



# ¡Les damos la bienvenida!

¿Comenzamos?

Certificados oficialmente por



**CODERHOUSE**

Esta clase va a ser

- grabada



# Clase en vivo

¿Comenzamos?

Certificados oficialmente por



**CODERHOUSE**

Semana 2. Data Engineering

# Introducción al Data Engineering

Certificados oficialmente por

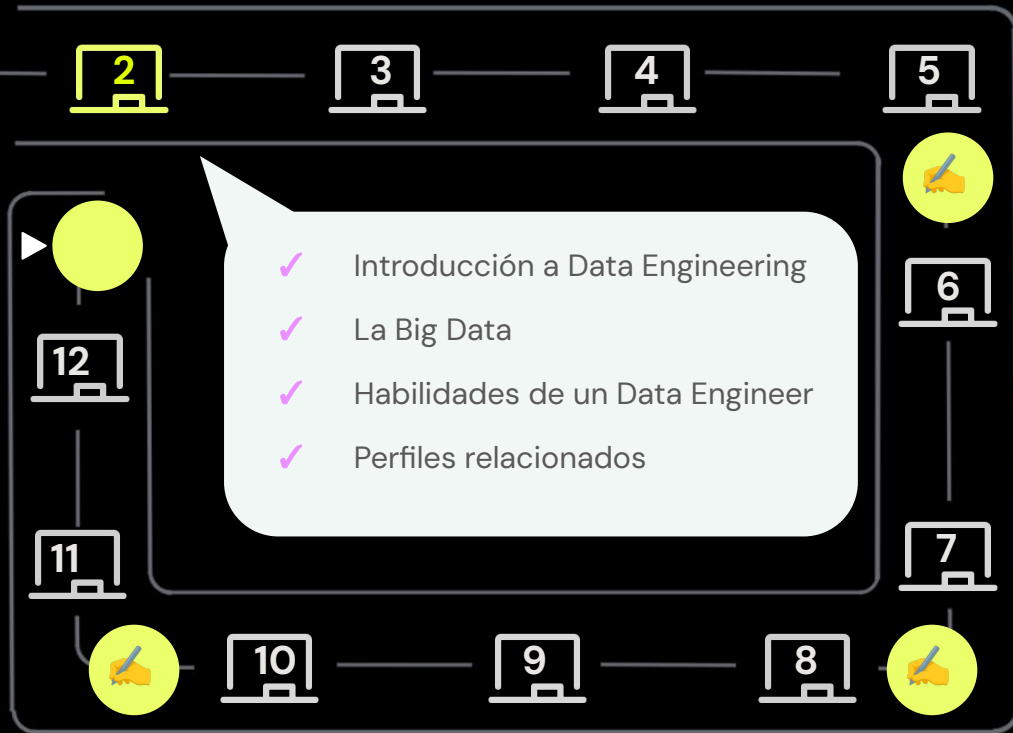


**CODERHOUSE**

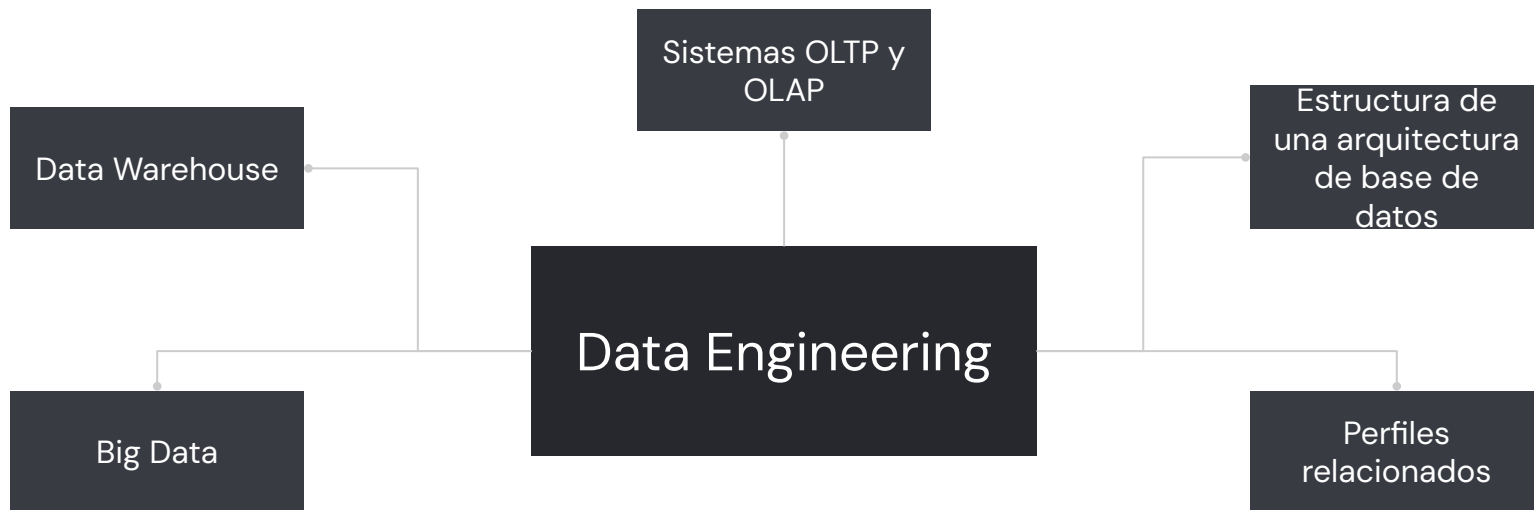
# Roadmap

Inicio:  
semana 1 –  
bienvenida  
y nivelación

Meta: entrega  
de Proyecto  
Final



## MAPA DE CONCEPTOS



# Objetivos de la sesión

- Comprender las diferentes versiones del Big Data.
- Comprender las 3 Vs del Big Data.
- Realizar una puesta en común del microdesafío S2.
- Poner en práctica los conceptos abordados en los contenidos on demand de la semana 2.



# Puesta en común microdesafío

¡Vamos a recuperar lo trabajado durante la semana!

Duración: **10 minutos.**





PUESTA EN COMÚN – MICRODESAFÍO

# Escape Room

Respecto a la experiencia que se tuvo con el microdesafío:

- ✓ ¿Las preguntas les parecieron fáciles o difíciles?
- ✓ ¿Qué pregunta supuso un mayor desafío?
- ✓ ¿Cómo lograron resolverlo?

[Acceso a la consigna completa](#) 🚀



# Big Data

Las 3 Vs



VIDEO N°2.1 – El origen del Big Data y su popularidad

# Recuperamos el tema de las 3Vs

Como se mencionó en el video el Big Data cumple con 3 requerimientos, conocidas como las 3V:

1. Volumen
2. Velocidad
3. Variedad

Las cuales veremos con mayor detalle ahora.



# Velocidad

La información llega continuamente. La gran velocidad con la que llega y se adquiere información permite entender cada vez más a los datos.

▶ Este tema se relaciona con el video 2.1 El origen del Big Data y su popularidad

Fuente de imagen: [Saffu](#)





# Volumen

La Big Data se compone por grandes cantidades de información almacenada en distintos repositorios. Para el 2020 estamos hablando de cerca de 40 zettabytes que representa un incremento de aproximadamente 300 veces respecto a 2005. Muchas compañías cuentan con Terabytes o incluso Petabytes de información en servidores.

▶ Este tema se relaciona con el video 2.1 El origen del Big Data y su popularidad

Fuente de imagen: [Krzysztof Kowalik](#)



**CODERHOUSE**

# Variedad

La Big Data contiene la información que utilizan compañías de todo tipo y puede tener múltiples fuentes y distintas formas como:

- ✓ Estructurada
- ✓ Semiestructurada
- ✓ No estructurada

▶ Este tema se relaciona con el video 2.1 El origen del Big Data y su popularidad

Fuente de imagen: [Ruslan Bardash](#)



**CODERHOUSE**

Las 5 Vs



# 5V's

Usualmente analizar Big Data implica hablar de 5V's en vez de 3V's añadiendo:

- ✓ **Veracidad:** lo cual se refiere a información de calidad y que realmente represente al fenómeno de estudio

- ✓ **Valor:** se refiere a la importancia que generan los datos para poder tomar decisiones, generar cambios y adaptarse al entorno

# Introducción a Data Engineering



# Para pensar

¿Qué es para ustedes un Data Engineer?  
¿Qué hace un Data Engineer?

Contesta mediante Kahoot

# El trabajo de un Data Engineer

# Trabajo de un Data Engineer

Los ingenieros de datos trabajan en una gran variedad de entornos de trabajo con el fin de crear sistemas que permitan la recopilación, administración y conversión de datos sin procesar en información utilizable para que la puedan interpretar y utilizar otros roles enfocados a Data y toma de decisiones dentro de la organización.

▶ Este tema se relaciona con el video 2.2 El origen del DE en Facebook y Airbnb.

Fuente de imagen: [ThisisEngineering](#)



**CODERHOUSE**

# Trabajo de un Data Engineer

El objetivo principal de los DE consiste en hacer que los datos sean accesibles para que las organizaciones puedan usarlos para evaluar y optimizar su desempeño.

▶ Este tema se relaciona con el video 2.2 El origen del DE en Facebook y Airbnb.

Fuente de imagen: [ThisisEngineering](#)



**CODERHOUSE**

# Skills de un Data Engineer

| Habilidad (Skill)                   | Descripción   |
|-------------------------------------|---|
| Conocimiento de sistemas            | Buen conocimiento de sistemas de bases de datos (SQL y NoSQL)                           |
| Manejo de lenguajes de programación | Manejo de lenguajes de programación como Python, R, Julia, Scala, C, Java, Golang, etc) |
| Comunicación                        | Habilidades comunicativas e interpersonales   |
| Lógica                              | Capacidad lógica para resolución de problemas   |
| Bases de Ciberseguridad             | Conocimiento en sistemas de seguridad para datos  |
| Bases matemáticas                   | Conocimientos básicos en matemáticas y estadística                                      |

# Perfiles relacionados con un Data Engineer





VIDEO N°2.3 – El DE con Analysts y Scientists

# Recuperamos el tema de Perfiles

En cualquier organización que utilice los datos para la toma de decisiones se requiere de lo que se conoce como Data Teams que se conforman de tres roles principales:

1. Data Analyst
2. Data Scientist
3. Data Engineer

Las cuales veremos con mayor detalle ahora.

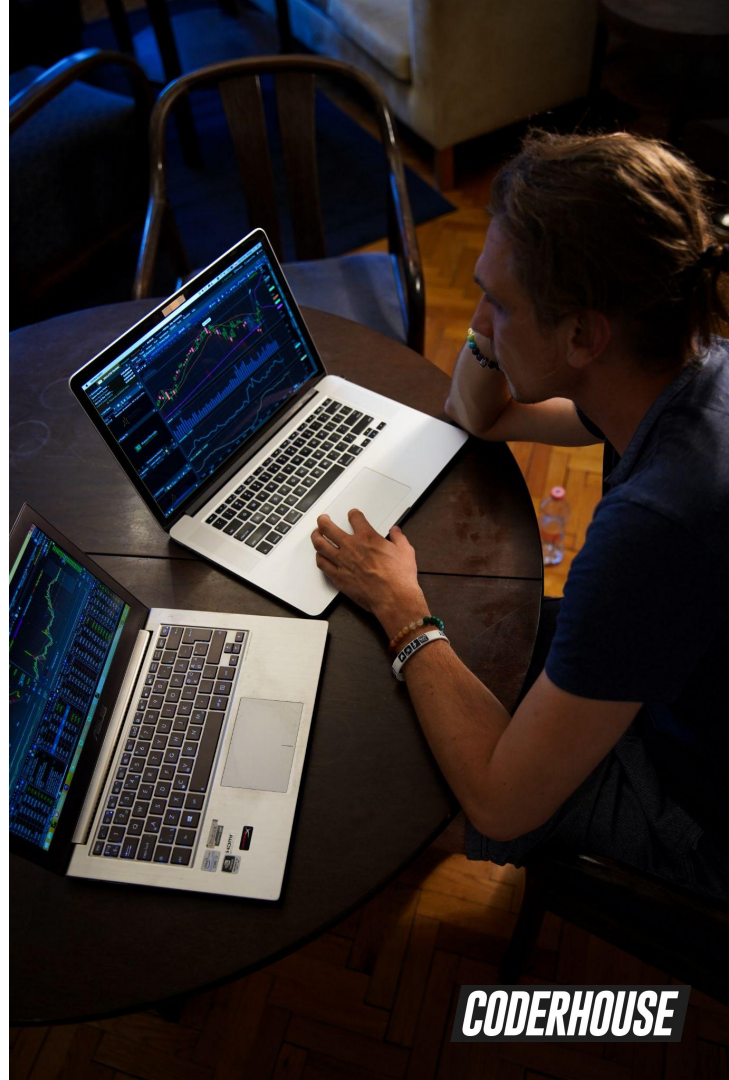


# Data Analyst

Utilizan datos para generar reportes y análisis asociados a datos. A diferencia de los Data Scientists y Data Engineers usualmente interactúan con datos sencillos de manipular que se han limpiado y transformado previamente. Sus análisis son de tipo descriptivo, diagnóstico, predictivo y prescriptivo. Pueden ser responsables de mantener dashboards, generar reportes y preparar visualizaciones de datos que ayudan a la toma de decisiones

▶ Este tema se relaciona con el video 2.3 El DE con Analysts y Scientists

Fuente de imagen: [Adam Nowakowski](#)



**CODERHOUSE**

# Data Scientist

Juegan un rol fundamental en el equipo de analítica. Estas personas necesitan de tener habilidades en matemática, estadística, programación y uso de herramientas (modelamiento estadístico, Machine Learning, Deep Learning e Inteligencia Artificial) para poder generar insights (conclusiones relevantes) sobre los datos.

▶ Este tema se relaciona con el video 2.3 El DE con Analysts y Scientists

Fuente de imagen: [Annie Spratt](#)



**CODERHOUSE**

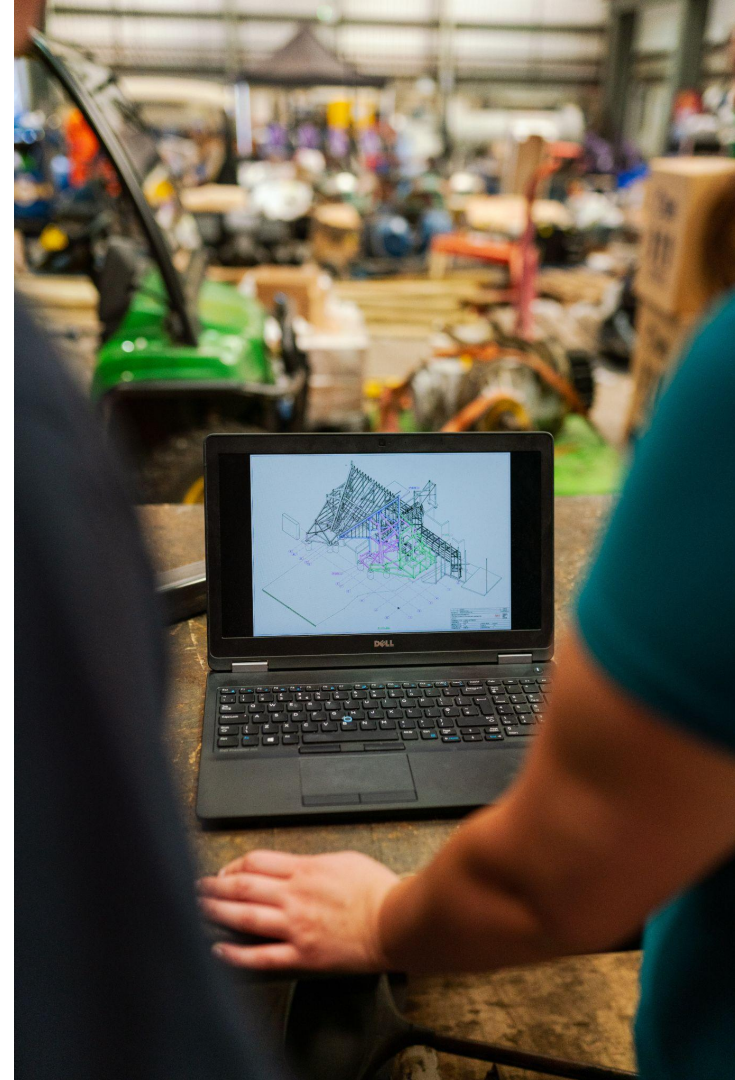


# Data Engineer

Tienen el papel de diseñar, crear y mantener datos que se necesiten usar para proyectos asociados a data. Ellos trabajan de la mano con los data scientists y los data analysts. La mayoría del trabajo asociado a los Data Engineers se relaciona con la preparación de la infraestructura, recopilación e integración de datos de diversas fuentes que permite el funcionamiento del equipo de data

▶ Este tema se relaciona con el video 2.3 El DE con Analysts y Scientists

Fuente de imagen: [ThisisEngineering RAEng](#)



# Otros perfiles importantes

Además existen otros roles que son importantes para orientar a los Data Teams como:

- ✓ **Data manager:** se encargan de desarrollar y orientar sistemas orientados a datos para satisfacer las necesidades de una organización

- ✓ **Director equipo data:** se encargan de diseñar y desarrollar sistemas usados para manejo de datos como pueden ser Data Warehouses, Data Stores, y otros sistemas de Business Intelligence. También se encargan de mantener y monitorear los sistemas de datos con el fin de asegurarse que cumplen con los requerimientos de la organización



# Break

¡En 10 minutos volvemos!

# Sistemas OLTP y OLAP



VIDEO N°2.4 – Sistemas OLTP vs OLAP

# Recuperamos el tema de OLTP Y OLAP

Como vimos en el video existen dos tipos de sistemas para el manejo del Big Data:

1. OLTP (Online Transaction Processing)
2. OLAP (Online Analytical Processing)

Los cuales veremos con mayor detalle ahora.







## Ejemplo en vivo

Analizaremos el funcionamiento de consultas asociadas a sistemas OLTP y OLAP tomando como base el funcionamiento de un call center. Tomaremos como referencia tres tablas [agentes](#), [calls](#) y [customers](#) junto con el siguiente [código](#)

Duración: **20 minutos**

# Paso a paso



**Momento 1** Creación de tablas calls, agents y customers para dar cumplimiento con el requerimiento que se necesita teniendo en cuenta la información que se provee



**Momento 2** OLTP: Almacenar transacciones que se dieron en un día particular en la empresa de Call Center. Esto traducido a sentencias SQL



**Momento 3** OLAP: extraer información para cada agente de: **duración de venta más rápida, a qué cliente fue y el nombre del agente.** Todo esto para poder otorgar bonos a los agentes.



**Momento 4** Interpretación de resultados en términos de tiempo en cada consulta

# Principales diferencias

| Item                    | OLTP                                  | OLAP  |
|-------------------------|---------------------------------------|---|
| <b>Data</b>             | Data en movimiento constante          | Data histórica principalmente                         |
| <b>Foco</b>             | Modificaciones en operaciones diarias | Extraer conclusiones de datos para toma de decisiones |
| <b>Fuentes de datos</b> | Transacciones                         | Distintas OLTPs                                       |
| <b>Transacciones</b>    | Cortas                                | Largas  |
| <b>Tiempo</b>           | Corto                                 | Largo   |
| <b>Querys</b>           | Simples                               | Complejos   |

# Principales diferencias

| Item          | OLTP                         | OLAP                                 |
|---------------|------------------------------|--------------------------------------|
| Normalización | Si                           | No                                   |
| Integridad    | Se debe mantener             | No se ve afectada                    |
| Métodos       | Bases de datos tradicionales | Data Warehouses                      |
| Usuarios      | Empleados, DBA, cajeros      | Analistas de negocio, CEOs, gerentes |
| Espacio       | Menos                        | Más                                  |
| Backup        | Regular                      | Rara vez                             |

# Estructura básica de una **arquitectura de base de datos**



Video N°2.7 - Arquitectura bases de datos

# Recuperamos el tema de Arquitectura

Como se aprecia en la infografía, la estructura de la arquitectura de base de datos tiene características como:

- ✓ 5 componentes principales.
- ✓ Se distinguen cuatro tipos de arquitecturas usualmente.
- ✓ Los modos de acceso a las bases de datos.



**CODERHOUSE**



# Cuestionario

**¿Te gustaría comprobar tus conocimientos de la clase?**

Compartimos a través del chat de Zoom el enlace a un breve **cuestionario**.

Duración: **10 minutos**



# Para pensar

¿Qué tanto creen que va a crecer el mundo de los datos?

**¿Verdadero o falso?**

Contesta mediante el chat de Zoom





## Para pensar

¿Llegará a haber un tope para la cantidad de información que almacenemos y “controlemos”? ¿Tendrá alguna consecuencia?

Contesta mediante el chat de Zoom

# ¿Preguntas?

Les invitamos a dejar sus  
preguntas en el chat



DATO CURIOSO

## ¿Sabías que...?

¡Para el 2025 se espera que la cantidad de Data generada sea de 463 exabytes! Un exabyte es una unidad de medida de almacenamiento de datos que equivale a  $10^{18}$  bytes.

[¡Lee más sobre esto!](#)



Lo que sabemos es una gota de agua, **lo que ignoramos es el océano.**

– Isaac Newton.

**¡Buen trabajo! 🧐**



**Te compartimos los  
siguientes  
recomendaciones**



## MATERIAL AMPLIADO

✓ [¿Cuáles son las 5 V's del Big Data?](#) |

CRISTINA JUAN

**Breve descripción:** Breve artículo expositivo de las 5 V's del Big Data

✓ [The 6 Functions of a Data Team](#) |

Sisense Team

**Breve descripción:** Artículo en inglés sobre los participantes en un equipo de Data.

✓ [Data Engineer job description](#) |

Resources for employers

**Breve descripción:** Breve artículo en inglés que describe la posición de un Data Engineer

✓ [Database Architecture Introduction](#) |

mongodb

**Breve descripción:** Breve artículo en inglés que describe la estructura de una arquitectura de base de datos



## MATERIAL AMPLIADO

✓ [OLTP](#) | IBM Cloud Education

**Breve descripción:** Artículo en inglés que habla sobre el sistema OLTP

✓ [OLAP](#) | IBM Cloud Education

**Breve descripción:** Artículo en inglés que habla sobre el sistema OLAP

✓ [OLAP vs. OLTP: What's the Difference?](#) |

Tanmay Sinha, Program Director, Db2  
Portfolio

**Breve descripción:** Artículo en inglés que describe las diferencias entre estos dos sistemas





## MATERIAL AMPLIADO

✓ [OLAP vs OLTP](#) | Explain By Example

**Breve descripción:** Video en inglés explicativo sobre las diferencias, ejemplo sencillo en escuela

[Amazon AWS](#) | IBM Cloud Education

**Breve descripción:** Fundamentos para desplegar Bases de datos en nube.

✓ [Sisence](#) |

### La evolución del Data Team

**Breve descripción:** Artículo que describe por qué los equipos de datos se están volviendo tan importantes para el éxito de las empresas.

# Resumen de la clase hoy

- ✓ Big Data como concepto
- ✓ Data Team
- ✓ Perfiles asociados a Data Team
- ✓ Roles y habilidades de un DE
- ✓ Arquitectura de bases de datos
- ✓ Sistemas OLTP y OLAP

# Manual de práctica

Recuerda que pueden desarrollar las ejercitaciones propuestas en el **manual de práctica**.

Consta de diferentes actividades resueltas en formato video que recibirás junto con toda la información necesaria para resolverlas por tu cuenta.

Para visualizarlo [ingresa aquí](#)

**Opina y valora**  
esta clase

**Muchas gracias.**

**#DemocratizandoLaEducación**



**¡Gracias por estudiar  
con nosotros!**