

**UNIVERSIDADE FEDERAL DE UBERLÂNDIA BACHARELADO EM
GESTÃO DA INFORMAÇÃO**

THASSIO FERREIRA ZARUZ

**ADOÇÃO DE SERVIÇO DE COMPUTAÇÃO EM NUVEM NA ÁREA DE BI EM
EMPRESA DE CALL CENTER**

UBERLÂNDIA – MG

2022

THASSIO FERREIRA ZARUZ

**ADOÇÃO DE SERVIÇO DE COMPUTAÇÃO EM NUVEM NA ÁREA DE BI EM
EMPRESA DE CALL CENTER**

Monografia apresentada ao Curso de Graduação em Gestão da Informação, da Universidade Federal de Uberlândia, como exigência parcial para a obtenção do título de Bacharel.

Orientador: José Eduardo Ferreira Lopes

UBERLÂNDIA – MG

2022

THASSIO FERREIRA ZARUZ

**ADOÇÃO DE SERVIÇO DE COMPUTAÇÃO EM NUVEM NA ÁREA DE BI EM
EMPRESA DE CALL CENTER**

Monografia apresentada ao Curso de Graduação em Gestão da Informação, da Universidade Federal de Uberlândia, como exigência parcial para a obtenção do título de Bacharel.

Orientador: José Eduardo Ferreira Lopes

Prof.

Prof.

Prof.

Uberlândia, 23 de Janeiro de 2022

RESUMO

O objetivo deste relato técnico foi de descrever a implementação de um serviço de computação em nuvem em uma empresa do setor de *Call Center*, visando o melhor custo-benefício para a organização, tendo em vista o setor em que está inserida. Para isso foi realizada a migração de determinados dados para a nuvem e, após processados, retornados à infraestrutura da empresa, com o intuito de ter maior controle sobre os custos de geração de relatórios. Os resultados obtidos mostram que foi possível melhorar a performance dos processos de ETL utilizando o *BigQuery*, um dos serviços do *Google Cloud Platform*, e também diminuir os custos com infraestrutura local, proporcionando ganhos financeiros e competitivos para a empresa, visto que esta passa a prestar serviços de maior qualidade e menor preço.

Palavras-chave: **computação em nuvem, *BigQuery*, ETL**

ABSTRACT

The objective of this technical report was to describe the implementation of a *cloud computing* service in a call center company, aiming at the best cost benefit for the organization, in view of the sector in which it is inserted. For this purpose, certain data was migrated to the cloud and, after being processed, returned to the company's infrastructure, in order to have greater control over the costs of reporting. The results obtained show that it was possible to improve the performance of ETL processes using *BigQuery*, one of *google cloud Platform services*, and also reduce local infrastructure costs, providing financial and competitive gains for the company, since it starts to provide higher quality and lower price services.

Keywords: *cloud computing*, *BigQuery*, ETL

SUMÁRIO

Sumário

<u>1 INTRODUÇÃO</u>	7
<u>2 REFERENCIAL TEÓRICO</u>	8
<u>2.1 CLOUD COMPUTING</u>	8
<u>2.2 INTEGRAÇÃO DE DADOS</u>	9
<u>2.3 BANCO DE DADOS</u>	10
<u>2.4 FERRAMENTAS</u>	10
<u>2.4.1 SQL SERVER</u>	10
<u>2.4.2 SQL SERVER INTEGRATION SERVICES</u>	11
<u>2.4.3 PENTAHO</u>	11
<u>2.4.4 GOOGLE CLOUD PLATFORM</u>	11
<u>2.4.5 GOOGLE BIGQUERY</u>	12
<u>3 SITUAÇÃO PROBLEMA</u>	13
<u>4 INTERVENÇÃO ADOTADA</u>	14
<u>5 RESULTADOS OBTIDOS</u>	20
<u>6 CONSIDERAÇÕES FINAIS</u>	21
<u>7 REFERÊNCIAS</u>	22

1 INTRODUÇÃO

A Computação em Nuvem é um conceito que surgiu a pouco tempo, não possuindo total consenso em sua definição. Porém ela pode ser compreendida de acordo com a previsão feita pelo cientista americano John McCarthy do Massachusetts Institute of Technology (MIT) em 1961,

“se os computadores, da forma como eu imagino, se tornarem os computadores do futuro, então a computação poderá ser organizada como um serviço público, assim como a telefonia é... Cada assinante pagará apenas pelos recursos que ele realmente utilizará, mas ele terá acesso a todos os recursos oferecidos pelas linguagens de programação de um grande sistema... Alguns assinantes poderão oferecer serviços a outros assinantes. A computação como um serviço público poderá ser base de uma nova e importante indústria”.

Então no século seguinte, a previsão feita por John McCarthy tanto se tornou realidade como uma necessidade para empresas dos mais variados setores, já que esta é uma tecnologia que fornece redução de riscos comparado aos servidores *on-premise*, maior gerenciabilidade de custos, uma vez que estes são por demanda e maior escalabilidade e segurança dos dados (IMASTERS, 2022; CARISSIMI, 2022).

A computação em nuvem consiste em um mecanismo que possibilita o acesso *on-demand* de diversos serviços de computação como servidores, redes, armazenamento de dados, aplicativos e outros, permitindo que o usuário gerencie os serviços adquiridos de acordo com sua necessidade. Desde que exista conexão à internet, os serviços de *cloud computing* podem ser utilizados a partir de servidores remotos, estes que possuem a premissa de serem provisionados com rapidez e mínimo de envolvimento por parte do provedor de serviço e com facilidade de configuração e gerenciamento por parte do usuário (TAURION, 2009; SUNYAEV, 2020).

A partir disso é possível entender que a *Cloud Computing* se tornou uma ferramenta que as empresas utilizam para melhorar e aumentar sua produtividade, reduzir custos e também para manter o negócio em um ambiente de rápida e constante evolução tecnológica, permitindo também que a empresa tenha total foco em seu produto, deixando as áreas de apoio serem fornecidas por empresas especializadas, economizando em equipes de sustentação e infraestrutura. Tendo

segurança de que a infraestrutura de Tecnologia da Informação seja projetada para suportar e otimizar o negócio, garantindo confiabilidade e escalabilidade (SUNYAEV, 2020).

Diante das características apresentadas da computação em nuvem, entende-se que a utilização desta tecnologia, permite às empresas tratar suas necessidades com mais eficiência, reduzindo custos e aumentando a rapidez de processos, ao comparar com soluções internas (CENTRALSERVER, 2023).

Assim o objetivo deste relato tecnológico é descrever a reestruturação de um processo feito em arquitetura interna passando a utilizar serviços em nuvem, sendo mais específico, a utilização do poder de processamento de um serviço em nuvem em uma empresa de Contact Center, utilizando-o da forma mais adequada às necessidades e premissas da empresa.

Cabe destacar que, este projeto, foi desenvolvido pela equipe de Business Intelligence (BI) da própria empresa, onde este pesquisador faz parte e atuou diretamente na solução apresentada.

2 REFERENCIAL TEÓRICO

Para fundamentar este relato, foi necessário discutir alguns conceitos e premissas de *cloud computing*, processo de integração de dados e sistemas de Banco de Dados. A partir desses conceitos foi necessário elencar ferramentas que forneçam os serviços de cada um deles, sendo elas: *SQL Server Integration Services (SSIS)*, Pentaho, *SQL Server*, *Google Big Query* e *Google Cloud Storage (GCS)*, sendo que esses dois últimos fazem parte do serviço de nuvem disponibilizado pelo *Google Cloud Platform (GCP)*.

2.1 CLOUD COMPUTING

Com a constante evolução da tecnologia computacional e das telecomunicações o acesso à internet se popularizou, tornando a ideia de ter arquivos e dados armazenados, assim como programas instalados em nosso próprios computadores ultrapassada, e isso afeta principalmente as organizações corporativas, as quais adotaram um sistema onde existem um servidor local contendo as aplicações e dados empresariais que são acessados por terminais através de uma rede, possibilitando maior segurança e eficiência nos processos da empresa, porém, isso obriga as

empresas a terem equipes especializadas para manter, manter e sustentar esses servidores (CENTRALSERVER, 2017).

Esse cenário levou a popularização de um conceito que surgiu em 1961, criado pelo professor John McCarthy, que posteriormente seria chamado de *cloud computing* (computação em nuvem), o termo foi criado pelo então CEO do Google, em 2007, ele foi usado para referenciar a forma em que a empresa gerenciava seus *Data Centers* (RITTINGHOUSE; RANSOME, 2009).

“Embora possa parecer revolucionário, o conceito de Computação em Nuvem é um passo evolutivo na eterna busca pelo compartilhamento e conseqüentemente maior aproveitamento dos recursos computacionais” (TAURION, 2009, p. 24).

Assim, a *Cloud Computing*, também conhecida como Computação em Nuvem é tida como “um conjunto de serviços de rede, que proporciona escalabilidade, qualidade de serviço, infraestrutura barata de computação sob demanda, que pode ser acessado de uma forma simples” (SUNYAEV, 2020, p. 198), permitindo que empresas e pessoas possam acessar serviços de qualidade por um preço acessível, que são mantidos por organizações especializadas.

2.2 INTEGRAÇÃO DE DADOS

A integração de dados é um procedimento que consiste no agrupamento inteligente de dados de diversas fontes, a fim de produzir uma informação e disponibilizá-la para um usuário, o objetivo desse procedimento é gerar informações coerentes às necessidades de cada organização e torná-las acessíveis (TIBCO, 2022).

Um dos maiores desafios para as organizações é a extração de conhecimento dos dados que descrevem o ambiente em que operam, e a cada dia esses dados aumentam em quantidade, vindo das mais variadas fontes e formatos. É necessário que cada empresa possibilite uma forma para que seus funcionários e clientes capturem valor desses dados. Isso significa que as organizações devem ser capazes de reunir dados relevantes onde quer que estejam, para fins de suporte aos relatórios da organização e aos processos de negócios (FMS2, 2022).

Na atualidade existem alguns tipos de tecnologias de integração de dados como a Integração de Informações Corporativas (EII) e Replicação de Dados Corporativos (EDR), a mais popular

entre elas Extrair, Transformar, Carregar (ETL). Essa tecnologia consiste na extração de dados de uma ou mais fontes de origem, após isso esses dados passam por transformações, nesta parte são aplicadas regras de negócio e transformações de tipo de dados, em seguida eles são carregados em outra fonte de dados, onde será permitido seu consumo pelos clientes da empresa (ASTERA, 2022; SAS, 2022).

2.3 BANCO DE DADOS

Muito antes da invenção dos computadores, a humanidade já tinha o hábito de armazenar informações e deixar registros contendo informações relevantes aos seus semelhantes. Com o passar das décadas isso deixou de ser um hábito e passou a ser uma necessidade, tanto social como empresarial. A criação e o uso efetivo dos computadores pelas organizações juntamente com o crescimento da capacidade de armazenamento, levou ao desenvolvimento de dois modelos de dados, em rede e hierárquico, criados por Charles Bachman em 1964 e só então em 1970, Edgar Frank Codd propôs o modelo de dados relacional, surgindo a ideia do banco de dados (DEVMEDIA, 2022).

Na ciência computacional, banco de dados é definido como uma coleção estruturada de dados. Os dados armazenados em um banco são organização de forma a permitir agilidade da busca e na recuperação por um computador, ou seja, não há nada além de uma simples coleção de itens, existem diversas formas de banco de dados como o relacional, hierárquico e orientado a objetos, cada um tendo suas particularidades e situações em que são mais adequados (MANOVICH, 2001).

Os bancos são gerenciados por Sistemas de Gerenciamento de Banco de Dados (SGBD) onde trabalha de acordo com o tipo de banco onde opera, mas todos eles permitem a utilização de linguagens que permitem que o usuário acesse seus dados, dois exemplos de SGBD são o SQL Server, um SGBD relacional criado pela Microsoft e o MongoDB, criado pela empresa de mesmo nome é um banco relacionado a objetos, a linguagem mais utilizada e conhecida é a *Structure Query Language*, ou simplesmente SQL, sendo nos bancos relacionais (DEVMEDIA, 2022).

2.4 FERRAMENTAS

2.4.1 SQL SERVER

O SQL Server é um SGBD desenvolvido e mantido pela Microsoft, como o nome sugere ele utiliza-se do padrão de linguagem SQL. Foi disponibilizado pela primeira vez em 1988, teve a empresa Sybase como responsável pelo seu desenvolvimento, mas posteriormente passou a pertencer a Microsoft (PORTALGSTI, 2022).

2.4.2 SQL SERVER INTEGRATION SERVICES

O *SQL Server Integration Services* ou apenas SSIS é um componente de software desenvolvido para realizar processos de integração de dados. É chamado de componente de software pois é adquirido juntamente do *SQL Server*. Também tendo a Microsoft como sua proprietária ele permite ótima integração com o *SQL Server*, possuindo um *driver* específico para se comunicar com o banco (MICROSOFT, 2023).

De acordo com o próprio site da Microsoft

“O SSIS (MicrosoftSQL Server Integration Services) é uma plataforma usada para a criação de soluções de integração de dados de alto desempenho, incluindo pacotes ETL (extração, transformação e carregamento) para data warehouse. O SSIS inclui ferramentas gráficas e assistentes para criação e depuração de pacotes; tarefas para execução de funções de fluxo de trabalho como, por exemplo, operações de transferência de arquivos – *File Transfer Protocol* - FTP, execução de instruções SQL e envio de mensagens de e-mail; fontes de dados e destinos para extração e carregamento de dados; transformações para limpeza, agregação, junção e cópia de dados; um serviço de gerenciamento, SSISDB, para administração de execução e armazenamento de pacotes; e APIs (interfaces de programação de aplicativo) para programação do modelo de objeto do *Integration Services*.” (MICROSOFT, 2023).

2.4.3 PENTAHO

O Pentaho é uma ferramenta de BI *open source* isto é, de biblioteca aberta, permitindo que qualquer pessoa possa acessar, modificar e redistribuir seu código. Ele foi desenvolvido na linguagem Java e surgiu com o desejo de ser uma ferramenta completa e gratuita de BI, hoje em dia é a ferramenta *open source* de BI mais utilizada no mundo (DEV MEDIA, 2022).

2.4.4 GOOGLE CLOUD PLATFORM

Como é descrito no site do Google Cloud Platform (GCP) “O *Google Cloud* consiste em um conjunto de recursos físicos (computadores e unidades de disco rígido) e recursos virtuais, como máquinas virtuais (VMs), localizados nos *data center* do google por todo o mundo. Cada local do data center está em uma região” (GOOGLE, 2022).

Essa plataforma fornece acesso a diversas ferramentas e funcionalidades, duas delas que foram utilizadas no projeto deste relato são o *Google Cloud Storage* (GCS) e o *Google BigQuery*, ou só *BigQuery*.

O GCS é uma ferramenta de armazenamento em nuvem, que permite o armazenamento de arquivos dos mais variados formatos e o compartilhamento dos mesmos, isso em uma infraestrutura que combina desempenho e a escalabilidade do GCP, com segurança e recursos de compartilhamento. Já o BigQuery é um serviço de armazenamento de baixo custo para análises, ele é uma ferramenta *NoOps*, ou seja, não exige uma infraestrutura pelo seu gerenciamento.

2.4.5 GOOGLE BIGQUERY

O BigQuery é um serviço de armazenamento de dados totalmente gerenciável e sem servidor que não requer uma infraestrutura responsável para gerenciar o banco de dados. Ele possui um mecanismo onde os recursos de processamento e memória não são fixos, ou seja, eles são alocados de acordo com a necessidade de cada consulta, onde esses recursos são divididos em vários *slots*, com a intenção de diminuir o tempo de execução de uma requisição feita pelo usuário, e partir daí é feito a cobrança de acordo com a quantidade de *slots* utilizados, o tempo necessário para gerar o resultado, e a quantidade de memória utilizada na requisição. Para acessar os dados armazenados, é utilizada a linguagem SQL, sendo possível consultar *terabytes* em segundos, isso é possível devido ao seu mecanismo de análise distribuída e escalonável (GOOGLE, 2023).

Ferramentas	Descrição	Função
SQL Server	Ferramenta de banco de dados relacional mantida pela Microsoft.	Gravar dados de produção dentro da

		infraestrutura da empresa.
SSIS	Ferramenta de integração de dados, é adquirida junto a licença do SQL Server, sendo uma ferramenta feita com objetivo de ter melhor integração com este banco.	Ler e gravar os dados que estão na infraestrutura da empresa.
Pentaho	Ferramenta de integração de dados de código aberto.	Migrar dados para serviço de nuvem.
BigQuery	Serviço de análise e armazenamento de <i>terabytes</i> do GCP.	Utilizado para processar os dados.
GCS	Serviço de armazenamento de arquivos do GCP.	Utilizado para migrar dados para o BigQuery

No quadro acima é feito um compilado das ferramentas utilizadas no projeto fazendo uma breve descrição de cada uma delas e qual seu objetivo no projeto relatado.

3 SITUAÇÃO PROBLEMA

A organização em questão é um Contact Center com sede em Uberlândia, Minas Gerais. Prestando serviços de relacionamento com clientes, vendas, *back office* e cobrança para todo o Brasil. A empresa fundada em 2006 é um dos principais *players* do mercado em seu setor de atuação.

Devido a área de atuação da organização, os dados são os principais insumos para realização de seus serviços e, com o desenvolvimento da empresa, o volume de dados também teve crescimento, fazendo com que a arquitetura dos processos de integração de dados adotadas de

antemão, não conseguissem suportar a quantidade de dados que devem processar em tempo hábil, necessitando de melhorias.

Como parte dos contratos entre a empresa e seus clientes, preveem gerar relatórios com os resultados de seus serviços prestados em determinada frequência, a falta deles ocasiona em perda de confiança dos clientes e em multas para a organização, fazendo com que a área de BI, responsável pela geração dos relatórios, seja ofensora do crescimento empresarial, devido ao aumento de despesas desnecessárias e também diminuindo a confiança dos clientes.

Os processos de BI são baseados no método *Extract, Transform e Load* (ETL), onde os dados são extraídos de diversas fontes, sofrem transformações que os possibilitem ser analisados e em seguida são carregados em um sistema que disponibilize-os para os usuários. No processo utilizado para este relato, o principal ofensor é a operação ROW_NUMBER do SQL Server, que duplica o tempo de execução do ETL, necessária para criação de diversas visões nos relatórios solicitados pelos clientes, esta função retorna uma enumeração de linhas de acordo com uma partição e uma ordenação definida pelo usuário, operação esta que exige muito recurso de memória e processamento do banco de dados.

Assim, o relato aqui apresentado tem como situação problema a dificuldade da empresa de Contact Center em processar e disponibilizar os dados das diversas operações dos diversos clientes em tempo hábil e a um custo razoável, para que eles possam gerar os seus relatórios e produzir as suas informações para suportar o processo de tomada de decisão. Este fato é importante pois a área de BI não é uma área de geração de receita diretamente para a empresa, sendo uma área meio e não uma área fim. Portanto a busca pela melhor relação custo-benefício entre preço e performance dos processos de BI se tornou essencial.

4 INTERVENÇÃO ADOTADA

Mediante a situação problema apresentada, a intervenção proposta foi a utilização do serviço de banco de dados em nuvem do Google Cloud Platform (GCP) o BigQuery, levando em consideração que os dados após serem processados devem ser disponibilizados para consultas em variados relatórios, com a premissa de que cada relatório conectado ao banco de dados possa ser atualizado diversas vezes ao dia gerando uma requisição e, conseqüentemente, uma cobrança.

Assim, foi decidido que o BigQuery seja utilizado apenas como um processador, para assim manter o controle de custos.

A organização possui toda sua infraestrutura de banco de dados baseada nos serviços da Microsoft, utilizando o SQL Server como sua ferramenta de gerenciamento de banco de dados, então os dados de produção teriam que sair de um banco SQL Server, migrar para o BIGQuery e em seguida retornar para os bancos de dados internos da empresa, para que os relatórios acessem os dados e sejam disponibilizados para os clientes.

Para a migração dos dados para o serviço em nuvem, foram cotados 3 possíveis ferramentas sendo elas, SSIS, Pentaho e a biblioteca Pandas da linguagem Python, cada uma com suas vantagens e desvantagens. A primeira opção foi a utilização do Python, que possui a biblioteca Pandas, que permite a leitura dos dados no SQL Server e a gravação deles diretamente no BigQuery, possuindo duas desvantagens, sendo que a primeira delas é a conexão com o banco de dados da empresa não ser nativa, ou seja, o *driver* utilizado na conexão é um genérico não tendo uma boa performance e a segunda desvantagem é o uso de memória por parte da própria linguagem Python, podendo impactar em outros processos que são executados no mesmo servidor. Outra opção seria utilizar Integration Services, uma ferramenta de ETL desenvolvida pela Microsoft, por isso possui boa integração com o banco SQL Server, porém para levar os dados para o BIGQuery primeiro é necessário gerar um arquivo no formato Comma Separated Value (CSV) e armazená-lo no GCS, para depois ser lido e carregado pelo BigQuery. Por fim, a última solução cotada foi o uso do Pentaho, outra ferramenta de ETL que, assim como o Python não possui integração nativa com o SQL Server, porém tem a capacidade de gerar nativamente arquivos com extensão AVRO, que são arquivos com dados serializados no formato binário compacto e esquema no formato JSON, utilizando menos memória que os arquivos CSV, utilizados pelo SSIS, sendo que este último também possui a capacidade de gerar arquivos AVRO, porém como não é nativa seria necessário expandir a licença do SSIS utilizada.

Com as vantagens e desvantagens apresentadas, foi decidido utilizar o SSIS e o Pentaho visando aproveitar as vantagens de cada uma das ferramentas. O processo arquitetado para migração dos dados para a nuvem foi pensado de forma que seja performático e barato, portanto,

visando a performance, a leitura dos dados de produção foi feita pelo SSIS e o carregamento no BigQuery foi feito pelo Pentaho.

A arquitetura da solução adotada para o referido problema é apresentada na Figura 1. Em seguida, são descritas as partes da arquitetura e as funcionalidades de cada uma das ferramentas utilizadas.

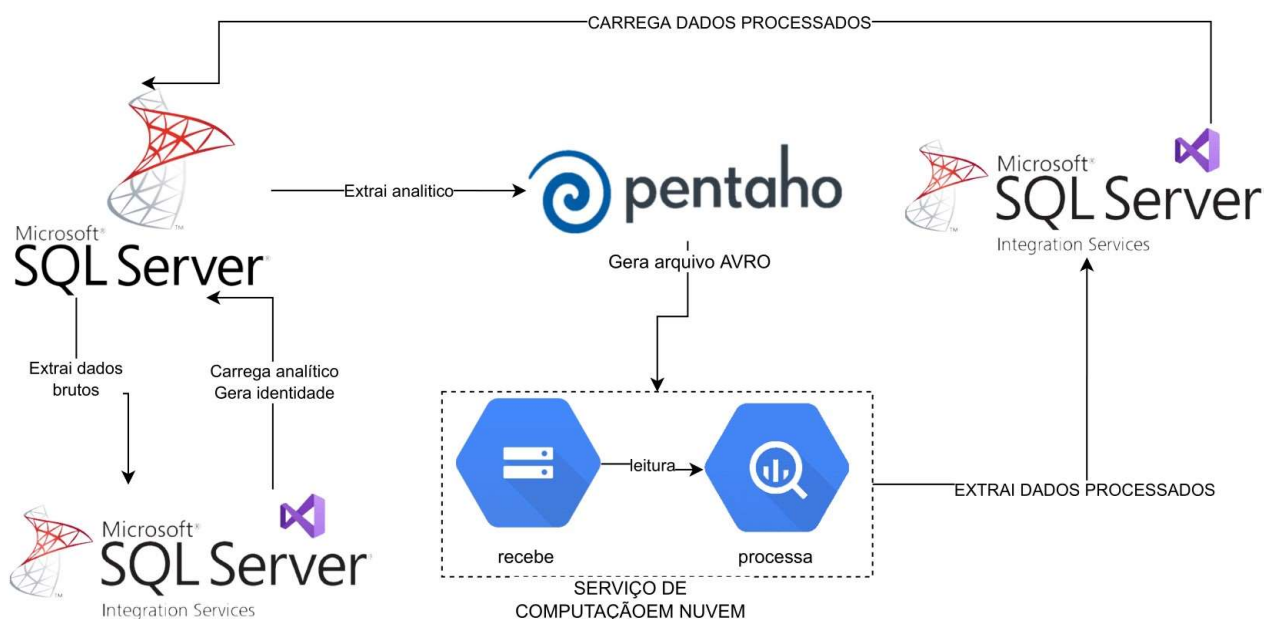


Figura 1 – Esquema da Solução Adotada
Fonte: Autoria própria

A primeira parte do processo consiste em um pacote de ETL feito no SSIS, nele são executadas 4 rotinas (*procedures*) que extraem os dados de produção do dia vigente, gerando dados analíticos que são inseridos em uma tabela contendo 256 colunas, sendo que uma delas é uma coluna incremental utilizada como identidade e outras 16 colunas que nesta etapa contém o valor NULL e a posteriori serão geradas pela função ROW_NUMBER do SQL no BigQuery, essas colunas são utilizadas para criar as tabelas fato com as visões de relatório solicitadas pelo cliente. Vale ressaltar que o processo é executado ao final do dia quando a tabela não será mais utilizada e também não haverá mais dados de produção, por isso é possível carregar os dados diretamente na tabela analítica e com os valores NULL. Após a geração e inserção dos dados na última tarefa do ETL é chamado o procedimento armazenado “sp_start_job” nativo do SQL Server que instrui o

SQL Server Agent a iniciar um trabalho imediatamente, no caso o trabalho executa um comando no CMD do servidor que executa uma tarefa do Pentaho. A seguir é apresentada Figura 2 do ETL que realiza esta etapa.



Figura 2: Esquema do Job ETL Utilizado

Fonte: Autoria própria

Na tarefa apresentada na Figura 3, é apresentada um *job* ETL construído no Pentaho Data Integration. Nele existe duas transformações principais que permitem a leitura dos dados analíticos necessários e a gravação deles no GCS, sendo a terceira transformação utilizada para chamar outro *job* ETL do SSIS, também por meio do procedimento “sp_start_job”.

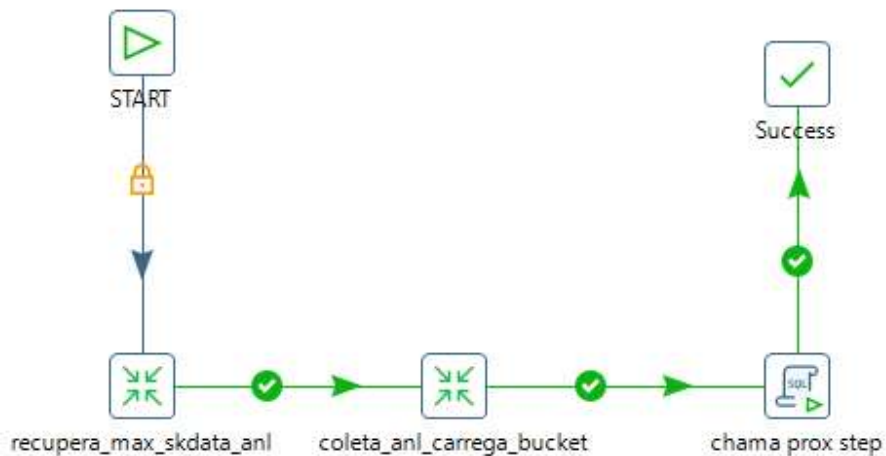


Figura 3: Job Principal

Fonte: Autoria Própria

A transformação “recupera_max_skdata_anl” é apresentada abaixo na Figura 4, nela o dia vigente é recuperado e armazenado em uma variável, que será utilizada na próxima transformação.



Figura 4: Transformação recupera_max_skdata_anl

Fonte: Autoria própria

Em seguida na transformação “coleta_anl_carrega_bucket” a variável contendo o dia vigente é recuperada e a partir dela são lidas as colunas da tabela analítica que são utilizadas para fazer o ROW_NUMBER onde o dia for o mesmo da variável e, depois são gravadas em um arquivo AVRO dentro do GCS, com isso somente 18 das 256 colunas são levadas para o serviço de nuvem, diminuindo o volume de dados, conseqüentemente, o custo. A Figura 5 demonstra a transformação citada.



Figura 5: Transformação coleta_anl_carrega_bucket
Fonte: Autoria própria

Assim com os dados já no GCS, a última tarefa do Pentaho chama outra Job com um procedimento ETL feito no SSIS, representada na Figura 6. Ele é dividido em três partes, fazendo toda a inserção, manipulação e extração dos dados que estão na nuvem e por último, a atualização no SQL Server das informações que foram calculadas. Na primeira parte do processo, uma tabela no BigQuery é esvaziada (truncada) e em seguida carregada com os dados do arquivo AVRO contendo as informações do dia vigente. Na segunda divisão do processo ocorre a execução da query contendo os cálculos utilizados para construir os relatórios, onde está a função ROW_NUMBER, ao final da query o resultado obtido por ela é gravado em outra tabela também situada no BigQuery. Vale ressaltar que essas duas etapas executam processos dentro do serviço em nuvem, possibilitando um melhor monitoramento por meio dos logs gerados pelo SSIS. Por fim, na última parte do Job, uma tabela do SQL Server é truncada e depois carregada com os dados calculados na etapa anterior e além dos dados calculados, também é gravado a coluna identidade da tabela analítico que, por meio dela, as colunas que anteriormente foram carregadas na tabela analítico com o valor NULL sofrem uma atualização, tendo agora o valor que possibilite a criação das tabelas fato nas visões necessárias.

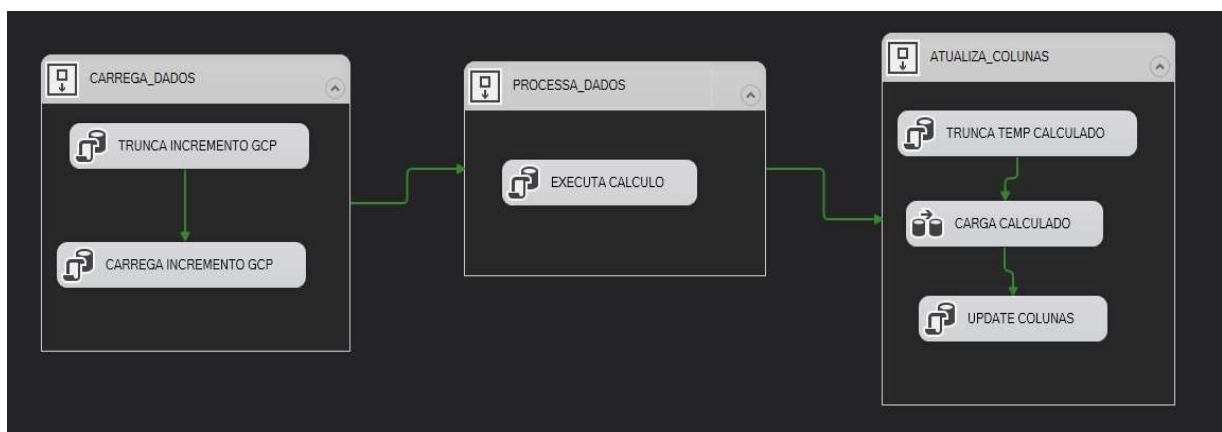


Figura 6: Transformação Final

Fonte: Autoria própria

Sugiro acrescentar um parágrafo conclusivo do item

5 RESULTADOS OBTIDOS

A arquitetura de processamento de dados utilizando a *cloud computing* possuía o objetivo de diminuição do tempo de execução do processamento dos dados analíticos e a restrição de não aumentar os custos já existentes. Tendo o objetivo sendo atingido e restrição em relação aos custos cumprida o projeto relatado cumpriu seus objetivos.

O processo realizado anteriormente era totalmente hospedado em uma máquina Windows Server de um servidor da própria empresa, esta máquina possui licença de SQL Server para dois pares de processadores, onde cada uma tem o valor de R\$ 900 mensais e 72 gigabytes de memória RAM que não gera custo recorrente para a empresa, a licença do Windows Server custa em média R\$ 173 por mês, tendo variação devido a cotação do valor ser em dólar. Qualquer máquina criada no servidor da empresa gera um custo de infraestrutura de R\$ 600, assim o processo como um todo gera um custo médio de R\$ 2.573 para a empresa. Com a adoção do processamento em nuvem, a licença do SQL Server em outra máquina já existente, teria acréscimo de um par de processadores, logo, essa máquina que possuía um custo médio de R\$ 4.373, passando a ter custo médio de R\$ 5.273. O processamento dos dados utilizando o BigQuery teve custo médio de R\$ 90,00 podendo variar com a cotação do dólar. Com objetivo de evitar custos não planejados foi feita uma projeção de custos, para isso foi utilizado um dia atípico onde ocorreu um pico de volume de dados, esse volume foi considerado para todos os dias do mês e em seguida foi feito o cálculo do ROW_NUMBER feito no processo, nesse estudo o valor mensal chegou em R\$ 200, considerando esse valor o custo médio total mensal do processo foi de R\$ 1.100, tendo uma economia de R\$ 1.473 mensais. Vale deixar claro que o preço considerado para a conversão dos valores em dólar foi de R\$ 5,20, valor cotado no dia do estudo.

Em relação a performance do processo, anteriormente ocorria muita variação no tempo de execução variando entre 3 e 9 horas, gerando *locks* e lentidão no servidor. Com a adoção da arquitetura utilizando os serviços do GCP, o tempo de execução variou entre 2 horas e 4 horas,

sendo que o maior tempo de execução foi registrado em um teste onde o volume de dados a serem processados foi aumentado, chegando a quase dobrar a quantidade de dados do mês.

Com a melhora de performance do processo além do ganho monetário direto com a economia média de R\$ 1.473 mensais, foi possível notar outros ganhos, principalmente competitivos. Com a diminuição de atrasos, ocorreu também um menor número de reclamações dos clientes, tanto internos como externos, insatisfeitos com os serviços da área e, conseqüentemente, menos multas pagas e também um ganho competitivo, já que a empresa passa a ser reconhecida pela qualidade de seus serviços prestados, aumentando a chance de a empresa ser escolhida por outros possíveis clientes. Outro ganho foi com a alocação de recurso humano, tendo em vista que era dispendido o tempo de vários funcionários para garantir a entrega dos relatórios no prazo acordado, agora sendo possível que estes foquem seus esforços em outras atividades.

Alguns pontos de atenção levantados no projeto foram em relação a dificuldade de manutenção no mesmo, já que o processo está dividido em várias partes e em vários ambientes, o conhecimento necessário é maior e a execução de reprocesso se torna mais trabalhosa.

6 CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi relatar a criação de uma arquitetura e migração de um processo para serviços em nuvem, a fim de melhorar a performance do processo e diminuir seus custos.

É possível dizer que os objetivos foram alcançados, já que esta forma de processar os dados foi implementada, não só na operação citada neste relato, mas assim como em outras que a empresa presta serviços.

Como dito na introdução do trabalho, este pesquisador compunha a equipe de Business Intelligence da empresa de Call Center, e teve participação em todas as etapas do desenvolvimento do projeto. Isso trouxe grande valor, de forma que foi possível aprofundar os conhecimentos práticos sobre a área de dados visto em teoria no curso Gestão da Informação.

Os resultados obtidos superaram as expectativas da empresa, visto que a confiança e satisfação de seus clientes aumentaram e seus custos foram reduzidos. O próximo passo é a adequação desta arquitetura para outras operações de forma que atenda aos requisitos de um dos clientes.

7 REFERÊNCIAS

ASTERA, **Técnicas e tecnologias comuns de integração de dados explicadas**. Disponível em: <<https://www.astera.com/pt/tipo/blog/técnicas-de-integração-de-dados/>>. Acesso em: 11 dez 2022.

Carissimi, Alexandre. (2015). **Desmistificando a Computação em Nuvem**. 2022. Disponível em: <https://www.researchgate.net/publication/301298378_Desmistificando_a_Computacao_em_Nuvem>. Acesso em: 10 dez 2022.

CentralServer, **Computação em Nuvem: entenda por que é essencial para as empresas?**. Disponível em: <<https://blog.centralserver.com.br/computacao-em-nuvem-entenda-por-que-ja-e-uma-realidade-nas-empresas/#:~:text=Aumento%20da%20competitividade&text=Mais%20agilidade%2C%20flexibilidade%2C%20praticidade%20e,recurso%20essencial%20para%20o%20sucesso.>>. Acesso em 29 jan 2023.

CentralServer, **Computação em Nuvem: por que é uma ótima opção para seu negócio?**. Disponível em: <<https://blog.centralserver.com.br/computacao-em-nuvem-e-uma-otima-opcao/>>. Acesso em 29 jan 2023.

DEVMEDIA, **A História dos Banco de Dados**. Disponível em: <<https://www.devmedia.com.br/a-historia-dos-banco-de-dados/1678>>. Acesso em 10 dez 2022.

DEVMEDIA, **Conceitos Fundamentais de Banco de Dados**. Disponível em: <<https://www.devmedia.com.br/conceitos-fundamentais-de-banco-de-dados/1649>>. Acesso em: 12 dez 2022.

DEVMEDIA, **Pentaho BI**. Disponível em: <<https://www.devmedia.com.br/pentaho-bi-conhecendo-a-ferramenta-open-source-bi/32095>>. Acesso em: 4 dez 2022.

FM2S, **Integração de dados: O que é? Por que ter?**. Disponível em: <<https://www.fm2s.com.br/blog/o-que-e-integracao-de-dados>>. Acesso em: 29 jan 2023.

GOOGLE, **Visão geral do Google Cloud**. Disponível em: <<https://cloud.google.com/docs/overview?hl=pt-br>> Acesso em: 4 dez 2022.

GOOGLE. **BigQuery**. 2022. Disponível em: <<https://support.google.com/cloud/answer/9113366?hl=pt-BR>>. Acesso em: 22 jan 2023.

IMASTERS. **UMA BREVE HISTÓRIA DA CLOUD COMPUTING**. 2022. Disponível em: <<https://imasters.com.br/cloud/uma-breve-historia-da-cloud-computing>>. Acesso em: 10 dez 2022.

MANOVICH, Lev. The Database. In: **The Language of New Media**. Massachusetts: MIT Press, 2001.

MICROSOFT, **SSIS**. Disponível em: <<https://learn.microsoft.com/pt-br/sql/integration-services/ssis-how-to-create-an-etl-package?view=sql-server-ver16>>. Acesso em: 10 dez 2022.

NETRIN, **Uso de dados nas empresas: como isso se tornou valioso e relevante para a tomada de decisão nos negócios**. Disponível em: <<https://netrin.com.br/como-o-uso-de-dados-se-tornou-valioso-e-relevante-para-as-empresas/>>. Acesso em: 4 dez 2022.

O banco de dados (Tha Database, Tradução de Camila Vieira). In: **A Linguagem da Nova Mídia** (2001). Revista Eco Pós. ISSN 2175-8689. Arte, Tecnologia e Mediação, V.18, N.1, ano de 2015. Universidade Federal do Rio de Janeiro. Disponível em: <https://edisciplinas.usp.br/pluginfile.php/4403037/mod_resource/content/1/O%20banco%20de%20dados.pdf>. Acesso em: 10 dez 2022.

ORACLE, **O que é um Banco de Dados**. Disponível em: <<https://www.oracle.com/br/database/what-is-database/>>. Acesso em: 10 dez 2022.

PORTALGSTI, **O que é SQL Server?**. Disponível em: <<https://www.portalgsti.com.br/sql-server/sobre/>>. Acesso em: 12 dez 2022.

RITTINGHOUSE, J. W., RANSOME, F. J. **Cloud Computing: Implementation, Management and Security**. CRC PRESS, 2009.

SAS. (2015), **ETL**. Disponível em: <https://www.sas.com/pt_br/insights/data-management/o-que-e-etl.html#:~:text=ETL%20é%20um%20tipo%20de,para%20construir%20um%20data%20warehouse.>>. Acesso em: 11 dez 2022.

SILBERSCHATZ, Abraham, KORTH, Henry F. e SUDARSHAN, S. **Sistema de Banco de Dados**. Editora Campus. 5a Edição, 2006.

TAURION, Cezar. **Cloud computing: computação em nuvem: transformando o mundo da tecnologia da informação**. Rio de Janeiro: Brasport, 2009.

TIBCO, **O que é integração de dados?**. Disponível em: <<https://www.tibco.com/pt-br/reference-center/what-is-data-integration>>. Acesso em: 10 dez 2022.