase	90
Modelo de Datos	90
Arquitectura	91
Primera Manipulación de HBase	93
HBase API	96
Carga de Archivos	98
Procesamiento de Datos con Spark	99

BIG DATA PROFESSIONAL CERTIFICATE



Big Data y sus Tecnologías



BDPC™ Versión 092021

CertiProf®

Plan

- Introducción
- Big Data/Datos Masivos
 - Una definición
 - ¿Por qué?
 - Aplicaciones
- Herramientas y Tecnologías
 - Sistemas de Archivos Distribuidos
 - Algoritmos Distribuidos
 - Sistemas de Bases de Datos Distribuidas
 - Sistemas de Organización
 - Dark Data
- Conclusión

Introducción

- Ingeniero informático - Universidad de Sherbrooke
- Más de años 15 de experiencia
 - Arquitectura de software para sistemas distribuidos en red
 - Sistemas de procesamiento automático del lenguaje natural
 - Genómica / Bioinformática
 - Consulta
 - "Big data"
- Ingeniero Senior en Hopper
- Utilicemos los datos para ayudar a nuestros usuarios a tomar decisiones de consumo mejor informadas