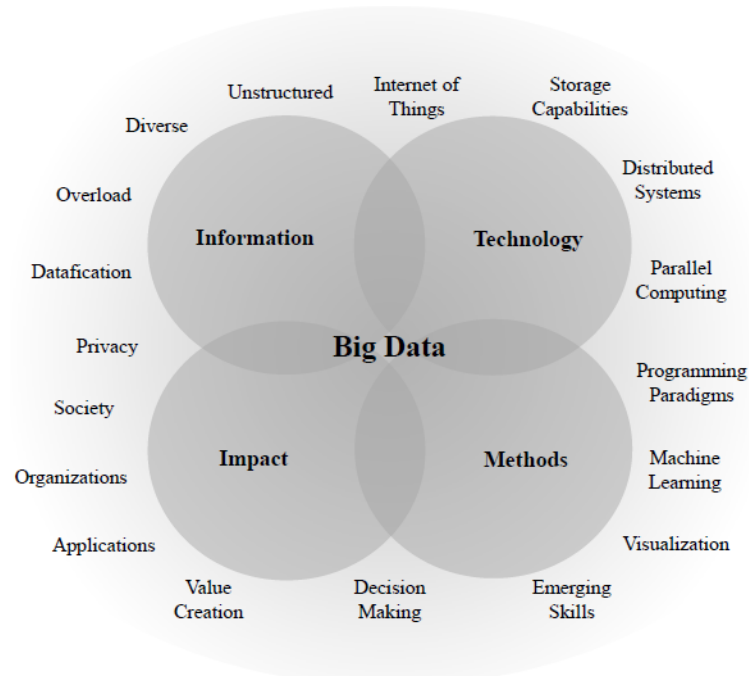


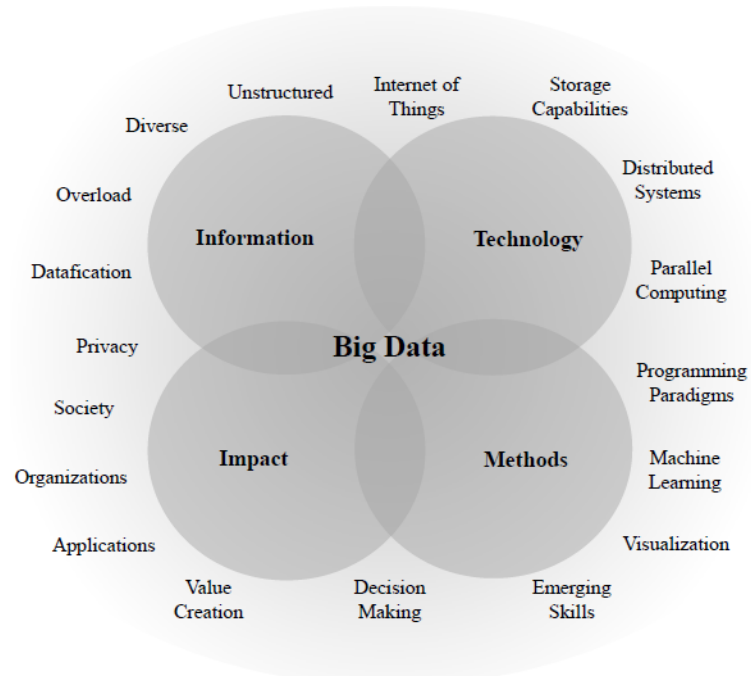
# Presentación asignatura Sistemas Almacenamiento Big Data

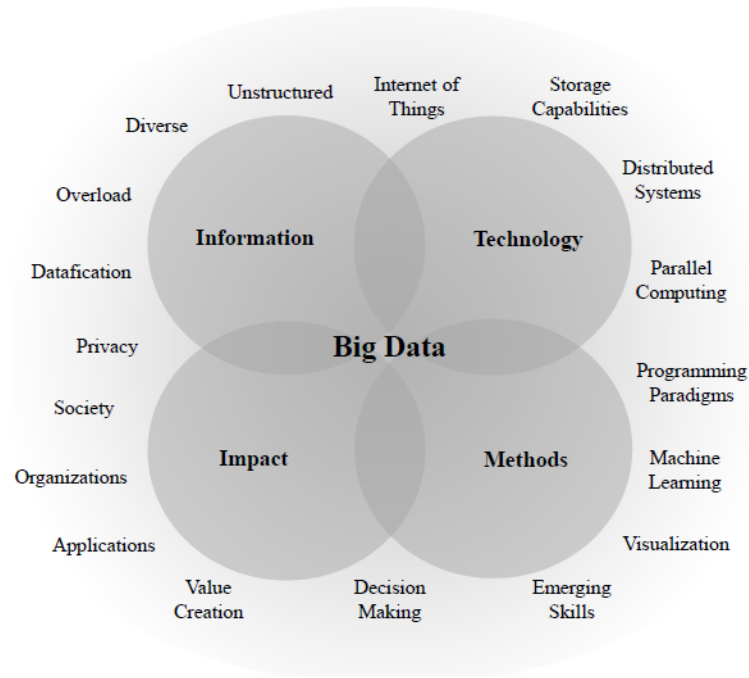
### Pablo Suárez-Otero

- Profesor de Ingeniería Informática
- Investigador especializado en la calidad de bases de datos NoSQL

Correo: [pablo.suarez.o@](mailto:pablo.suarez.o@professional.universidadviu.com)professional.universidadviu.com







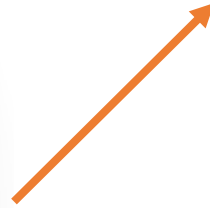
cassandra

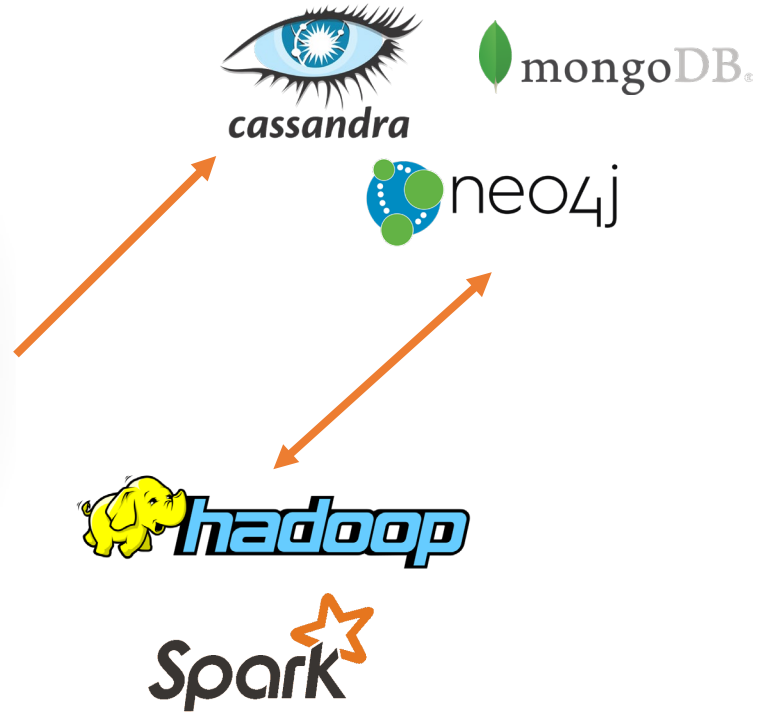
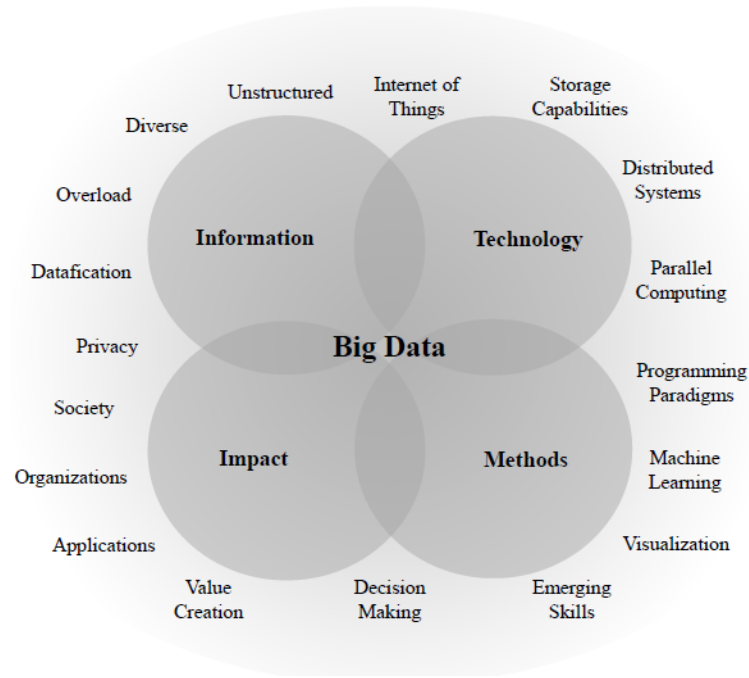


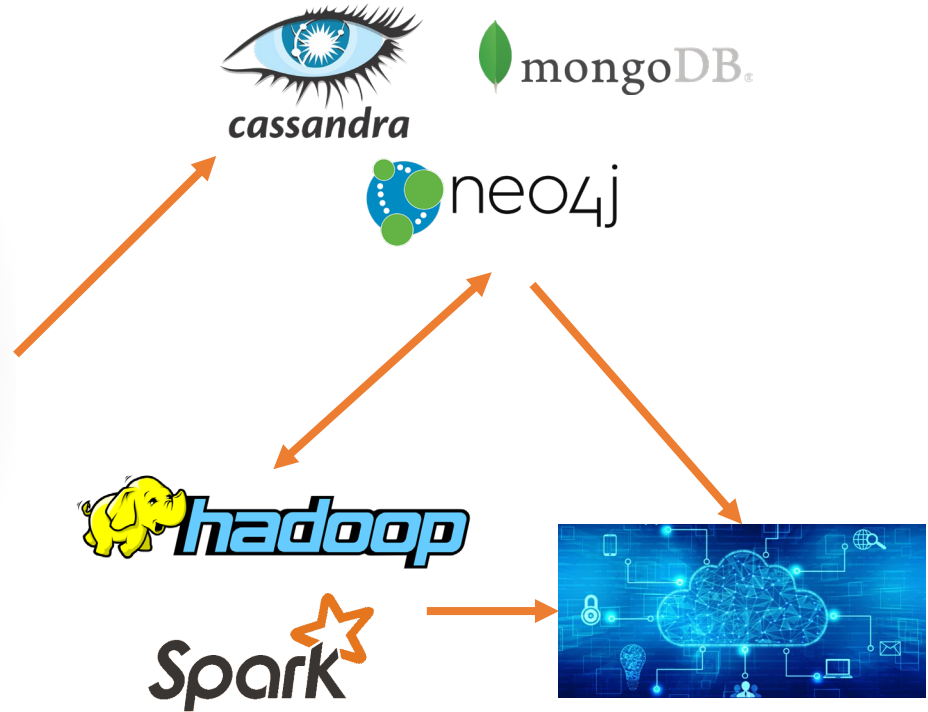
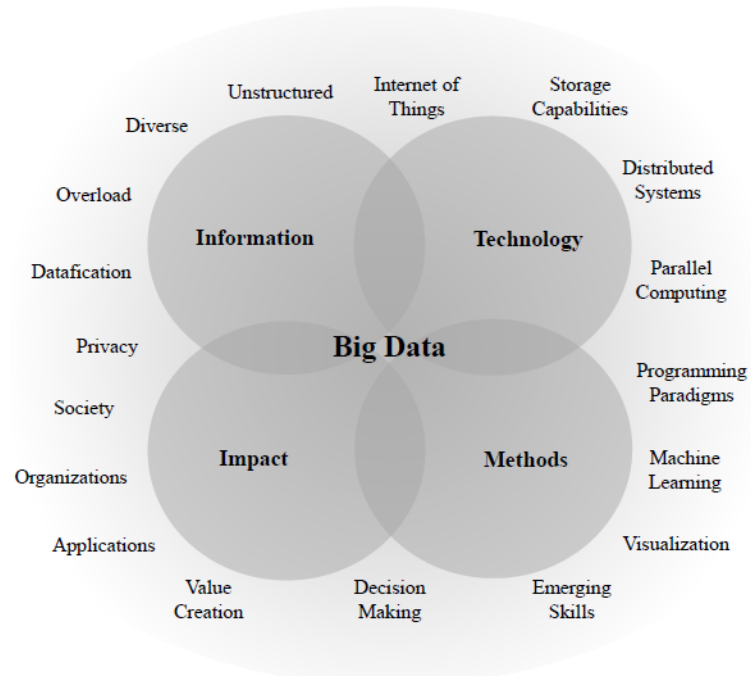
mongoDB®



neo4j







- Introducción a los conceptos de Big Data
- Beneficios manejo de Big Data
- Casos de estudio
- Almacenamiento en la nube
- Retos almacenamiento en la nube



- Características almacenamiento Big Data
- Problemas sistemas tradicionales
- Sistemas destinados a Big Data: NoSQL y NewSQL
- Propiedades ACID
- Teorema CAP
- Propiedades BASE
- BBDD orientadas a clave-valor
- BBDD orientadas a documentos
- BBDD orientadas a grafos
- BBDD orientadas a columnas

- Modelado conceptual
- Puesta en marcha de base de datos NewSQL
- Uso de base de datos NewSQL
- Datos vs necesidades negocio
- Modelado BBDD orientada a documentos.
- Modelado BBDD orientada a columnas
- Soluciones comerciales
- Técnicas de ingestión de datos
- Técnicas de almacenamiento de datos por tecnología
- Planes de consultas
- Integración con Hadoop

### Sesión 1: Modelado de Base de datos Cassandra

- Basándonos en un modelo conceptual y consultas aprenderemos como crear el esquema de una base de datos orientada a columnas.
- Usaremos Cassandra como la tecnología a usar en la asignatura.

### Sesión 2: Puesta en marcha de BBDD Cassandra y uso de Python

- Configuración y creación de tablas Cassandra
- Comunicación con programa Python
- Inserción, borrado y actualización de datos manual

### Sesión 3: Consulta e ingesta de datos

- Como procesar la información con Python de forma masiva
- Como insertar datos en batch
- Uso de dsbulk para migración de datos

En esta actividad se procederá a crear un esquema de base de datos Cassandra atendiendo a las necesidades del negocio. Para poder realizar la tarea se proveerá la siguiente información:

- Descripción del problema
- Modelo conceptual normalizado
- Descripción de las consultas a realizar

Se deberá subir un documento con la siguiente información:

- Tablas creadas para satisfacer a las consultas. Se debe incluir la información de las claves primarias, así como la relación de cada columna con su atributo en el modelo conceptual
- Justificación redactada de por qué se ha elegido dicho modelo, explicando las ventajas que tiene el diseño con respecto a otros diseños válidos para almacenar la información requerida.

### PARTE 2 (10%)

Realización del script para implementar el esquema obtenido durante el desarrollo de la parte 1. **La obtención de un esquema incorrecto en la parte 1 NO será motivo de penalización en la parte 2.**

### PARTE 3 (40%)

Se pide configurar un entorno de una base de datos Cassandra con los siguientes componentes:

- Base de datos instalada en sistema operativo Linux o Docker. Se recomienda usar la máquina virtual provista por la VIU, pero se pueden usar otros entornos que el alumno desee, incluido Windows con Docker.
- Programa Python que realice las consultas y operaciones que modifiquen los datos según las instrucciones de la actividad.

La base de datos a implementar tendrá el esquema desarrollado para la Actividad 1.

Usaremos obligatoriamente Cassandra y Python.

- Versión de Cassandra a usar: 4.1 o 5.0
- Versión de Python: 3.13.0 (última disponible a 15 de octubre de 2023)

Sistema operativo:

- Máquina de Amazon Workspaces (ver correos coordinador para solicitar acceso) o
- Ordenador personal: se requiere en este caso tener permisos de administrador y ser capaz de tener Docker en el caso de Mac y Windows. Se recomienda que todo el mundo pruebe esta opción salvo que ya se tenga acceso a la máquina de Amazon Workspaces

Software opcional pero altamente recomendable:

- Docker: Para usar Cassandra se recomienda encarecidamente usar Docker con el objetivo de instalar un contenedor con Cassandra (existen manuales en el campus). Existe la posibilidad de instalar Cassandra de forma nativa en Linux
- En Windows es necesario habilitar Hyper-V

Software opcional:

- Datastax Devcenter: IDE que se puede usar para creación de tablas Cassandra de forma interactiva. El problema reside en que requiere de una versión de Java 1.8.0\_151 o anterior. Si se usa una máquina Windows se pueden tener diferentes versiones instalados y configurar Devcenter para usar la 1.8.0\_151.
  - **SE PUEDEN REALIZAR LAS PRACTICES SIN IDE**, ya que solo se usará para la creación de las tablas lo cual se puede hacer de forma manual con los comandos de creación de tablas. El tiempo que se tardará en hacerlas sin IDE es probablemente inferior a hacerlas con IDE si hay problemas con la instalación.
- Python IDE: Utilizaré Pycharm en las clases pero cada alumno es libre de utilizar el IDE que desee o de no usar IDE si así lo considera. Lo único que se exige es que se suba un fichero .py en la parte 3 de la actividad

- Tras la impartición del Tema 2 se abrirán temas en el foro con casos de uso en los que se debata sobre qué base de datos habría que usar como mejor elección.
- En estas discusiones se valorará más la participación y las razones expuestas que la elección más óptima sin razonar. Si no se razona la respuesta dada se puede valorar dicha respuesta como 0 aun habiendo elegido la base de datos correcta.
- Se valorará positivamente la participación activa en el debate, teniendo que participar como mínimo en la mitad de las discusiones para optar a más del 50% de la calificación máxima de la actividad.



- Está disponible un test de autoevaluación tras el visionado de un vídeo en el que se introduce a la base de datos Neo4J:

<https://www.youtube.com/watch?v=NcNU7kcYXDM>

- Hay 45 minutos de tiempo límite una vez que se comienza el test.
- Se podrá realizar desde las 00h del 31 de octubre hasta las 23:59 del 1 de noviembre.

## Actividades y fechas de la asignatura

Actividad	Fecha
Presentación + Tema 1: Introducción al Big Data	Inicio: 22 de octubre
Tema 2: Sistemas de almacenamiento	Inicio: 24 de octubre
Actividades Foro Debate: Discusión mejor tipo BBDD	Inicio: 24 de octubre
Laboratorio 1: Esquema base de datos orientadas a columnas Cassandra	Inicio: 29 de octubre (18:00h)
Actividades Portfolio 1: Creación del esquema del proyecto	Entrega 18 de noviembre 23:59 CET
Tema 3: Modelado y técnicas para Big Data	Inicio: 31 de octubre
Laboratorio 2: Manipulación de datos de BBDD Cassandra a través de Python	Inicio: 5 de noviembre
Actividades Portfolio 2: Implementación del esquema, procesamiento de datos, consultas	Entrega 18 de noviembre 23:59 CET
Laboratorio 3: Ingesta de datos en BBDD Cassandra y Laboratorio 4: Migración de datos	7 de noviembre
Laboratorio 4: Migración de datos, Tema 1: Cloud y Tutoría grupal	8 de noviembre

## Pesos de la evaluación

Actividad	Peso
Actividades Foro	10%
Actividades Portfolio 1	40%
Evaluación continua: Auto-evaluación Vídeo	10%
Prueba final	40%

**Importante: Hay que aprobar el portfolio y el examen por separado. En caso de no tener una de las dos partes aprobadas, la nota máxima sería 4, Suspenso.**

**Test auto-evaluación Vídeo y Actividades de Foro solo disponible 1ª convocatoria.  
En 2ª convocatoria se guardará la calificación obtenida en la 1ª**

### Convocatoria ordinaria

TAREA	Fecha
Examen (Alternativa 1)	14 de noviembre a las 12h CET
Examen (Alternativa 2)	14 de noviembre a las 20h CET
Portfolio	18 de noviembre a las 23:59h CET

### Convocatoria extraordinaria

TAREA	Fecha
Examen (Alternativa 1)	25 de marzo de 2024 a las 12h CET
Examen (Alternativa 2)	25 de marzo de 2024 a las 20h CET
Portfolio	25 de marzo de 2024 antes de las 23:55h CET

- Existen dos alternativas de examen en cada convocatoria.
- Solo se puede realizar una de las dos alternativas.
- En el caso de que un alumno haga ambos exámenes, se suspenderá la convocatoria y no se corregirá ninguno de los dos.

Pablo Suárez-Otero

Correo: [pablo.suarez.o@](mailto:pablo.suarez.o@professional.universidadviu.com)professional.universidadviu.com

¿Preguntas?