

---

# Máster en Business Intelligence & Data Science

Edición 2017/2018



Universidad  
de Alcalá

---

**Asignatura:** Procesamiento de Streams

**Módulo:** Paralelización de datos

**Profesor(es):** Borja Moreno Pozo, [borja.moreno@strategybigdata.com](mailto:borja.moreno@strategybigdata.com)

## OBJETIVOS

El objetivo del módulo es presentar las principales soluciones tecnológicas presentes en el mundo de open source. Junto con su presentación se explicarán sus arquitecturas y casos de usos.

En concreto, los objetivos del módulo serán:

1. Conocer las principales herramientas utilizadas para procesamiento de streams.
2. Aprender la arquitectura de las distintas soluciones.
3. Explotación de distintas fuentes de datos para enriquecer los proyectos.
4. Aprender a identificar cuales herramientas son las idóneas para las distintas casuísticas.

## REQUISITOS PREVIOS

Es necesario tener una base conceptual del funcionamiento de Spark.

## METODOLOGÍA

La asignatura se va a plantear para dando un fuerte contenido teórico. Para el contenido teórico, se explicará la arquitectura particular, casos de uso e implementación de las herramientas. Para poder afianzar los contenidos teóricos, se tendrá una también una parte práctica. Se realizarán ejercicios en clase, entregables y discusiones sobre los casos de uso.

## PROGRAMA

El programa se estructura en los bloques que se describen a continuación.

### **Bloque 1:** “Apache Spark Streaming”

- ¿Qué es Apache Spark Streaming?
- Arquitectura.
- Ejemplos.

### **Bloque 2:** “Apache Kafka”

- ¿Qué es Kafka?
- Arquitectura.
- Ejemplos.

### **Bloque 3:** “Apache Flink”

- ¿Qué es Flink?
- Arquitectura.

### **Bloque 4:** “Arquitecturas”

- Arquitectura lambda.
- Arquitectura kappa.

## EVALUACIÓN

La evaluación se realiza mediante Pruebas de Evaluación Continua (PEC). La siguiente tabla las describe. Las fechas de publicación, entrega y publicación de calificaciones se hará disponible a través del Aula Virtual.

Prueba	Tipo (indicar tipo de prueba, práctica, test, etc. Se recomienda ejercicios prácticos o casos prácticos)	Peso
PEC1	Practica individual: ejercicios prácticos usando spark streaming	20%
PEC2	Practica individual: ejercicios prácticos usando kafka	20%
PEC3	Práctica individual: ejercicio práctica Kafka y spark streaming	50%
PEC4	Práctica individual: ejercicio práctica de arquitectura	10%

## BIBLIOGRAFÍA OBLIGATORIA

- Documentación oficial de apache spark streaming: <https://spark.apache.org/streaming/>
- Documentación oficial de apache kafka: <https://kafka.apache.org/>
- Documentación oficial de apache flink: <https://flink.apache.org/>

## BIBLIOGRAFÍA RECOMENDADA

- <http://milinda.pathirage.org/kappa-architecture.com/>, Milinda Pathirage.
- <https://mapr.com/developercentral/lambda-architecture/>, MapR (Developer Central)

## PROFESORADO

**Borja Moreno Pozo** es Ingeniero Industrial en la Universidad Politécnica de Madrid con un Master en Big Data and Business Analytics en el CIFF. Mi carrera profesional, se desarrolla primero en HPE como Big Data Consultant y en actualidad Big Data Engineer en Strategy Big Data.

LinkedIn: <https://www.linkedin.com/in/borja-moreno-pozo-3547a177>

Mail: [borja.moreno@strategybigdata.com](mailto:borja.moreno@strategybigdata.com)