**Data WareHouse**

Define-se Data Warehouse, são repositórios centrais que armazenam dados de uma ou mais fontes heterogêneas. Armazéns de dados são ferramentas analíticas construídas para apoiar a tomada de decisões para usuários de relatórios em vários departamentos, o Data Warehouse torna-se uma biblioteca de informações que podem ser recuperadas e analisadas para uma melhor tomada de decisão. Trabalha para criar um sistema único e unificando os dados, como repositório central ele armazena os metadados. (ACERVO LIMA, 2022)

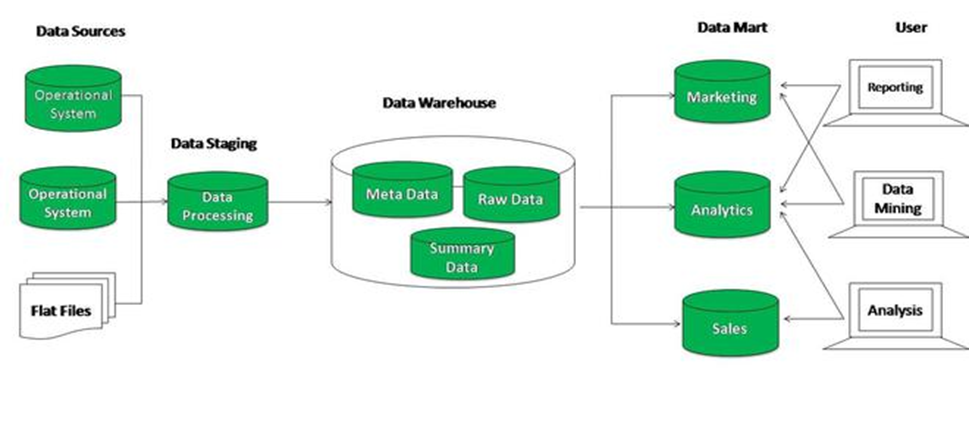
**Figura 1 – Fluxo de dados por meio de arquitetura do Data Warehouse**

Fonte: ACERVO LIMA, 2022

Atualmente com a expansão da tecnologia em nuvem reduziu-se o esforço e custos na construção de armazenamento de dados para as empresas. Com isso os Data Warehouse estão deixando de ser data centers físicos para Data Warehouse baseados em nuvens. Algumas organizações ainda operam de forma tradicional, mas o futuro do armazenamento está computação em nuvem.

Existem várias ferramentas disponíveis no mercado, algumas com armazenamento de dados em nuvem ou híbrida. Portanto, cabe a empresa identificar qual a melhor solução de acordo com os seus requisitos para implantação do seu sistema de armazenamento de dados. A seguir cinco ferramentas e suas devidas características, porém existem outras ferramentas disponíveis no mercado.

**Figura 1 – Fluxo de dados por meio de arquitetura do Data Warehouse**

****

**Amazon Redshift**

Amazon Redshift é uma das ferramentas de DW mais classificadas por usuários online. E como o nome sugere, Redshift faz parte dos serviços de web, um data Warehouse em escala de petabytes totalmente gerenciado baseado em nuvem pela Amazon Company. Iniciando com gigabytes de dados e escala para petabytes ou mais. Se integra a diversas ferramentas de ETL (Extração, Transformação e Carregamento), além de ferramentas de BI (Business Intelligence). Isso permite o uso de dados para acumular novos insights para empresas e clientes. (ACERVO LIMA, 2022)

Caracteriza-se por: Automatizar tarefas administrativas, como gerenciar, monitorar e dimensionar o data Warehouse. E permite a execução de consultas não estruturadas, o que otimiza o tempo, conforme mostrado na figura a seguir.

**Figura 2 - Arquitetura de Data Warehouse do Amazon Redshift**

Diagrama

Descrição gerada automaticamente

O principal componente da arquitetura de um data warehouse do Amazon Redshift é o “cluster”, composto por um ou mais nós de computação. O serviço permite que analistas executem consultas em questão de segundos, isso porque a ferramenta continua atualizando o pool de dados, justamente para que as conexões possam ser reutilizadas ao reaplicar ao informações de drives com falhas e substituindo-as quando necessário. (DOCS.AWS.AMAZON, 2023).

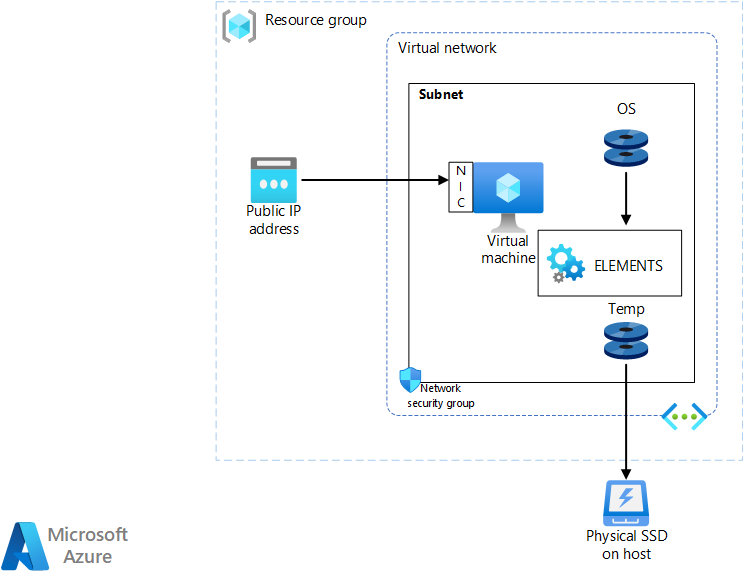
**Microsoft Azure**

Lançada pela Microsoft em 2010, Azure é uma plataforma de computação em nuvem pública que oferece infraestrutura como serviço (IaaS), plataforma como (PaaS) e o software como serviço (SaaS), é um provedor de serviços de computação em nuvem para construir, testar, implementar e gerenciar aplicativos e serviços por meio de data centers gerenciados pela Microsoft.

O Azure oferece mais de 200 produtos e serviços em nuvem, como análise de Dados, Computação Virtual, Armazenamento, Rede Virtual, Gerenciador de Tráfego da Internet, Sites, Serviços de Mídia, Serviços Móveis, Integração etc. O Azure oferece conexões cruzadas, incluindo redes virtuais (VPNs), caches, redes de entrega de conteúdo (CDNs) e conexões ExpressRoute para melhorar a usabilidade e o desempenho. Fornece uma base segura em toda a infraestrutura física e segurança operacional com serviço de hospedagem na web gerenciado que ajuda na construção de aplicativos e serviços e APIs Restful da web.

Como seus aplicativos mais populares as máquinas virtuais ou contêiners, oferece uma variedade de planos que visa atender requisitos de qualquer aplicativo desde pequenos até em escala global. (ACERVO LIMA, 2022).

**Figura 3 - Arquitetura Microsoft Azure com configuração de nó único**



**Google BigQuery**

O BigQuery é um data warehouse corporativo econômico e complemente sem servidor oferecida pelo Google. Com serviços de “machine learning” e “BI “integrados que funcionam em várias nuvens e escalonamento de acordo com os dados. Pode-se criar dados estruturados, semiestruturados e não estruturados em escala global diretamente no BigQuery. Usa-se SQL simples em suas consultas. É eficiente para gerar insights através das informações coletadas., seu sistema interativo de indexação do Big Query permite consultas muito rápidas e completas, torna-se interessante para empresas que utilizam seus próprios dados e que lidam com diversos tipos de gerenciamento de dados ente suas equipes. Os dados podem ser analisados em tempo real para obter informações atualizadas,

Permite analisar petabytes de dados em uma velocidade eficiente, além de possuir uma ótima relação entre custo e benefício (CLOUD.GOOGLE.COM, 2023).

**Figura 4 - Arquitetura Google Cloud no BigQuery em lotes e streams**

Diagrama

Descrição gerada automaticamente

Fonte: CLOUD.GOOGLE.COM, 2023

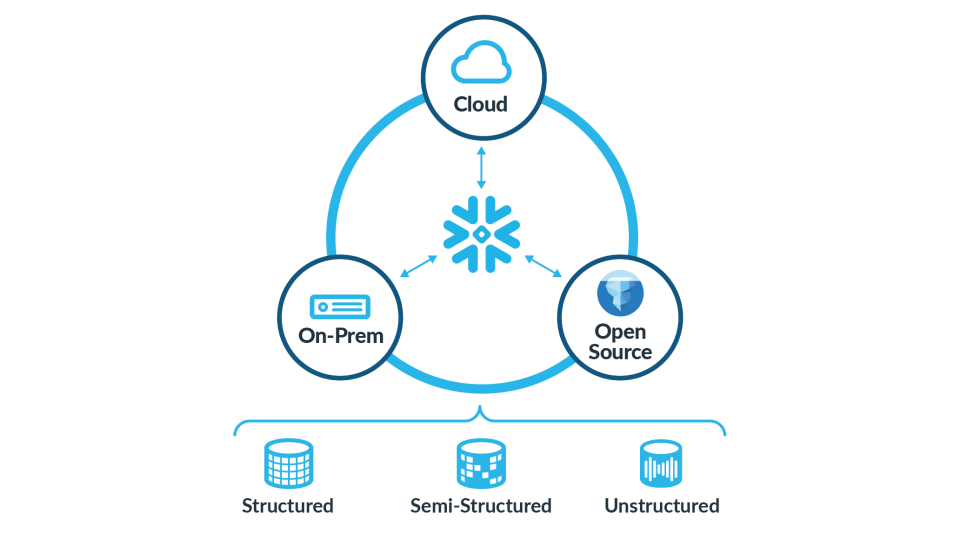
**Snowflake**

Uma plataforma baseada em armazenamento de dados baseado em nuvem construído sobre os Amazon Web Services ou a infraestrutura de nuvem do Microsoft Azure, oferece

serviços para dados estruturados e semiestruturados com o poder de computação dinâmico e escalonável, eficiente e de fácil manuseio, e combinados a um mecanismo de consulta SQL com cobranças baseadas estritamente no seu uso, sua arquitetura permite armazenamento e a computação sejam escalonados separadamente, portanto os clientes podem usar e pagar pelo serviço individualmente.

Sua nuvem possui natureza elástica, o que significa que uma grande quantidade de dados pode ser armazenada e várias consultas pode ser executada simultaneamente. O recurso possibilita que dados estruturados e semiestruturados combinados podem ser carregados no banco de dados em nuvem sem se tornar uma categoria fixa. (ACERVO LIMA, 2022).

**Figura 5 - Arquitetura Data Warehouse Snowflake**



Fonte: SNOWFLAKE.COM, 2023

**PostgreSQL**

É um sistema de gerenciamento banco de dados de código aberto mais avançado e flexível, que pode servir como banco relacional simples ou de séries temporais e até mesmo como uma solução para armazenamento de dados eficiente e de baixo custo. Pode-se integrá-lo a várias ferramentas e aplicativos externos para mineração de dados e relatórios. Extremamente estável, apoiado por mais de vinte anos de desenvolvimento comunitário que contribui para altos níveis de resiliência, integridade e correção, os procedimentos e as funções podem ser criados em várias linguagens de programação, como “PL”, “pgSQL”, “Python” etc.

**Figura 6 – Arquitetura do Data Warehouse PostgreSQL**

Interface gráfica do usuário, Diagrama

Descrição gerada automaticamente

Fonte: POSTGRESQL.ORG, 2023