

SOLUÇÃO PROPOSTA

Esse processo é do tipo Batch com uma carga diária em horário específico.
Serão levados arquivos do tipo CSV do ambiente on premise para a cloud.
Toda a tratativa será realizada com a stack da Google Cloud.

Um job será agendado na crontab para executar a transmissão de dados via script Python gerando logs dos eventos.
Os dados serão guardados em storage frio como estratégia de dataguard dos dados.
Os dados serão disponibilizados integralmente para a companhia, sem embutir regras de negócio.
Solução automática e escalável.

As camadas de ingestão no Data Lake serão:

- Stage (raw data);
- Work (joins);
- Target (data consumption);
- Metrics (métricas).

O pipeline será automatizado, com deploy em ambientes de DEV, QA e PRD.

O resultado será visualizado em ferramenta de BI, Data Studio.

O pipeline será monitorado e um e-mail será disparado acionando a área responsável em caso de falha.

Para não estender muito, o alerta não foi contemplado, apenas o Log e o relatório será exibido de maneira simples, sem técnicas de BI como Star Schema/SnowFlake ou gráficos.

Simple MER

BILL OF MATERIALS

tube_assembly_id

component_id_1

quantity_1

component_id_2

quantity_2

component_id_3

quantity_3

component_id_4

quantity_4

component_id_5

quantity_5

component_id_6

quantity_6

component_id_7

quantity_7

component_id_8

quantity_8

PRICE QUOTE

tube_assembly_id

supplier

quote_date

annual_usage

min_order_quantity

bracket_pricing

quantity

cost

COMP BOSS

component_id

component_type_id

type

connection_type_id

outside_shape

base_type

height_over_tube

bolt_pattern_long

bolt_pattern_wide

groove

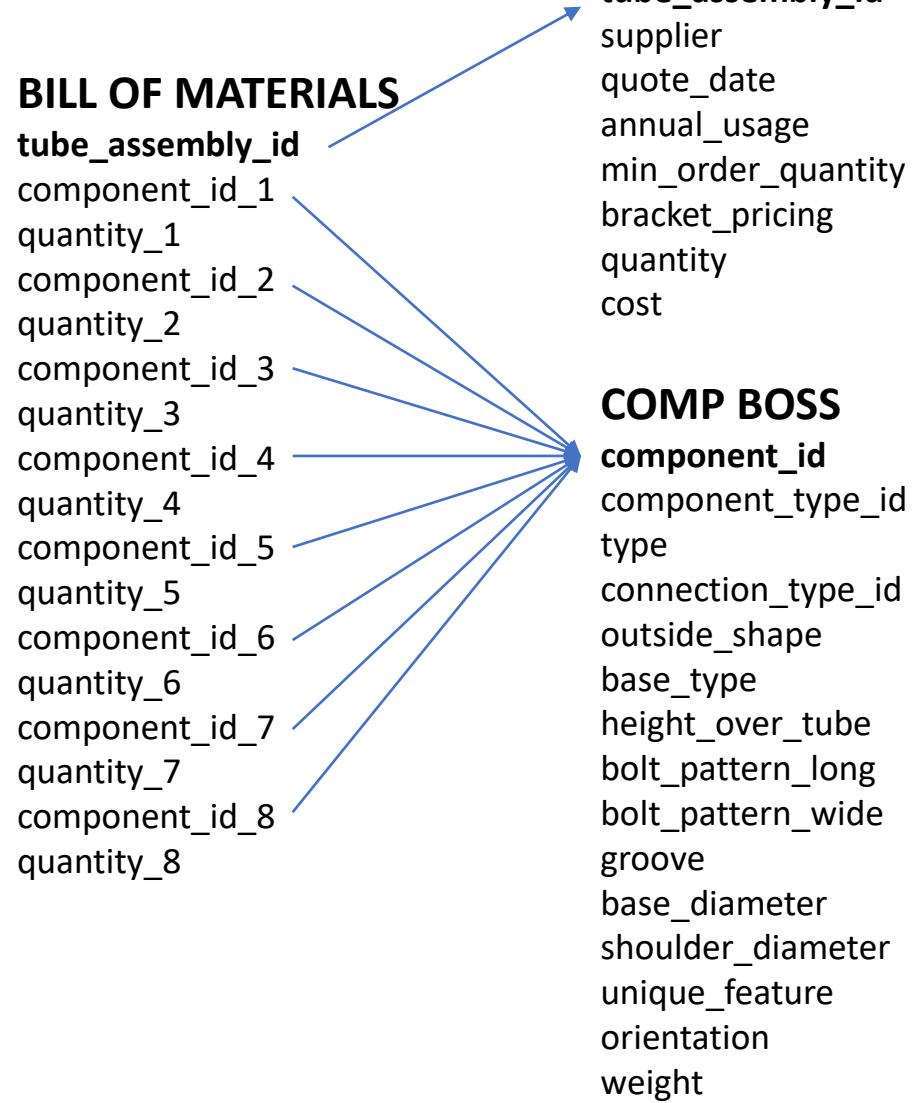
base_diameter

shoulder_diameter

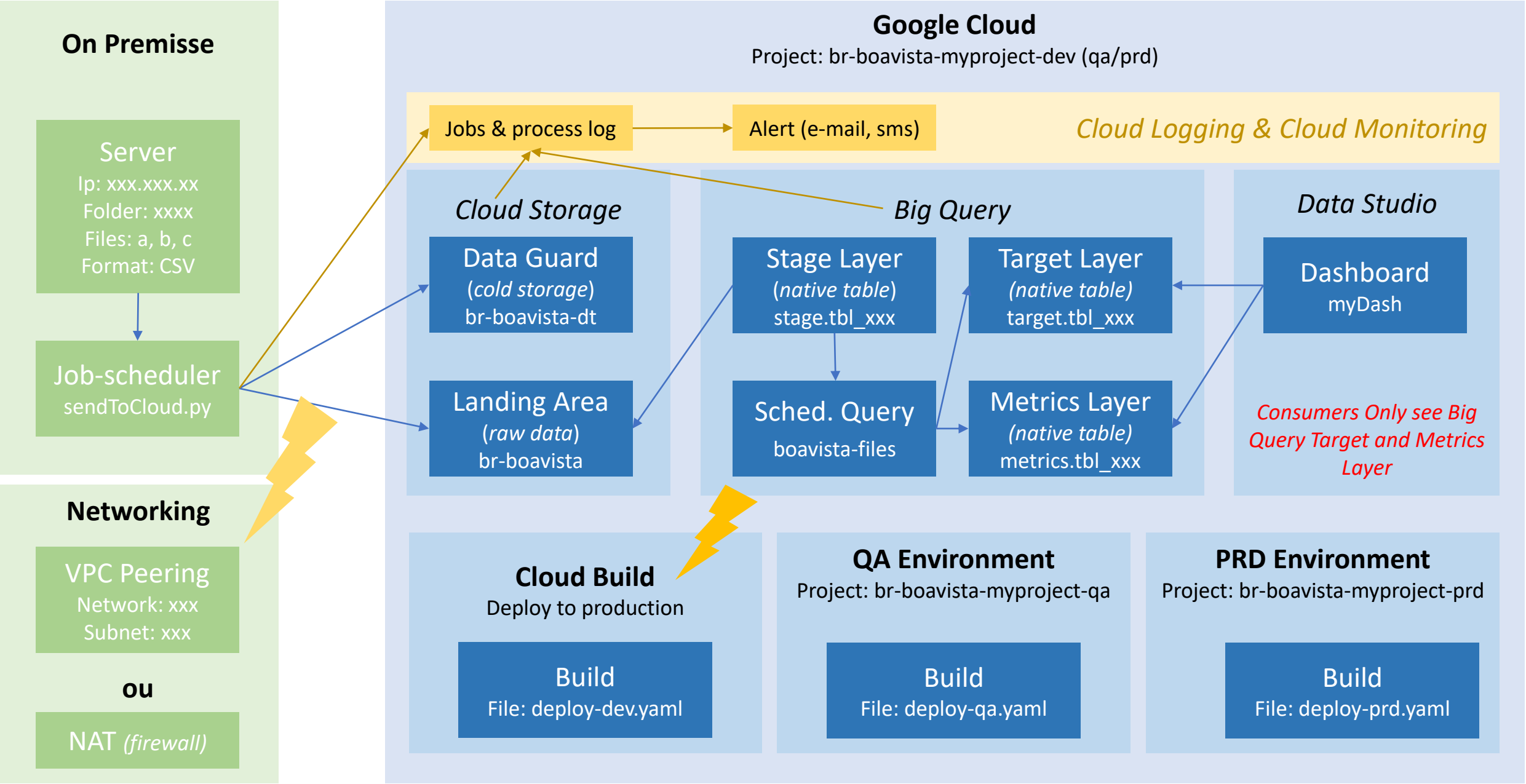
unique_feature

orientation

weight



Arquitetura Batch – from on premisses to cloud



Evidências – set up conta de serviço

Permissions for project "marcio-pessoal"

These permissions affect this project and all of its resources. [Learn more](#)










View By: MEMBERS ROLES ☐ Include Google-provided role grants ?

Filter

Filter table

?

|||

<input type="checkbox"/>	Type	Member ↑	Name	Role	Analyzed permissions (excess/total) ?	Inheritance
<input type="checkbox"/>		674382484593-compute@developer.gserviceaccount.com	Compute Engine default service account	Editor	 3432/3436 ▼	
<input type="checkbox"/>		674382484593@cloudbuild.gserviceaccount.com		Cloud Build Service Account	?	
<input type="checkbox"/>		674382484593@cloudservices.gserviceaccount.com	Google APIs Service Agent ?	Editor	?	
<input type="checkbox"/>		br-boavista-marcioteste@marcio-pessoal.iam.gserviceaccount.com	br-boavista-marcioteste	BigQuery Admin BigQuery Data Transfer Service Agent Storage Admin	? ? ?	

Evidências – logs do script Python

Google Cloud Platform

marcio-pessoal

Search products and resources

Operations Logging

Logs Explorer

Logs Dashboard

Logs-based Metrics

Logs Router

Logs Storage

Log fields

Search fields and values

RESOURCE TYPE

Global21

SEVERITY

Info16

Error3

Default2

Log Explorer

OPTIONS

REFINE SCOPE

Project

SHARE LINK

1/17/21, 11:48 PM - 1/18/21, 2:17 AM

PAGE

New features are available in the Logs Explorer.

Dismiss

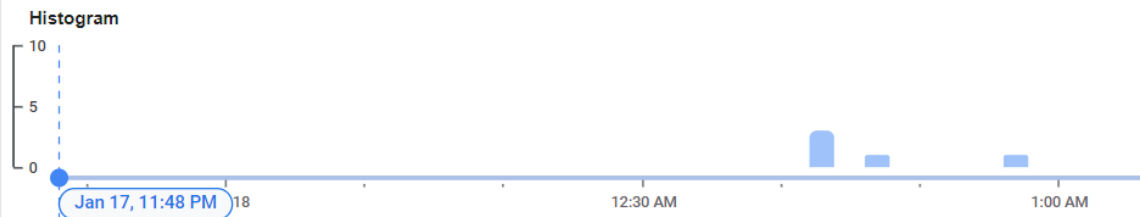
Query preview

logName="projects/marcio-pessoal/logs/br-boavista"

Save

Stream logs

Histogram



Query results

Jump to now

Actions

SEVERITY	TIMESTAMP	BRT	SUMMARY
> i	2021-01-18 01:11:41.121	BRT	File [comp_boss.csv] uploaded to Bucket [br-boavista].
> i	2021-01-18 01:11:43.747	BRT	File [price_quote.csv] uploaded to Bucket [br-boavista].
> i	2021-01-18 01:18:05.261	BRT	Bucket [br-boavista] created.
> i	2021-01-18 01:18:06.914	BRT	Bucket [br-boavista-dt] created.
> i	2021-01-18 01:18:09.553	BRT	File [data/bill_of_materials.csv] uploaded to Bucket [br-boavista-dt].
> i	2021-01-18 01:18:10.693	BRT	File [data/comp_boss.csv] uploaded to Bucket [br-boavista-dt].
> i	2021-01-18 01:18:12.642	BRT	File [data/price_quote.csv] uploaded to Bucket [br-boavista-dt].
> i	2021-01-18 01:18:14.308	BRT	File [data/bill_of_materials.csv] uploaded to Bucket [br-boavista].
> i	2021-01-18 01:18:15.383	BRT	File [data/comp_boss.csv] uploaded to Bucket [br-boavista].
> i	2021-01-18 01:18:17.514	BRT	File [data/price_quote.csv] uploaded to Bucket [br-boavista].

To view more results, expand the time range for this query.

Extend time by: 1 hour

Edit time

Evidências – Storage, Datalake layers, Bigquery datasets e tables

Google Cloud Platformmarcio-pessoal

Storage

Storage browser

CREATE BUCKETDELETEREFRESH

FilterFilter buckets

Bucket sorting and filtering are available in the Storage browser. Now you can filter your buckets.

Name	Created
br-boa-vista-dt	Jan 17, 2021, 5:4
br-boavista	Jan 17, 2021, 5:4
dataprep-staging-0f6c4c24-3417-4d54-9809-1c1fa5d1f0...	Aug 26, 2020, 11:

Google Cloud Platformmarcio-pessoal

BigQuery

Query history

Saved queries

Job history

Transfers

Scheduled queries

Reservations

BI Engine

Resources

Search for your tables and datasets

marcio-pessoal

br_boavista_metrics

br_boavista_stage

csv_bill_of_materials

csv_comp_boss

csv_price_quote

br_boavista_target

br_boavista_work

marcio-pessoal

br_boavista_metrics

br_boavista_stage

br_boavista_target

br_boavista_work

Query editor

1 select count(1) from br_boavista_stage.csv_price_quote;

RunSave querySave viewSchedule queryMore

Query results

SAVE RESULTSEXPLORE DATA

Query complete (0.1 sec elapsed, 0 B processed)

Job informationResultsJSONExecution details

Row	f0_
1	30214

Evidências – Scheduled queries

Google Cloud Platform

marcio-pessoal

Search products and resources

BigQuery


Scheduled queries

CREATE SCHEDULED QUERY

Filter transfer configs



	Display name	Source	Schedule (UTC)	Region	Destination dataset	Next scheduled <div></div>	Actions
<div></div>	Schedule Bill of Materials	Scheduled Query	every 24 hours	us		January 18, 2021 at 11:21:00 PM UTC-3	<div></div>
<div></div>	Schedule Comp Boss	Scheduled Query	every 24 hours	us		January 18, 2021 at 11:31:00 PM UTC-3	<div></div>
<div></div>	Schedule Price Quote	Scheduled Query	every 24 hours	us		January 18, 2021 at 11:46:00 PM UTC-3	<div></div>


Evidências – Data Studio, relatório em tabela







br-boavista



Arquivo Editar Exibir Inserir Página Organizar Recurso Ajuda


 Compartilhar


 Visualizar





 Adicionar uma página










 Adicionar dados

 Adicionar um gráfico

 Adicionar um controle



Tema e layout

	tube_assembly_id	Record Count
1.	TA-04788	14
2.	TA-04750	12
3.	TA-00199	10
4.	TA-01597	9
5.	TA-02227	9
6.	TA-08313	9
7.	TA-19094	9
8.	TA-04762	9
9.	TA-19025	9
10.	TA-16293	9
11.	TA-05812	9
12.	TA-18998	9
13.	TA-05467	9
14.	TA-02308	9
15.	TA-09575	8
16.	TA-19966	8
17.	TA-08606	8
18.	TA-05696	8

1 - 100 / 8855 < >

	tube_assembly_id	quantity_1
1.	TA-02684	6
2.	TA-03795	5
3.	TA-17058	5
4.	TA-04473	4
5.	TA-04678	4
6.	TA-11307	4
7.	TA-14192	4
8.	TA-11310	4

1 - 100 / 21198 < >

	component_id	Record Count
1.	C-0899	1
2.	C-0796	1
3.	C-0161	1
4.	C-0513	1
5.	C-0643	1
6.	C-0243	1
7.	C-1065	1
8.	C-0387	1

1 - 100 / 147 < >

Tema e layout

TEMA

LAYOUT

Tema atual

Default

Personalizar

Default

Edge

Default

Edge

Extrair tema da imagem

Evidências – Data Studio, acesso as fontes de dados do BQ

br-boavista

Arquivo Editar Exibir Inserir Página Organizar Recurso Ajuda


+ Adicionar uma página ↩ ➡ 🖱️ + Adicionar dados + Adicionar um gráfico ➦ + Adicionar um controle <> 🖼️ 📐 🔗 🔄 Tema e layout

tube_assembly_id	Record Count	tube_assembly_id	quantity_1
1. TA-04788	14	1. TA-02684	6

← Adicionar dados ao relatório

Credenciais dos dados: Proprietário ✕

 Use o BigQuery BI Engine para que seus relatórios do BigQuery sejam carregados com mais rapidez. [Ver mais](#)

 **BigQuery**
Por Google

O BigQuery é um serviço de armazenamento de dados de baixo custo, totalmente gerenciado e em escala de petabytes criado pelo Google para a análise de dados. A consulta e o processamento de dados realizados por esse serviço são pagos, e as cobranças são feitas no cartão de crédito registrado no projeto de faturamento.

[SAIBA MAIS](#) [INFORMAR UM PROBLEMA](#)

MEUS PROJETOS	Projeto	Conjunto de dados	Tabela
PROJETOS COMPARTILHADOS	marcio-pessoal	br_boavista_metrics	tbl_bill_of_materials
CONSULTA PERSONALIZADA	My First Project	br_boavista_stage	tbl_comp_boss
	My First Project	br_boavista_target	tbl_price_quote
CONJUNTOS DE DADOS PÚBLICOS	My First Project	br_boavista_work	
	My Project		

Cancelar

Adicionar