

Trabalho de Infraestrutura Google BigQuery

Instituto Infnet

Aluno: Charles de Araujo Melo

1. Escolha 5 conceitos fundamentais sobre o Google BigQuery e os descreva.

Análises:

Além de armazenar, o BigQuery também é utilizado para executar análises nos conjuntos de dados que são mantidos na nuvem.

Os insights gerados são obtidos com rapidez e em tempo real por meio de consultas SQL, que buscam em fontes de dados nativos, externos, em várias nuvens ou em conjuntos de dados públicos.

Ambiente gerenciável:

À medida que novos dados entram, o próprio sistema recebe e os organiza automaticamente, sem precisar de ajustes manuais como expandir capacidade de armazenamento ou reservar uma nova unidade.

Preserva os dados armazenados:

O BigQuery replica as informações contidas no banco em diferentes zonas, evitando a perda de dados por falha nas máquinas ou falhas zonais.

Criptografia:

É um recurso que faz a proteção dos dados enquanto eles são transferidos entre as zonas, evitando que hackers possam decodificar as informações contidas.

Eficiência na execução de consultas:

O sistema do BigQuery opera com uma velocidade que otimiza as cargas de trabalho de consulta, preservando a agilidade das respostas.

2. Descreva como é a arquitetura do Google BigQuery.

O BigQuery é um serviço escalável serverless (sem servidor), altamente disponível e com petabytes, que permite executar consultas SQL complexas rapidamente. Ele permite que você se concentre na análise, em vez de lidar com a infraestrutura. A ideia de hardware é completamente abstraída e não é visível para nós, nem mesmo como máquinas virtuais.

3. Apresente exemplos de utilização do Google BigQuery em bases NoSQL e SQL?

Com a transformação digital integrada à rotina das empresas, um imenso volume de dados é gerado diariamente, o qual precisa ser processado para gerar insights. Nesse cenário, o Google BigQuery é a solução que serve de apoio às corporações na estruturação dos seus bancos de dados.

4. Descreva os principais benefícios em utilizar o Google BigQuery.

Um dos principais benefícios é a análise em tempo real. Pelo fato de você estar usando toda a infraestrutura do Google, a parte de stream, de captura de dados do BigQuery funciona como uma performance excelente.

Isso permite que você possa, por exemplo, capturar dados em tempo real para garantir, por exemplo, resultados atualizados de forma muito mais rápida, de tal maneira que você possa construir consultas ou relatórios praticamente online, com informações online.

5. O que é um pipeline de dados?

O pipeline de dados permite que dados de fontes diferentes sejam integrados com eficiência. Por meio dele, é possível analisar dados relativos ao comportamento do cliente-alvo, que pode ser o consumidor final, automação de processos, jornadas do comprador e experiências do cliente.

6. Dê 2 (dois) exemplos de aplicações onde os pipelines de dados são utilizados em seu dia-a-dia.

Contabilidade mensal como transações em bancos e utilização do netflix.

7. Selecione uma base de dados pública brasileira para utilizar neste exercício. Você pode baixá-la em algum formato que desejar (ex.: formato .csv). Informe onde e como você conseguiu os seus dados. Explique se são estruturados ou não estruturados. Cada linha/registro em seu banco de dados corresponde a quais informações? Cada registro possui quantas colunas associadas e quais atributos elas representam? Qual o tamanho do banco de dados escolhido?

Selecionei uma base de dados chamada Vendas E-Commerce, essa base de dados é um projeto meu de uma loja virtual feita em Wordpress. Os dados são estruturados divididos em cinco Datasets, cada linha e registro corresponde a um dado especificado e estruturado ao negócio.

8. Formule pelo menos 2 perguntas sobre sua base de dados. O que você quer saber sobre os dados que escolheu?

1. Quais os estados a empresa possui clientes?
2. Quantos clientes por estados a empresa possui?

9. Formule uma hipótese sobre o que você acha que vai encontrar quando filtrar e analisar seus dados.

A empresa deve possuir clientes nos 26 estados do Brasil

10. Crie uma nova variável a partir de outras variáveis da base de dados que te auxilie na avaliação de sua hipótese.

DF_Charles

11. Importe a sua base de dados na infraestrutura Google BigQuery. Inclua em seu relatório a forma que você realizou a importação.

Google Cloud

Projeto Charles

SANDBOX Configure o faturamento

Explorer

+ ADICIONAR

Ampliar pesquisa para todos os projetos.

MOSTRAR APENAS RECURSOS COM ESTRELA

projeto-charles-381423

Conexões externas

Vendas

Criar tabela

Pressione **F11** para sair do modo tela cheia

Origem

Criar tabela de
Fazer upload

Selecionar arquivo *
Produto.csv

Formato do arquivo
CSV

Destino

Projeto *
projeto-charles-381423

Conjunto de dados *
Vendas

Tabela *
Produto

Letras Únicode, marcas, números, conectores, traços ou espaços são permitidos.

Tipo de tabela
Tabela nativa

Esquema

☒ Detectar automaticamente

O esquema será gerado automaticamente.

Configurações de particionamento e de cluster

Particionamento
Sem particionamento

CRIAR TABELA CANCELAR

Importei minha base dados contendo 5 datasets na infraestrutura Google BigQuery fazendo o upload

Google Cloud

Projeto Charles

Pesquise (/) recursos, documentos, produtos e muito mais

Pesquisa

SANDBOX Configure o faturamento para fazer upgrade para a experiência completa do BigQuery. Saiba mais

DISPENSAR ATIVAR

DISPENSAR FAZER UPGRADE

Explorador

+ ADICIONAR

Ampliar pesquisa para todos os projetos.

MOSTRAR APENAS RECURSOS COM ESTRELA

projeto-charles-381423

Conexões externas

Vendas

Categoria

Clientes

Items

Ordens

Produto

MOSTRAR MAIS

Vendas

Informações do conjunto de dados

ID do conjunto de dados
projeto-charles-381423.Vendas

Criado
29 de mar. de 2023, 19:21:50 UTC-3

Validade da tabela
padrão

Última modificação
29 de mar. de 2023, 19:21:50 UTC-3

Local dos dados
US

Descrição

Compilação padrão

Janela de viagem no tempo
7 dias

Não diferencia maiúsculas e minúsculas
false

Rótulos

Tags

HISTÓRICO PESSOAL HISTÓRICO DO PROJETO

ATUALIZAR

seu o de

Não precisei utilizar códigos no pré processamentos de dados por conta da infraestrutura possuir interface.

13. Inclua em seu relatório o código fonte necessário para definir e executar um pipeline que implemente, na ordem correta, todos os passos de pré-processamento que você escolheu para analisar sua base de dados.

3.

```
SELECT DISTINCT (
state) as Estados, Count(state) AS Quantidade_Estado
FROM `trabalho-charles.Vendas.Clientes`
group by state
```

14. Insira em seu relatório um esquema que represente o funcionamento de seu pipeline de dados.

Comence sua teste gratuito ou faça login em crédito. Não se preocupe, você não será cobrado se seus créditos acabarem. Saiba mais

Google Cloud

Projeto Charles

Pesquise (/) recursos, documentos, produtos e muito mais

Pesquisa

SANDBOX

Configure o faturamento para fazer upgrade para a experiência completa do BigQuery. Saiba mais

DISPENSAR

FAZER UPGRADE

Explorer

+ ADICIONAR

Clientes

Q Digite para pesquisar

Ampliar pesquisa para todos os projetos.

Mostrar apenas recursos com estrela

projeto-charles-381423

Conexões externas

Vendas

Categoria

Clientes

Items

Ordens

Produto

Mostrar mais

Clientes

ESQUEMA

DETALHES

Visualizar

LINHAAGEM

Linha	id	created_at	first_name	last_n
1	36	2018-01-26 03:29:13 UTC	Mariana	Góes
2	73	2018-02-03 07:37:38 UTC	Cristiano	Almei
3	84	2018-11-01 15:39:40 UTC	Carol	Bueno
4	95	2018-10-01 11:35:59 UTC	Mariana	Rosa
5	0	2017-11-01 14:45:41 UTC	Marta	Jesús
6	27	2018-11-08 11:12:47 UTC	Luan	Lima
7	87	2018-08-16 20:25:23 UTC	Beatriz	Abreu
8	92	2017-10-16 16:29:15 UTC	Beatriz	Bueno
9	4	2018-08-06 07:24:16 UTC	Romário	Teixe
10	21	2018-02-21 15:35:55 UTC	Diego	Freire
11	32	2018-07-22 11:59:27 UTC	Ana	Camp
12	58	2018-09-13 03:13:02 UTC	Rafael	Monti
13	77	2018-11-30 10:29:55 UTC	Ana	Rodri
14	89	2017-12-07 19:24:54 UTC	Daniela	Pires
15	54	2018-01-25 02:11:25 UTC	Diego	Teixe
16	40	2018-08-17 13:21:08 UTC	Gabriela	Nasci
17	61	2018-04-30 18:36:56 UTC	Francisco	Abreu
18	24	2018-07-25 09:34:12 UTC	Jair	Barro
19	48	2018-05-29 04:22:54 UTC	Clarissa	Albuq
20	63	2018-05-05 17:52:26 UTC	Calque	Batist

Resultados por página: 50 1 - 50 de 100

Histórico pessoalHistórico do projeto

*Consulta... a 3

EXECUTAR

Consulta concluída

1 SELECT * FROM projeto-charles-381423.Vendas.CLIENTES LIMIT 5

Pressione Alt+F1 para consultar as opções de acessibilidade

Salvar resultados

Resultados da...INFORMAÇÕES DO JOBRESULTADOSJSONDETALHES DA EXECUÇÃO

Linha	id	created_at	first_name	last_n
1	36	2018-01-26 03:29:13 UTC	Mariana	Góes
2	73	2018-02-03 07:37:38 UTC	Cristiano	Almei
3	84	2018-11-01 15:39:40 UTC	Carol	Bueno
4	95	2018-10-01 11:35:59 UTC	Mariana	Rosa
5	0	2017-11-01 14:45:41 UTC	Marta	Jesús

Atualizar

15. Exporte os seus dados processados em formato .csv e importe em um software de visualização. Se possível, você também pode integrar diretamente o Google BigQuery com uma ferramenta de visualização

Índice

Seção

(x)

+ Código + Texto

✓ [4] from google.colab import auth
auth.authenticate_user()
print('Authenticated')

Authenticated

✓ [5] #Montando o Drive
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

✓ [6] %load_ext google.colab.data_table

✓ [7] # Realizar queries diretamente do meu projeto
%%bigquery --project trabalho-charles
SELECT * FROM `trabalho-charles.Vendas.Cientes` LIMIT 5

Job ID 432446d9-4d71-47ba-8de5-a2aa95b57d5e successfully executed: 100%

Downloading: 100%

1 to 5 of 5 entries

Filter

?

index	id	created_at	first_name	last_name	email	cell_phone	country	state	street	number	additional
0	36	2018-01-26 03:29:13+00:00	Mariana	Góes	mariana@meu_email.com	9 7324-4293	Brasil		null	NaN	Apto 25
1	73	2018-02-03 07:37:38+00:00	Cristiano	Almeida	cristiano@usuario.com	9 2630-9907	Brasil		null	NaN	Conjunto 24
2	84	2018-11-01 15:39:40+00:00	Carol	Bueno	carol@meu_email.com	9 3760-2211	Brasil		null	NaN	Conjunto 26
3	95	2018-10-01 11:35:59+00:00	Mariana	Rosa	mariana@usuario.com	9 3139-2145	Brasil		null	NaN	
4	0	2017-11-01 14:45:41+00:00	Marta	Jesus		9 9102-7834	Brasil	Acre	null	NaN	Conjunto 16

Show 25 per page

✓ [11] %%bigquery DF_Charles --project trabalho-charles
SELECT * FROM `trabalho-charles.Vendas.Cientes` LIMIT 20

Job ID 5b42cf6a-dc22-4177-bcd7-43da58c2dca2 successfully executed: 100%

✓ 0s conclusão: 15:56

Arquivo importado no Colab

Índice

Seção

(x)

<>

+ Código + Texto

DF_Charles

1 to 20 of 20 entries

Filter

index	id	created_at	first_name	last_name	email	cell_phone	country	state	street	number	additional
0	36	2018-01-26 03:29:13+00:00	Mariana	Góes	mariana@meu_email.com	9 7324-4293	Brasil			NaN	Apto 25
1	73	2018-02-03 07:37:38+00:00	Cristiano	Almeida	cristiano@usuario.com	9 2630-9907	Brasil			NaN	Conjunto 24
2	84	2018-11-01 15:39:40+00:00	Carol	Bueno	carol@meu_email.com	9 3760-2211	Brasil			NaN	Conjunto 26
3	95	2018-10-01 11:35:59+00:00	Mariana	Rosa	mariana@usuario.com	9 3139-2145	Brasil			NaN	
4	0	2017-11-01 14:45:41+00:00	Marta	Jesus		9 9102-7834	Brasil	Acre		NaN	Conjunto 16
5	27	2018-11-08 11:12:47+00:00	Luan	Lima	luan@teste.com	9 2529-8197	Brasil	Acre	Praça 86 do Estado Acre	303.0	Apto 27
6	87	2018-08-16 20:25:23+00:00	Beatriz	Abreu	beatriz@usuario.com	9 2254-6798	Brasil	Acre	Rua 95 do Estado Acre	482.0	Conjunto 16
7	92	2017-10-16 16:29:15+00:00	Beatriz	Bueno	beatriz@teste.com	9 4435-2960	Brasil	Acre	Rua 71 do Estado Acre	978.0	Apto 24
8	4	2018-08-06 07:24:16+00:00	Romário	Teixeira		9 3093-6522	Brasil	Bahia	Praça 56 do Estado Bahia	191.0	Apto 12
9	21	2018-02-21 15:35:55+00:00	Diego	Freire	diego@teste.com	9 6686-4094	Brasil	Bahia	Praça 4 do Estado Bahia	906.0	Apto 20
10	32	2018-07-22 11:59:27+00:00	Ana	Campos	ana@teste.com	9 1932-9841	Brasil	Bahia	Avenida 52 do Estado Bahia	402.0	
11	58	2018-09-13 03:13:02+00:00	Rafael	Rodrigues	rafael@meu_email.com	9 6550-3557	Brasil	Bahia	Praça 94 do Estado Bahia	647.0	Apto 25
12	77	2018-11-30 10:29:55+00:00	Ana	Monteiro	ana@teste.com	9 3871-4132	Brasil	Bahia	Praça 86 do Estado Bahia	604.0	Apto 30
13	89	2017-12-07 19:24:54+00:00	Daniela	Pires	daniela@usuario.com	9 6666-2265	Brasil	Bahia	Rua 31 do Estado Bahia	441.0	
14	54	2018-01-25 02:11:25+00:00	Diego	Teixeira	diego@exemplo.com	9 5247-1737	Brasil	Pará	Rua 8 do Estado Pará	722.0	Conjunto 16
15	40	2018-06-17 13:21:00+00:00	Gabriela	Nascimento	gabriela@teste.com	9 5382-5772	Brasil	Amapá	Avenida 85 do Estado Amapá	476.0	
16	61	2018-04-30 18:36:56+00:00	Francisco	Abreu	francisco@exemplo.com	9 6754-1775	Brasil	Amapá	Rua 72 do Estado Amapá	50.0	Apto 12
17	24	2018-07-25 09:34:12+00:00	Jair	Barroso	jair@teste.com	9 6679-6120	Brasil	Ceará	Rua 86 do Estado Ceará	770.0	Apto 21
18	48	2018-05-29 04:22:54+00:00	Clarissa	Albuquerque	clarissa@exemplo.com	9 7403-4273	Brasil	Ceará		NaN	Apto 12
19	63	2018-05-05 17:52:26+00:00	Caíque	Batista	caique@teste.com	9 8818-5886	Brasil	Ceará	Rua 76 do Estado Ceará	239.0	Apto 19

Show 25 per page

DF_Charles.describe()

1 to 8 of 8 entries

Filter

index	id	number
count		20.0
mean		53.25
std		29.5026760606024
min		0.0
25%		30.75
50%		56.0
75%		78.75
max		95.0

0s conclusão: 15:56

Arquivo importado no Colab

16. Utilizando a ferramenta de visualização, crie gráficos (no mínimo dos gráficos, um de barras e um de dispersão) um suportem as suas conclusões com relação às hipóteses investigadas.

+ Código + Texto

3s

%bigquery --project trabalho-charles

SELECT DISTINCT

state

FROM `trabalho-charles.Vendas.Clientes` LIMIT 100

Job ID 9dcde498-b746-4d8e-bb3d-bec18e9263aa successfully executed: 100%

Downloading: 100%

1 to 25 of 26 entries

Filter

index	state
0	
1	Acre
2	Bahia
3	Pará
4	Amapá
5	Ceará
6	Goiás
7	Piauí
8	Alagoas
9	Paraná
10	Roraima
11	Sergipe
12	Amazonas
13	Maranhão
14	Rondônia
15	Pernambuco
16	São Paulo
17	Mato Grosso
18	Minas Gerais
19	Rio de Janeiro
20	Santa Catarina
21	Espírito Santo
22	Distrito Federal
23	Rio Grande do Sul
24	Mato Grosso do Sul

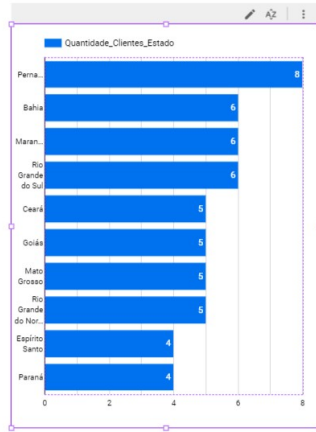
Show 25 per page

0s conclusão: 18:05

Clientes por Estados

Estados	Quantidade_Clientes_Estado
1. Pernambuco	8
2. Bahia	6
3. Maranhão	6
4. Rio Grande do Sul	6
5. Ceará	5
6. Goiás	5
7. Mato Grosso	5
8. Rio Grande do Norte	5
9. Espírito Santo	4
10. Paraná	4
11. São Paulo	4
12. Minas Gerais	4
13. Rio de Janeiro	4
14. Acre	4
15. Mato Grosso do Sul	4
16. Piauí	3
17. Alagoas	3
18. Santa Catarina	3
19. Distrito Federal	3
20. Amapá	2

1 - 26 / 26 < >



Data da última atualização: 05/04/2023 16:56:26

Gráfico

CONFIGURAÇÃO

Intervalo de escalas personalizado (automático)

☐ Escala de registro

Eixo Y

☐ Mostrar título do eixo

Grade

☐ Roboto

Tamanho da fonte d...

12px

Automático

Legenda

☐ Roboto

Automático

Alinhamento

Esquerdo

Máximo de linhas

5

Segundo plano e borda

☐

Dados

Pesquisa

Clientes - 05/04/2023, 16:55

188 Estados

123 Quantidade_Clientes_Estado

123 Record Count

Dados

Propriedades

Adicionar um campo
Adicionar um parâmetro
Adicionar dados

Status do período de teste gratuito: R\$ 2.355,98 de crédito e 90 dias restantes. Com uma conta completa, você tem acesso ilimitado a todos os recursos do Google Cloud Platform. DISPENSAR ATIVAR

Google Cloud Trabalho Charles Pesquise (/) recursos, documentos, produtos e muito mais Pesquisa

Explorer + ADICIONAR Clientes *Consulta... a 2

Ampliar pesquisa para todos os projetos. **MOstrar APENAS RECURSOS COM ESTRELA**

- trabalho-charles
 - Conexões externas
 - Vendas
 - Categoria
 - Clientes
 - Items
 - Ordens
 - Produto

Clientes CONSULTA COMPARTILHAR ATUALIZAR

ESQUEMA DETALHES **VISUALIZAR** LINHAGEM

Linha	id	country	state
1	1293	Brasil	null
2	1907	Brasil	null
3	1211	Brasil	null
4	1145	Brasil	null
5	1834	Brasil	Acre
6	1197	Brasil	Acre
7	1798	Brasil	Acre
8	1960	Brasil	Acre
9	1522	Brasil	Bahia
10	1094	Brasil	Bahia
11	1841	Brasil	Bahia
12	1557	Brasil	Bahia
13	1132	Brasil	Bahia
14	1265	Brasil	Bahia
15	1737	Brasil	Pará
16	1772	Brasil	Amapá
17	1775	Brasil	Amapá
18	1120	Brasil	Ceará
19	1273	Brasil	Ceará
20	1886	Brasil	Ceará
21	1149	Brasil	Ceará
22	1668	Brasil	Ceará

Resultados da... SALVAR RESULTADOS

INFORMAÇÕES DO JOB **RESULTADOS** JSON DETALHES DA EXECUÇÃO

Linha	state
1	null
2	Acre
3	Bahia
4	Pará
5	Amapá
6	Ceará
7	Goiás
8	Piauí
9	Amapá

Resultados por página: 50 1 - 50 de 100 ATUALIZAR

HISTÓRICO PESSOAL HISTÓRICO DO PROJETO

17. Por fim, escreva um texto em seu relatório fazendo uma análise final, tendo em vista os resultados obtidos. Responda às perguntas que fez no início do exercício e discuta se sua hipótese foi confirmada ou refutada.

A hipótese estava errada, nos 26 estados do Brasil a empresa possui clientes em 21 estados como pode ser observado nas consultas feitas.

Resultados da consulta

SALVAR RESULTADOS EXPLORAR DADOS

INFORMAÇÕES DO JOB	RESULTADOS	JSON	DETALHES DA EXECUÇÃO	GRÁFICO DE EXECUÇÃO	PRÉ-VISUALIZAÇÃO
Linha	Estados				Quantidade_Estado
1	null				0
2	Acre				4
3	Bahia				6
4	Pará				1
5	Amapá				2
6	Ceará				5
7	Goiás				5
8	Piauí				3
9	Alagoas				3
10	Paraná				4
11	Roraima				2
12	Sergipe				2
13	Amazonas				2
14	Maranhão				6
15	Rondônia				1
16	Pernambuco				8
17	São Paulo				4
18	Mato Grosso				5
19	Minas Gerais				4
20	Rio de Janeiro				4
21	Santa Catarina				3

Carregar mais