

 UNIVERSIDAD DE MEDELLIN	DISEÑO MICROCURRICULAR	Código: F-GAC-03D
		Versión: 01
		Edición: 22/05/2017
Nombre del Programa	DIPLOMADO EN BIG DATA	
Facultad articulada:	Ingenierías	
Tipo de Evento:	Diplomado	
Duración:	108 horas	
<p>Objetivo General: Aprender a almacenar y acceder a grandes volúmenes de datos, que provienen diferentes fuentes para la toma de desiciones.</p>		
<p>Objetivos específicos:</p> <ul style="list-style-type: none"> ✓ Conocer la diferentes herramientas y tecnologías disponibles para el desarrollo de proyectos de Big Data. ✓ Realizar un consumo básico de información ✓ Usar bases de datos No SQL ✓ Crear, administrar y usar flujos de datos ✓ Usar bases de datos relacionales en Big Data ✓ Optimizar el almacenamiento, ✓ Usar grafos ✓ Realizar una introducción a análisis predictivo 		
<p>Conocimientos Previos Requeridos: Base de datos relacionales, Lenguaje SQL, programación (cualquier lenguaje)</p>		
ESTRUCTURA		
Tema Principal	Subtemas	Nro. Horas
1 Módulo	INTRODUCCIÓN A BIG DATA	6
	• Introducción a Big Data	
	• Evolución de los datos	
	• Evolución de la plataforma (OpenSource)	
	• Las 4 V	
	• Ciclo de vida de la información	
	• Ecosistema de Big Data	
	• Ciclo de vida de los datos	

		<ul style="list-style-type: none"> • Casos de Uso • Aplicaciones para seguros 	
2	Módulo	<p style="text-align: center;">CONSUMO BÁSICO DE DATOS</p> <ul style="list-style-type: none"> • Introducción a HDFS • Introducción a MapReduce (MRv2: YARN) • Pig • Sqoop • Interfases REST • Lab1: Análisis de datos no estructurados con MapReduce (Hacemos un lab de desarrollo con MapReduce o solo usamos un .jar?) • Lab2: Análisis de datos usando Pig • Lab3: Usando Sqoop 	8
3	Módulo	<p style="text-align: center;">BASE DE DATOS NOSQL</p> <ul style="list-style-type: none"> • Introducción a HBase (NoSql) • operaciones CRUD • Almacenamiento • HIVE • Phoenix • Lab1: Carga, CRUD y consulta de datos con HBase • Lab2: Consultas usando HIVE • Lab3: Consultas usando Phoenix 	16
4	Módulo	<p style="text-align: center;">INTRODUCCIÓN AL ANÁLISIS EN FLUJO DE DATOS</p> <ul style="list-style-type: none"> • Generación de eventos • Consumo de eventos • Creación de topología • Creación de consultas • Creación de salida hacia PowerBI • Visualización de eventos agregados en tiempo real 	8
5	Módulo	<p style="text-align: center;">ANÁLISIS DE FLUJOS DE DATOS</p> <ul style="list-style-type: none"> • Colección, agregación y movimiento de eventos Usando Flume • Spark con Phyton y Scala • Storm • Kafka • Lab1: Usando Flume • Lab2: Usando Spark para análisis de Flujos de Datos • Lab3: Usando Storm • Lab4: Usado Kafka 	16
6	Módulo	<p style="text-align: center;">BASE DE DATOS RELACIONALES SOBRE BIG DATA</p> <ul style="list-style-type: none"> • Introducción • Kudu • Impala • HCatalog 	16

		<ul style="list-style-type: none"> • PolyBase • Lab1: Kudu • Lab2: Impala • Lab3: HCatalog • Lab4: PolyBase 	
7	Módulo	<p style="text-align: center;">ALMACENAMIENTO, WORKFLOW</p> <ul style="list-style-type: none"> • Introducción a la optimización de almacenamiento • Avro • Gzip • HBase • Oozie • Lab 1: Avro y Gzip • Lab2: Oozie • Lab3: HBase 	8
8	Módulo	<p style="text-align: center;">GRAFOS Y ANÁLISIS PREDICTIVO</p> <ul style="list-style-type: none"> • Análisis de grafos con Pegasus • Introducción a análisis predictivo • RHadoop • Mahout • Lab1: Pegasus • Lab2: RHadoop • Lab3: Mahout 	16
9	Módulo	<p style="text-align: center;">MICROSOFT R SERVER</p> <ul style="list-style-type: none"> • Introducción a lenguaje R • R Server • Lab1: Introducción a lenguaje R • Lab2: Desarrollo de algoritmos en R Server 	16