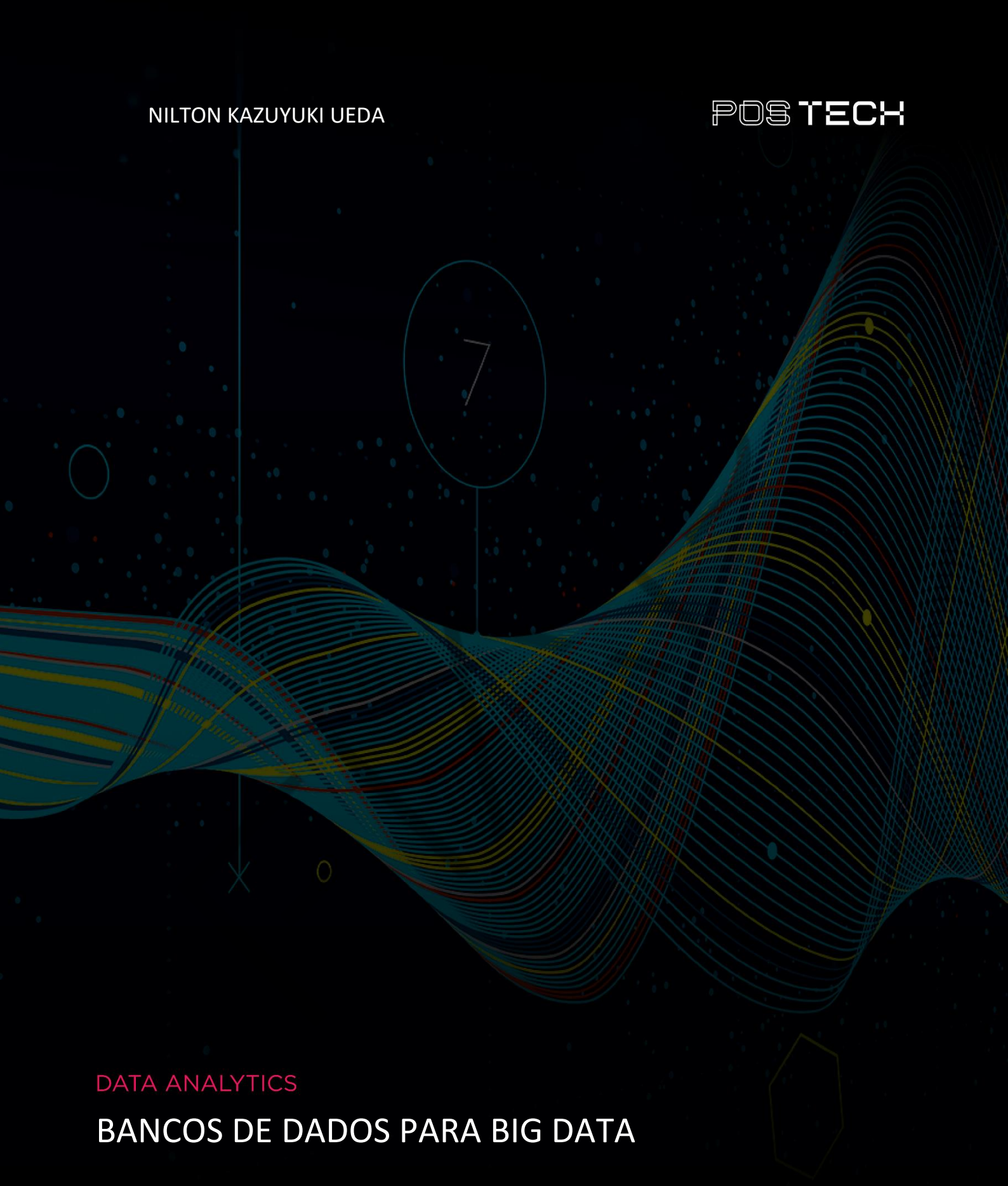
NILTON KAZUYUKI UEDA



DATA ANALYTICS

BANCOS DE DADOS PARA BIG DATA

AULA 02

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

SUMÁRIO

O QUE VEM POR AÍ? 3
HANDS ON 4
SAIBA MAIS 5
O QUE VOCÊ VIU NESTA AULA? 10
REFERÊNCIAS 11

FIAP

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

O QUE VEM POR AÍ?

Nesta aula, exploraremos como essa poderosa ferramenta pode simplificar e aprimorar a gestão de dados em escala empresarial. No cenário atual, a quantidade de dados gerados diariamente é colossal, tornando-se um desafio gerenciar e extrair insights valiosos dessa imensa quantidade de informações. É aí que a computação em nuvem e o Google BigQuery entram em cena. Com sua capacidade de processar, armazenar e analisar grandes volumes de dados de maneira eficiente e escalável, o BigQuery se tornou uma das soluções mais populares para a disponibilização de base de dados na nuvem.

Você aprenderá desde o básico, até as funcionalidades mais avançadas do Google BigQuery, permitindo que você explore todo o potencial dessa plataforma. Abordaremos temas como a criação de projetos e datasets, o carregamento de dados, a definição de esquemas e a execução de consultas SQL poderosas.

Além disso, exploraremos como otimizar consultas, compartilhar dados com segurança, integrar o BigQuery a outras ferramentas do ecossistema do Google Cloud Platform e aproveitar recursos avançados, como aprendizado de máquina e análise preditiva.

Não é necessário conhecimento prévio do Google BigQuery, apenas curiosidade e interesse em aprender sobre essa tecnologia revolucionária.

Prepare-se para mergulhar em um mundo de possibilidades com o Google BigQuery. Ao final deste curso, você terá adquirido conhecimentos sólidos para disponibilizar, explorar e extrair valor de suas bases de dados na nuvem, capacitando sua organização a tomar decisões embasadas em dados e impulsionar o crescimento e a inovação.

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

HANDS ON

Para encontrar os materiais utilizados no nosso Hands On, acesse o github da disciplina.

Foque em entender tudo o que os códigos transmitem, replique-os em sua máquina local, e teste com muita dedicação para que o aprendizado seja definitivo!

FIAP

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

SAIBA MAIS

ENTENDENDO MELHOR O FUNCIONAMENTO DO GOOGLE CLOUD PLATFORM

Agora que você já entendeu o poder que o BigQuery tem, vamos nos aprofundar no uso dos recursos completos da solução de nuvem do Google.

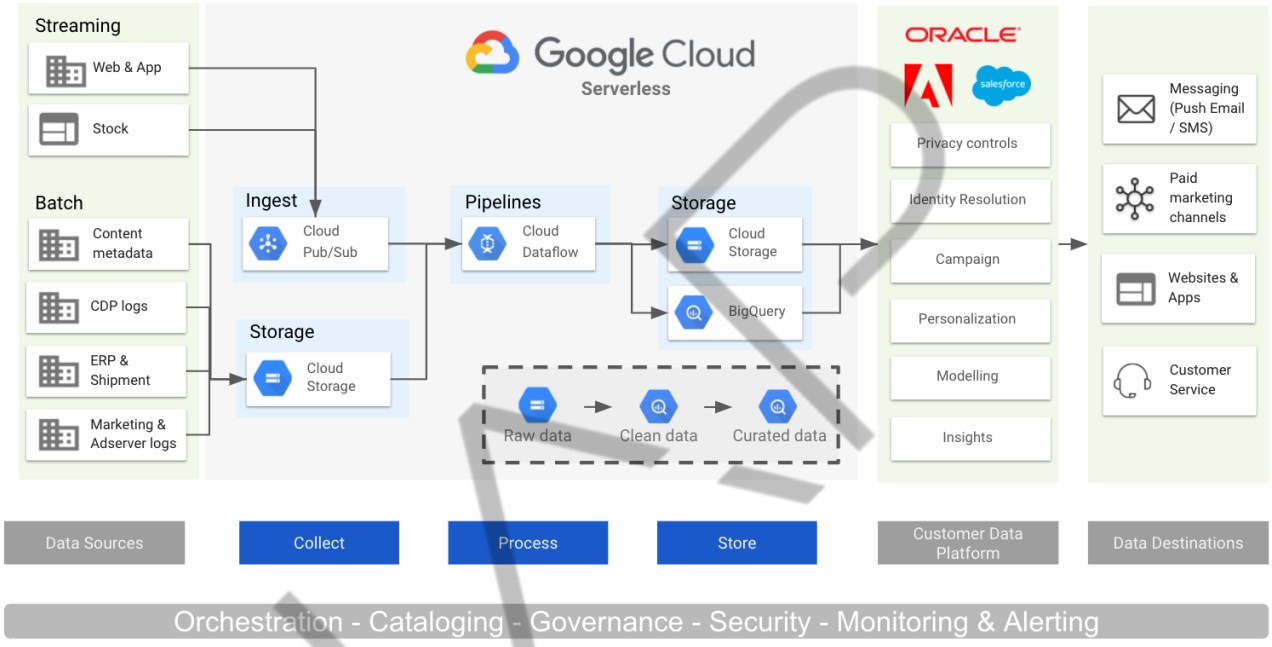


Figura 1 – Google Cloud Platform
Fonte: Hendrik Fleury (2022)

A figura 1 – “Google Cloud Platform” tem o conceito de Arquitetura de Dados Moderna focada principalmente na Google Cloud Platform. Podemos notar claramente o papel que o BigQuery desempenha neste tipo de solução; o motor de consulta de dados massivamente com alto poder de processamento.

Aqui está uma explicação do funcionamento completo do BigQuery integrado com essas soluções nativas em nuvem:

- Google Cloud Storage (GCS):
    - O GCS é um serviço de armazenamento de objetos altamente escalável no GCP.
    - Você pode carregar arquivos de dados diretamente no GCS, seja de fontes internas ou externas ao GCP.

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

    - O BigQuery pode ler dados armazenados no GCS como uma fonte de dados para processamento e análise.

- Carregamento de dados no BigQuery:
    - Após armazenar seus arquivos de dados no GCS, você pode carregá-los no BigQuery.
    - O BigQuery oferece várias opções de carregamento, incluindo carregamento de arquivos CSV, JSON, Avro, Parquet, entre outros.
    - Você pode carregar dados diretamente do GCS para o BigQuery usando comandos SQL ou APIs do BigQuery.
- Preparação de dados no BigQuery:
    - O BigQuery permite executar transformações nos dados carregados por meio de consultas SQL.
    - Você pode realizar operações de filtragem, agregação, junção de tabelas, criação de novas colunas e outros tipos de manipulação de dados.
    - Essas transformações podem ser aplicadas aos dados carregados ou a tabelas já existentes no BigQuery.
- Google Dataflow:
    - O Google Dataflow é um serviço de processamento de dados em lote e em streaming no GCP.
    - Você pode usar o Dataflow para executar transformações avançadas nos dados antes de carregá-los no BigQuery.
    - O Dataflow é integrado ao BigQuery, permitindo que você processe dados em escala e, em seguida, carregue-os diretamente no BigQuery para análise.
- Análise de dados no BigQuery:
    - Após o carregamento e a preparação dos dados, você pode executar consultas SQL avançadas no BigQuery para analisar os dados armazenados.

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

- O BigQuery é projetado para oferecer consultas rápidas e escaláveis, permitindo analisar grandes conjuntos de dados em questão de segundos ou minutos.
- Você pode usar recursos como janelas temporais, funções analíticas, agregações, entre outros recursos avançados do SQL, para extrair insights dos seus dados.

- Looker:
  - O Looker é uma plataforma de business intelligence (BI) e visualização de dados.
  - Ele pode ser integrado ao BigQuery para criar painéis interativos, relatórios e visualizações dos dados armazenados no BigQuery.
  - O Looker oferece recursos avançados de criação de dashboards, geração de relatórios e compartilhamento de insights com a equipe.

Essa é uma visão geral do funcionamento completo do BigQuery integrado com as soluções nativas em nuvem oferecidas pelo Google, como o GCS, o Dataflow, o Looker e outros. Essas integrações permitem que você carregue, prepare, analise e visualize dados de maneira eficiente e escalável na nuvem, facilitando a obtenção de insights valiosos a partir dos seus dados

REALIZANDO EXPLORAÇÕES E ANÁLISE DE DADOS AVANÇADA COM SQL NO GOOGLE BIGQUERY

O Google BigQuery é um serviço de armazenamento e análise de dados na nuvem que suporta consultas SQL avançadas. Ele permite executar consultas sofisticadas para extrair informações úteis dos dados armazenados no BigQuery. Aqui estão alguns exemplos de consultas SQL avançadas no BigQuery:

1. Consultas com JOIN

Você pode usar junções para combinar dados de várias tabelas com base em colunas em comum. Por exemplo, suponha que você tenha duas tabelas: "Orders" e

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

"Customers", e você queira obter detalhes de pedidos juntamente com as informações do cliente correspondente. Você pode executar a seguinte consulta:

```
SELECT o.order_id, o.order_date, c.customer_name
FROM Orders o
JOIN Customers c ON o.customer_id = c.customer_id
```

2. SUB-QUERY

As subconsultas são aninhadas dentro de uma consulta principal. Elas podem ser usadas para realizar cálculos intermediários ou filtrar resultados com base em condições específicas. Por exemplo, você pode querer obter todos os clientes que fizeram mais de 10 pedidos. A consulta pode ser assim:

```
SELECT customer_id, customer_name
FROM Customers
WHERE customer_id IN (
 SELECT customer_id
 FROM Orders
 GROUP BY customer_id
 HAVING COUNT(order_id) > 10)
```

3. Consultas com AGGREGATION

O BigQuery oferece várias funções de agregação que permitem calcular estatísticas ou resumos dos dados. Por exemplo, você pode usar a função SUM para calcular a soma de valores em uma coluna, ou COUNT para contar o número de registros que atendem a uma determinada condição. Aqui está um exemplo que calcula a receita total por categoria de produto:

```
SELECT product_category, SUM(revenue) AS total_revenue
FROM Sales
```

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

```
GROUP BY product_category
```

4. Consultas com função janela (window functions)

As consultas com janela permitem realizar cálculos em um conjunto de linhas que estão relacionadas a uma linha específica. Por exemplo, você pode querer calcular a média móvel de uma série temporal. Aqui está um exemplo que calcula a média móvel de 7 dias para as vendas diárias:

```
SELECT sale_date, sales_amount,
AVG(sales_amount) OVER (ORDER BY sale_date ROWS BETWEEN 6 PRECEDING AND CURRENT ROW) AS moving_average
FROM DailySales
```

Esses são apenas alguns exemplos de consultas SQL avançadas que você pode executar no Google BigQuery. Ele possui recursos avançados adicionais, como partições de tabela, clustering e suporte a dados geoespaciais, que permitem otimizar ainda mais suas consultas e análises.

FIAP

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

O QUE VOCÊ VIU NESTA AULA?

Entendemos melhor o funcionamento do Google Cloud Plataform e todos seus principais componentes para construção de uma arquitetura de dados voltada para nuvem. Além disso, nos aprofundamos ainda mais dentro dos conceitos e práticas de uso de SQL dentro do Google BigQuery.

O que achou desta aula? Conte-nos na comunidade do Discord! Te esperamos lá para ajudar com dúvidas, anúncios de lives e muito mais.

FIAP

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

REFERÊNCIAS

Google BigQuery (SQL) 101, 2021. Disponível em: <https://medium.com/basedosdados/bigquery-101-8b39da1ce52b>. Último acesso em: 19 jun 2023.

Modern Data Warehousing with BigQuery (Cloud Next '19), 2019. Disponível em: <https://www.youtube.com/watch?v=eOQ3YJKgvHE>. Último acesso em: 19 jun 2023.

BigQuery Documentação Oficial, [s.d.]. Disponível em: <https://cloud.google.com/bigquery/>. Último acesso em: 19 jun 2023.

FLEURY, Jan. 3 **Best practices to design and operate CDP architectures on Google Cloud Platform,** 2022. Disponível em: <https://www.crystalloids.com/insights/3-best-practices-to-design-and-operate-cdp-architectures-on-google-cloud-platform>. Último acesso em: 19 jun 2023.

THALLAM, Rajesh. **New Blog Series - BigQuery Explained: An Overview (2020)**. Disponível em: <https://medium.com/google-cloud/bigquery-explained-overview-357055ecfda3>. Último acesso em: 19 jun 2023.

WAIBEL, XINRAN. **BigQuery Best Practices,** 2020. Disponivel em: <https://medium.com/google-cloud/bigquery-best-practices-9452c294c9d9>. Último acesso em: 19 jun 2023.

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

PALAVRAS-CHAVE

Palavras-chave: Cloud, Data Lake, Data Warehouse, Data Lakehouse, SQL, Google Cloud, BigQuery, Pipeline, Integração de Dados.

FIAP

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com

PDF exclusivo para Igor Torves - rm353785
igor.torves.sl@hotmail.com