

---

# Máster en Business Analytics y Big Data

## Edición Executive 2015



---

**Asignatura:** Motores de búsqueda

**Módulo:** Motores de búsqueda

**Coordinador:** Miguel-Angel Sicilia, [msicilia@uah.es](mailto:msicilia@uah.es)

**Profesor:** Luis Polanco, [luis.polanco@campusciff.net](mailto:luis.polanco@campusciff.net)

## OBJETIVOS

### ¿Cuáles son los objetivos de este bloque?

Formar al alumno en técnicas de indexación y búsqueda de información mediante el uso de motores de búsqueda.

Durante las sesiones de contenido conceptual el alumno adquirirá una visión global teórica sobre motores de indexación y búsqueda. Responderemos a preguntas sobre qué son, para que se utilizan, cuál es su papel en entornos Big Data, qué beneficios tienen su uso así como factores claves, herramientas y técnicas comunes que se aplican durante procesos de indexación de datos.

Durante las sesiones de contenido práctico, el alumno desarrollará habilidades y conocimientos prácticos sobre uno de los motores de búsqueda más populares del mercado: ElasticSearch.

### ¿Qué aprenderá el alumno en esta asignatura?

Conoceremos qué es un motor de búsqueda y para qué se utiliza

Analizaremos los motores de búsqueda más populares en el mercado actual

Comprenderemos donde encajan los motores de búsqueda en el universo de Big Data

    Cuando usar una base de datos documentales vs. motor de búsqueda

    Hablabamos sobre los casos de uso más importantes para este tipo de componentes

Instalaremos y configuraremos un motor de búsqueda: ElasticSearch.

Publicaremos datos en ElasticSearch y los indexaremos

Trabajaremos sobre los diferentes tipos de búsquedas

Crearemos un cluster y analizaremos las arquitecturas de despliegue

Analizaremos las diferentes herramientas de monitorización para operaciones

## METODOLOGÍA

### Sesiones presenciales

3h de contenido conceptual sobre motores de búsqueda – hablaremos, haciendo uso de ayudas visuales, sobre motores de búsqueda: qué son, para qué se utilizan, cómo encajarlos en el mundo de big data y analizaremos las tendencias del mercado actual para estas tecnologías.

7h de contenido práctico donde trabajaremos sobre ElasticSearch: instalaremos ElasticSearch en nuestras máquinas y llevaremos a cabo procesos de configuración, indexación y búsquedas.

### Trabajos a llevar a cabo

6 ejercicios prácticos a completar individualmente

1 ejercicio práctico a completar en grupo

## MATERIALES

**Plan de Lección (PL):** guía para el facilitador que contiene el desarrollo de cada una de las lecciones con sugerencias metodológicas y con referencias cruzadas a otras asignaturas ya impartidas, a material de distribución y ayudas visuales.

**Ayudas Visuales (formato PowerPoint) (AV):** son ayudas visuales con los contenidos principales del Curso que el facilitador va proyectando, complementando con la información existente en el Plan de lecciones.

**Material técnico de referencia:** sección que incluye un conjunto de publicaciones técnicas en texto completo que complementan los contenidos del Curso.

**Material de evaluación (ME):** Documentos que presentan los instructivos y formatos guías de los talleres que se deben realizar en algunas lecciones y otros materiales de apoyo

## PROGRAMA

### Teoría 3h

#### Sesión 1 – Motores de Búsqueda (2h)

#### Teoría – Introducción a los motores de búsqueda

¿Qué es un motor de búsqueda?

Big Data y el papel de los motores de búsqueda

¿Cuáles son los componentes básicos de cualquier motor de búsqueda?

¿Cómo funcionan los motores de búsqueda?

Estudio de casos reales

Actividades:

Clase teórica

Materiales:

Ayudas audiovisuales (formato PowerPoint)

## **Sesión 2 –ElasticSearch – Conceptos Básicos**

Teoría - Conceptos básicos (1h)

Introducción a ElasticSearch

- ¿Qué es ElasticSearch?
- Evolución Genética
- Características principales de ES

Conceptos básicos

- Terminología de ES
- Arquitectura de servidor
- Operaciones automáticas
- Índices y búsquedas

El cluster API

Los logs de ElasticSearch

Herramientas complementarias

- KOPF
- Sense
- Kibana

Actividades:

2h de clase teórica

Ejercicios personales

Materiales:

Ayudas visuales (formato PowerPoint)

Guía paso a paso de instalación de ES

ElasticSearch software - <http://elastic.co/>

Ejercicio personal – instalación de ES

## **Sesión 3 – Instalación y configuración de ES**

Instalación y configuración de ElasticSearch

- Arrancando y parando ES
- Ficheros y carpetas de configuración

Materiales:

Ayudas visuales (formato PDF)

Guía paso a paso de instalación de ES

ElasticSearch software - <http://elastic.co/>

Ejercicio personal – instalación de ES

## **Sesión 4 – Arquitectura del API de ES**

¿Cómo podemos interactuar con ES?

Tipos de llamadas REST

Llamadas básicas – GET/DELETE + URL + Query

Llamadas avanzadas - PUT/POST + URL + Payload

Rest APIs expuestas por ElasticSearch

Materiales

Ayudas visuales (formato PowerPoint)

ElasticSearch software - <http://elastic.co/>

ElasticSearch reference - [aquí](#)

## **Sesión 5 –Indexación y carga de datos (2h)**

Práctica – Analizadores y herramientas de carga de datos (2h)

Analizadores

Herramientas de carga de datos - logstash

Carga de datos masivas – standard vs. bulk

Actividades

1h de clase práctica

Ejercicios personales suministrados – Indexación y carga de datos

Materiales

Ayudas visuales (formato PowerPoint)

ElasticSearch software - <http://elastic.co/>

ElasticSearch reference - [aquí](#)

[Grok debugger](#)

### **Sesión 6 – El Document API (2h)**

CRUD, consultas básicas y ordenación

DSL - El lenguaje de consulta de ElasticSearch

Operaciones CRUD en ElasticSearch

Operaciones de búsqueda básicas

Indexación y carga de datos

- Creación de índices
- Creación de mapping y templates
- Propiedades y atributos

Valores especiales en un documento

Actividades

Ejercicios compartidos durante la clase

Ejercicio personal – CRUD y búsquedas

Materiales

Ayudas visuales (formato PDF)

ElasticSearch reference - [aquí](#)

Sense/[JSONLint](#)

### **Sesión 7 – Consultas avanzadas (2h)**

Práctica – Búsquedas avanzadas (2h)

Operaciones de búsqueda avanzadas

- Búsquedas para componentes Type Ahead,
- Boosted - Mejorando la calidad de las respuestas
- Facetas y sugerencias

Operaciones de filtrado y ordenación

Arquitecturas de despliegue/escalabilidad

- Clustering
- Replication
- Sharding

Cosas a considerar para mejorar el performance de tu motor de búsqueda

Actividades

Ejercicios compartidos durante la clase

Ejercicio personal – CRUD y búsquedas

Materiales

Ayudas visuales (formato PowerPoint)

ElasticSearch software - <http://www.elasticsearch.org/>

ElasticSearch reference - [aquí](#)

KOPF - <https://github.com/lmenezes/elasticsearch-kopf>

Kibana - <http://www.elasticsearch.org/overview/kibana/>

## **REFERENCIAS**

<http://www.elasticsearch.org>

<https://github.com/elasticsearch/elasticsearch>

<http://www.elasticsearch.org/guide/en/elasticsearch/reference/current/index.html>  
<http://logstash.net/>  
<http://www.elasticsearch.org/overview/kibana/>  
<http://www.elasticsearch.org/guide/en/elasticsearch/rivers/current/>  
<http://lucene.apache.org/solr/resources.html>  
<http://support2.microsoft.com/ph/15080>

## EVALUACIÓN

La evaluación del curso responde a la necesidad de verificar que cada alumno alcanza el nivel deseado en la materia consiguiendo así cumplimentar los aspectos técnicos que deben ser reforzados.

El cumplimiento del desempeño del curso será evaluado siguiendo los siguientes principios:

- Evaluación permanente a través de la interacción durante todo el Curso
- Evaluación de la apreciación diaria de los participantes sobre cada una de las lecciones y sobre el curso como conjunto.
- Evaluación de los diferentes ejercicios planteados así como mediante el trabajo conjunto de los participantes durante las clases prácticas

### Niveles de consecución de los objetivos

Objetivo	Nivel alto	Nivel medio	Nivel bajo
Motores de búsqueda	Comprender el modelo de indexación en motores de búsqueda modernos.	Saber las diferencias entre un motor de búsqueda y una base de datos tradicional	Comprender cuál es la necesidad de negocio que resuelve un motor de búsqueda
Instalación y configuración de ES	Comprender al detalle el modelo de instalación, configuración y ejecución de ES como cluster	Configurar ES como servicio en una máquina y ser capaz de arrancarlo y pararlo sin problemas	Saber instalar y ejecutar Elasticsearch manualmente en una máquina
Indices y RESTFul API	Configuración de índices haciendo uso de modelos de mapping adecuados para cada caso de uso.	Ejecución de consultas básicas sobre índices en ES	Altas, bajas y modificaciones de documentos en un servidor de ES
Herramientas de utilidad en ES	Comprensión avanzada de logstash y su modelo de carga de datos	Uso básico de logstash y comprensión adecuada de herramientas de gestión de ES	Configuración, instalación y ejecución de las herramientas de utilidad de ES
Arquitectura de producción de ES	Comprensión avanzada del modelo de sharding y replica de ES	Uso básico de herramientas de gestión para el análisis del estado del cluster	Configuración básica de un cluster de ES

### Modelo de evaluación

Ejercicio	Peso
Instalación y configuración de Elasticsearch	20%

Creación de Índices y Mappings	20%
Subida de datos con BulkAPI	20%
Cálculos con valores geográficos	20%
Índices autocomplete	20%

## PROFESORADO

**Luis Polanco** –Arquitecto de Big Data y SCRUM Master en Accenture

Luis trabaja como arquitecto de Big Data para Accenture. En este rol, Luis colabora directamente con diferentes clientes a nivel global comprendiendo y dando respuestas a sus necesidades de negocio desde un punto de vista técnico y de implementación.

Enfocado principalmente en el uso de nuevas tecnologías, Luis lidera técnicamente el equipo de Rapid Development. Un grupo de trabajo Agile/SCRUM de desarrolladores profesionales enfocados al delivery de aplicaciones basadas en tecnologías de Big Data.

Durante el año pasado, Luis y su equipo han puesto en producción diferentes proyectos que utilizan ElasticSearch, Storm o Spark así como bases de datos MongoDB o Cassandra.

Actualmente, Luis está diseñando la plataforma de Big Data para Accenture Enterprise, una solución basada en Hortonworks distribuida sobre Amazon Cloud para el consumo masivo de eventos de negocio integrada en tiempo real con indexación sobre ElasticSearch.

Académicamente, Luis es Ingeniero Superior en Informática especializado en Ingeniería del Software, está certificado como Accenture Senior Technology Architect, como SCRUM Master por la Scrum Alliance y tiene el certificado de Lean IT.