# Anteproyecto Trabajo Fin de Grado

### Título: “Estudio y Evaluación de Sistemas Big Data de Tratamiento de Información”

### Autor: Enrique Ruiz García

### Tutor: Santiago Pavón Gómez

### Titulación: Grado en Ingeniería de Tecnologías de Servicios de Telecomunicación

### Departamento: Ingeniería Sistemas Telemáticos

### Centro: Escuela Técnica Superior de Ingenieros de Telecomunicación de Madrid

## Introducción

En el contexto actual de las tecnologías de la información toma gran importancia poder manejar la gran cantidad de información que circula por internet de una manera rápida y eficiente para poder prestar servicios de la manera más personalizada posible. Toda esta información habitualmente llamada Big Data, supone un gran activo para las empresas ya que les permite obtener una gran cantidad de información de sus clientes, lo cual les permite ofrecerles soluciones más especializadas lo que supone una ventaja competitiva. Big Data, entonces, alude a un conjunto de datos cuyo tamaño constituye un desafío a la capacidad de la mayoría de los softwares utilizados para capturar, gestionar y procesar la información dentro de un lapso tolerable.

El Big Data requiere una gran capacidad de almacenamiento, búsqueda, compartición, análisis y visualización, y al haber surgido de manera reciente las tecnologías para su tratamiento están la mayoría de ellas desarrollándose. El presente trabajo va a indagar en estas tecnologías desarrolladas para el tratamiento de Big Data.

## Objetivos

El Big Data presenta grandes dificultades para su manejo, es por ello que se necesitan herramientas que se encarguen de analizar la inmensa cantidad de datos de una manera rápida, es por ello que el trabajo se va a centrar en la tecnología Apache Spark para el manejo de grandes cantidades de información. Una de las principales razones por las que vamos a analizar esta tecnología es porque puede analizar la información de una manera mucho más rápida de como lo hacen herramientas similares.

Con la realización del trabajo también se busca la familiarización con el framework Play2 el cual se utiliza para el desarrollo rápido de aplicaciones web con Java y Scala, el cual está basado en el modelo MVC.

Por tanto el objetivo principal del trabajo va a ser realizar una aplicación utilizando Play2 en la cual podamos manejar, e interpretar sus datos mediante Apache Spark.

## Casos de estudio

El procesado de la información de Big Data es algo que ha surgido hace poco tiempo y por lo tanto existen muchas líneas de investigación para su gestión y tratamiento. En el presente trabajo se va a analizar Apache Spark, una herramienta desarrollada en el lenguaje de programación Scala, que ha sido desarrollada desde su inicio buscando soportar en memoria algoritmos iterativos que se pudiesen desarrollar sin escribir un conjunto de resultados cada vez que se procesaba un dato.

Spark tiene un framework integrado para implementar análisis avanzados que incluye la librería MLlib, el motor gráfico GraphX, Spark Streaming, y la herramienta de consulta Shark. Esta plataforma asegura a los usuarios la consistencia en los resultados a través de distintos tipos de análisis.

En este trabajo se va a estudiar la construcción de aplicaciones web utilizando el framework Play2 con Scala, así como su despliegue unido a Spark.

## Metodología y plan de trabajo

El trabajo se va a desarrollar en tres fases. La primera de ellas va a consistir en el análisis de la documentación tanto de Apache Spark, del lenguaje Scala y del framework Play2 y la realización de los primeros tutoriales básicos para poder empezar con el grueso del trabajo. La segunda fase del trabajo va a consistir en el diseño de una aplicación creada a través de Play2 en la cual vamos a realizar el análisis de la información con Apache Spark. La tercera fase consistirá en llevar a cabo el análisis de datos y el estudio de estos para observar en ellos las ventajas que nos produce el manejo de Apache Spark.

Para el desarrollo se va a utilizar un ordenador personal en el cual se instalaran los diferentes frameworks.

## Bibliografía

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

http://ampcamp.berkeley.edu/3/exercises/overview-of-the-exercises.html

<http://www.playframework.com/documentation/2.3.x/Home>

<http://typesafe.com/platform>