

---

# **Máster en Business Analytics y Big Data**

## **Edición 2015 - intensiva**



---

**Asignatura:** ESCALABILIDAD EN ALMACENES DE DATOS  
**Módulo:** BIG DATA/GESTIÓN DE DATOS  
**PROFESOR:** Javier Rodríguez Rodríguez,

## OBJETIVOS

El objetivo general de la asignatura es el de adquirir los conocimientos y habilidades necesarias para comprender la problemática de escalar almacenes de datos NoSQL. Nos centraremos para ello en MongoDB y Cassandra como ejemplos paradigmáticos de cómo se persigue escalabilidad con diferentes estilos arquitectónicos.

Los objetivos concretos de la asignatura son los siguientes resultados del aprendizaje:

1. Comprensión de los conceptos básicos que se asocian al rendimiento y escalabilidad de almacenes de datos.
2. Interiorizar el impacto de los índices en el rendimiento de consultas.
3. Comprender la repercusión de la replicación en la disponibilidad.
4. Comprensión del impacto del particionamiento y redundancia de datos en la escalabilidad.
5. Ser capaz de diseñar índices para colecciones de documentos en MongoDB y estudiar su impacto en el rendimiento.
6. Comprender y ensayar las particularidades de la arquitectura de replicación y *sharding* de MongoDB.
7. Comprender y ensayar la arquitectura de clustering de Cassandra.

## METODOLOGÍA

Se usará una combinación de descripción teórica de los conceptos objetivo complementada con ejercicios prácticos que permitan a los alumnos experimentar de primera mano los conocimientos adquiridos.

Los ejercicios se realizarán en instalaciones virtualizadas con entornos preconfigurados de MongoDB y Cassandra.

Es conveniente resaltar que la asignatura no pretende formar en habilidades prácticas propias de un administrador de sistemas NoSQL sino que está diseñada para comprender la complejidad inherente a instalaciones complejas de bases de datos NoSQL y mostrar las herramientas necesarias para conseguir verdadera escalabilidad en ese tipo de productos.

## PROGRAMA

### **Sesión 1:** Rendimiento y escalabilidad con MongoDB

Actividades: Diseño de índices y verificación de impacto. Implantación redundante de MongoDB.

Materiales: VM con MongoDB 3 y Ops Manager

### **Sesión 2:** Escalabilidad con Cassandra

Actividades: Uso de CCM para simular un clúster de Cassandra en entornos monopuesto.

Materiales: VM con Cassandra 2.4 y CCM

## PROFESORADO

Javier Rodríguez Rodríguez (<https://es.linkedin.com/in/mezco>) es físico de formación e informático de profesión. Es el responsable del área de Arquitectura e Integración en Gfi España. Trabaja en el diseño e implementación de arquitecturas complejas con múltiples tecnologías y paradigmas, desde los clásicos SOA y BPM a los más modernos enfoques de API Management, microservicios y Big Data.