

---

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

**Edición 2015**



---

## **Syllabus del Programa**

## Contenidos y Competencias

### Estructura General

El programa se divide en cuatro bloques que responden a las diferentes áreas competenciales que se plantean como objetivos de aprendizaje.

- El Bloque I incluye los conceptos de business case del Big Data y cómo la analítica de negocio encaja dentro de la transformación del negocio mediante los datos. También se incluye una actualización en las metodologías ágiles que son propias de este tipo de tecnologías aplicadas.
- El Bloque II incluye el desarrollo y estudio de casos para cada una de las áreas clave del negocio, y pretende contextualizar y aplicar las competencias y conocimientos del resto de los bloques en la práctica.
- El Bloque III desarrolla las competencia de la profesión del data scientist, utilizando entornos y técnicas de data science aplicados al negocio, desde la programación estadística hasta la aplicación del aprendizaje automático.
- El Bloque IV abarca las competencias tecnológicas necesarias para el planteamiento, diseño y desarrollo de proyectos de Big Data. Esto incluye la persistencia (tecnologías NoSQL), la paralelización de datos (Hadoop, Spark), así como la adquisición de datos (streaming, sourcing Web).

Bloque I	Habilidades y profesión	ECTS
	Módulos introductorios/business case	5
	Gestión de proyectos	2
Bloque II	Innovación y datos	ECTS
	Módulos específicos de áreas de negocio	6
	Trabajo Fin de Máster	7
Bloque III	Data science	ECTS
	Técnicas de análisis	12
	Herramientas de análisis	6
Bloque IV	Tecnologías de Big Data	ECTS
	Gestión de datos	12
	Paralelización de datos	6
	Adquisición de datos	4

El módulo de introducción pretende introducir el business case del Big Data para las organizaciones y cómo se puede extraer valor o ventaja competitiva del mismo. Para ello también se introducen los principales indicadores y factores medibles que son fundamentales en la transformación del negocio mediante el proceso de analítica.

Desde el punto de vista de la profesión, se introduce el concepto de data science, la tipología de los estudios data science y una primera introducción al toolset del data scientist como nuevo rol profesional.

Finalmente, se analizan los proyectos de Big Data desde la perspectiva ética y legal, teniendo en cuenta fundamentalmente la privacidad y los usos legítimos de los datos.

Módulo	Descriptoros	Principales herramientas
Big Data: conceptos, retos y oportunidades	Analítica e inteligencia de negocio, valor del Big Data	Casos prácticos
Data science: métodos y herramientas	Data science como profesión, entornos de data science, tipología de estudios, concepto de modelo predictivo	IPython, R
Medición y analítica para el negocio	KPIs, indicadores, estrategia	Casos prácticos
Aspectos legales, éticos y regulatorios	Datos personales, privacidad, condiciones de uso, usos éticos de la información	Casos prácticos

El módulo de gestión de proyectos busca capacitar para trabajar en un entorno ágil de analítica, comprendiendo los principios y técnicas fundamentales de las metodologías ágiles. Como complement, se desarrollan competencias para el uso de repositorios de artefactos necesarios tanto en el desarrollo como en el trabajo colaborativo del data scientist.

Módulo	Descriptoros	Principales herramientas
Métodos ágiles	Métodos ágiles, agile analytics, Lean Learning,	Scrum, MVI
Configuración y colaboración	Repositorios, entornos de colaboración	Git, IPython

El módulo de herramientas de análisis profundiza en herramientas de data science y en las tareas fundamentales de preparación y limpieza de datos. También desarrolla las competencias de programación estadística y con series de datos en estos entornos que son fundamentales para el paso a técnicas más avanzadas.

<b>Módulo</b>	<b>Descriptor</b>	<b>Principales herramientas</b>
Entornos de análisis de datos	Data wrangling, data cleaning	IPython, R
Programación y estadística para analistas de datos	Técnicas estadísticas, inferencia, regresión, modelos probabilísticos	IPython, R

El módulo de técnicas de análisis desarrolla el uso de las herramientas de análisis con el objetivo de cubrir todo el ciclo de data science hasta llegar al objetivo de los modelos predictivos que sirven a los propósitos del negocio. Se desarrolla el uso y evaluación de las técnicas computacionales fundamentales en aprendizaje automático, minería de texto y redes sociales. Se contextualiza todo el trabajo técnico en el ciclo de descubrimiento de conocimiento que incluye a la minería de datos.

<b>Módulo</b>	<b>Descriptor</b>	<b>Principales herramientas</b>
Modelos predictivos	Modelos predictivos, evaluación	IPython, R
Aprendizaje automático	Clasificación, clustering, reglas de asociación	IPython, R
Procesamiento del lenguaje y la minería de texto	Parsing, reconocimiento, opinión mining	NLTK
Análisis de redes sociales	Redes sociales, centralidad, prestigio, subredes, cores	Gephi
Minería de datos	KDD, data mining, visualización	Pentaho, D3

El módulo de bases de datos no convencionales introduce la tipología de NoSQL y los principales modelos, así como los factores principales de diseño de modelos de datos y consultas subyacentes a estas tecnologías. En un segundo módulo, se profundiza en la distribución horizontal y despliegue de estas tecnologías en clusters.

Se trata también el diseño de soluciones de streaming para datos en tiempo real y las técnicas de diseño de soluciones para ese contexto.

Finalmente, se introducen los motores de indexación y sus lenguajes de consulta, así como su uso como motores de persistencia.

Módulo	Descriptor	Principales herramientas
Bases de datos no convencionales	NoSQL, persistencia políglota, modelos basados en consultas	Cassandra, MongoDB, Riak, Neo4j
Procesamiento de streams	Streams, diseño de sinks	Storm, Kafka
Escalabilidad en almacenes de datos	Sharding, database clusters, CAP theorem	Cassandra, MongoDB, Riak, Neo4j, Hadoop
Motores de indexación	Indexación, ranking	Lucene, Solr, ElasticSearch

El módulo desarrolla competencias para la extracción de datos mediante APIs o scraping, así como el diseño de soluciones de crawling masivas.

También se desarrollan las competencias para tratar con datos RDF y triplestores en el marco de la Web of Data y sus estándares asociados, incluyendo el uso de terminologías en la Web of Data y el diseño de técnicas de inferencia.

Módulo	Descriptor	Principales herramientas
La Web de los Datos	Linked Data, RDF, ontologías, Schema.org, inferencia	SPARQL, Protege, Fuseki
Extractores de datos	Scraping, crawling, REST, OAuth, ETL	Scrapy, Nutch, Pentaho Kettle

Este módulo introduce los principales modelos de procesamiento de datos paralelos sobre Hadoop y Spark, para después profundizar en otros elementos asociados a la gestión de datos como catálogos sobre almacenes distribuidos o transformaciones de lenguajes de consulta relacionales, que enriquecen las tecnologías base.

Módulo	Descriptor	Principales herramientas
Modelos de paralelización de datos	MapReduce, Pangoor, HBASE, data clusters, RDD	Hadoop, Spark y relacionados
Gestión de datos paralelos	Ecosistemas Hadoop (HIVE, Pig, Sqoop) y Spark	Hadoop, Spark y relacionados

El Temario del Master está enunciado de forma indicativa. Por ende, CIFI se reserva el derecho de realizar modificaciones o ajustes, con el objetivo de mejorar la calidad del Programa de Formación.