Contenido

[1. Autores del trabajo, planificación y entrega 3](#_Toc444537686)

[1.1 Autores 3](#_Toc444537687)

[1.2 Planificación 3](#_Toc444537688)

[1.3 Entrega 3](#_Toc444537689)

[2. Descripción del tipo de tecnología 3](#_Toc444537690)

[3. Fuentes de información (documentos) 3](#_Toc444537691)

[3.1 Fuentes sobre el tipo de tecnología en general 4](#_Toc444537692)

[3.1.1 Fuente de información 1 sobre el tipo de tecnología en general 4](#_Toc444537693)

[3.1.2 Fuente de información 2 sobre el tipo de tecnología en general 4](#_Toc444537694)

[3.1.n Fuente de información n sobre el tipo de tecnología en general 4](#_Toc444537695)

[3.2 Fuentes sobre la tecnología específica A 4](#_Toc444537696)

[3.2.1 Fuente de información 1 sobre la tecnología específica A 4](#_Toc444537697)

[3.2.2 Fuente de información 2 sobre la tecnología específica A 4](#_Toc444537698)

[3.2.n Fuente de información n sobre la tecnología específica A 4](#_Toc444537699)

[3.3 Fuentes sobre la tecnología específica B 4](#_Toc444537700)

[3.3.1 Fuente de información 1 sobre la tecnología específica B 4](#_Toc444537701)

[3.3.2 Fuente de información 2 sobre la tecnología específica B 4](#_Toc444537702)

[3.3.n Fuente de información n sobre la tecnología específica B 4](#_Toc444537703)

[4. Fuentes de información (cursos no gratuitos) 4](#_Toc444537704)

[4.1 Cursos no gratuitos sobre el tipo de tecnología en general 4](#_Toc444537705)

[4.1.1 Curso no gratuito 1 sobre el tipo de tecnología en general 4](#_Toc444537706)

[4.1.2 Curso no gratuito 2 sobre el tipo de tecnología en general 4](#_Toc444537707)

[4.1.n Curso no gratuito n sobre el tipo de tecnología en general 4](#_Toc444537708)

[4.2 Cursos no gratuitos sobre la tecnología específica A 4](#_Toc444537709)

[4.2.1 Curso no gratuito 1 sobre la tecnología específica A 4](#_Toc444537710)

[4.2.2 Curso no gratuito 2 sobre la tecnología específica A 4](#_Toc444537711)

[4.2.n Curso no gratuito n sobre la tecnología específica A 4](#_Toc444537712)

[4.3 Cursos no gratuitos sobre la tecnología específica B 5](#_Toc444537713)

[4.3.1 Curso no gratuito 1 sobre la tecnología específica B 5](#_Toc444537714)

[4.3.2 Curso no gratuito 2 sobre la tecnología específica B 5](#_Toc444537715)

[4.3.n Curso no gratuito n sobre la tecnología específica B 5](#_Toc444537716)

[5. Fuentes de información (cursos gratuitos) 5](#_Toc444537717)

[5.1 Cursos gratuitos sobre el tipo de tecnología en general 5](#_Toc444537718)

[5.1.1 Curso gratuito 1 sobre el tipo de tecnología en general 5](#_Toc444537719)

[5.1.2 Curso gratuito 2 sobre el tipo de tecnología en general 5](#_Toc444537720)

[5.1.n Curso gratuito n sobre el tipo de tecnología en general 5](#_Toc444537721)

[5.2 Cursos gratuitos sobre la tecnología específica A 5](#_Toc444537722)

[5.2.1 Curso gratuito 1 sobre la tecnología específica A 5](#_Toc444537723)

[5.2.2 Curso gratuito 2 sobre la tecnología específica A 5](#_Toc444537724)

[5.2.n Curso gratuito n sobre la tecnología especifica A 5](#_Toc444537725)

[5.3 Cursos gratuitos sobre la tecnología específica B 5](#_Toc444537726)

[5.3.1 Curso gratuito 1 sobre la tecnología específica B 5](#_Toc444537727)

[5.3.2 Curso gratuito 2 sobre la tecnología específica B 5](#_Toc444537728)

[5.3.n Curso gratuito n sobre la tecnología especifica B 5](#_Toc444537729)

[6. Ayudas para estudiar las tecnologías 5](#_Toc444537730)

[7. Recursos para implementar las tecnologías 6](#_Toc444537731)

[7.1 Recursos para implementar la tecnología A 6](#_Toc444537732)

[7.1.1 Recursos no gratuitos para implementar la tecnología A 6](#_Toc444537733)

[7.1.1 Recursos no gratuitos para implementar la tecnología A 6](#_Toc444537734)

[7.2 Recursos para implementar la tecnología B 6](#_Toc444537735)

[7.2.1 Recursos no gratuitos para implementar la tecnología B 6](#_Toc444537736)

[7.2.1 Recursos no gratuitos para implementar la tecnología B 6](#_Toc444537737)

[8. Conclusiones 6](#_Toc444537738)

# 1. Autores del trabajo, planificación y entrega

## 1.1 Autores

En este apartado se debe indicar el número de grupo y los nombres de los autores, poniendo en primer lugar al coordinador del grupo.

## 1.2 Planificación

En este apartado se debe incluir un enlace (URL) compartido a la planificación del trabajo utilizando una herramienta online de diagramación Gantt (por ejemplo, GanttPro, versión gratuita).

En este enlace hay un ejemplo de una posible planificación del trabajo, que se puede usar como referencia, pero indicando los nombres reales de las tecnologías y de los participantes.

[Ejemplo planificación](https://app.ganttpro.com/shared/token/b0b82a0da290d4dcc93d8813795ad00093b8c583b346f796b38148ef71895eb1#!/app/home).

Hay que tener en cuenta que, como puede verse en el ejemplo, cada participante del grupo debe tener asignadas tareas que sumen al menos 15 horas. El peso de este trabajo en la calificación total de la asignatura es de un 10%, por tanto requiere de una dedicación de 15 horas del total de 150 horas de la asignatura.

## 1.3 Entrega

En este apartado debe incluirse un enlace (URL) a un repositorio en GitHub o en BitBucket creado para el trabajo.

En dicho repositorio debe encontrarse, al menos los siguientes archivos en la rama máster:

* Trabajo terminado: del trabajo terminado con el nombre TG1\_final.ocx
* Presentación del trabajo: TG1\_final.pptx

Dichos archivos serán los que se tendrán en cuenta para la calificación del trabajo.

# 2. Descripción del tipo de tecnología

En este apartado se debe indicar el tipo de tecnología en general y las tecnologías específicas sobre las que trata el trabajo.

El objetivo del trabajo es ponerse en la situación de una persona ya titulada en el Grado en Sistemas de Información, y que desea actualiza sus conocimientos sobre dichas tecnologías.

En este documento se debe recoger toda la información que se ha recopilado para poder aprender la teoría y práctica de dichas tecnologías, así como las ayudas que existen para poder financiar su estudio o su implementación en empresas u otras organizaciones.

# 3. Fuentes de información (documentos)

En los sub-apartados de este apartado se deben indicar documentos de interés para aprender sobre el tipo de tecnología en general, y sobre cada una de las tecnologías elegidas.

Sobre cada documento se debe

## 3.1 Fuentes sobre el tipo de tecnología en general

### 3.1.1 Fuente de información 1 sobre el tipo de tecnología en general

### 3.1.2 Fuente de información 2 sobre el tipo de tecnología en general

### 3.1.n Fuente de información n sobre el tipo de tecnología en general

## 3.2 Fuentes sobre la tecnología específica A

### 3.2.1 Fuente de información 1 sobre la tecnología específica A

### 3.2.2 Fuente de información 2 sobre la tecnología específica A

### 3.2.n Fuente de información n sobre la tecnología específica A

## 3.3 Fuentes sobre la tecnología específica B

### 3.3.1 Fuente de información 1 sobre la tecnología específica B

### 3.3.2 Fuente de información 2 sobre la tecnología específica B

### 3.3.n Fuente de información n sobre la tecnología específica B

# 4. Fuentes de información (cursos no gratuitos)

## 4.1 Cursos no gratuitos sobre el tipo de tecnología en general

### 4.1.1 Curso superior en Big Data ([www.camaramadrid.es](http://www.camaramadrid.es))



El objetivo de este curso está enfocado a:

* Conocer y entender los conceptos básicos de BI y cómo evoluciona hasta llegar al Big Data.
* Conocer y entender los conceptos básicos de Big Data.
* Despertar y arrancar iniciativas en el ámbito del Big Data y su aplicación en la empresa y el negocio.
* Guiar a los asistentes en la puesta en marcha de un proyecto partiendo del Big Data.
* Conocer y entender qué es Business Analytics, qué problemas de negocio resuelve y su aplicación.
* Conocer nuevos retos analíticos y cómo las redes sociales y en general internet, intervienen en ellos.

**Programación del curso:**

INTRODUCCIÓN A LA INTELIGENCIA DE NEGOCIO

* Conceptos básicos
* Arquitectura y componentes principales
* Ejemplos prácticos en distintos sectores
* Herramientas y tendencias actuales.

BIG DATA, UN BUEN COMPLEMENTO PARA BI

* Conceptos básicos y tipos
* Tecnología para no iniciados, Hadoop, NoSQL...
* Cómo complementa a la inteligencia de negocio tradicional
* Del "data science" al "data intelligence"
* Como afrontar un proyecto
* Aplicaciones prácticas de Big Data
* Legalidad y reputación corporativa.

TEORÍA DE LA DECISIÓN

* Concepto
* Business Analytics
* Introducción a la Estadística
* Aplicación práctica de la minería de datos
* Customer Analytics
* Análisis para Segmentación, Captación, Fidelización/ Retención, Búsqueda de Gemelos, Valor de Cliente

EFECTIVIDAD COMERCIAL

* Modelos NBA (Next Best Action/ Offer), Escenarios What if,
* Previsión de la Demanda
* Optimización de Precios
* Decisiones Basadas en Eventos
* Escucha, Analiza, Actúa
* Cadena de Valor en ámbitos Big Data
* Impacto de las redes sociales
* Impacto de la Información Audiovisual
* Convergencia de la información estructurada y no estructurada
* Casos prácticos

ALGUNOS PLAYERS EN BUSINESS DISCOVERY

* QlikView
* Tableau
* Ejemplos prácticos: Internet of Things (IoT, wearables)
* Nuevos dispositivos
* Nuevas Reglas del juego
* Cómo aplicarlas en beneficio del negocio: Interés Empresarial en Big Data
* Limitaciones y ventajas
* Perfiles

MASTER CLASS

* Aplicación en Retail
* Aplicación en Seguros
* Aplicación Banca
* Aplicación en AAPP

**Duración:**

180 horas lectivas

**Precio:**

Precio total del curso: 2100 €

**Lugar de impartición:**

Cámara Oficial de Comercio e Industria de Madrid  
IFE   
c/ Pedro Salinas, 11   
28043 Madrid

### 4.1.2 Máster en Arquitectura Big Data ([www.kschool.com](http://www.kschool.com))



Dos de los perfiles más solicitados en el ámbito del Big Data son los Arquitectos y Desarrolladores, por lo que esta es la oportunidad perfecta para todos aquellos profesionales con perfiles técnicos de especializarse con un Máster en Big Data y aprovechar la situación actual de demanda de las empresas.

El objetivo de este Máster en Big Data es que los alumnos conozcan en profundidad los diferentes paradigmas de procesamiento en sistemas Big Data y dominen las principales tecnologías y su utilización para el diseño de arquitecturas escalables adaptadas a cada proyecto.

**Programación del curso:**

MÓDULO 1: INTRODUCCIÓN

* Paradigmas de procesamiento en Big Data
* Economía del dato
* Innovación y Big Data

MÓDULO 2: BATCH PROCESSING

* Hadoop
* Pig
* Hive
* Sqoop
* Flume
* Spark Core
* Spark 2.0.

MÓDULO 3: STREAMING PROCESSING

* Fundamentos de Streaming Processing
* Spark Streaming
* Kafka
* Storm
* Samza
* Pulsar y Apache Apex
* Implementación de un sistema real-time

MÓDULO 4: SISTEMAS NOSQL

* HBase
* Cassandra
* MongoDB
* Neo4J
* Redis
* Berkeley DB

MÓDULO 6: SISTEMAS DE COMPUTACIÓN HÍBRIDOS

* Arquitectura Lambda
* Arquitectura Kappa
* Apache Flink e implementaciones prácticas
* Druid
* ElasticSearch
* Logstash
* Kibana

MÓDULO 7: CLOUD COMPUTING

* Amazon Web Services
* Google Cloud Platform

MÓDULO 8: ADMINISTRACIÓN DE SISTEMAS BIG DATA

* ·Administración e Instalación de clusters: Cloudera y Hortonworks
* Optimización y monitorización de servicios
* Seguridad: Apache Knox, Ranger y Sentry

MÓDULO 9: ANÁLISIS AVANZADO DE DATOS

* Introducción al Data Science
* Data Analysis en Python: NumPy, Pandas, Scikit-learn, Matplotlib
* Machine Learning en Big Data: Apache Mahout y Spark MLib
* Análisis estadístico con R

MÓDULO 10: VISUALIZACIÓN DE DATOS

* Herramientas de visualización: Tableau y CartoDB
* Librerías de visualización: D3, Leaflet, Cytoscape

**Duración:**

230 horas – 7 meses.

**Precio:**

6500 €

**Lugar de impartición:**

Aula KSchool

Calle José Picón, 31

28028 – Madrid

### 4.1.3 Máster Big Data ([www.cemadig.com](http://www.cemadig.com))



El Big Data es una de las áreas clave de mayor trascendencia en el futuro inmediato, que las empresas necesitan abordar para optimizar sus procesos de negocio, y el Master Big Data es una oportunidad.

El Master Big Data proporciona los conocimientos técnicos necesarios para diseñar proyectos e identificar oportunidades de negocio, con el manejo de las herramientas, determinación de procesos y técnicas de análisis en las distintas etapas de gestión de un proyecto de datos.

Este Master Big Data combina una sólida formación teórica con un enfoque práctico, e integra un compendio de avanzadas disciplinas tecnológicas con el objeto de implantar la cultura de los datos en la estrategia empresarial.

El objetivo de este máster es que el alumno que lo termine sea capaz de:

* Tratar, almacenar y gestionar grandes cantidades de datos de diferentes formatos y estructuras
* Extraer, organizar y analizar la información obteniendo conclusiones relevantes para determinar predicciones
* Identificar oportunidades de negocio que se transformen en innovadores productos y servicios

Para ello aprenderá entre otras cosas a:

* Manejar los modelos de datos NoSQL
* Programar en R
* Crear, gestionar y realizar consultas una base de datos en MongoDB
* Conocer, configurar y procesar con Hadoop y comprender las posibilidades de procesamiento que ofrece.
* Aprender Python y utilizar las librerías para análisis de datos

**Programación:**

Bases de datos SQL

* Introducción a las Bases de datos
* Diseño de base de datos
* Implementación del modelo relacional con SQL

Bases de tatos NoSQL

* Introducción a las bases de datos NoSQL
* Diferentes bases de datos NoSQL
* Teorema de CAP
* MongoDB vs Casandra
* Mongo DB
* Introducción a Mongo DB
* Operaciones CRUD
* Búsqueda avanzada con Find
* Proyección, índices y agregación, cursores
* Modelos de datos

Business Intelligence

* Que es business intelligence
* Fuentes de Business intelligence
* Data Marts
* Modelo dimensional
* Modelo semántico OLAP
* Microsoft SqlServer Integration Services
* Modelo tabular y expresiones DAX
* Modelo de Cubos
* MDX

Arquitecturas paralelas de Big Data y Cloud Computing

* Recopilando información Internet of Things y servicios cloud computing
* Conectividad con fuentes de datos heterogéneas
* Almacenando información y análisis en tiempo real
* Visualización y análisis de la información

Fundamentos de Estadística

* Estadística descriptiva
* Fenómenos aleatorios y probabilidad
* Inferencia Estadística

Machine Learning

* Machine Learning con R
* Machine Learning con Python

Visualización avanzada

* EDA
* Análisis Exploratorio de Datos, Dashboards, Infografías, 3D, Realidad aumentada
* Tipos de gráficos.
* Herramientas de visualización: Tableau, Qlikview, R + D3.
* Interactividad y manipulación directa
* Proceso y desarrollo de una visualización

Big data en la empresa

* Casos reales de soluciones BigData y Machine Learning aplicados a diferentes departamentos de una organización
* Sistemas de Soporte a la Decisión
* Marketing y Ventas
* Gestión Económica Financiera
* Operaciones y Logística

Enterprise Data Science

* La empresa orientada al Dato:
* Características.
* Niveles de Madurez
* Tipología de las organizaciones y aplicación del “Data Science”
* Creación de equipos de “científicos de datos”.
* ¿Cómo estructurar un proyecto Data Science en la empresa?

Text Mining y redes sociales

* La importancia de las redes sociales como generadores de información
* Acceso a las APIs mas importantes: Acceso a los datos.
* Limpieza de los datos y análisis de texto pasa su uso.
* Análisis del sentimiento de los mensajes en redes sociales.
* Visualizaciones: Representación de la información en las redes sociales
* Código reproducible.

Open Data

* Fundamentos de open data
* Búsqueda e incorporación de datos a proyectos.
* Open data gubernamental
* Explorando otras fuentes de datos abiertos
* Principales algoritmos abiertos

Modelos de emprendimiento en empresas de Big Data

* Puesta en marcha de tu idea de negocio en 12 semanas con Lean Startup.
* Diseño del Modelo de Negocio.
* Aplicación del proceso de Desarrollo de Clientes.
* Metodologías ágiles en el desarrollo de productos o servicios.

**Duración:**

1.500 horas (60 ECTS).

Fecha de inicio del máster: Octubre 2017

Fecha de finalización del máster: Junio 2018

Horario de clases presenciales:

* Viernes, de 16.00 a 21.00 h.
* Sábados, de 9.00 a 14.00 h.

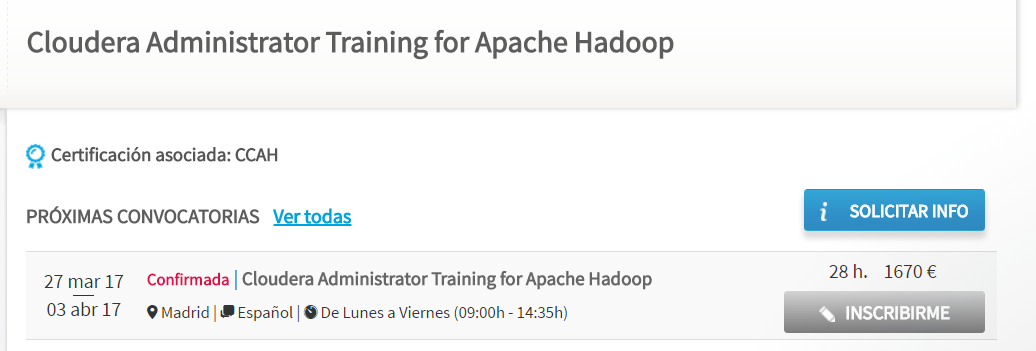
**Precio:**

Presencial: 5500 €

Online: 3800 €

## 4.2 Cursos no gratuitos sobre Hadoop

### 4.2.1 Cloudera Administrator training for Apache Hadoop ([www.pue.es](http://www.pue.es))



Cloudera fue la primera compañía en comercializar Apache Hadoop® y desarrollar soluciones de nivel empresarial construidas sobre esta poderosa tecnología de código abierto.

El curso de administrador de Cloudera para Apache Hadoop ofrece un conocimiento global de todas las medidas necesarias para operar y mantener un clúster Hadoop. Desde la instalación y la configuración a través de balanceo de carga y tuning, es la mejor preparación para los desafíos del mundo real que enfrentan por los administradores de Hadoop

PUE es Training Partner oficial de Cloudera, autorizado por dicha multinacional para impartir formación oficial en tecnologías Cloudera.

PUE está también acreditado y reconocido para realizar servicios de consultoría y mentoring en la implementación de soluciones Cloudera en el ámbito empresarial con el consiguiente valor añadido en el enfoque práctico y de negocio del conocimiento que se traslada en sus cursos oficiales.

El objetivo de este curso es que la persona que lo finalice sepa:

* Funciones de Cloudera Manager que facilitan la gestión de clústeres, agregación de logs, gestión de configuración, administración de recursos, informes, alertas y administración de servicios.
* Fundamentos de la arquitectura de YARN, MapReduce, Spark y HDFS.
* Determinar el hardware y la infraestructura correctos para el clúster.
* Configuración y despliegue del clúster necesarios para integrarlo en el data center.
* Cómo cargar datos en el clúster desde archivos generados dinámicamente mediante Flume y desde RDBMS utilizando Sqoop.
* Configurar FairScheduler para proporcionar acuerdos de nivel de servicio para varios usuarios de un clúster.
* Mejores prácticas para preparar y mantener Apache Hadoop en producción.
* Trouleshooting, diagnóstico, tinning y resolución de problemas en Hadoop.

Tras la realización del presente curso oficial el alumno estará preparado para superar el examen de certificación Cloudera Certified Administrator for Apache Hadoop (CCAH).

**Programación:**

**Historia de Apache Hadoop**

* ¿Por qué Hadoop?
* Componentes principales de Hadoop
* Conceptos fundamentales

**Instalación del Clúster Hadoop**

* Fundamentos de un Cluster Management Solution
* Funciones de Cloudera Manager
* Instalación de Cloudera Manager
* Instalación de Hadoop (CDH)

**El Hadoop Distributed File System (HDFS)**

* Características de HDFS
* Lectura y escritura de archivos
* Consideraciones acerca de memoria de NameNode
* Aspectos generales de seguridad en HDFS
* Empleo del Namenode Web UI
* Empleo del File Shell de Hadoop

**MapReduce y Spark en YARN**

* Funciones de Framework computacionales
* YARN: El gestor de recursos del cluster
* Conceptos de MapReduce
* Conceptos de Apache Spark
* Ejecución de Computational Frameworks en YARN
* Exploración de las aplicaciones de YARN Applications a través de la Web UIs y de Shell
* Agregación de los logs de YARN

**Configuración de Hadoop y registros de Daemon**

* Elementos de Cloudera Manager para la gestión de configuraciones
* Localizar configuraciones y aplicar cambios de configuración
* Gestión de instancias de Role y añadir servicios
* Configuración del servicio HDFS
* Configuración de los logs de Hadoop Daemon
* Configuración del servicio YARN

**Obtención de datos en HDFS**

* Ingestión de datos desde fuentes de recursos externos con Flume
* Ingestión de datos desde bases de datos relacionales con Sqoop
* REST Interfaces
* Buenas prácticas para la importación de datos

**Planificación de un cluster Hadoop**

* Consideraciones generales de planificación
* Elección correcta de Hardware
* Opciones de Virtualización
* Consideraciones de red
* Configuración de nodos

**Instalación y configuración de Hive, Pig e Impala**

* Hive
* Impala
* Pig

**Clientes Hadoop incluidos en Hue**

* ¿Qué es un cliente de Hadoop?
* Instalación y configuración de clientes Hadoop
* Instalación y configuración de Hue
* Autorizaciones y autenticación Hue

**Configuración avanzada de un cluster**

* Parámetros avanzados de configuración
* Configuración de puertos Hadoop
* Configuración de HDFS para la organización en rack
* Configuración de HDFS para obtención de alta disponibilidad

**Seguridad Hadoop**

* ¿Por qué es importante la seguridad en Hadoop?
* Conceptos del sistema de seguridad de Hadoop
* Qué es Kerberos y cómo funciona
* Securización de un clúster Hadoop Cluster con Kerberos
* Otros conceptos de seguridad

**Gestión de recursos**

* Configuración de cgroups con Static Service Pools
* El Fair Scheduler
* Configuración de Dynamic Resource Pools
* Configuraciones de CPU y memoria YARN
* Impala Query Scheduling

**Mantenimiento de un cluster**

* Chequeo del estado de HDFS
* Copia de datos entre clústers
* Añadir y eliminar de nodos en el clúster
* Rebalanceo del Cluster
* Directorio de Snapshots
* Actualización del clúster

**Solución de problemas y monitorización de un cluster**

* Sistema general de monitorización
* Monitorización de clústers Hadoop
* Solución de problemas habituales en el clúster de Hadoop
* Errores habituales en la configuración

**Precio:**

1670 €

**Duración:**

28 horas

**Lugar de impartición**

PUE Madrid

c/ Arregui y Aruej, 25-27.

Madrid, 28007.

### 4.2.2 Curso Desarrollador Hadoop ([www.formacionhadoop.com](http://www.formacionhadoop.com))



Se trata de un curso online de 40 horas enfocado a los desarrolladores que quieran aprender a construir potentes aplicaciones de procesamiento de datos utilizando la tecnología de Apache Hadoop.

Una vez finalizado el curso, el alumno deberá realizar el examen de certificación para acreditar que ha superado el curso adquiriendo los conocimientos necesarios. El examen se compone de preguntas tipo test. Para superar la certificación el alumno debe obtener más de un 70% de preguntas correctas. Una vez superado, el alumno recibe un título acreditativo de que ha superado el curso.

**Programación:**

– Conceptos Básicos de Hadoop

– Desarrollo de Aplicaciones MapReduce

– API MapReduce

– Resolución de problemas con MapReduce

– Ecosistema Hadoop: Oozie,Hue,Hive y Pig

**Requisitos:**

– Conocimientos básicos en programación (Java)

– Conocimientos básicos de Linux/Unix

– Comprensión de los conceptos informáticos comunes

– No se requiere conocimientos previos de Hadoop

**Materiales:**

– Temario completo del máster en formato pdf

– Ejercicios prácticos (cuaderno de ejercicios, soluciones y datasets)

– Material privado complementario para facilitar el aprendizaje(vídeos y documentos)

– Examen de certificación de Formación Hadoop

– Canal de Youtube con vídeo tutoriales para facilitar el aprendizaje

– Bolsa de Empleo

**Tutorías:**

El alumno dispone de varios métodos para solventar sus dudas:

* Tutorías vía chat desde el aula virtual.
* Contacto con el tutor del curso vía email.
* Tutorías vía Skype.

**Duración:**

40 horas

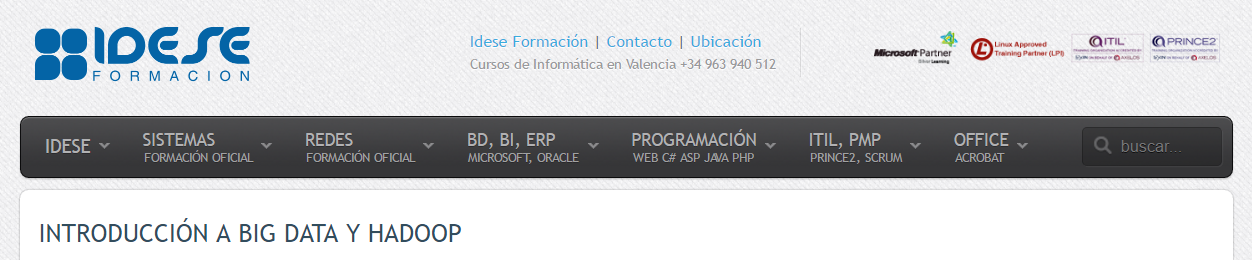
**Precio:**

199 €

**Lugar de impartición:**

Online

### 4.2.3 Introducción a Big Data y Hadoop



En este curso se hace una introducción al ecosistema de aplicaciones de Hadoop para Big Data, incluyendo la programación map/reduce y la programación para Amazon EC2.

**Programación:**

Módulo 1: Big Data y el ecosistema Hadoop

Módulo 2: Almacenamiento de datos en Hadoop

Módulo 3: Procesamiento de datos con map/reduce

Módulo 4: Personalización de map/reduce

Módulo 5: Construcción de aplicaciones fiables usando map/reduce

Módulo 6: Automatización de datos con OOZIE

Módulo 7: Manejo de OOZIE

Módulo 8: Características avanzadas de OOZIE

Módulo 9: Hadoop y el tiempo real

Módulo 10: Seguridad en Hadoop

Módulo 11: Ejecución de aplicaciones Hadoop en AWS

**Duración:**

16 horas

**Precio:**

720 €

**Lugar de impartición:**

Valencia

Marques de Zenete 18 bajo,

46002, Valencia, España

## 4.3 Cursos no gratuitos sobre la tecnología específica B

### 4.3.1 Curso no gratuito 1 sobre la tecnología específica B

### 4.3.2 Curso no gratuito 2 sobre la tecnología específica B

### 4.3.n Curso no gratuito n sobre la tecnología específica B

## 4.4 Curso no gratuito sobre la tecnología específica A y B

Queríamos mencionar este curso, ya que, aunque no se trata de un curso de una de las dos tecnologías en concreto, sí que es un curso sobre las dos tecnologías específicas que hemos elegido,

### 4.4.1 Curso de Big Data con Apache Hadoop y Apache Spark



Durante el curso se estudiará en profundidad las principales funcionalidades de Apache Hadoop y Apache Spark en cuestiones relativas al almacenamiento, intercambio y procesamiento y seguridad de los datos entre otras, a través de numerosas prácticas durante la formación. Al finalizar el curso, conocerás todos los componentes del ecosistema Hadoop que hoy se utilizan en proyectos Big Data (Hive, HDFS, Pig, etc.), y serás capaz de utilizar Apache Spark para el análisis de datos a gran escala, tanto en modo offline como en tiempo real con Spark Streaming.

**Programación:**

**1. INTRODUCCIÓN A TECNOLOGÍAS BIG DATA, APACHE HADOOP Y APACHE SPARK**

* ¿Qué es Big Data y qué hay de nuevo?
* Principales tecnologías Big Data.
* Apache Hadoop: plataforma de almacenamiento y procesamiento de datos.
* Apache Spark: procesamiento de datos.

**2. APACHE HADOOP: CORE.**

* Almacenamiento de datos en Hadoop: HDFS.
* Gestión de recursos y tareas: YARN.
* Instalación, administración y monitorización del cluster con Apache Ambari.
* Seguridad y gobierno de datos.

**3. INGESTA Y TRANSFORMACIÓN DE DATOS EN APACHE HADOOP.**

* Apache Sqoop: intercambio de datos con bases de datos relacionales.
* Apache Flume y Apache Kafka: ingesta de datos.
* Apache Pig: transformación de datos en Hadoop.
* Herramientas ETL: Informatica PowerCenter, Pentaho Kettle.

**4. ACCESO Y EXPLOTACIÓN DE DATOS EN APACHE HADOOP.**

* MapReduce: “antiguo” framework de procesamiento de datos.
* Apache HBase: NoSQL sobre Hadoop.
* Apache Hive: SQL sobre Hadoop.
* Otras tecnologías de SQL sobre Hadoop: Cloudera Impala, Apache Drill, Apache Phoenix.
* Procesamiento de eventos en tiempo real con Apache Storm.

**5.- APACHE SPARK**

* Introducción a Apache Spark. Spark vs MapReduce vs Tez
* Introducción a Scala.
* Conceptos básicos, RDDs, transformaciones y acciones.
* Desarrollo de aplicaciones con Spark.

**6.- SPARK STREAMING**

* Introducción a Spark Streaming
* Casos de uso. Apache Storm vs Spark Streaming.
* Desarrollo de aplicaciones realtime con Spark Streaming.

**Duración:**

18 horas

* Fecha de inicio:**Lunes 17 de abril de 2017**
* Primera semana**Lunes, miércoles y jueves de 18:30 a 21:30**
* Segunda semana:**Lunes, miércoles y jueves de 18:30 a 21:30**
* Nº semanas: **2**
* Fin de curso: **Jueves 27 de abril a las 21:30**

**Precio:**

320 €

**Lugar de impartición:**

New Horizons,

Paseo de la Castellana, 163, cercano el metro de Cuzco,

Madrid.

# 5. Fuentes de información (cursos gratuitos)

## 5.1 Cursos gratuitos sobre el tipo de tecnología en general

### 5.1.1 Curso gratuito 1 sobre el tipo de tecnología en general

### 5.1.2 Curso gratuito 2 sobre el tipo de tecnología en general

### 5.1.n Curso gratuito n sobre el tipo de tecnología en general

## 5.2 Cursos gratuitos sobre la tecnología específica A

### 5.2.1 Curso gratuito 1 sobre la tecnología específica A

### 5.2.2 Curso gratuito 2 sobre la tecnología específica A

### 5.2.n Curso gratuito n sobre la tecnología especifica A

## 5.3 Cursos gratuitos sobre la tecnología específica B

### 5.3.1 Curso gratuito 1 sobre la tecnología específica B

### 5.3.2 Curso gratuito 2 sobre la tecnología específica B

### 5.3.n Curso gratuito n sobre la tecnología especifica B

# 6. Ayudas económicas para estudiar las tecnologías

El apartado 6 del TG1 se refiere a posibles ayudas económicas que se puede conseguir (becas, descuentos, etc.) para aprender las tecnologías.

# 7. Recursos para implementar las tecnologías

En el apartado 7, hay que indicar recursos gratuitos y no gratuitos. Los recursos se refieren a entornos de desarrollo, librerias de software, etc. necesarios para implementar prototipos.

## 7.1 Recursos para implementar Hadoop

### 7.1.1 Recursos gratuitos para implementar la tecnología A

**Herramienta Hadoop**



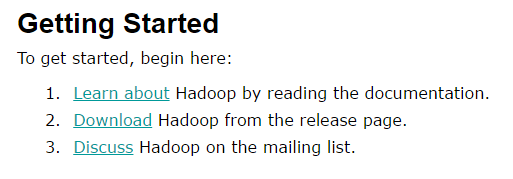
El principal recurso que necesitamos es el propio software Apache Hadoop. Este framework nos lo podemos descargar desde la página web:

<http://hadoop.apache.org/>

El proyecto ya incluye los siguientes módulos:

* Hadoop Common
* Hadoop Distributed File System (HDFS™)
* Hadoop YARN
* Hadoop MapReduce

Todos los recursos necesarios para su implementación se pueden conseguir desde la web anteriormente mencionada (tutoriales, documentación, herramientas, ayuda, etc)



### 7.1.2 Recursos no gratuitos para implementar la tecnología A

**Servidores y personal cualificado**

Código abierto no significa que todo sea gratuito. La implementación de Hadoop localmente requiere el uso de servidores, así como de expertos en Hadoop, para configurarlos, adaptarlos y mantenerlos.

**Microsoft Azure HDInsight.**

Esta herramienta da la opción de utilizar Hadoop en la nube:

¿Por qué usar Hadoop en la nube?

**La nube ahorra tiempo y dinero:** Un servicio en la nube permite poner en marcha un clúster de Hadoop en cuestión de minutos sin costo inicial alguno.

**La nube es flexible y escala con rapidez:** En la nube de Microsoft Azure, solo paga por el almacenamiento y los servicios de proceso que utilice, cuando los utilice. Puede poner en marcha un clúster de Hadoop, analizar los datos y apagarlo para detener el contador.

**Velocidad gracias a la nube:** Cree un clúster de Hadoop en cuestión de minutos y agregue nodos a petición. La nube ofrece a las organizaciones un tiempo de amortización inmediato.

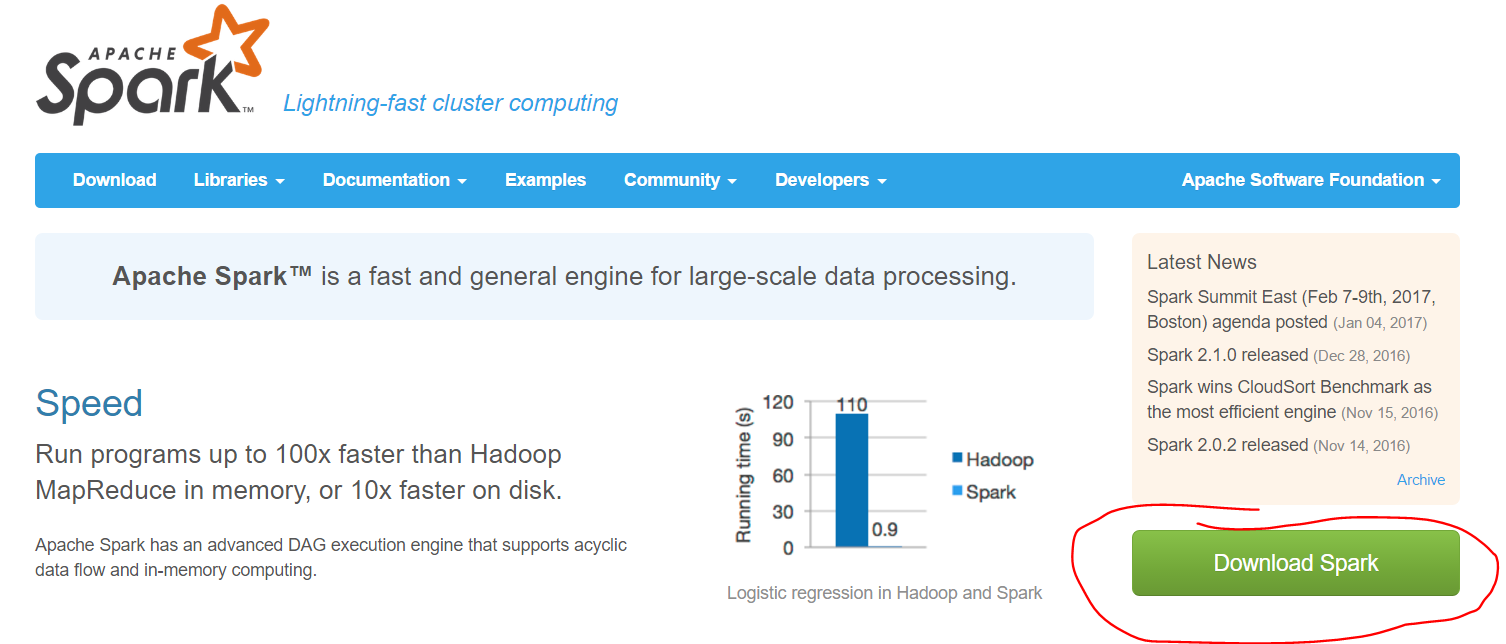
Microsoft Azure HDInsight es un servicio en la nube de Azure basado en Apache Hadoop al 100 %. Ofrece todas las ventajas de Hadoop, además de la capacidad de integración con Excel, clústeres de Hadoop locales y el ecosistema de software y servicios empresariales de Microsoft.

## 7.2 Recursos para implementar la tecnología B

### 7.2.1 Recursos gratuitos para implementar la tecnología B

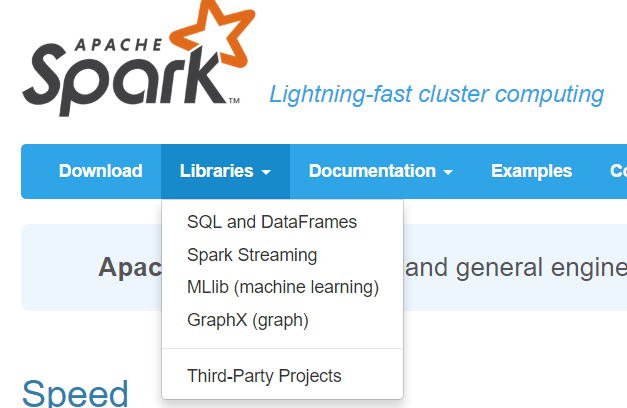
El principal recurso a la hora de la implementación es la propia herramienta Apache Spark. Se podrá descargar desde la web:

<http://spark.apache.org/>

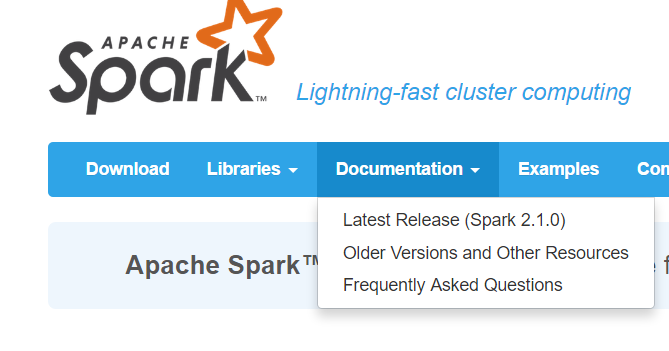


Librerías

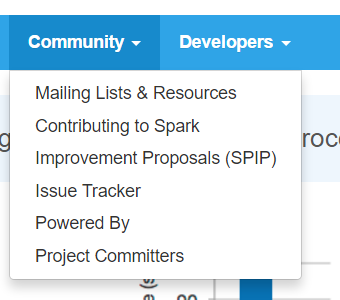
Todas las librerías que pudiésemos necesitar las podremos consultar desde aquí:



Así como toda la documentación a consultar:



También está a nuestro alcance toda una comunidad:



### 7.2.2 Recursos no gratuitos para implementar la tecnología B

No hemos encontrado ningún recurso de pago necesario para una posible implementación, más allá de los cursos no gratuitos de los que ya hemos hablado en los anteriores puntos. Al igual que en la anterior tecnología podríamos mencionar los servidores y posibles expertos en esta tecnología. Esto se debe a que ambos productos son proyectos de Apache.org. Esto significa que son productos open source de software libre.

# 8. Conclusiones

Aunque hay quienes ven estos dos frameworks como competidores en el espacio de big data, no es tan fácil hacer una comparación Spark vs Hadoop. Hacen muchas cosas igual, pero hay algunas áreas donde ambos no se superponen.

A la hora de buscar fuentes de información sobre estas tecnologías, debido a la gran demanda de conocimiento que existe sobre el big data, hemos podido comprobar la multitud de cursos y masters que se ofrecen de este tecnología.

Esto conlleva una contrapartida, ya que al ser un tema tan demandado es muy difícil encontrar cursos con una buena relación calidad / precio. Desde nuestra posición de estudiantes, en la cual lo que más nos pueden llegar a interesar son los cursos gratuitos, hemos podido observar que son casi inexistentes. En cambio nos será de gran ayuda la gran documentación existente de ambas herramientas y la gran comunidad de usuarios que hay detrás, aportando conocimiento. Esto tiene mucho que ver que ambas sean de código abierto.