



Leonardo Delgado de Oliveira
Agnaldo do Carmo Carcereri Souza
Gabriel Oliveira Costa
Igor Dias Ferreira

FOLHA DE APROVAÇÃO

**MODELAGEM DE UM DATAWAREHOUSE SOBRE DADOS DE PREVISÃO DO
TEMPO DA AMAZÔNIA**

Após a exposição dos discentes **Leonardo Delgado de Oliveira, Agnaldo do Carmo Carcereri Souza, Gabriel Oliveira Costa, Igor Dias Ferreira**, sobre a realização do artigo, a Banca Examinadora, composta pelos professores abaixo identificados, reuniu-se e aprovou o presente artigo que, por atender aos requisitos estabelecidos, pelo curso de Bacharelado em Sistemas de Informação da Faculdade Metodista Granbery, para obtenção do título de Bacharel em Sistemas de Informação.

Juiz de Fora, 06 de Dezembro de 2022.


Ricardo Silva Campos – Orientador
Faculdade Metodista Granbery


Marco Antônio Pereira de Araújo – Examinador
Faculdade Metodista Granbery

JUIZ DE FORA
2022

Modelagem de um *Data Warehouse* sobre dados de previsão do tempo da Amazônia

Leonardo Delgado de Oliveira

Agnaldo do Carmo Carcereri Souza

Gabriel Oliveira Costa

Igor Dias Ferreira

Orientador: Ricardo Silva Campos

Resumo

A análise de dados vem se destacando cada vez mais na atualidade, principalmente no que se refere a facilitar tomadas de decisão e torná-las mais eficazes. Não basta apenas coletar e armazenar um grande volume de dados, também é preciso saber realizar o tratamento e análise, a fim de se obter métricas e relatórios necessários para o negócio, seja ele de finalidade comercial ou não.

O objetivo do projeto é a modelagem de um *Data Warehouse* (DW) no modelo dimensional do estrela, que será construído utilizando como base os dados históricos de previsão do tempo da Amazônia e durante o processo apontar alguns dos benefícios dessa abordagem em relação a tradicional em planilhas nas quais os dados são originalmente disponibilizados pelo Ministério da Defesa.

O projeto visa não só a modelagem de um DW e exemplificação de métricas, mas também busca fomentar pesquisas futuras e o uso dessa tecnologia, disponibilizando um modelo adequado para análise de dados relacionados a cenários climáticos do bioma, uma vez que a Amazônia detém grande relevância no cenário mundial, especialmente no que tange a ecologia e clima. A construção do DW será realizada utilizando dados disponibilizados pelo Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM), disponibilizados entre Junho de 2018 e Agosto de 2022.

Palavras-Chave: *Data Warehouse*, Previsão do Tempo da Amazônia, Modelo Dimensional Estrela, Análise de Dados.

Abstract

Data analysis has been increasingly highlighted today, especially with regard to facilitating decision-making and making them more effective. It is not enough just to collect and store a large volume of data, it is also necessary to know how to process and analyze it in order to obtain metrics and reports necessary for the business, whether for commercial purposes or not.

The objective of the project is the modeling of a Data Warehouse (DW) in the dimensional model of the star, which will be built using as a base the historical data of weather forecast of the Amazon and during the process to point out some of the benefits of this approach in relation to the traditional one in spreadsheets in which the data is originally made available by the Ministério da Defesa.

The project aims not only at modeling a DW and exemplifying metrics, but also seeks to encourage future research and the use of this technology, providing an adequate model for data analysis related to climatic scenarios in the biome, since the Amazon has great relevance on the world stage, especially with regard to ecology and climate. The construction of the DW will be carried out using data provided by the Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM), made available between June 2018 and August 2022.

Keywords: Data Warehouse, Amazon Weather Forecast, Star Schema Model, Data Analysis.

1-Introdução

Em razão de uma sociedade globalizada, onde as mudanças ocorrem de maneira cada vez mais acelerada e diante do cenário dinâmico e exigente atual do mercado, a informação se tornou vital para que as empresas possam se manter de forma competitiva e sobreviver no mercado. Devido a esse aumento, as organizações necessitam de velocidade nas análises para definir suas estratégias. Neste caso, o uso do *Data Warehouse* (DW), utilizado por Sistemas de Processamento Analítico Online (OLAP), sendo um componente da área de inteligência de negócios, pode otimizar consultas com um grande número de dados que demandam muito esforço computacional (TORRES *et al.*, 2011).

Existem cenários em que esses recursos são utilizados para auxiliar organizações cujo foco não é comercial, como institutos de pesquisa ou instituições governamentais, as tomadas de decisão mais rápidas e eficazes fornecidas por um DW junto a um sistema OLAP possibilitam a otimização no uso de recursos (MELO, 2003).

Em vista disso, temos um caso para a necessidade de análise de um grande volume de dados, o histórico de previsão do tempo da Amazônia, que são utilizados para controle, monitoramento e comparação com outras informações, por exemplo às referentes a queimadas no mesmo período de tempo. A coleta destes dados é realizada e disponibilizada pelo Centro Gestor e Operacional do Sistema de Proteção da Amazônia (CENSIPAM) que é um órgão do Ministério da Defesa.

Sendo assim, o estudo em questão tem sua relevância quando levantamos a quantidade de dados gerados a cada mês. Logo, é fundamental tratá-los e analisá-los utilizando uma ferramenta adequada e que atue como facilitador para lidar com esses dados, como por exemplo um DW, eliminando anomalias que tabelas comuns do Excel, como as em que os dados são disponibilizados, podem deter.

Posto isto, utilizaremos os dados de previsão do tempo da Amazônia para modelar um DW que será construído adotando a abordagem de modelagem dimensional estrela. Com isso, espera-se poder facilitar futuras pesquisas e consultas em cima destes dados, com o propósito de definir uma modelagem de DW mais apropriada para ser, posteriormente, consumida por ferramentas OLAP em cenários onde se busca obter tomadas de decisão mais ágeis referentes aos dados fornecidos.

1.1 - Objetivos

1.1.1- Objetivo Geral

O intuito do projeto é a construção de um DW em modelo dimensional estrela adequado para receber os dados fornecidos pelo CENSIPAM, os quais apresentam o histórico de previsão do tempo da Amazônia Legal. Dessa forma, tendo o intuito de facilitar futuras pesquisas disponibilizando um modelo apropriado para serem feitas as análises deste grande volume de dados através do consumo de ferramentas OLAP, e fomentar o uso do DW no lugar de planilhas convencionais, dadas suas vantagens.

1.1.2- Objetivos Específicos

Este projeto tem como foco a modelagem de uma base dimensional estrela, um DW apropriado para receber posteriormente os dados referentes às previsões de tempo da Amazônia Legal, mas não se limitando apenas a eles, podendo absorver dados de outras bases que contenham informações relevantes acerca do bioma, e devido às vantagens no uso do DW em relação a planilha disponibilizada pelo órgão CENSIPAM, incentivar o uso da tecnologia, sendo uma pesquisa facilitadora para trabalhos futuros, uma vez que a Amazônia demonstra vulnerabilidade às variabilidades e mudanças do sistema climático, tornando fundamentais análises e monitoramentos dessa natureza (ROCHA *et al.*, 2012).

1.2- Justificativa

Atualmente, a inteligência de negócios e o suporte à decisão não se limitam às fronteiras dos negócios comerciais, tais estratégias estão sendo utilizadas em diversas áreas do conhecimento. Seja em instituições de ensino, governamentais, privadas com ou sem fins lucrativos, onde houver demanda de um processo decisório, sistemas desse tipo serão necessários. Nesta categoria de sistemas facilitadores a análise de dados e suporte a decisão, destaca-se o *Data Warehouse*, que possibilita a geração de dados integrados e históricos permitindo que decisões sejam tomadas com base em fatos e não em especulações e, embora já exista há décadas, ainda não há uma metodologia formal para sua implementação (MELO, 2003).

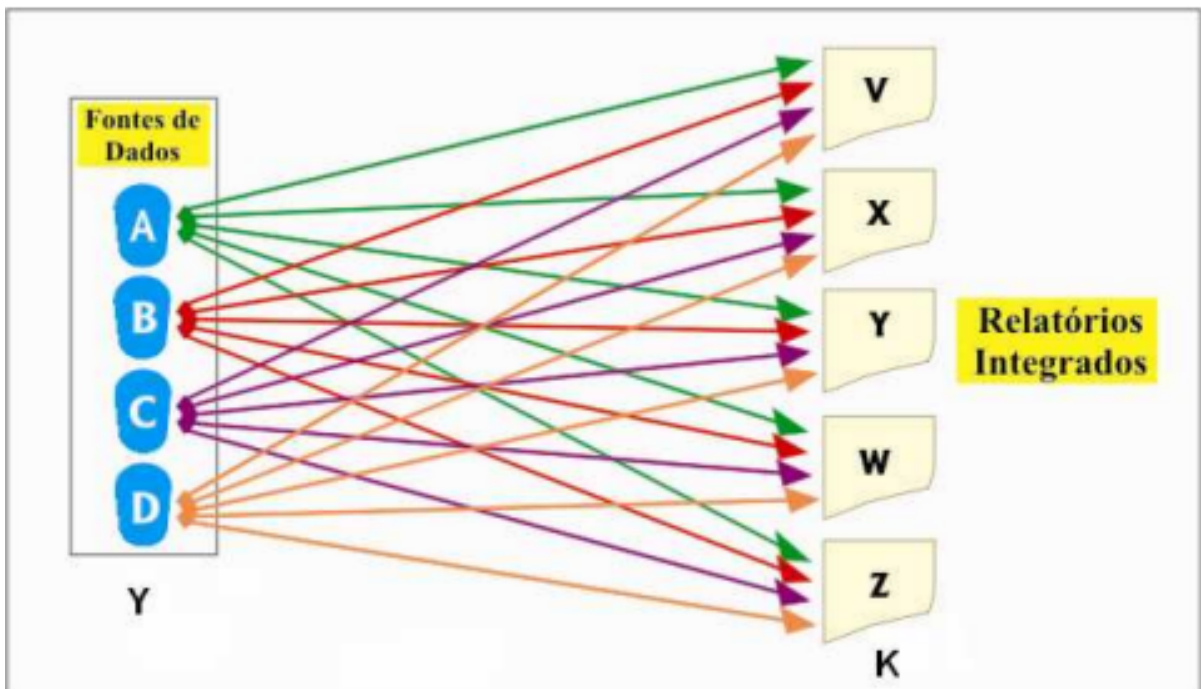
A Amazônia é a única grande extensão contínua de floresta tropical úmida do mundo, possuindo uma área de aproximadamente 7 milhões de km², desempenha um papel crucial nas trocas de energia, umidade e massa entre a superfície continental e a atmosfera, prestando assim serviços ambientais fundamentais para a manutenção do clima regional e global. Diante desse cenário, é notável que os dados fornecidos da temperatura do bioma são de grande relevância, mais especificamente no que refere a questões ambientais e ecológicas, já que esses dados podem ser indicadores diretamente ligados ao desmatamento (ROCHA *et al.*, 2012).

Dentre os benefícios de um DW podemos destacar o fácil e rápido acesso à informação, no meio empresarial essa vantagem ganha relevância, por exemplo, em rotinas de extração de relatórios, já para institutos que queiram obter resultados analíticos, ela aparece na obtenção prática de métricas necessárias para dar continuidade em suas pesquisas, inclusive economizando recursos no processo (CALDAS *et al.*, 2006).

O custo de desenvolvimento envolvido em produzir tais resultados é alto, e usualmente não podem ser reaproveitados, pois as necessidades de dados mudam constantemente, além de levar vários meses para ficar pronto. Utilizando o *Data Warehouse*, esse processo de extração de resultados se simplifica, porque os dados já se encontram integrados, ou seja, fornecem uma fonte única, pulando diretamente para a etapa de construção dos relatórios ou resultados. Dessa forma o que antes demoraria semanas ou meses passa a ser entregue em horas ou, no máximo, dias (CALDAS *et al.*, 2006).

Considerando que a base de dados de previsão do tempo da Amazônia seja uma das Y fontes de dados e K o número de resultados ou relatórios que deseja-se obter a partir destas bases, o custo total de extração convencional é de $Y \times K$, pois para cada resultado deverão ser tratadas e consultadas todas as bases, conforme apresentado na Figura 1 (CALDAS *et al.*, 2006).

Figura 1: Custo dos resultados sem um DW ($Y \times K$).



Fonte: CALDAS, 2006.

Já o custo com o DW sendo utilizado deverá cair para $Y + K$, pois os dados já estão no DW, inclusive as métricas relevantes para os resultados e relatórios permitindo o salto para a etapa de construção destes documentos, como ilustra a Figura 2 (CALDAS *et al.*, 2006).

Figura 2: Custo dos resultados com um DW ($Y + K$).



Fonte: CALDAS, 2006.

Dadas as vantagens apresentadas em utilizar o DW e a importância dos dados que serão usados como base para o estudo, o presente trabalho pretende realizar uma análise exploratória sobre o processo de modelagem de um DW e definir a construção que apresenta uma estrutura apropriada para receber em um modelo dimensional estrela capaz de receber os dados disponibilizados pelo CENSIPAM, bem como de outras bases caso seja a vontade de futuros usuários, além disso serão adicionadas ao diagrama métricas genéricas como exemplo, para que em trabalhos futuros fique exposta uma direção de como inserir estes dados de acordo com eventuais necessidades.

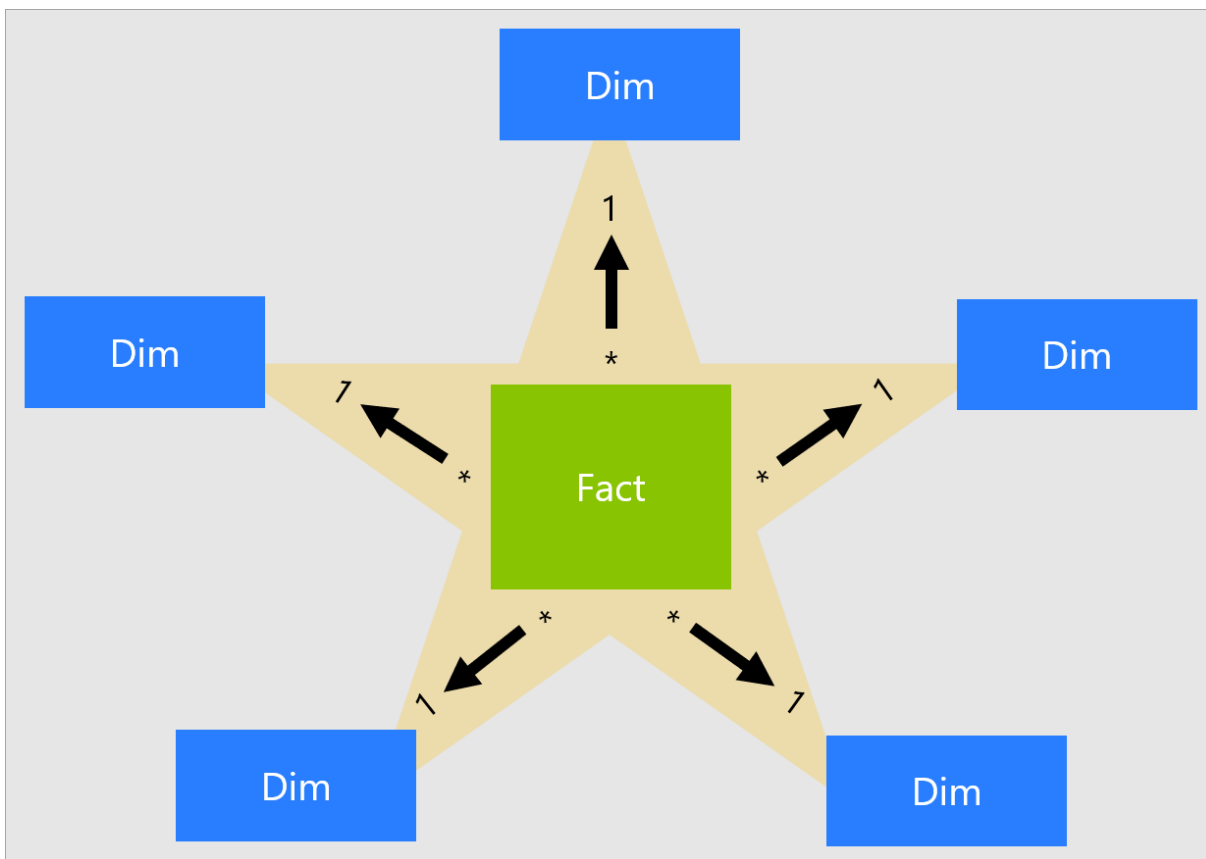
2- Referencial Teórico

Atualmente, a forma como organizamos os dados para maximizar a utilização dos mesmos, pode se tornar um problema. O gerenciamento de dados se tornou uma atividade extremamente importante dentro de uma organização, podendo facilitar não só as consultas a estes, bem como a facilitar na geração de relatórios para uma gestão estratégica. Organizações que possuem um grande volume de dados armazenados, têm cada vez mais optado por uma nova forma de armazenar e gerir esses dados, os *Data Warehouses* (RASLAN; CALAZANS, 2014).

Um *Data Warehouse*, pode ser descrito como uma coleção de dados em um ambiente orientado a assunto, integrado, não volátil e variante no tempo, com foco em auxiliar nas tomadas de decisão (INMON, 2002), tudo em uma única fonte, ainda que os dados estejam espalhados em diversos sistemas. Isso permite que as empresas identifiquem padrões e possam se posicionar estrategicamente no mercado, adaptando os meios utilizados para transformar e acessar à carga de dados, para recuperação, análise e extração. Essas operações vão depender de como os dados foram armazenados. Nos *Data Warehouses*, esses dados são armazenados em forma de cubo. Duas modelagens principais se destacam no armazenamento dimensional, sendo eles: Estrela e Floco de Neve (RASLAN; CALAZANS, 2014).

O esquema do tipo estrela possui um diagrama em que sua forma se assemelha a uma estrela, irradiando de um ponto central.

Figura 3: Modelo dimensional estrela.



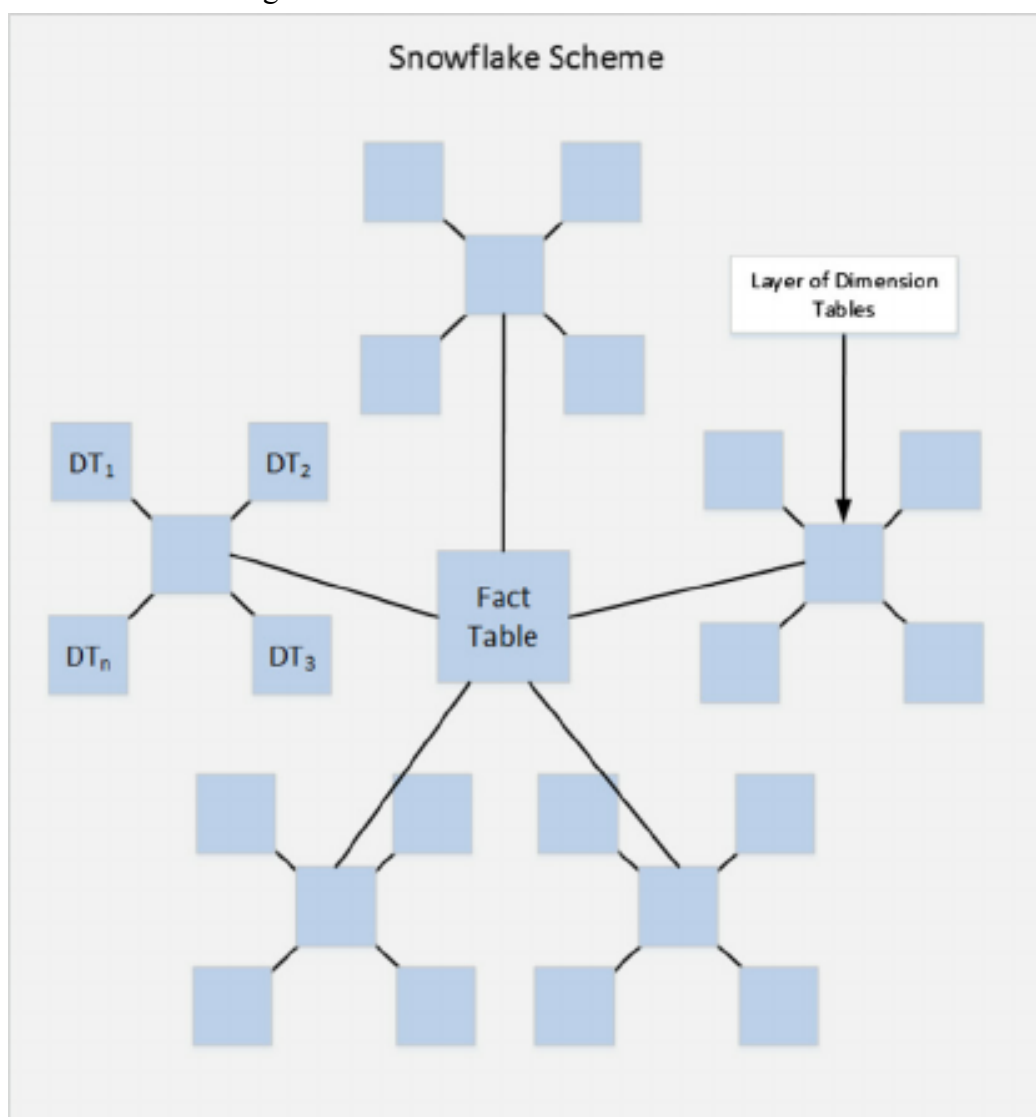
Fonte: MICROSOFT, 2022.

O centro da estrela consiste em uma grande tabela, chamada tabela fatos, enquanto as pontas da Estrela são as tabelas nomeadas dimensões. As tabelas de dimensão estão associadas à tabela de fatos, por meio de chaves estrangeiras. Essas tabelas dimensionais, não possuem relacionamento umas com as outras, estão conectadas apenas à tabela de fatos. Entre as vantagens do esquema do tipo estrela, estão o mapeamento direto entre as entidades de negócio, que são analisados por usuários finais, além do design e desempenho otimizados para consultas em bases dimensionais que utilizam o modelo estrela (RASLAN; CALAZANS, 2014).

O modelo dimensional distingue fatos de atributos, um atributo normalmente é um dado qualitativo que é conhecido previamente, já um fato é uma observação do mercado, ou uma métrica, em sua maioria os fatos são numéricos, embora alguns possam conter texto também. Os fatos na tabela fatos devem ser aditivos, por conseguinte ser a soma dos fatos ao longo de todas as dimensões, mas além destes temos também fatos semi-aditivos que são aqueles aditivos somente em relação a algumas dimensões específicas e os não aditivos como um fato textual qualitativo (BRUZAROSCO, 2000).

O modelo Floco de Neve é semelhante ao Estrela, partindo da definição de fatos e dimensões, de acordo com os requisitos de negócio, visualmente é similar a um floco de neve, com a tabela fato em sua região central e as dimensões ao seu redor, a semelhança provém porque ocorre a aplicação da terceira forma normal sobre as entidades dimensão, ou seja, este modelo é o produto da decomposição de uma ou mais dimensões. Por ser normalizado há menos redundância de valores em uma tabela, poupando armazenamento, mas em contrapartida requerendo mais conexões ao se realizar consultas, consequentemente tornando-as mais lentas (KIMBALL; ROSS, 2002).

Figura 4: Modelo dimensional floco de neve.



Fonte: Adaptado de Hughes (2015).

Após a modelagem de um DW é feito o processo de ETL (Extração, Transformação, Carregamento) dos dados, permitindo que em seguida possam ser consumidos por uma aplicação OLAP, que resolve problemas de síntese, análise e consolidação de dados, pois é o

processamento analítico *online* dos dados. Mantendo uma estrutura de dados amigável e eficiente, amplia a capacidade de visualização das informações possibilitando a consulta a partir de diferentes perspectivas, uma vez que a visualização é realizada em dados agregados, e não em dados operacionais, dado o objetivo da aplicação OLAP em apoiar os usuários finais a tomar decisões estratégicas (ANZANELLO, 2007).

Diante da necessidade apresentada do uso de *Data Warehouses*, e da importância dos dados de previsão do tempo da Amazônia, será analisada a composição dos dados fornecidos e feita a modelagem destes em um modelo dimensional estrela, possibilitando que este trabalho atue como auxiliador para estudos futuros sobre condições climáticas e ambientais sobre o bioma, uma vez que o DW fornece uma maior eficiência na construção e consumo de indicadores capazes de facilitar nas tomadas de decisão pelas ferramentas OLAP.

3- Metodologia

Devido a natureza exploratória do projeto, foi realizada uma pesquisa em que durante o processo buscou-se atingir seus objetivos reforçando cada etapa por fundamentações bibliográficas, almejando construir um diagrama de um modelo dimensional, um DW e apresentar os passos dessa modelagem, algo semelhante a uma sequência de instruções, que pode ser seguida de acordo com a necessidade específica de cada estudo posterior, incentivando o uso dessa tecnologia dadas suas vantagens para com o modelo em que dados são disponibilizados.

Inicialmente, foi definido que os dados de previsão do tempo da Amazônia seriam a base para a construção do DW, a escolha se baseou na relevância desses dados, uma vez que o presente trabalho busca também que o modelo construído possa ser reaproveitado em estudos futuros ou até em pesquisas de áreas de conhecimento específico, temos por exemplo estudos de Rocha et al. (2012) e Melo (2003) que trabalharam com dados de natureza similar além de destacarem a relevância do bioma no cenário nacional no que tange a tópicos climáticos e ambientais.

Esse conjunto de dados pode ser obtido através do Portal Brasileiro de Dados Abertos¹, são disponibilizados pelo CENSIPAM, órgão do Ministério da Defesa, e apesar do nome de Previsão do Tempo da Amazônia, na verdade se tratam de um histórico das previsões de períodos retroativos, em sua maioria os dados são mensais, com exceção de alguns arquivos que agrupam o histórico em trimestres ou semestres, os dados utilizados neste estudo

¹ Disponível em: <<https://dados.gov.br/dataset/previsao-climatica-do-censipam>>. Acesso em: 01 de outubro de 2022

são os referentes ao período de junho de 2018 até agosto de 2022, todos disponibilizados em formato CSV.

Outros fatores determinantes na escolha da base foi o volume disponibilizado, aproximadamente 2mb de dados por mês, ainda que não serão realizadas consultas ou coletadas métricas de desempenho, este é um fator relevante quando se busca destacar as vantagens no uso de um DW, uma base com poucos dados deteriora a razão do uso da tecnologia, relatórios ou resultados podem ser rapidamente extraídos de bases pequenas, o que poderia afetar negativamente a reusabilidade do modelo, além disso, as colunas das tabelas fornecidas são facilmente agrupadas e categorizadas mesmo para quem não é especialista nas áreas do conhecimento que trabalham com estes dados, o que facilita a etapa de modelagem do DW.

Primeiramente foi realizada uma validação para confirmar que em todos os arquivos as colunas das tabelas eram coincidentes, para que não faltassem colunas no diagrama construído, evitando assim que durante a etapa de ETL, não houvesse uma coluna correspondente para fazer o carregamento direto dos dados da fonte original para o DW. Após a validação, concluímos que todos os arquivos possuem exatamente as mesmas 10 colunas, podemos ver na Tabela 1, que consiste de uma parte da planilha com os dados referentes ao mês de fevereiro do ano de 2022:

Tabela 1: Colunas e algumas linhas dos dados de outubro de 2019.

Data	Cidade	Condicao_de_Tempo_Prevista	Temperatura_Tendencia	Temperatura_Minima(°C)	Temperatura_Maxima(°C)	Vento_Velocidade_Minima(Km/h)	Vento_Velocidade_Maxima(Km/h)	Vento_Direcao	Vento_Intensidade
01/10/2019	Abaetetuba - PA	Claro a parcialmente nublado	Estavel	24	34	45	95	E - NE	Calmos/Fracos
02/10/2019	Abaetetuba - PA	Claro a nublado com pancadas de chuva isoladas entre o l Estavel		23	33	50	95	E - N	Fracos/Moderados
03/10/2019	Abaetetuba - PA	Claro a parcialmente nublado	Estavel	23	34	45	95	SE - E	Fracos/Moderados
04/10/2019	Abaetetuba - PA	Claro a nublado	Estavel	24	34	45	95	E - NE	Fracos/Moderados
05/10/2019	Abaetetuba - PA	Claro a nublado com pancadas de chuva e trovoadas a tar Estavel		24	34	50	95	E - NE	Fracos/Moderados
06/10/2019	Abaetetuba - PA	Claro a nublado com pancadas de chuva e trovoadas a tar Estavel		24	34	45	95	E - NE	Fracos/Moderados
07/10/2019	Abaetetuba - PA	Poucas nuvens a nublado com pancadas de chuva e trovo Estavel		23	33	50	95	E - N	Fracos/Moderados
08/10/2019	Abaetetuba - PA	Claro a parcialmente nublado com pancadas de chuva e tr Estavel		25	33	55	100	NE	Fracos/Moderados
09/10/2019	Abaetetuba - PA	Parcialmente nublado com possibilidade de chuva em are Estavel		26	33	55	100	NE	Fracos/Moderados
10/10/2019	Abaetetuba - PA	Parcialmente nublado com possibilidade de chuva em are Estavel		25	34	55	100	NE	Fracos/Moderados
11/10/2019	Abaetetuba - PA	Parcialmente nublado a nublado com pancadas de chuva l Estavel		25	33	50	95	NE	Fracos/Moderados
12/10/2019	Abaetetuba - PA	Claro a parcialmente nublado com pancadas de chuva e tr Estavel		25	34	45	95	NE	Fracos/Moderados
13/10/2019	Abaetetuba - PA	Parcialmente nublado a nublado com pancadas de chuva l Estavel		25	33	50	95	NE	Fracos/Moderados
15/10/2019	Abaetetuba - PA	Claro a nublado com pancadas de chuva isoladas a tarde Estavel		23	34	50	95	E - NE	Fracos/Moderados
16/10/2019	Abaetetuba - PA	Claro a nublado com pancadas de chuva isoladas a tarde Estavel		23	34	50	95	E - NE	Fracos/Moderados
17/10/2019	Abaetetuba - PA	Claro a nublado com pancadas de chuva e trovoadas isola Estavel		23	33	50	95	E - NE	Fracos/Moderados

Fonte: Adaptado do Portal Brasileiro de Dados Abertos (2022).

Como podemos ver na Tabela 1, temos as seguintes colunas:

- Data
- Cidade
- Condicao_de_Tempo_Prevista
- Temperatura_Tendencia
- Temperatura_Minima(°C)
- Temperatura_Maxima(°C)
- Vento_Velocidade_Minima(Km/h)

- Vento_Velocidade_Maxima(Km/h)
- Vento_Direcao
- Vento_Intensidade

Um DW é um repositório de dados em que uma de suas características é ter variáveis em relação ao tempo, ou seja, há um reconhecimento de que o desempenho do negócio é medido e comparado cronologicamente (BRUZAROSCO, 2000), no nosso caso podemos considerar a oscilação de temperatura média anual, por exemplo. Dessa forma, uma dimensão praticamente obrigatória em um DW seria a correlata ao tempo, no nosso caso vamos chamá-la de Data, em que dada a natureza dos dados disponibilizados o menor nível de granularidade possível é dia.

Dado que as dimensões são coleções de atributos textuais, ou seja, dados não aditivos que são altamente correlacionados entre si, podemos identificar através das colunas das tabelas mais 3 dimensões possíveis para o modelo do DW, sendo elas Temperatura, Vento e Localidade, em todas as dimensões foram adicionadas chaves primárias artificiais, com o objetivo exclusivo de recuperar o dado da dimensão (BRUZAROSCO, 2000).

Da mesma maneira em que pode se observar uma hierarquia na coluna Data encontrada nas planilhas, há também algo semelhante na coluna Cidade, dado que ela apresenta não somente a cidade em que ocorre determinado fato, mas também o estado, sendo assim, concluímos que durante a possível etapa de ETL destes dados para o DW, essa coluna seria fragmentada, onde a granularidade mais baixa seria cidade, depois seguida estado e por fim país, que mesmo não constando na tabela, os dados são referentes a Amazônia Legal, portanto o país referido será sempre o Brasil.

Em seguida, para determinar a composição da tabela fatos central do DW, a qual vamos dar o nome de Tempo, temos as chaves primárias de cada dimensão, uma vez que essa tabela é justamente a síntese do relacionamento existente entre as diversas dimensões, e naturalmente temos também fatos, observamos que as colunas:

- Temperatura_Minima(°C)
- Temperatura_Maxima(°C)
- Vento_Velocidade_Minima(Km/h)
- Vento_Velocidade_Maxima(Km/h)

Apesar de serem correlacionadas com as dimensões Temperatura e Vento respectivamente, podem ser consideradas como fatos semi-aditivos, uma vez que entregam

métricas que podem ser associadas e agrupadas a algumas das outras dimensões, exemplo a temperatura máxima no dia X na cidade Y.

Como o objetivo desse estudo é a construção de um modelo genérico acima dos dados escolhidos, e o foco é apenas nas etapas da modelagem, vamos adicionar algumas métricas escolhidas de maneira arbitrária, apenas para exemplificar o modo em que podemos definir a estrutura da tabela fatos, naturalmente usuários do modelo podem adaptá-lo às necessidades específicas do seu negócio ou estudo.

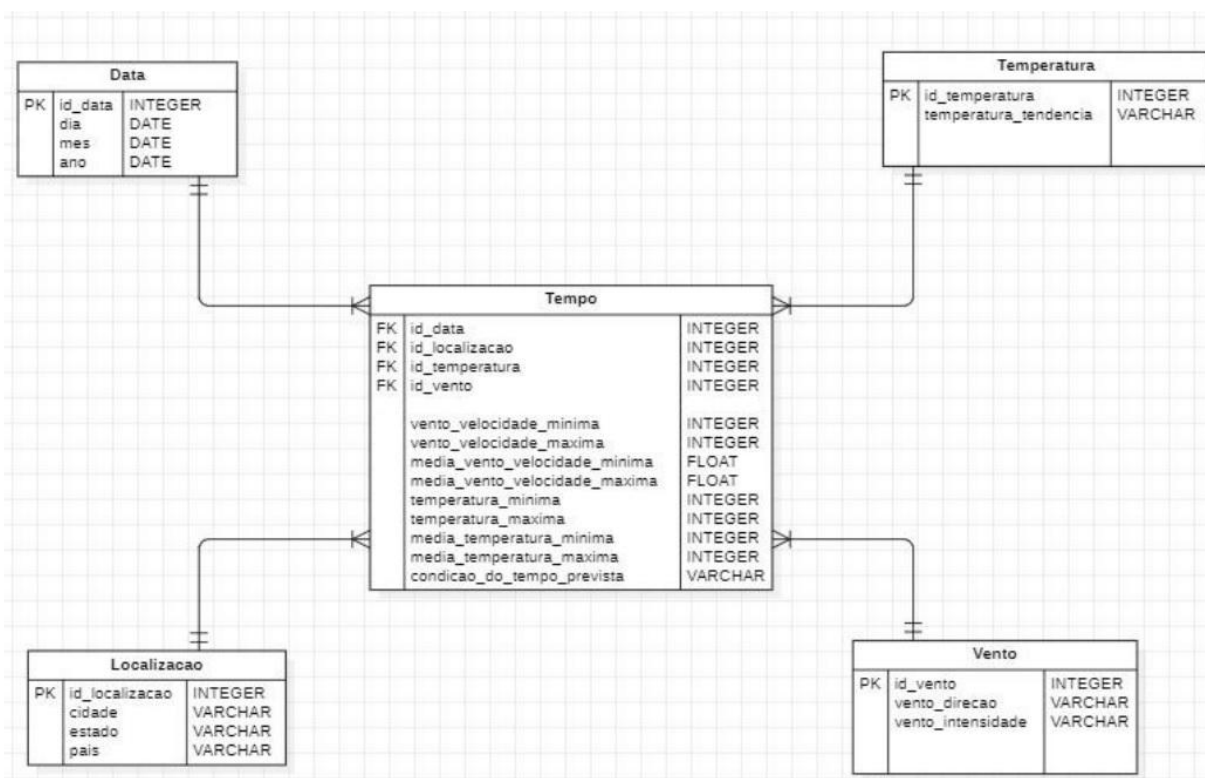
Definimos então algumas métricas como exemplo que podiam ser extraídas dessa base de dados, sendo elas a média de temperatura máxima, média de temperatura mínima, média vento velocidade máxima e média vento velocidade mínima, além destes, temos um fato aditivo textual, o campo representado pela coluna `Condicao_de_Tempo_Prevista` que é um fato aditivo pois é a conclusão resultante de todas as informações encontradas nas dimensões da estrela. Após definir esses fatos, podemos seguir com a construção do diagrama estrela e a distribuição das colunas em suas respectivas dimensões.

4- Resultados

4.1 - Modelo Estrela

Seguindo os passos descritos anteriormente e utilizando a ferramenta (StarUML), que permite a construção de diagramas, chegamos ao modelo dimensional estrela do DW construído a partir dos dados de Previsão do Tempo da Amazônia, representado pelo diagrama na Figura 5:

Figura 5: Modelo dimensional estrela do DW.



5- Considerações Finais

A partir do diagrama resultante e principalmente do processo de modelagem que foi realizado, podemos dizer que apesar do esforço inicial consideravelmente maior para a construção de um DW em relação a apenas tratar os dados no seu formato original e consumi-los, a fácil inserção de métricas e a reusabilidade do DW permite que a longo prazo haja uma economia de tempo e gasto de recursos, sua superioridade fica evidente ao observar que outras fontes poderiam popular esse DW apenas adicionando campos ou dimensões, e até mesmo reaproveitando tratativas do processo de ETL que o carregaria com os dados.

A base de dados utilizada, que detém o histórico de Previsão do Tempo da Amazônia é de grande relevância para o cenário climático mundial, a modelagem realizada neste trabalho possibilita o reaproveitamento para estudos mais aprofundados acima destes dados e pode auxiliar em novos estudos direcionados ao processo de construção de modelos dimensionais.

Para trabalhos futuros, pretendemos construir o DW em uma aplicação OLAP, utilizando o diagrama estrela modelado, e realizar o processo de ETL para popular o DW, buscando inclusive outras fontes de dados que sejam semelhantes às fornecidas pelo CENSIPAM, chegando em um ambiente que já esteja apto a prover um consumo de dados onde possam ser comparadas métricas de desempenho na velocidade de consultas em relação às planilhas fornecidas inicialmente.

Destacamos que a modelagem do DW é o meio de um processo e não o fim, porém a modelagem dimensional estrega grande responsabilidade para as aplicações de DW, sua metodologia, ou seja, seus elementos, técnicas e processo de desenvolvimento, aumentam a probabilidade de sucesso do projeto. Esse modelo gera uma interface gráfica intuitiva para a base de dados, facilitando o entendimento dos usuários, além de apoiar significativamente as tomadas de decisões uma vez que entrega métricas com maior eficácia. E no cenário atual onde informação e o tempo são recursos valiosos, a ferramenta que acelera a entrega deles pode ser uma grande vantagem para quem opta por utilizá-la.

Referências Bibliográficas

ANZANELLO, C. A. OLAP Conceitos e Utilização. **UFRGS-Instituto de Informática–Universidade Federal do Rio Grande do Sul**, Rio Grande do Sul, Porto Alegre, p. 1-7, 2007. Disponível em: <<https://www.klebermota.eti.br/wp-content/OLAP.pdf>> Acesso em 21 de agosto de 2022.

BRUZAROSCO, C. D.; CASTOLDI, V. A.; PACHECO, S. C. R. Criando data warehouse com o modelo dimensional. **Acta Scientiarum**, Florianópolis, p. 1389-1397, 2000. Disponível em: <https://www.researchgate.net/profile/Roberto-Pacheco-5/publication/277203169_Criando_data_warehouse_com_o_modelo_dimensional/links/568d1e7708aeecf87b218304/Criando-data-warehouse-com-o-modelo-dimensional.pdf> Acesso em 02/11/2022.

CALDAS, M. P. K.; SCANDELARI, L.; KOVALESKI, J. L. Aplicações sobre uma Data Warehouse no ambiente das organizações e suas vantagens. **XIII SIMPEP**, Bauru, p. 1-8, 2006. Disponível em: <https://simpep.feb.unesp.br/anais/anais_13/artigos/931.pdf> Acesso em 29 de outubro de 2022.

HUGHES, G; DOBBINS, C. The utilization of data analysis techniques in predicting student performance in massive open online courses (MOOCs). **Research and practice in technology enhanced learning**, v. 10, n. 1, p. 1-18, 2015. Disponível em: <<https://telrp.springeropen.com/articles/10.1186/s41039-015-0007-z>>. Acesso em: 24 de agosto de 2022.

INMON, W. H. Building the Data Warehouse. **Communications Of The Acm**, v. 39, n. 11, p. 49-50, 1 nov. 2002. Disponível em: <<http://dx.doi.org/10.1145/240455.240470>>. Acesso em: 10 de maio de 2022.

KIMBALL, R; ROSS, M. **The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling**. 2. ed. Canada: John Wiley And Sons, 2002. 447p.

MELO, A. H. S. **Uma estratégia de Desenvolvimento de Datawarehouse geográfico com integração híbrida aplicada ao monitoramento de queimadas na Amazônia**. 2003. 101p. Dissertação de Mestrado. Centro de tecnologia e geociências, Universidade Federal de Pernambuco, Recife. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/5569/1/arquivo7012_1.pdf>. Acesso em: 15 de outubro de 2022.

Microsoft Corporation, Understand star schema and the importance for Power BI. Disponível em <<https://docs.microsoft.com/en-us/power-bi/guidance/star-schema>>. Acesso em: 19 de agosto de 2022.

Portal Brasileiro de Dados Abertos, Previsão do Tempo da Amazônia. Disponível em: <<https://dados.gov.br/dataset/previsao-climatica-do-censipam>>. Acesso em: 01 de outubro de 2022.

RASLAN, D. A.; CALAZANS, A. T. S. Data Warehouse: conceitos e aplicações. **Universitas Gestão e TI**, Brasília, v. 4, n. 1, p. 25-37, jan/jun, 2014. Disponível em: <<https://www.publicacoesacademicas.uniceub.br/gti/article/download/2612/2400>>. Acesso em: 10 de maio de 2022.

ROCHA, M. V.; CORREIA, F. W. S.; FIALHO, E. S. A Amazônia frente às mudanças no uso da Terra e do clima global e a importância das áreas protegidas na mitigação dos impactos: um estudo de modelagem numérica da atmosfera. **ACTA Geográfica**, Boa Vista, p.31-48, 2012. Disponível em: <<http://revista.ufr.br/actageo/article/view/1092/865>>. Acesso em: 15 de outubro de 2022.

StarUML. Disponível em: <<https://staruml.io/>>.

TORRES FILHO, H. G.; CLERICUZI, A. Z.; SOUZA, K. J. S.; BIONE, B. K. A. Business Intelligence no ambiente empresarial do Banco do Brasil. **VIII Simpósio de excelência em gestão e tecnologia**, p. 1-13, 2011. Disponível em: <<https://www.aedb.br/seget/arquivos/artigos11/32514469.pdf>>. Acesso em: 22 de maio de 2022.