



**Instituto Superior
de Engenharia**

Politécnico de Coimbra

Sistemas de Informação II

2024/25 Trabalho Prático – Meta 1

Bruno Martins - 2022147149
João Rosa - 2022131973

Índice

1. Resumo.....	3
2. Introdução.....	4
3. Contextualização.....	5
3.1 A Evolução dos Data Warehouses.....	5
3.2 A Revolução da Nuvem.....	6
4. Estado da Arte sobre Data Warehousing.....	7
5. Serviços de Data Warehouse na Nuvem.....	8
5.1 Amazon Web Services (AWS) - Amazon Redshift.....	8
5.2 Microsoft Azure - Azure Synapse Analytics.....	9
5.3 Google Cloud Platform (GCP) - BigQuery.....	10
6. Comparação entre AWS, Azure e GCP.....	10
6.1 Custo.....	11
6.2 Escalabilidade.....	12
6.3 Facilidade de uso.....	13
6.4 Performance.....	14
6.5 Segurança.....	15
7. Estudos de Caso.....	16
7.1. Amazon Redshift: McDonald's.....	16
7.2. Azure Synapse Analytics: The Coca-Cola Company.....	17
7.3. Google BigQuery: Spotify.....	18
8. Conclusão.....	19

1. Resumo

Este relatório aborda a comparação das arquiteturas em nuvem das principais plataformas de Data Warehouse — Amazon Web Services (AWS), Microsoft Azure e Google Cloud Platform (GCP). A migração de Data Warehouses para a nuvem proporciona vantagens como elasticidade, redução de custos com infraestrutura e maior acessibilidade. O objetivo deste estudo é comparar os serviços Amazon Redshift, Azure Synapse Analytics e Google BigQuery, destacando as suas principais funcionalidades, benefícios, limitações, e exemplos de uso bem-sucedido. Além disso, este relatório oferece uma avaliação detalhada de desempenho, custo e escalabilidade, com o intuito de tomar a decisão da plataforma mais adequada às necessidades específicas de cada caso. Estudos de caso e dados práticos complementam a análise comparativa, facilitando a tomada de decisões estratégicas. Da figura 1 a 6 mostramos os logótipos das diferentes empresas a analisar durante o trabalho.



Figura 1 - AWS



Figura 2 - Amazon Redshift



Figura 3 - Microsoft Azure



Figura 4 - Azure Synapse Analytics



Google Cloud Platform

Figura 5 - GCP



Google BigQuery

Figura 6 - Google BigQuery

2. Introdução

Os Data Warehouses (DWs) são fundamentais para a gestão de grandes volumes de dados em organizações. Tradicionalmente, essas soluções eram implementadas em servidores locais, exigindo grandes investimentos em hardware, manutenção, e mão de obra especializada para gerir esses sistemas complexos. Com o surgimento das soluções em nuvem, houve uma transformação na forma como os DWs são construídos e mantidos, tendo sido benéfico a nível de escalabilidade, flexibilidade e redução de custos operacionais.

As plataformas de cloud computing, como AWS, Azure e GCP, oferecem serviços especializados e robustos para Data Warehousing que permitem o processamento de grandes volumes de dados com alta eficiência e custos controlados.

Este relatório foca-se na comparação das soluções de Data Warehousing em nuvem oferecidas por essas três plataformas: Amazon Redshift, Azure Synapse Analytics e Google BigQuery. A análise cobre as suas funcionalidades técnicas, vantagens, limitações e integrações com outras ferramentas de análise, com o objetivo de fornecer uma visão abrangente para ajudar a escolher a melhor solução de acordo com as necessidades. Além disso, o relatório explora os avanços tecnológicos na área, incluindo o uso de machine learning e análises em tempo real. A figura 7 é meramente ilustrativa e pretende representar um Data Warehouse.

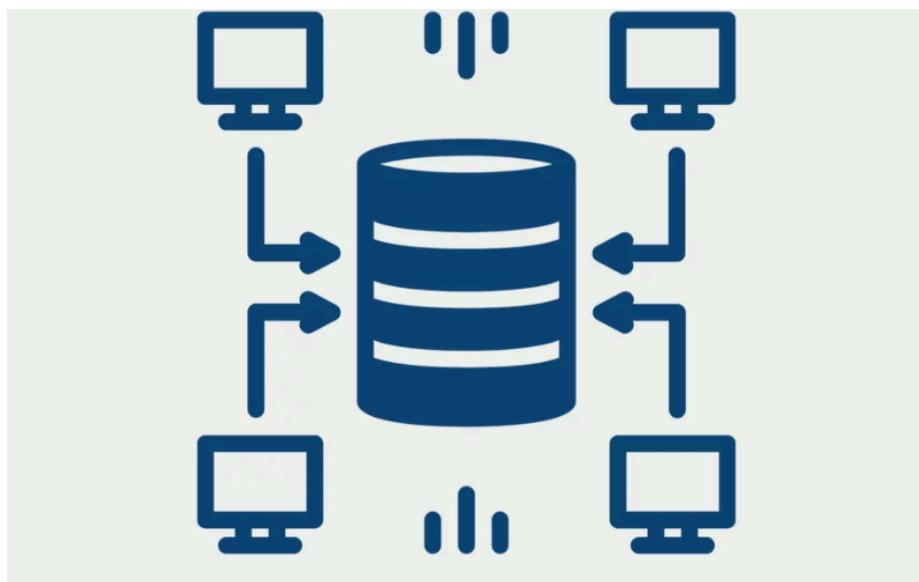


Figura 7- Data Warehouse

3. Contextualização

Neste capítulo iremos contextualizar um pouco a base do tema escolhido para a realização deste relatório, Data Warehouse.

3.1 A Evolução dos Data Warehouses

O conceito de Data Warehousing surgiu no final da década de 1980, com o intuito de centralizar dados provenientes de diversas fontes empresariais, facilitando a criação de relatórios e análises que apoiassem a tomada de decisões. Nos primeiros anos, empresas de grande porte implementaram Data Warehouses em sistemas locais, armazenando dados em servidores próprios, o que trazia desafios consideráveis, como:

- **Custos elevados:** Manter servidores físicos, infraestrutura e equipas técnicas para gerir o sistema.
- **Falta de escalabilidade:** A capacidade de armazenamento e processamento era limitada pela infraestrutura física.
- **Manutenção complexa:** A gestão de grandes volumes de dados exigia a participação de equipas especializadas, além da necessidade constante de atualizações e reparos em servidores.

Na figura 8 é mostrada como é a infraestrutura física, atualmente de Data Warehouse.



Figura 8 - Data Warehouse Físico

3.2 A Revolução da Nuvem

Com a computação em nuvem, ilustrada pela figura 9, as empresas passaram a adotar serviços de Data Warehousing baseados em cloud. Essas soluções trazem vantagens significativas:

- **Escalabilidade elástica:** A capacidade de aumentar ou diminuir os recursos de computação e armazenamento de acordo com a demanda.
- **Redução de custos:** As empresas pagam apenas pelos recursos que utilizam, eliminando a necessidade de investimentos pesados em infraestrutura física.
- **Acesso global:** As plataformas em nuvem podem ser acessadas de qualquer local, facilitando a colaboração global e o trabalho remoto.
- **Integração com machine learning e análise preditiva:** As soluções em nuvem integram-se facilmente com outras ferramentas de análise avançada, ampliando o objetivo das suas aplicações.

As três plataformas principais, AWS, Azure e GCP, competem diretamente neste espaço, oferecendo serviços especializados para Data Warehousing que atendem às necessidades de diversos setores, desde tecnologia até serviços financeiros. Neste relatório, iremos analisar e comparar as principais características e funcionalidades dessas plataformas, abordando o impacto de cada uma na transformação digital das empresas.



Figura 9 - Data Warehouse Cloud

4. Estado da Arte sobre Data Warehousing

Um **Data Warehouse (DW)** integra dados de múltiplas fontes, permitindo às organizações tomar decisões baseadas em informações agregadas e confiáveis. Ao contrário das bases de dados operacionais, que são otimizadas para operações diárias, os Data Warehouses são projetados para consultas rápidas e análise de dados históricos.

A arquitetura tradicional de um DW segue três camadas principais:

1. **ETL (Extração, Transformação e Carregamento)** – Processo que prepara os dados das fontes para análise.
2. **Armazenamento centralizado** – Onde os dados são organizados e armazenados, geralmente em esquemas estrela ou floco de neve.
3. **Ferramentas de análise** – Interfaces e ferramentas de BI que permitem a exploração dos dados.

Nos últimos anos, os Data Warehouses evoluíram com a adoção de soluções em **nuvem**, oferecendo maior elasticidade, menor custo e escalabilidade. Além disso, surgiram novas abordagens, como o **armazenamento em colunas** e a arquitetura **MPP (Processamento Paralelo Massivo)**, que melhoraram o desempenho das consultas.

Desafios Atuais

Os principais desafios ainda em estudo incluem:

- Integração de dados heterogêneos.
- Garantia da qualidade dos dados.
- Crescimento exponencial de dados e custos associados.
- Segurança e conformidade com regulamentações.

Tendências Emergentes

Recentemente, o Data Warehousing tem se integrado a tecnologias como **Data Lakes** para armazenar dados não estruturados e **análises em tempo real**, além de fusões com ferramentas de **machine learning** para análises preditivas.

5. Serviços de Data Warehouse na Nuvem

Neste capítulo serão demonstrados e explicados cada um dos serviços e algumas vantagens e desvantagens dos mesmos.

5.1 Amazon Web Services (AWS) - Amazon Redshift

O **Amazon Redshift** é um dos serviços mais populares de Data Warehouse em nuvem, utilizado por empresas de todos os tamanhos para processamento massivo de dados e análises avançadas. É construído sobre uma arquitetura distribuída de clusters, onde os dados são divididos entre múltiplos nós de computação.

Características principais:

- **Armazenamento em colunas:** O Redshift utiliza um formato de armazenamento em colunas que otimiza a leitura de grandes volumes de dados, tornando as consultas mais rápidas.
- **Consultas paralelas massivas (MPP):** A arquitetura MPP permite que várias consultas sejam executadas em paralelo, reduzindo o tempo de resposta.
- **Escalabilidade automática:** O Redshift permite adicionar e remover nós de computação dinamicamente, à medida que as necessidades de processamento mudam.
- **Deep integration:** O Redshift integra-se de forma nativa com outros serviços da AWS, como o **S3** (para armazenamento de dados em grandes volumes) e **AWS Glue** (para ETL automatizado).

Casos de uso: O Redshift é amplamente utilizado em setores como e-commerce, marketing digital e serviços financeiros, onde a análise de grandes quantidades de dados em tempo real é fundamental para a personalização de ofertas e otimização de campanhas.

Vantagens:

- Grande capacidade de processamento e alta performance em consultas de grandes volumes de dados.
- Integração profunda com o ecossistema AWS, permitindo a utilização de serviços como **SageMaker** para machine learning.
- Boa relação qualidade - preço.

Limitações:

- A complexidade de configuração e gestão pode ser um obstáculo para empresas com equipas técnicas menos experientes.
- Custos elevados quando há grande necessidade de escalabilidade para volumes muito grandes de dados.

5.2 Microsoft Azure - Azure Synapse Analytics

O **Azure Synapse Analytics** (anteriormente conhecido como Azure SQL Data Warehouse) é a solução da Microsoft para Data Warehousing na nuvem. Destaca-se pela sua capacidade de unir dados operacionais e analíticos numa única plataforma.

Características principais:

- **Análises integradas:** O Synapse integra a análise de dados com ferramentas como o **Power BI** e o **Azure Machine Learning**, permitindo uma análise avançada sem necessidade de mover dados entre diferentes plataformas.
- **Suporta queries SQL e Spark:** O Azure Synapse suporta tanto consultas SQL quanto o processamento com Spark, facilitando a análise de dados estruturados e não estruturados.
- **Escalabilidade sob demanda:** O Synapse ajusta automaticamente os recursos de computação e armazenamento com base nas necessidades do workload.

Casos de uso: O Synapse é amplamente adotado em setores como saúde, educação e serviços públicos, onde é necessário realizar análises profundas sobre dados demográficos, registo de pacientes ou desempenho acadêmico.

Vantagens:

- Integração nativa com ferramentas Microsoft como o **Power BI**, facilitando a criação de relatórios e dashboards interativos.
- Capacidade de processar grandes volumes de dados em tempo real com alta performance.
- Elasticidade que permite expandir e contrair recursos de acordo com as necessidades do negócio.

Limitações:

- Algumas funcionalidades avançadas, como consultas Spark, podem exigir maiores conhecimentos técnicos.
- O custo pode se tornar elevado em cenários de uso intensivo de computação.

5.3 Google Cloud Platform (GCP) - BigQuery

O **Google BigQuery** é um Data Warehouse serverless, conhecido pela sua simplicidade de uso e capacidade de escalar automaticamente sem a necessidade de configuração de servidores. Foi projetado para análises rápidas de grandes volumes de dados e é amplamente utilizado por empresas de tecnologia e startups.

Características principais:

- **Serverless:** O BigQuery é totalmente gerido pela Google, não sendo necessária a configuração manual de servidores ou clusters.
- **Análise em tempo real:** Suporta consultas rápidas em dados de streaming, permitindo que as empresas façam análises em tempo real.
- **Machine Learning integrado:** O BigQuery oferece modelos de machine learning que podem ser treinados diretamente na plataforma, sem necessidade de mover dados para outros sistemas.

Casos de uso: O BigQuery é amplamente utilizado por empresas que precisam de escalabilidade automática para grandes volumes de dados, como startups de tecnologia, empresas de media e companhias aéreas que precisam processar e analisar dados de clientes em tempo real.

Vantagens:

- Facilidade de uso com uma interface intuitiva e consultas rápidas em SQL.
- Escalabilidade automática sem necessidade de gerenciamento de recursos.
- Integração com ferramentas Google como o **Google Analytics** e **Data Studio**, facilitando a visualização de dados.

Limitações:

- O custo de consultas pode ser elevado para empresas com grandes volumes de dados.
- Pode ser menos personalizável em comparação com AWS ou Azure para empresas que necessitam de controlo total sobre a infraestrutura.

6. Comparação entre AWS, Azure e GCP

Neste capítulo serão exemplificados vários temas e as diferenças entre as 3 empresas em cada um deles.

6.1 Custo

Neste capítulo abordamos o tema da escalabilidade nos 3 serviços das diferentes empresas.

AWS Redshift:

- O **Redshift** utiliza um modelo de precificação baseado no **número de nós** e no **tempo de utilização**. Cada nó tem um custo que varia de acordo com o tipo e a capacidade de processamento. Existem instâncias otimizadas para diferentes tipos de cargas de trabalho, como armazenamento denso ou computação intensiva.
- Para empresas que necessitam de grande **escalabilidade horizontal** e formam grandes clusters, o custo pode aumentar rapidamente à medida que mais nós são adicionados.
- No entanto, estão disponíveis **descontos para compromissos a longo prazo**, como a reserva de nós por períodos de 1 ou 3 anos, o que pode ajudar a reduzir os custos em até 75%.
- A estrutura de custos também inclui o armazenamento de dados, que é cobrado à parte caso sejam utilizados serviços adicionais da AWS, como o **Amazon S3** para arquivamento de dados.

Azure Synapse Analytics:

- No **Synapse Analytics**, o custo é determinado com base na utilização de **recursos de computação** e **armazenamento** de forma separada, o que permite maior flexibilidade para ajustar o orçamento ao workload. Para cargas de trabalho mais pequenas, o modelo "pay-as-you-go" ajuda a pagar apenas pelos recursos efetivamente utilizados.
- Um dos principais pontos fortes do Azure é a **elasticidade no ajuste dos recursos**, permitindo que as empresas alocam mais ou menos capacidade de computação conforme necessário, ajustando o orçamento de forma dinâmica.
- Para empresas já integradas no **ecossistema Microsoft**, o Synapse oferece uma vantagem competitiva, uma vez que a integração com ferramentas como o **Power BI** permite análise e visualização de dados em tempo real, o que pode reduzir os custos operacionais para quem já possui licenças dessas ferramentas.

Google BigQuery:

- O **BigQuery** segue um modelo de custo "pay-as-you-go", em que as empresas pagam pelas **consultas executadas** e pelo **volume de dados armazenados**. Este modelo é altamente flexível e benéfico para empresas com cargas de trabalho imprevisíveis ou sazonais, já que não há necessidade de provisionar infraestrutura antecipadamente.
- Contudo, este modelo pode tornar-se caro para empresas que realizam consultas frequentes a grandes volumes de dados, especialmente se os dados estiverem distribuídos em várias regiões e a latência da rede for um fator.
- O custo pode também incluir a transferência de dados entre outros serviços da Google Cloud ou para fora da plataforma, algo a ter em conta para empresas que dependem de múltiplas plataformas ou que integrem o BigQuery com outros sistemas de cloud computing.

6.2 Escalabilidade

Neste capítulo abordamos o tema da escalabilidade nos 3 serviços das diferentes empresas.

AWS Redshift:

- O Redshift oferece **escalabilidade tanto vertical como horizontal**, permitindo que as empresas adicionem ou removam **nós de computação** conforme necessário. A **escalabilidade vertical** permite aumentar a capacidade de um nó existente, enquanto a **escalabilidade horizontal** envolve a adição de mais nós ao cluster.
- No entanto, o processo de escalabilidade pode ser **manual**, o que significa que os utilizadores precisam de planejar a adição ou remoção de recursos de forma proactiva, o que pode ser desafiador para empresas sem uma equipa técnica especializada. Dependendo da configuração, pode ser necessário realocar ou redistribuir dados para otimizar o desempenho.

Azure Synapse Analytics:

- O Synapse Analytics destaca-se pela sua **escalabilidade sob demanda**, o que significa que os recursos de computação e armazenamento são automaticamente ajustados com base nas necessidades do workload.
- Esta flexibilidade de escalabilidade torna-o uma excelente solução para empresas que enfrentam picos sazonais de carga de trabalho ou que precisam de ajustar os recursos rapidamente, sem necessidade de intervenção manual. A capacidade de escalar automaticamente oferece uma **experiência mais fluida**, eliminando a necessidade de ajustar manualmente os nós ou clusters.

Google BigQuery:

- **Serverless** por natureza, o BigQuery escala automaticamente conforme as necessidades da aplicação, sem que o utilizador tenha de se preocupar com a configuração ou gestão da infraestrutura.
- Esta escalabilidade automática é ideal para empresas que necessitam de uma **expansão imediata**, sem necessidade de planeamento prévio, facilitando a resposta a picos inesperados de carga de trabalho. Este modelo é adequado tanto para startups em rápido crescimento como para grandes corporações que necessitam de escalabilidade instantânea para processar grandes volumes de dados.

6.3 Facilidade de uso

Neste capítulo abordamos o tema da facilidade de uso nos 3 serviços das diferentes empresas.

AWS Redshift:

- O Redshift requer uma **configuração inicial complexa**, especialmente na criação e gestão dos clusters de nós, o que pode ser um desafio para empresas sem uma equipa técnica experiente. Além disso, a **otimização de consultas** e o ajuste da performance dos clusters podem exigir conhecimentos especializados para garantir que os recursos são utilizados de forma eficiente.
- No entanto, a Amazon disponibiliza ferramentas como o **Redshift Advisor**, que recomenda ajustes automáticos para melhorar o desempenho e sugere mudanças no design de tabelas e consultas.

Azure Synapse Analytics:

- O **Synapse Analytics** é considerado **intuitivo** para utilizadores que já estão familiarizados com o ecossistema Microsoft. A interface é amigável, especialmente para quem usa o Power BI, facilitando a criação de dashboards e relatórios sem necessidade de exportar dados.
- A sua **curva de aprendizagem** aumenta para utilizadores que pretendam tirar partido de funcionalidades avançadas, como o uso de **Apache Spark** para análise de dados não estruturados. A integração com múltiplas linguagens (SQL, Spark, etc.) pode exigir um conhecimento técnico mais profundo.

Google BigQuery:

- O BigQuery é reconhecido pela sua **simplicidade** e interface intuitiva, especialmente para **consultas SQL**. A natureza **serverless** elimina a necessidade de configurar ou gerir servidores, tornando-o fácil de usar para empresas que não dispõem de uma equipa técnica dedicada à infraestrutura.
- Esta simplicidade torna o BigQuery uma solução acessível para empresas de tecnologia e startups que pretendem focar-se na análise de dados sem se preocuparem com a gestão da infraestrutura subjacente.

6.4 Performance

Neste capítulo abordamos o tema da performance nos 3 serviços das diferentes empresas.

AWS Redshift:

- O Redshift oferece **elevada performance** em consultas de grandes volumes de dados, principalmente devido à sua arquitetura de **armazenamento em colunas** e suporte a **processamento paralelo massivo (MPP)**, que permite processar várias consultas em simultâneo.
- A performance, no entanto, depende da **correcta configuração e optimização** dos nós e da distribuição dos dados entre os clusters.

Azure Synapse Analytics:

- O Synapse oferece **elevada performance** para análises em tempo real, suportando tanto **consultas SQL** como o processamento de dados em **Spark**. A sua capacidade de gerir cargas de trabalho mistas (dados estruturados e não estruturados) torna-o uma plataforma versátil para empresas que necessitam de análise avançada de dados.
- O desempenho do Synapse pode ser melhorado com ajustes automáticos de recursos durante picos de utilização, minimizando a latência.

Google BigQuery:

- O **BigQuery garante alta performance sem necessidade de configuração manual**. A sua infraestrutura escalável permite processar grandes volumes de dados em tempo real, sendo ideal para cargas de trabalho que requerem consultas rápidas e eficientes.
- No entanto, a performance pode variar conforme a **complexidade dos dados** e a quantidade de recursos disponíveis no momento da consulta.

6.5 Segurança

Neste capítulo abordamos o tema da segurança nos 3 serviços das diferentes empresas.

AWS Redshift:

- O Redshift oferece **encriptação de dados em repouso e em trânsito**, utilizando **chaves geridas pela AWS** ou pelos próprios clientes. Também suporta **integração com AWS Identity and Access Management (IAM)** para controlo de permissões de acesso.
- Está em conformidade com regulamentos como o **GDPR** e o **HIPAA**, sendo adequado para empresas que processam dados sensíveis ou regulamentados.

Azure Synapse Analytics:

- O Azure Synapse beneficia das **políticas de segurança robustas** do Azure, incluindo a integração com o **Azure Active Directory (AAD)**, que permite a autenticação multifator e o controlo detalhado de permissões.
- Oferece **encriptação avançada** e cumpre com padrões de conformidade internacionais, como **ISO/IEC 27001**, **GDPR** e **FedRAMP**.

Google BigQuery:

- O BigQuery segue os **padrões de segurança da Google Cloud**, suportando **encriptação de dados** e oferecendo conformidade com normas como **ISO/IEC 27001**, **SOC** e **GDPR**. A Google gera a infraestrutura automaticamente, aplicando patches e mantendo a segurança da plataforma sem intervenção do utilizador.

7. Estudos de Caso

Neste capítulo serão descritos os casos de sucesso da implementação das diferentes plataformas em empresas de sucesso.

7.1. Amazon Redshift: McDonald's

A gigante do fast-food, McDonald's, recorre ao Amazon Redshift, como ilustra a figura 10, para analisar dados de vendas, desempenho de restaurantes e comportamento dos clientes. A plataforma permite a consolidação de dados de diversas fontes, como sistemas de ponto de venda (POS), aplicações móveis e inquéritos de satisfação, proporcionando uma visão abrangente das operações.

Benefícios para a McDonald's:

- **Personalização de Ofertas:** Através da análise das preferências dos clientes, a McDonald's consegue personalizar ofertas e promoções, aumentando a relevância para o consumidor.
- **Otimização de Operações:** O Redshift permite monitorizar o desempenho de cada restaurante em tempo real, identificar estrangulamentos e otimizar a gestão de stocks, aumentando a eficiência operacional.
- **Melhoria da Experiência do Cliente:** Através da análise de dados como tempo de espera e pedidos mais frequentes, a McDonald's consegue identificar áreas de melhoria na experiência do cliente.

Exemplo Concreto:

O McDonald's, através do Redshift, identifica que num determinado restaurante existe um pico de vendas de McFlurries às 16h. Com esta informação, o gerente pode reforçar a equipa nesse período, assegurando um serviço mais rápido e eficiente, e até mesmo oferecer promoções de McFlurries nesse horário para aumentar as vendas.



Figura 10 - :Amazon Redshift: McDonald's

7.2. Azure Synapse Analytics: The Coca-Cola Company

A Coca-Cola utiliza o Azure Synapse Analytics para integrar dados de diferentes departamentos, como vendas, marketing, produção e distribuição. A plataforma permite a criação de um "gémeo digital" da cadeia de abastecimento, possibilitando a otimização da produção, a previsão da procura e a gestão de stock em tempo real.

Benefícios para a Coca-Cola:

- **Simulação de Cenários:** O Synapse permite simular o impacto de diferentes decisões na cadeia de abastecimento, como o lançamento de um novo produto ou a alteração de um fornecedor, permitindo uma tomada de decisão mais informada.
- **Previsão da Procura:** Através da análise de dados históricos e tendências de mercado, a Coca-Cola consegue prever a procura futura de produtos e ajustar a produção de forma eficiente.
- **Otimização da Distribuição:** A monitorização do transporte de produtos em tempo real permite identificar e resolver problemas na cadeia de abastecimento, garantindo entregas mais rápidas e eficientes.

Exemplo Concreto:

Com o Azure Synapse Analytics, a Coca-Cola pode analisar dados de vendas de diferentes regiões e identificar um aumento na procura de bebidas sem açúcar durante o verão. Com base nesta informação, a empresa pode ajustar a produção e distribuição, garantindo a disponibilidade destes produtos nos locais e períodos de maior procura. Na figura 11 é mostrada a relação entre a Coca Cola e a Microsoft.



Figura 11 - Azure Synapse Analytics: The Coca-Cola Company

7.3. Google BigQuery: Spotify

O Spotify utiliza o Google BigQuery, como ilustra a figura 12, para analisar dados de streaming de música e preferências dos utilizadores. A plataforma permite processar grandes volumes de dados em tempo real, como histórico de músicas ouvidas, playlists criadas e artistas favoritos.

Benefícios para o Spotify:

- **Personalização de Recomendações:** O BigQuery permite analisar os gostos e o histórico de reprodução de cada utilizador, oferecendo recomendações de músicas, artistas e playlists mais relevantes.
- **Identificação de Tendências:** Através da análise das preferências musicais dos utilizadores em diferentes regiões, o Spotify consegue identificar novas tendências e artistas em ascensão.
- **Otimização do Catálogo:** A análise da popularidade de músicas e artistas permite otimizar o catálogo e as playlists oferecidas aos utilizadores.

Exemplo Concreto:

O Spotify, através do BigQuery, identifica que utilizadores que ouvem música clássica também demonstram interesse em jazz. Com base nesta informação, a plataforma pode recomendar artistas de jazz a utilizadores que apenas ouviam música clássica, expandindo os seus horizontes musicais.



Figura 12 - Google BigQuery: Spotify

8. Conclusão

A análise comparativa entre Amazon Web Services (AWS), Microsoft Azure e Google Cloud Platform (GCP) demonstra que cada uma dessas plataformas de Data Warehouse em nuvem apresenta **vantagens e desvantagens específicas**, dependendo das necessidades da empresa e da sua infraestrutura existente. O **Amazon Redshift** é ideal para organizações que já utilizam o ecossistema AWS e precisam de uma solução robusta com alta performance e escalabilidade manual. Por outro lado, o **Azure Synapse Analytics** é uma escolha excelente para empresas que utilizam o ecossistema Microsoft, oferecendo uma integração perfeita com ferramentas como Power BI e Azure Machine Learning, além de suporte para análises avançadas.

Por sua vez, o **Google BigQuery** destaca-se pela sua simplicidade e escalabilidade automática, sendo uma solução atrativa para empresas que desejam uma abordagem serverless e que procuram análises rápidas e eficientes sem a necessidade de gerenciar a infraestrutura. Contudo, o custo elevado de consultas frequentes em grandes volumes de dados pode ser um ponto de atenção.

A escolha da plataforma ideal deve ser feita com base em fatores como o volume de dados, a necessidade de escalabilidade, o ecossistema tecnológico já utilizado pela empresa e os recursos disponíveis para a gestão de dados.

9. Referências

1. "Cloud Data Warehouse - Amazon Redshift - AWS". Amazon Web Services, Inc., aws.amazon.com/redshift/.
2. "Spotify Case Study | Google Cloud". Google Cloud, cloud.google.com/customers/spotify.
3. "O McDonald's impulsiona a transformação digital na nuvem na AWS". Amazon Web Services, Inc., aws.amazon.com/pt/solutions/case-studies/mcdonalds.
4. "BigQuery enterprise data warehouse". Google Cloud, cloud.google.com/bigquery.
5. "The Coca-Cola Company and Microsoft announce five-year strategic partnership to accelerate cloud and generative AI initiatives - Stories". Stories, news.microsoft.com/2024/04/23/the-coca-cola-company-and-microsoft-announce-five-year-strategic-partnership-to-accelerate-cloud-and-generative-ai-initiatives.
6. "Cloud Data Warehouse Comparison: Redshift vs BigQuery vs Azure vs Snowflake for Real-Time Workloads". Striim, www.striim.com/blog/cloud-data-warehouse-comparison-redshift-vs-bigquery-vs-azure-vs-snowflake-for-real-time-data.
7. Smine, Sadok. "Big Data Solutions: BigQuery, Redshift, and Azure Synapse Analytics". Medium, medium.com/@sadoksmine8/big-data-solutions-bigquery-redshift-and-azure-synapse-analytics-4e842692a9f9.