

# CS2032 - Cloud Computing (Ciclo 2024-2)

## Data Science

### Semana 5 - Taller 3: Ingesta de datos en S3

---

ELABORADO POR: GERALDO COLCHADO

# Contenido

Ingesta de datos en S3

1. **Objetivo del taller 3**
2. Concepto: Ingesta de datos
3. Ejercicio 1: Subir archivo a S3 con python
4. Ejercicio 2: Subir archivo a S3 con contenedor
5. Ejercicio 3: Ejercicio propuesto
6. Cierre

# Objetivo del taller 3:

## Ingesta de datos en S3

---

- Aprender a subir un archivo a S3 con python
- Aprender a subir un archivo a S3 con python desde un contenedor

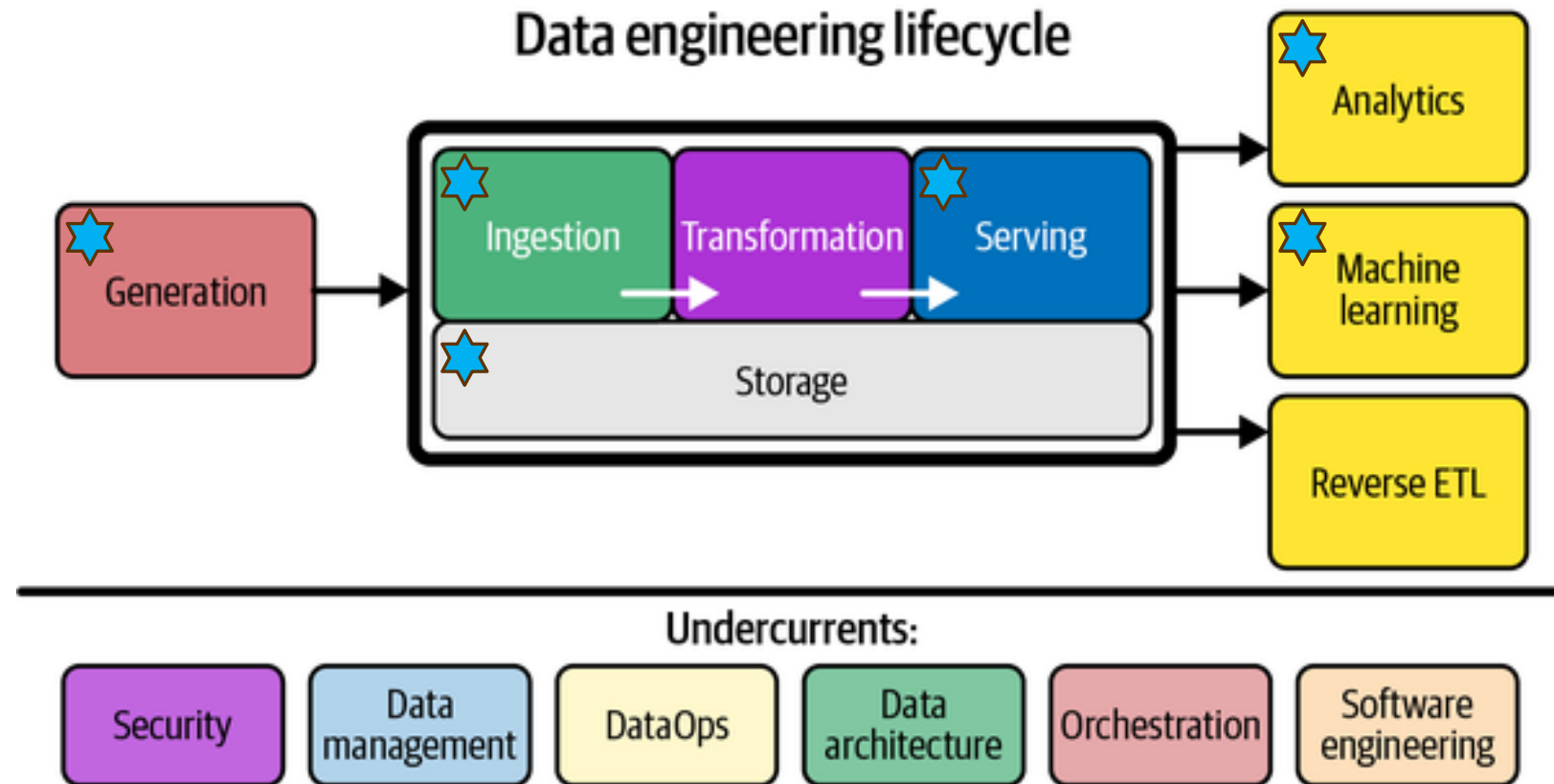
# Contenido

Ingesta de datos en S3

1. Objetivo del taller 3
2. **Concepto: Ingesta de datos**
3. Ejercicio 1: Subir archivo a S3 con python
4. Ejercicio 2: Subir archivo a S3 con contenedor
5. Ejercicio 3: Ejercicio propuesto
6. Cierre

# Concepto Ingesta de datos

*“La ingesta implica mover datos desde una fuente u origen al almacenamiento”*



# Concepto

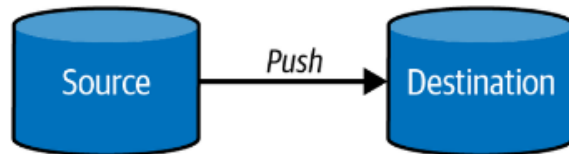
## Ingesta de datos - Estrategias

### Estrategias de ingesta

A la hora de obtener datos desde una fuente origen y llevarlos a un destino, podemos seguir tres planteamiento:

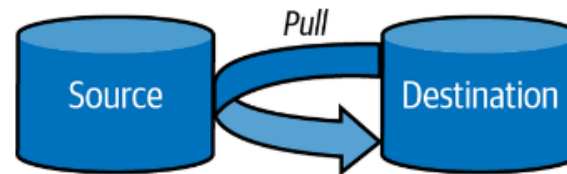
#### Push

El origen envía los datos al destino



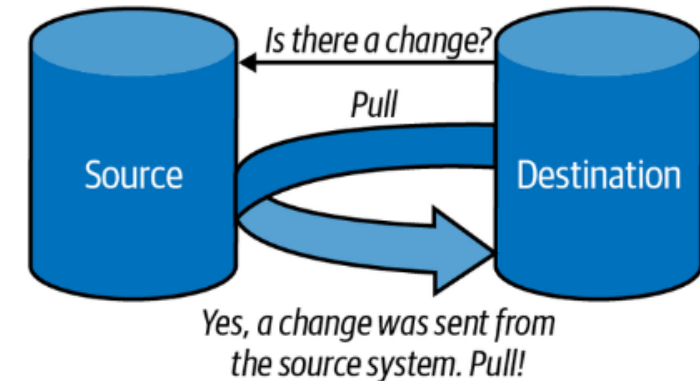
#### Pull

El destino recupera los datos del origen



#### Poll

El destino comprueba de forma periódica el origen para ver si ha habido cambios. Si hubiera cambios, hace un *pull* de los mismos.



# Contenido

Ingesta de datos en S3

1. Objetivo del taller 3
2. Concepto: Ingesta de datos
3. **Ejercicio 1: Subir archivo a S3 con python**
4. Ejercicio 2: Subir archivo a S3 con contenedor
5. Ejercicio 3: Ejercicio propuesto
6. Cierre

# Ejercicio 1:

## Subir archivo a S3 con python

- **Paso 1:** Cree un repositorio “ingesta01” en github y suba los archivos indicados por docente
- **Paso 2:** Ingrese a “MV Desarrollo” y en /home/ubuntu/ descargue repositorio con git clone
- **Paso 3:** Cree un nuevo bucket S3 para subir el archivo. Ejemplo: gcr-output-01
- **Paso 4:** En ingesta.py reemplace nombreBucket
- **Paso 5:** Configure el archivo /home/ubuntu/.aws/credentials
- **Paso 6:** Instale la librería boto3: pip install boto3
- **Paso 7:** Ejecute el programa ingesta.py
- **Paso 8:** Verifique en el bucket que se haya subido el archivo

```
~/ingesta01 (main) $ python3 ingesta.py  
Ingesta completada
```

Amazon S3 > Buckets > gcr-output-01

### gcr-output-01 Información


**Objetos** | Propiedades | Permisos

---

**Objetos (1) Información**

Los objetos son las entidades fundamentales que se explicita. [Más información](#)

🔍 *Buscar objetos por prefijo*

<input type="checkbox"/>	Nombre ▲	Tipo
<input type="checkbox"/>	 <u>data.csv</u>	csv



# Contenido

Ingesta de datos en S3

1. Objetivo del taller 3
2. Concepto: Ingesta de datos
3. Ejercicio 1: Subir archivo a S3 con python
4. **Ejercicio 2: Subir archivo a S3 con contenedor**
5. Ejercicio 3: Ejercicio propuesto
6. Cierre

# Ejercicio 2:

## Subir archivo a S3 con contenedor

- **Paso 1:** Elimine el archivo “data.csv” de bucket

- **Paso 2:** Analice Dockerfile

- **Paso 3:** Construya la imagen:  
docker build -t ingesta01 .

- **Paso 4:** Ejecute contenedor y verifique carga de archivo en bucket:

```
docker run -v /home/ubuntu/.aws/credentials:/root/.aws/credentials ingesta01
```

```
~/ingesta01 (main) $ docker run -v /home/ubuntu/.aws/credentials:/root/.aws/credentials ingesta01  
Ingesta completada
```

```
FROM python:3-slim  
WORKDIR /programas/ingesta  
RUN pip3 install boto3  
COPY . .  
CMD [ "python3", "./ingesta.py" ]
```

Amazon S3 > Buckets > gcr-output-01


### gcr-output-01 [Información](#)

Objetos	Propiedades	Permisos
---------	-------------	----------

**Objetos (1) [Información](#)**

Los objetos son las entidades fundamentales que se explicita. [Más información](#)

🔍 Buscar objetos por prefijo

<input type="checkbox"/>	Nombre ▲	Tipo
<input type="checkbox"/>	 <a href="#">data.csv</a>	csv

# Contenido

Ingesta de datos en S3

1. Objetivo del taller 3
2. Concepto: Ingesta de datos
3. Ejercicio 1: Subir archivo a S3 con python
4. Ejercicio 2: Subir archivo a S3 con contenedor
5. **Ejercicio 3: Ejercicio propuesto**
6. Cierre

# Ejercicio 3:

## Ejercicio propuesto - Ingesta de MySQL (Pull)

---

- Cree otro repositorio “ingesta02” a partir de “ingesta01”
- Modifique el programa “ingesta.py” para conectarse con una base de datos MySQL y leer todos los registros de una tabla y lo guarde en un archivo csv. Luego suba ese archivo csv a un bucket S3.
- Construya la imagen y ejecute el contenedor
- Muestre las evidencias en un pdf en el padlet indicado por el docente

# Contenido

Ingesta de datos en S3

1. Objetivo del taller 3
2. Concepto: Ingesta de datos
3. Ejercicio 1: Subir archivo a S3 con python
4. Ejercicio 2: Subir archivo a S3 con contenedor
5. Ejercicio 3: Ejercicio propuesto
6. **Cierre**

# Cierre:

## Ingesta de datos en S3 - Qué aprendimos?

---

- Aprender a subir un archivo a S3 con python
- Aprender a subir un archivo a S3 con python desde un contenedor

# Gracias

Elaborado por docente: Geraldo Colchado