Operaciones 4.0

- Introducción
- Proceso de negocio
- Arquitectura de datos (Solución)
- · Captura de datos
 - Servicio Pub-Sub
 - Diseño Dataflow
- Diccionario de datos
 - o Diccionario de datos material_measurements
 - o Diccionario de datos df batteries
 - o Diccionario de datos df material groups
 - Diccionario de datos df_materials
 - o Diccionario de datos df plants
 - Diccionario de datos df_ponds
 - o Diccionario de datos df_subsystems
 - o Diccionario de datos df warehouses
- Procedimientos almacenados y vistas
- Dashboard
 - o Proceso 1
 - o Proceso 2
- Incidentes
- Anexos

Introducción &

<< completar >>

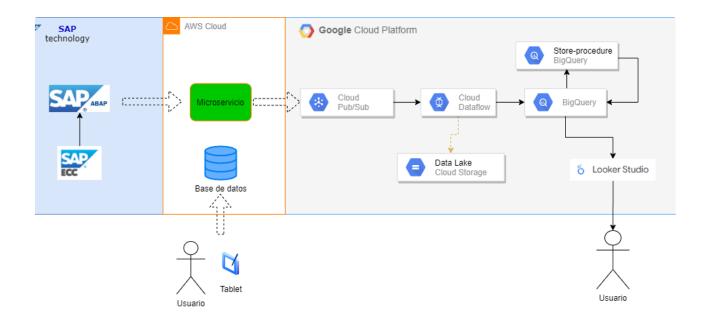
Proceso de negocio 🔗

<< completar >>

Arquitectura de datos (Solución) ∂

La información de este proyecto tiene principalmente dos fuentes de información. La primera relacionada con el formulario de ingreso de datos, el cual contiene información ingresada por el usuario final en la revisión de los tanques de la compañía, ingresando sus parámetros reales. La segunda fuente de información proviene desde datos del ERP de Gasco, en este caso SAP ECC.

Para SAP ECC, la información es extraída a través de un job(definir programa en SAP) el cual se ejecuta cada 3 a 5 minutos aprox. y envía la información por servicio a AWS, donde se almacena en una base de datos(definir que base de datos). Esta información es obtenida mediante Pub/Sub vía streaming. Los datos son transferidos a Bigquery donde se realiza una transformación de los datos para dejar la data disponible para la utilización en DataStudio.



3 Sugerencia de mejoras: Reemplazar el envió de mensajes desde microservicios delegando la función de extraer datos a otra herramienta ej: Kafka.

Captura de datos 🔗

Servicio Pub-Sub 🔗

Los datos enviados desde los microservicios son recibidos en los siguientes topic de Pub/Sub según correspondan. Se crea un tópico para cada ambiente(Desarrollo/Calidad/Productivo)

ID del tema	Clave de encriptación	Nombre del tema	Retención
operations-cylinder-counts	Google-managed	projects/analyticscicd/topics/operations-cylinder- counts	_
operations-cylinder-counts-dev	Google-managed	projects/analyticscicd/topics/operations-cylinder- counts-dev	7 días
operations-cylinder-counts-qa	Google-managed	projects/analyticscicd/topics/operations-cylinder- counts-qa	7 días
operations-cylinder-initial-stock	Google-managed	projects/analyticscicd/topics/operations-cylinder- initial-stock	_
operations-cylinder-initial-stock-dev	Google-managed	projects/analyticscicd/topics/operations-cylinder- initial-stock-dev	7 días
operations-cylinder-initial-stock-qa	Google-managed	projects/analyticscicd/topics/operations-cylinder- initial-stock-qa	7 días
operations-details-mseg	Google-managed	projects/analyticscicd/topics/operations-details- mseg	_
operations-details-mseg-dev	Google-managed	projects/analyticscicd/topics/operations-details- mseg-dev	7 días

operations-details-mseg-qa	Google-managed	projects/analyticscicd/topics/operations-details- mseg-qa	7 días
operations-material-measurements	Google-managed	projects/analyticscicd/topics/operations-material- measurements	31 días
operations-material-measurements- dev	Google-managed	projects/analyticscicd/topics/operations-material- measurements-dev	7 días
operations-material-measurements- qa	Google-managed	projects/analyticscicd/topics/operations-material- measurements-qa	7 días
operations-material-stocks	Google-managed	projects/analyticscicd/topics/operations-material- stocks	7 días
operations-material-stocks-dev	Google-managed	projects/analyticscicd/topics/operations-material- stocks-dev	_
operations-material-stocks-qa	Google-managed	projects/analyticscicd/topics/operations-material- stocks-qa	_

1 Se recomienda crear un topic por cada tabla de la base de datos donde se requiere streaming.

Los datos procesados y estructurados desde los topics se encuentran en las siguientes tablas de BigQuery:

```
1 #{project}:{dataset}.{tabla}
2
3 #Desarrollo
5 analyticscicd.operations_dev.df_material_measurements
6 analyticscicd.operations_dev.df_cylinder_counts
7 analyticscicd.operations_dev.df_details_mseg
8 analyticscicd.operations_dev.df_initial_stock_cylinders_count
9 analyticscicd.operations_dev.df_material_stocks
11 #Calidad
12
13 analyticscicd.operations_qa.df_material_measurements
14 analyticscicd.operations_qa.df_cylinder_counts
15 analyticscicd.operations_qa.df_details_mseg
16 analyticscicd.operations_qa.df_initial_stock_cylinders_count
17 analyticscicd.operations_qa.df_material_stocks
18
19 #Productivo
21 analyticscicd.operations.df_material_measurements
22 analyticscicd.operations.df_cylinder_counts
23 analyticscicd.operations.df_details_mseg
24 analyticscicd.operations.df_initial_stock_cylinders_count
25 analyticscicd.operations.df_material_stocks
```

Por otro lado, las siguientes tablas fueron cargadas de forma manual. Al existir alguna actualización estas debe ser ingresada de forma manual desde Bigquery.

```
1 #{project}:{dataset}.{tabla}
2
3 #Desarrollo
```

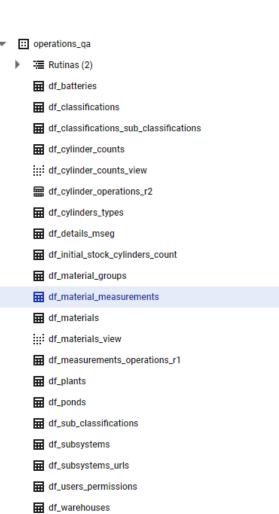
```
5 analyticscicd.operations_dev.df_batteries
 6 analyticscicd.operations_dev.df_classifications
7 analyticscicd.operations_dev.df_classifications_sub_classifications
8 analyticscicd.operations_dev.df_cylinders_types
9 analyticscicd.operations_dev.df_material_groups
10 analyticscicd.operations_dev.df_materials
11 analyticscicd.operations_dev.df_plants
12 analyticscicd.operations_dev.df_plants_location
13 analyticscicd.operations_dev.df_ponds
14 analyticscicd.operations_dev.df_sub_classifications
15 analyticscicd.operations_dev.df_subsystems
16 analyticscicd.operations_dev.df_subsystems_urls
17 analyticscicd.operations_dev.df_users_permissions
18 analyticscicd.operations_dev.df_warehouses
19 analyticscicd.operations_dev.df_zones
20
21 #Calidad
23 analyticscicd:operations_qa.df_batteries
24 analyticscicd.operations_qa.df_classifications
25 analyticscicd.operations_qa.df_classifications_sub_classifications
26 analyticscicd.operations_qa.df_cylinders_types
27 analyticscicd.operations_ga.df_material_groups
28 analyticscicd.operations_qa.df_materials
29 analyticscicd.operations_qa.df_plants
30 analyticscicd.operations_qa.df_plants_location
31 analyticscicd.operations_qa.df_ponds
32 analyticscicd.operations_qa.df_sub_classifications
33 analyticscicd.operations_qa.df_subsystems
34 analyticscicd.operations_qa.df_subsystems_urls
35 analyticscicd.operations_qa.df_users_permissions
36 analyticscicd.operations_qa.df_warehouses
37 analyticscicd.operations_qa.df_zones
38
39 #Productivo
40
41 analyticscicd:operations.df_batteries
42 analyticscicd.operations.df_classifications
43 analyticscicd.operations.df_classifications_sub_classifications
44 analyticscicd.operations.df_cylinders_types
45 analyticscicd.operations.df_material_groups
46 analyticscicd.operations.df_materials
47 analyticscicd.operations.df_plants
48 analyticscicd.operations.df_plants_location
49 analyticscicd.operations.df_ponds
50 analyticscicd.operations.df_sub_classifications
51 analyticscicd.operations.df_subsystems
52 analyticscicd.operations.df_subsystems_urls
53 analyticscicd.operations.df_users_permissions
54 analyticscicd.operations.df_warehouses
55 analyticscicd.operations.df_zones
```

Diseño Dataflow 🔗

Para procesar y estructurar los datos se diseñaron los siguiente pipeline:

Repositorio material measurements	gasco-operaciones-dataflow/src/material_measurements/
Repositorio cylinder_counts	gasco-operaciones-dataflow/src/cylinder_counts/
Repositorio material_stocks	gasco-operaciones-dataflow/src/material_stocks/
Repositorio initial stock	gascox/gasco-operaciones-2.0-dataflow
Repositorio details mseg	gascox/gasco-operaciones-2.0-dataflow

<< Actualizar documentación de tablas>>



df_zones

iiii view_cylinder_counts_det iiii view_material_measurements

Diccionario de datos 🔗

Diccionario de datos material_measurements ⊘

Field name	Туре	Mode	Collation	Policy Tags 🔞	Description
id	INTEGER	NULLABLE			
pond_id	INTEGER	NULLABLE			
material_group_id	INTEGER	NULLABLE			
pond_volume	FLOAT	NULLABLE			
pond_level	FLOAT	NULLABLE			
pond_pressure	FLOAT	NULLABLE			
pond_temperature	FLOAT	NULLABLE			
sample_temperature	FLOAT	NULLABLE			
sample_density	FLOAT	NULLABLE			
correct_density	FLOAT	NULLABLE			
total_weight	FLOAT	NULLABLE			
propane_percent	FLOAT	NULLABLE			
butane_percent	FLOAT	NULLABLE			
send_to_sgp	BOOLEAN	NULLABLE			
register_date	DATE	NULLABLE			
register_time	TIME	NULLABLE			
created_at	TIMESTAMP	NULLABLE			
updated_at	TIMESTAMP	NULLABLE			

Diccionario de datos df_batteries ⊘

Field name	Туре	Mode	Collation	Policy Tags ?	Description
id	INTEGER	NULLABLE			
description	STRING	NULLABLE			
plant_id	INTEGER	NULLABLE			

Diccionario de datos df_material_groups ⊘



Diccionario de datos df_materials ⊘



Diccionario de datos df_plants ⊘

Field name	Туре	Mode	Collation	Policy Tags ?	Description
id	INTEGER	NULLABLE			
description	STRING	NULLABLE			
sap_code	INTEGER	NULLABLE			
subsystem_id	INTEGER	NULLABLE			

Diccionario de datos df_ponds ⊘



Diccionario de datos df_subsystems ℰ



Diccionario de datos df_warehouses ℰ



Procedimientos almacenados y vistas ♂

Para la estructuración de los dashboard se crearon los siguientes procedimientos almacenados y Vistas los cuales se adjuntan a continuación.

Nombre	Tipo	Versión	Fecha	SqI
sp_measurements_oper ations_r1	Procedimiento almacenado			
sp_measurements_oper ations_r1	Procedimiento almacenado			
sp_cylinder_operations_ r2	Procedimiento almacenado			
sp_cylinder_operations_ r2	Procedimiento almacenado			

Dashboard *∂*

A continuación se documentan todos los dashboard desarrollados para el proyecto operaciones 4.0. Cada ambiente tiene su versión del dashboard.

Proceso 1 🔗

Ambiente	Nombre DashBoard	Link
DEV	DEV Ope01 - Panel Operaciones - Subsistemas	8 DEV Ope01 - Panel Operaciones - Subsistemas
QA	QA Ope01 - Panel Operaciones - Subsistemas	& QA Ope01 - Panel Operaciones - Subsistemas
PRD	Ope01 - Panel Operaciones - Subsistemas	ြို Ope01 - Panel Operaciones - Subsistemas

Detalle Proceso 1

Ambiente	Nombre Dashboard	Link
PRD	Ope01 - Panel Operaciones - Ss Maipú	ြ Ope01 - Panel Operaciones - Ss Maipú
PRD	Ope01 - Panel Operaciones - Ss Mejillones	6 Ope01 - Panel Operaciones - Ss Mejillones
PRD	Ope01 - Panel Operaciones - Ss Belloto	ျွဴ Ope01 - Panel Operaciones - Ss Belloto
PRD	Ope01 - Panel Operaciones - Ss Talca	გ Ope01 - Panel Operaciones - Ss Talca
PRD	Ope01 - Panel Operaciones - Ss Biobio	ြ Ope01 - Panel Operaciones - Ss Biobio
PRD	Ope01 - Panel Operaciones - Ss Osorno	ြို Ope01 - Panel Operaciones - Ss Osorno
PRD	Ope01 - Panel Operaciones - Ss Coyhaique	წ Ope01 - Panel Operaciones - Ss Coyhaique

Proceso 2 🔗

AMBIENTE	NOMBRE DASHBOARD	LINK DASHBOARD
DEV	DEV - Panel Operaciones 2 - Subsistemas	ြို့ DEV - Panel Operaciones 2 - Subsistemas
QA	QA - Panel Operaciones 2 - Subsistemas	8 QA - Panel Operaciones 2 - Subsistemas
PRD	Panel Operaciones 2 - Subsistemas	ြီ Panel Operaciones 2 - Subsistemas

Detalle Proceso 2

Ambiente	Nombre Dashboard	Link
PRD	Panel Operaciones 2 - Ss Maipu	8 Panel Operaciones 2 - Ss Maipu
PRD	Panel Operaciones 2 - Ss Mejillones	& Panel Operaciones 2 - Ss Mejillones
PRD	Panel Operaciones 2 - Ss Belloto	8 Panel Operaciones 2 - Ss Belloto
PRD	Panel Operaciones 2 - Ss Talca	% Panel Operaciones 2 - Ss Talca
PRD	Panel Operaciones 2 - Ss Biobio	8 Panel Operaciones 2 - Ss Biobio
PRD	Panel Operaciones 2 - Ss Osorno	& Panel Operaciones 2 - Ss Osorno
PRD	Panel Operaciones 2 - Ss Coyhaique	8 Panel Operaciones 2 - Ss Coyhaique

Para el acceso a los dashboard, existe una tabla mantenedora la cual define los permisos de acceso a los datos.

Las tablas en Bigquery asociadas a este proceso se listan a continuación:

- Desarrollo: analyticscicd.operations_dev.df_users_permissions
- Calidad: analyticscicd.operations_qa.df_users_permissions
- Productivo: analyticscicd.operations.df_users_permissions

Para actualizar la información de estas tablas, se utiliza el siguiente Script:



- Para este script se considera campo report:
 - report = 'Reporte 1'; corresponde al proceso 1 del proyecto operaciones 4.0
 - report = 'Reporte 2'; corresponde al proceso 2 del proyecto operaciones 4.0

Incidentes *⊘*

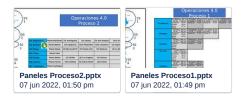
En esta sección se documentarán los incidentes que existan con este proyecto relacionado con la ingesta de datos.

1. Procesos Dataflow: Se ha determinado que los procesos en dataflow, los cuales están prendidos entre las 07 AM hasta las 08 PM, no logran cerrar de forma correcta, lo que implica mantener máquinas sobre-operando, aumentando el costo de utilización de los recursos en GCP.

Resolución: Se evidencia que la incidencia se produce solamente al iniciar por primera vez el dataflow, luego el proceso logra cerrar de forma correcta.

Anexos €

Paneles DataStudio



Presentación planta de trabajo

