Tokio.

CURSO DE

# Big Data




Acreditado por la UCAM*

UCAM
UNIVERSIDAD
CATÓLICA DE MURCIA

*Consulta condiciones.

# Módulo 0: Introducción a la Programación

## Unidad 1. Introducción a Python

- Python el nuevo desconocido
- Características básicas del lenguaje
- Programación orientada a objetos y excepciones
- Manipulación de datos

## Unidad 2: Introducción a la Estadística

- Variables estadísticas
- Tabla de frecuencias
- Gráficas estadísticas
- Principales parámetros estadísticos
- Estadística bidimensional
- Regresión lineal

## Unidad 3: Introducción a R

- Manejo de datos
- Gráficos estadísticos con R
- Programación con R
- Medidas descriptivas

### Unidad 4: Introducción a SQL

- Estructura básica de una consulta en SQL
- Tipos de consultas en SQL
- Funciones de agregación en SQL
- Consultas SQL sobre más de una tabla

## Modulo 1: Introducción al Big Data

### Unidad 1: Ecosistema Big Data

- Definición de componentes y arquitectura
- Disponibilidad, escalabilidad y resiliencia
- Introducción a Hadoop y MapReduce

### Unidad 2: Estrategias basadas en datos

- Cuadros de mando (Dashboards)
- Business Intelligence vs. Big Data

### Unidad 3: Entornos de procesamiento

- Cloud Computing
- Internet de las Cosas (IoT)

### Unidad 4: Casos de uso de Big Data: ejemplos en la industria

## Módulo 2. El dato y su ciclo de vida

### Unidad 1: Datos

- El dato
- Calidad del dato
- Derechos sobre los datos

### Unidad 2: Ciclo de vida del dato

- Fuentes de datos
- Extracción, Tratamiento y Carga
- Almacenamiento de datos
- Análisis para la explotación
- Visualización y storytelling para la explotación
  - La selección de los elementos visuales
- Toma de decisiones

## Módulo 3. Almacenamiento escalable de datos

### Unidad 1: Bases de datos distribuidas

- Tipo de escalabilidad
  - Escalabilidad vertical
  - Escalabilidad horizontal
- Bases de datos no distribuidas

- Teorema de CAP

### Unidad 2: NoSQL

- Bases de datos relacionales

### Unidad 3: NoSQL – Clave valor

### Unidad 4: NoSQL – Columnares

- Arquitectura
- Modelado de datos en Casandra
- Cassandra Query Languaje – CQL
- CQL – Creación de modelo de datos
  - Keyspace
  - Table

### Unidad 5: NoSQL – Orientada a documentos

### Unidad 6: NoSQL – Orientada a grafos

## Módulo 4. Arquitectura Big Data

### Unidad 1: El ecosistema Hadoop

- Introducción a Hadoop
- Herramientas del ecosistema Hadoop

### Unidad 2: Cluster y sistemas distribuidos (HDFS, MapReduce)

### Unidad 3: Análisis de datos con Hive y Pig

### Unidad 4: Procesamiento de datos con Spark

- Spark RDD (Resilient Distributed Datsets)
- Spark Streaming
- Spark SQL

## Módulo 5. Análisis para la explotación de datos

### Unidad 1: Perfiles de datos

- Científicos de datos
- Ingeniero de datos

### Unidad 2: Análisis exploratorio de datos

- Estadística descriptiva
- Distribución de los datos
- Exploración de datos categóricos y binarios
- Correlación
- Exploración de 2 o más variables

### Unidad 3: Técnicas de muestreo de datos

- Selección aleatoria

- Selección Bias
- Selección por distribución estadística

## Unidad 4: Contraste de hipótesis

- Testeo de muestras A/B
- Testeo de hipótesis
- Significancia estadística y P-Value
- P-Value

## Unidad 5: Regresión y predicción

- Regresión linear
- Regresión multilineal
- Interpretar los resultados de una regresión
- Predecir usando regresión

## Unidad 6: Aprendizaje supervisado

- Conceptos
- Algoritmos: árboles de decisión

## Unidad 7: Aprendizaje no supervisado

- Componentes principales
- Algoritmos: K-Means, clusters jerárquicos

### Unidad 8: Introducción al Deep Learning

- Conceptos fundamentales
- Redes neuronales

## Módulo 6. Presentación de proyectos Big Data y storytelling

### Unidad 1: Presentación de un proyecto Big Data

- La importancia del contexto
- La audiencia y su importancia

### Unidad 2: Componentes para la presentación de un proyecto Big Data

- Ideas de diseño

# Especialización Business Intelligence

## Módulo 1. Introducción al BI

### Unidad 1: Historia y evolución del BI

- ¿Qué es el BI?
- Componentes BI
- Fuentes de información

### Unidad 2: Conceptos básicos BI

- Herramientas/procesos básicos BI

### Unidad 3: Modelado de datos

- Modelo en estrella
- Modelo copo de nieve
- Modelo multidimensional

### Unidad 4: Aprovisionamiento de datos

- Datawarehouse
- Datamart
- Principales diferencias

## Módulo 2. ETL e introducción a la visualización

### Unidad 1: Procesos de extracción transformación y cargas y herramientas.

- Diseño de proceso ETL
- Principales herramientas del mercado y particularidades

### Unidad 2: Introducción a la visualización

- Introducción a la visualización de la información
- Beneficios de la visualización
- Categorías de la visualización

### Unidad 3: Herramientas y particularidades

- Principales herramientas del mercado y comparativa

# Módulo 3. Introducción al PowerBI

## Unidad 1: Comprensión básica de PowerBI

- ¿Qué es?
- ¿Para qué sirve?
- Interfaz

## Unidad 2: Conexión a orígenes de datos

## Unidad 3: Editor Power Query

- Formateo de datos
- Detección de errores
- Pivot

## Unidad 4: Creación de gráficos

## Unidad 5: Construcción de informe con filtros

- Vinculación y desvinculación en gráficos

## Unidad 6: Introducción a DAX (Data Analysis Expressions)

- Introducción a DAX
- Tablas y columnas calculadas, medidas simples
- Filtrado de filas y medidas avanzadas

## Unidad 7: Creación de métricas y gráficos avanzados

# Módulo 4. Introducción al Tableau: Visualización I

## Unidad 1: Tableau I

- Comprensión básica de Tableau
- Conexión a fuente/base de datos
- Combinación de datos
- Editar y guardar fuente de datos
- Dimensiónes y métricas

## Unidad 2: Tableau II

- Conversión tipo de datos
- Representación gráfica de los datos
- Filtrado de informes y características
- Creación jerarquías y drill down

## Unidad 3: Tableau III

- Creación campos calculados. (medidas y dimensiones)
- Creación de parámetros
- Combinación campos calculados y parámetros
- Descripciones emergentes

## Unidad 5. Tableau Avanzado: Visualización II

### Unidad 1: Tableau avanzado I

- Creación de grupos
- Creación de conjuntos
- Expresiones LOD

### Unidad 2: Tableau avanzado II

- Creación de dashboard
- Objetos dashboard
- Fomatos dashboard
- Interactividad dashboard

### Unidad 3: Tableau avanzado III

- Extensiones Tableau
- Configuración extensión
- Gráficos no nativos

### Unidad 4: Movilidad y colaboración

- Creación de historias
- Adaptación de informes para móvil y tablet
- Tableau online/Tableau server/Tableau public

# Especialización Apache Hadoop

## Módulo 1. Introducción al Big Data

### Unidad 1: Conceptos básicos

- Definición y conceptos
- Evolución de los datos

### Unidad 2: SQL vs. NoSQL

### Unidad 3: Preparación entorno de trabajo

## Módulo 2: Apache Hadoop (HDFS)

### Unidad 1: Conceptos básicos y arquitectura

- Conceptos fundamentales
- Arquitectura Hadoop

### Unidad 2: Lectura, escritura y replicación

### Unidad 3: Permisos y borrado de datos

### Unidad 4: Safemode, snapshots y gestión de caché

### Unidad 5: Profundizando con la Shell

## Módulo 3: Map reduce & yarn

### Unidad 1: Map reduce: el algoritmo

- Arquitectura
- Configuración

### Unidad 2: Implementación de algoritmos map reduce

### Unidad 3: Maneras de ejecutar el algoritmo

### Unidad 4: YARN: Arquitectura, gestión de colas y Zookeeper

## Módulo 4: Ecosistema Hadoop

### Unidad 1: Apache Hive

- Arquitectura
- Componentes

### Unidad 2: Apache Sqoop

- Arquitectura
- Componentes

## Unidad 3: Apache Pig

- Arquitectura
- Componentes
- Pig Latin

## Unidad 4: Apache Spark

- Arquitectura
- Componentes
- Tipos de datos
- Integración con Hive

## Unidad 5: Reporting con Zeppelin

- La importancia del reporting
- Creación de reportes con Zeppelin


Tokio.