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

Javier Cia del Toro, Alberto Moraga Fernández, Jorge Benítez Abad.

## 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).

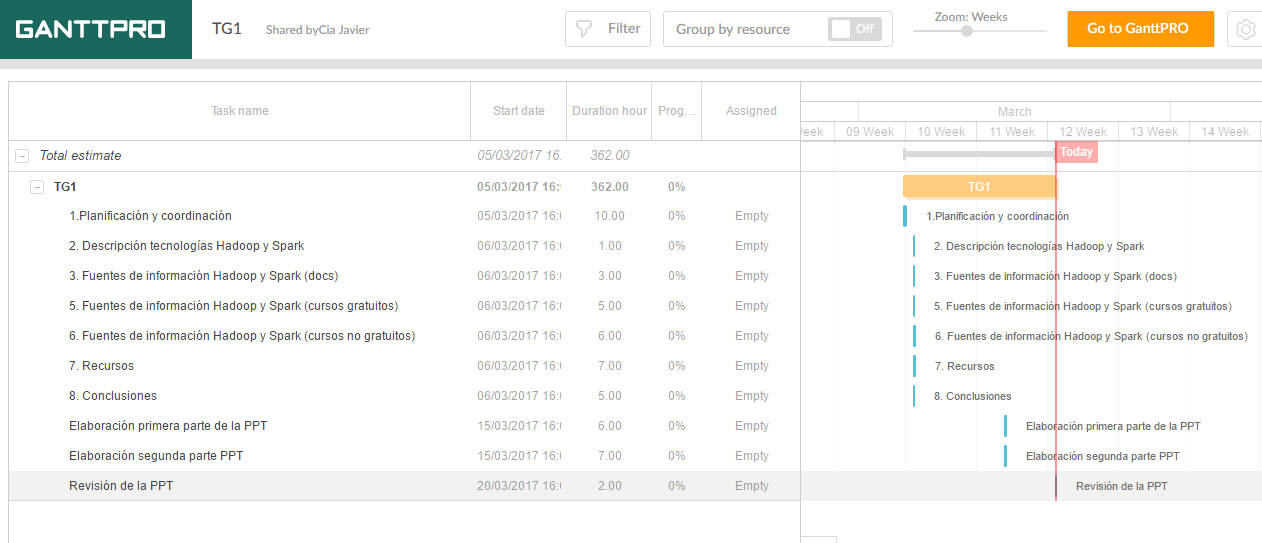
En este enlace hay un ejemplo de una posible planificación del trabajo, que se puede usar como referencia, pero indicando los nombres reales de las tecnologías y de los participantes.

[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.

**Respuesta:**

En cuanto a la planifiación que hemos realizado es la siguiente:



Como vemos lo hemos hecho con GanttPro, pero nos hemos encontrado con el problema de que no nos permitía trabajar por equipos con la herramienta gratuita, por lo tanto hemos indicado manualmente abajo a quien corresponde cada tarea:

-**Alberto Moraga:** Recursos, conclusiones y elaboración primera parte PPT.

**La suma total es de 18 horas.**

**-Javier Cia:** Planificación y Coordinación, descripción tecnologías Hadoop y Spark, fuentes de información Hadoop y Spark(docs) y revisión de la PPT.

**La suma total es de 17 horas.**

**-Jorge Benitez:** Fuentes de información Hadoop y Spark (cursos gratuitos), fuentes de información Hadoop y Spark (cursos no gratuitos), ayudas y elaboración segunda parte de la PPT.

**La suma total es de 18 horas.**

A continuación adjuntamos el [enlace](https://app.ganttpro.com/shared/token/0b7b9b1352aa16d0ac8c1a4b53fadd8a8b479a4431f26a386c8c9e8f857f1218#!/app/home) del proyecto para comprobar que lo hemos hecho.

## 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.

**Respuesta:**

Ajuntamos en este apartado el [enlace](https://github.com/javierCia1/TG1) al repositorio en GitHub que contendrá tanto la presentación como el presente documento que contiene las preguntas que debemos contestar.

# 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.

**Respuesta:**

El tipo de tecnologías sobre las que hemos decidido hacer el trabajo son dos tecnologías Big Data llamadas Hadoop y Spark.

El sistema Hadoop como tal tiene infinidad de funcionalidades y herramientas asociadas relacionadas con Big Data.

En resumen, Hadoop es un sistema de archivos para almacenar enormes cantidades de datos en un cluster maestro-esclavo y que tiene infinidad de herramientas que permiten interactuar con el:

* HDFS: Es el sistema de archivos como tal en el cual se almacenan los datos, está compuesto por el nodo maestro, que se encarga de almacenar los metadatos, y por una serie de nodos esclavos que se encargan de almacenar los datos.
* MapReduce: Es la herramienta propia de Hadoop para procesar los datos que están almacenados en HDFS, en este trabajo es bastante importante ya que Hadoop no va a competir con Spark en todos sus aspectos, sino principalmente en este (cuando expliquemos Spark veremos a que nos referimos).
* Flume: Sirve para mover grandes cantidades de logs a un respositorio central
* Hive: Sirve para organizar los datos de una manera que sean más accesibles para su procesamiento.
* Sqoop: Sirve para transferir datos desde bases de datos tradicionales (con datos estructurados) a HDFS, y viceversa.
* Y un amplio etcétera de herramientas que sirven para interactuar con HDFS.

En cambio, Spark es un framework orientado sola y exclusivamente al procesamiento de grandes cantidades de datos-

Debido a esto en lo que concretamente vamos a realizar este trabajo sobre Spark y Hadoop es en materia de procesamiento de grandes cantidades de datos (Big Data).

# 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 Big Data

### 3.1.1 Fuente de información 1 sobre Big Data

**Respuesta:**

**Título:** Hadoop: soluciones Big Data

**Autor:** Boris Lublinsky y Kevin T.Smith

**Fuente:** Buscador Google

Es un libro que se refiere a todas las tecnologías Big Data relacionadas con Hadoop, entre sistemas obviamente incluye el procesamiento de grandes cantidades de datos con herramientas como MapReduce y Spark.

### 3.1.2 Fuente de información 2 sobre Big Data

**Respuesta:**

**Título:** Big Data la revolución de los datos masivos.

**Autor:** Viktor Mayer-Schönberger, Kenneth Cukier, Antonio J. Iriarte Jurado

**Fuente:** Buscador Google

Es un libro introductorio que abarca los diferentes aspectos del Big Data.

### 3.1.3 Fuente de información 3 sobre Big Data

**Respuesta:**

**Título:** ¿Qué es el Big Data?

**Autor:** IBM (No pone autor especifico)

**Fuente:** Web de IBM

Es un artículo para gente que tenga conocmientos sobre informática tradicional, pero que no comprenda el significado y los conceptos principales del Big Data.

## 3.2 Fuentes sobre Hadoop (MapReduce)

### 3.2.1 Fuente de información 1 sobre Hadoop, MapReduce

**Respuesta:**

**Título:** Hadoop in Practice

**Autor:** Alex Holmes

**Fuente:** Google Scholar

Es un libro que mediante ejemplos explica el ecosistema Hadoop, el capítulo de principal interés (aunque todos lo son, ya que dan una introducción a MapReduce) es el número tres, ya que aporta información sobre programación MapReduce.

### 3.2.2 Fuente de información 2 sobre Hadoop, MapReduce.

### **Respuesta:**

**Título:** Introducción a Hadoop y su ecosistema

**Autor:** jcasanella, publicado en el blog TicOut

**Fuente:** Buscador de Google

Artículo que aporta información clara, concisa y simple(sobre todo para principiantes en la materia) acerca del ecosistema Hadoop y del procesamiento de datos en este.

### 3.2.3 Fuente de información 3 sobre Hadoop, MapReduce.

**Respuesta:**

**Título:** Hadoop MapReduce CookBook

**Autor:** Srinath Perera

**Fuente:** Google Books

Artículo enfocado a lectores con conocimientos acerca de programación en cualquier lenguaje que quieran aprender los conceptos de programación MapReduce para aplicarlos a entorno Hadoop.

## 3.3 Fuentes sobre Spark

### 3.3.1 Fuente de información 1 sobre Spark

**Respuesta:**

**Título:** Introducción a Apache Spark

**Autor:** Mario Macías, Mauro Gómez, Rubèn Tous y Jordi Torres

**Fuente:** Buscador Google

Se trata de un libro para empezar a programar en Spark sin mucho conocmiento previo en materia de Big Data.

### 3.3.2 Fuente de información 2 sobre Spark

**Respuesta:**

# Título: Learning Spark: Lightning-Fast Big Data Analysis

**Autor:** Holden Karau, Andy Konwinski, Patrick Wendell, Matei Zaharia

**Fuente:** Buscador Google

Es uno de los libros más conocidos para formarse en la tecnología Spark, acarca desde fundamentos de Spark hasta conceptos más complejos acerca de esta tecnología

### 3.3.3 Fuente de información 3 sobre Spark

**Respuesta:**

# Título: A Quick Guide to Spark Streaming

**Autor:** J. Scott.

**Fuente:** La revista SG Buzz me llevó a este artículo.

Es un artículo principalmente conceptual (aunque también entra en cierta materia tecnica) que ayuda a entender lo que es Spark y como funciona.

# 4. Fuentes de información (cursos no gratuitos)

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

### 4.1.1 Curso no gratuito 1 sobre el tipo de tecnología en general

### 4.1.2 Curso no gratuito 2 sobre el tipo de tecnología en general

### 4.1.n Curso no gratuito n sobre el tipo de tecnología en general

## 4.2 Cursos no gratuitos sobre la tecnología específica A

### 4.2.1 Curso no gratuito 1 sobre la tecnología específica A

### 4.2.2 Curso no gratuito 2 sobre la tecnología específica A

### 4.2.n Curso no gratuito n sobre la tecnología específica A

## 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

# 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

# 7. Recursos para implementar las tecnologías

## 7.1 Recursos para implementar la tecnología A

### 7.1.1 Recursos gratuitos para implementar la tecnología A

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

## 7.2 Recursos para implementar la tecnología B

### 7.2.1 Recursos gratuitos para implementar la tecnología B

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

# 8. Conclusiones