

**PLANETALIA**

Formación de Vanguardia desde 1999

# **Máster Certificado en Big Data y Minería de Datos**

**300 horas lectivas**

**100 horas proyectos**

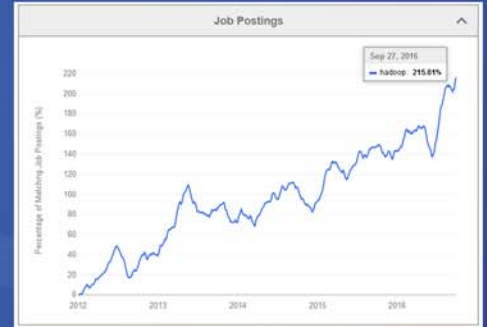
**12 certificaciones oficiales**

# Una apuesta segura

En el mundo actual, se genera constantemente una cantidad nunca antes vista de información, a todos los niveles. 7.000 twits por segundo en Twitter. 40.000 consultas por segundo en Google. 31 millones de mensajes en Facebook por cada minuto, por más de 1.000 millones de usuarios. 6.000 millones de horas de vídeo vistas al año en Youtube. 1 billón de fotos al año online.

BigData es el conjunto de tecnologías diseñadas para transmitir, almacenar, indexar y sobre todo, procesar y analizar volúmenes enormes de información muy heterogénea y poco estructurada en un tiempo. La enorme demanda de profesionales formados en estas tecnologías habla por sí sola:

- 2,147 ofertas de empleo a Noviembre de 2016
- Sueldos medios de 38,000€ - de los más altos en Informática
- Escasísima competencia - únicamente 5 candidatos por puesto



## 12 Certificaciones Oficiales

En este máster no sólo te formamos en todas las tecnologías Big Data más demandadas por las empresas, sino que además somos el único Máster que te prepara para 12 Certificaciones Oficiales, de forma que seas siempre el candidato número 1 en cualquier proceso de selección.

**cloudera**

Cloudera Certified Administrator for Apache Hadoop (CCA)  
CCA Spark and Hadoop Developer  
CCP Data Engineer

**mongoDB**

MongoDB Certified Developer, Associate Level  
MongoDB Certified DBA, Associate Level

**Cassandra**

Datastax Apache Cassandra Certified Developer  
Datastax Apache Cassandra Certified Administrator  
Datastax Apache Cassandra Certified Architect

Microsoft®  
**SQL Server**

MCSA - SQL Server 2012/2014

**tableau**  
SOFTWARE

Tableau Desktop 9.0 Qualified Associate

**IBM**

IBM Certified Specialist - SPSS Statistics Level I v2

IBM Certified Solution Advisor - Big Data & Analytics VI

# Temario Resumido

## Parte I - Introducción a BigData y al Análisis Estadístico

### MÓDULO 1 - Conceptos Fundamentales de Big Data

BigData  
4V: Volumen, Velocidad, Variedad, Veracidad  
Precursores - Google File System, Google BigTable, Dynamo  
BigData y las herramientas ETL  
BigData y la computación en la nube  
BigData y las Arquitecturas Orientadas a Servicios  
Análisis y Minería de Datos  
Algoritmos fundamentales de Minería de Datos  
Aprendizaje Profundo  
Sistemas Distribuidos y Paralelización de Algoritmos  
Teorema CAP.  
Sistemas SQL vs NoSQL

### MÓDULO 2 - Conceptos Fundamentales de Análisis Estadístico

Introducción a la estadística  
Los paquetes más extendidos: R, Tableau y SPSS  
Tipos de variables  
Distribuciones. Estadística descriptiva.  
Medidas de dispersión, forma y posición.  
Introducción a la probabilidad.  
Probabilidad condicionada. Árboles de decisión. Inferencia Bayesiana  
La Distribución binomial  
La distribución normal  
La distribución de Poisson  
Otras distribuciones frecuentes: Chi2, t de Student, Poisson.  
Teorema Central del Límite

### MÓDULO 3 - Inferencia Estadística y Visualización de Resultados

Conceptos básicos de inferencia estadística  
Diseño de experimentos. Reglas de validez. Errores.  
Estimadores y niveles de confianza  
Contraste de hipótesis  
Análisis multivariable. ANOVA  
Correlación  
Regresión Lineal  
Regresión No Lineal  
Pruebas no paramétricas  
Gráficos con R  
Gráficos con SPSS  
Gráficos con Tableau



## Parte II - Sistemas Relacionales y Data Warehousing

### MÓDULO 4 - Diseño y Modelado de Data Warehouses

Introducción a los DWH. Diferencias con los sistemas OLAP y OLTP  
Arquitecturas Data Warehouse. Data Marts  
Modelado Kimball o esquemas de estrella.  
Tablas de hechos y tablas de dimensiones  
Gestión de cambios dimensionales lentos  
Integración vía dimensiones conformes  
Gestión de jerarquías dimensionales  
Algunas técnicas avanzadas

### MÓDULO 5 - SQL Server 2014

Arquitectura general de SQL Server 2014  
Opciones de Instalación. Configuración de Instancias  
Tipos de datos  
Creación de bases de datos, tablas y vistas.  
Extensiones SQL para XML  
Extensiones SQL para datos geoespaciales  
Planes de Ejecución

### MÓDULO 6 - Implementación de Data Warehouses con SQL Server

Introducción a SSIS  
Flujo de datos en SSIS  
Flujo de control en SSIS. Tareas  
Fusión y Combinación de datos  
Búsquedas difusas  
Servicios de calidad de datos (DQS)  
Gestión de Errores y Eventos  
Carga incremental de data warehouses  
SQL Server Change Data Capture

### MÓDULO 7 - Minería de datos y reporting con SQL Server

Introducción a SSAS  
Bases de datos Multidimensionales y MDX  
Creación y modificación de cubos  
Predicción de Atributos Discretos  
Predicción de Atributos Continuos  
Segmentación y Clustering  
Sistemas Recomendadores  
Introducción a los Reporting Services(SSRS)  
Informes básicos - Listas, Tablas y Matrices  
Informes Avanzados - Gráficos, Indicadores y Mapas  
Subinformes

### MÓDULO 8 - Administración de SQL Server 2014

Monitorización de SQL Server  
Recopilación y Análisis de datos de rendimiento  
Auditoría y Traza de servidores  
Copia de seguridad y restauración  
Importación y Exportación de datos  
SQL Server y la seguridad  
Roles y permisos  
SQL Server en clusters  
Alta disponibilidad

# Temario Resumido



## Parte III - El Ecosistema Hadoop

### MÓDULO 9 - Apache Hadoop - Administración

- Introducción a Hadoop
- Arquitectura
- Planificación y despliegue de nodos Hadoop
- Clústers Hadoop
- Gestión y planificación de trabajos
- Gestión de Clusters
- Monitorización
- Optimización del rendimiento
- Solución de Problemas
- Apache Zookeeper
- Apache Ambari

### MÓDULO 10 - Importación y Exportación de datos

- Sqoop
- Flume
- Hadoop como sistema ETL

### MÓDULO 11 - Análisis de datos con Hadoop

- Pig y el lenguaje Pig Latin
- Hive
- HBase
- Gestión de trabajos con Oozie
- Apache Spark
- Apache Drill
- Cloudera Impala

### MÓDULO 12 - Desarrollo sobre Hadoop

- Acceso programático a HDFS
- Archivos de Secuencia
- Algoritmo Map Reduce
- API de Streaming
- Map Reduce desde Java
- Combinadores y Particionadores
- Job Counters y reporting
- Uso de la caché distribuida
- Pruebas unitarias con MRUnit
- Uso de tipos de datos propios
- Algoritmos fundamentales sobre Hadoop

### MÓDULO 13 - IBM InfoSphere BigInsights

- Introducción a la distribución BigInsights
- InfoSphere BigSheets
- IBM BigR - Integración entre R y BigInsights
- BigSQL 3.0
- BigInsight Analytics
- Introducción a Infosphere Streams

## Parte IV - Otras Tecnologías Big Data

### MÓDULO 14 - Apache Cassandra

- Arquitectura Básica
- Quórum de Lectura y Escritura. Reparación de datos
- MemTables, SSTables y Commit Logs. vNodes.
- Rutas de escritura y lectura.
- Instalación y configuración de clusters Cassandra
- Particionadores y Snitches
- Lenguaje de consultas CQL
- Compactación y compresión de datos.
- Acceso Programático a Cassandra
- Cassandra y PIG
- Cassandra y Hive
- Modelado de datos en Apache Cassandra

### MÓDULO 15 - MongoDB

- Introducción a MongoDB
- Operaciones Básicas con MongoDB
- Indexación y diseño de esquemas
- Desarrollo con MongoDB
- Modelado de datos
- Framework de Agregación
- Particionamiento (Sharding)
- Replicación
- Securización

### MÓDULO 16 - Neo4J

- Bases de datos orientadas a grafos
- Operativa básica de Neo4J
- El lenguaje de consultas Cypher
- Integración con aplicaciones. Traversal API
- Monitorización y ajuste del rendimiento.
- Escalabilidad
- Casos de uso

# Temario Resumido



## Parte V - Aplicaciones

### Herramientas de Minería de datos

- R sobre Hadoop
- Minería de datos con Apache Mahout
- Minería de datos con Weka
- RapidMiner
- KNIME
- Qlikview
- Orange

### Minería de datos textuales

- Indexación textual con Apache Solr
- Conceptos de análisis textual
- Vectorización de textos
- Extracción de información textual con IBM Biginsights Text Analytics
- Clasificación
- Análisis semántico latente
- Latent Dirichlet Allocation
- Clustering y Categorización de Texto

### Minería de datos con Python

- Introducción al lenguaje
- Estructuras de datos
- Programación funcional en Python
- Entrada/Salida
- Integración de R y Python
- Análisis del sentimiento en medios sociales
- Proceso de lenguaje natural con NTLK

## Docente



*El máster se imparte por un único docente, lo que permite una mayor homogeneidad en el tratamiento del temario, continuidad en el desarrollo, y una mayor atención personalizada a cada alumno al conocer el docente su evolución e intereses específicos.*

*Alexander Hristov dispone de más de 30 años de experiencia en el sector de las nuevas tecnologías, de los cuales los últimos 17 ha simultaneado la docencia con la consultoría y dirección de proyectos.*

*Alexander Hristov está certificado en todas las tecnologías que se imparten en el máster. En concreto, dispone de las siguientes certificaciones:*

**Cloudera Certified Administrator for Apache Hadoop (CCAH)**  
**IBM Big Data & Analytics Certified Solution Advisor**  
**IBM Cloud Computing Certified Solution Advisor**

**DataStax Certified Developer on Apache Cassandra**  
**MongoDB Certified Developer**  
**Project Management Professional (PMP)**  
**MCSA SQL Server**

# Fechas, Horarios y Modalidades

## *Modalidad Presencial*



Asistencia presencial al Máster en las aulas de Planetalia, con grupos muy reducidos.

- Turno de Mañanas: 9:00 a 14:00, Lunes a Viernes
- Turno de Tardes: 17:00 a 20:30, Lunes a Viernes

## *Modalidad Teleasistencia*

Asistencia al Máster mediante nuestro sistema de Aula Virtual, formando parte de los mismos grupos presenciales. Disfrutarás de todas las ventajas de las clases en directo, con las mismas posibilidades de interacción con el docente y el resto de alumnos, pero desde la comodidad de tu casa y sin desplazarte.

- Turno de Mañanas: 9:00 a 14:00, Lunes a Viernes
- Turno de Tardes: 17:00 a 20:30, Lunes a Viernes



## *Modalidad A Distancia*



La opción más flexible - total libertad de horarios y sin desplazamientos. Aprende a tu ritmo con los mismos materiales educativos, prácticas y ejercicios que los alumnos presenciales, disponiendo en todo momento del apoyo del docente.

# Nuestra Experiencia

Más de 16 años, más de 1.000 cursos impartidos, más de 45.000 horas lectivas



Indra



Atos

JAZZTEL



Banesto



INFORMATICA  
El Corte Inglés



KPMG

bankinter



Telefonica



Santander

Ibermática



Isdefe

BBVA Gobernalia

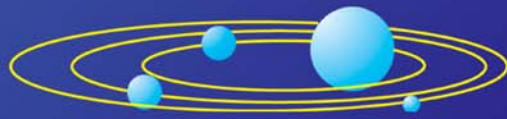


altran

CORREOS

pelayo

Todas las marcas y logotipos son propiedad de sus respectivos titulares, y se referencian aquí únicamente con propósito informativo, sin que dicha presencia implique apoyo, endorso o relación de ningún tipo.



# PLANETALIA



RIGHTS PROTECTED

c/Tórtola 10, Oficina 4ºB. 28019 Madrid  Oporto. Tel: 91 698 48 48  
Web : [www.planetalia.com](http://www.planetalia.com) E-mail : [info@planetalia.com](mailto:info@planetalia.com)

(c) 2016-2017 Planetalia SL. Todos los derechos reservados