Pontifícia Universidade Católica do Rio de Janeiro

Coordenação Central de Extensão

Letícia Vasconcellos Mendes

**Desenvolvimento de Data Warehouse para auxiliar na avaliação da taxa de ocupação de berços nos terminais**.

Trabalho de Conclusão de Curso de Pós-Graduação

**Curso BI Master 2019-2**

Professor Orientador: Anderson Silva do Nascimento

Rio de Janeiro

julho de 2021

**Agradecimentos**

Gostaria de agradecer a todos professores do curso de BI Master da PUC pelas excelentes aulas teóricas e práticas que agregaram muito conhecimento sobre os diversos assuntos abordados. Em especial, aproveito para registrar que a disciplina de BI (Business Intelligence) foi fundamental na minha evolução profissional, visto que já apliquei vários conceitos e ferramentas apresentadas no curso no meu trabalho.

**Resumo**

O objetivo deste trabalho é viabilizar a análise do tempo de atracação das embarcações em terminais que chegam aos portos para realizar operações dos produtos que transportam.

Os dados extraídos do sistema transacional que controla o movimento de operações e estadia das embarcações precisam ser tratados de forma que essa análise possa ser feita de forma objetiva.

Nesse sentido, a criação do DataWarehouse se faz necessária pois engloba as etapas necessárias para a análise, extração, transformação e carga dos dados de forma a serem facilmente consumidos por aplicações que permitem geração de gráficos e painéis que apoiam na tomada de decisão.

**Sumário**

[*1.1 Motivação* 5](#_Toc77879837)

[*1.2 Objetivos* 6](#_Toc77879838)

[*1.3 Descrição* 6](#_Toc77879839)

[2. Conceitos e modelo de dados 7](#_Toc77879840)

[*2.1 Modelo de dados transacional* 8](#_Toc77879841)

[*2.2 Modelo de dados dimensional* 9](#_Toc77879842)

[3. Tratamento e carga dos dados 11](#_Toc77879843)

[*3.1 Modelo transacional* 11](#_Toc77879844)

[3.2 Carga modelo dimensional 12](#_Toc77879845)

[4. Resultados 16](#_Toc77879846)

[6. Conclusões 21](#_Toc77879847)

[6.1. Trabalhos futuros 22](#_Toc77879848)

[Referências Bibliográficas 23](#_Toc77879849)

1. ***Introdução***

As embarcações chegam ao terminal de um porto e atracam em um dos possíveis pontos de atracação (berços), geralmente chamados de píer, aonde realizam uma série de operações. Dentre elas, destacamos as operações de atracação e desatracação, que justamente determinam o tempo de ocupação daquele ponto de atracação (berço).

## *1.1 Motivação*

Este projeto visa automatizar o cálculo do tempo de atracação e para isso engloba o tratamento de regras que envolvem cada detalhe das operações de atracação e desatracação, considerando também como requisito principal que a análise deve ser feita por mês.

A embarcação pode atracar e desatracar mais de uma vez por mês no mesmo ponto de atracação e também pode ficar atracada por mais de um mês. Cada período deve ser contabilizado separadamente para que o tempo de ocupação possa ser calculado corretamente. Além disso, caso a desatracação ocorra em um mês posterior ao mês da atracação, deve-se calcular o tempo de atracação por cada mês, visto que a análise é sempre feita mês a mês.

A embarcação pode ser de três tipos: navios, barcaças e empurrador e rebocador, mas essa classificação não importa na análise final, sendo necessária a agregação de todos os tipos de embarcação na solução final.

Diante da complexidade relatada acima, fez-se necessário um projeto de DataWareHouse que transformasse os dados de forma a viabilizar a análise final desejada, que é o tempo que o berço fica ocupado por mês.

## *1.2 Objetivos*

Uma das maiores preocupações da empresa é saber se os pontos de atracação estão sendo subutilizados, sendo assim esse trabalho tem dois grandes objetivos.

* Criar uma estrutura de dados consistente, que seja atualizada de forma automática mensalmente.
* Desenvolver um dashboard para melhor acompanhamento dos tempos de ocupação dos berços e auxiliar nas tomadas de decisões da empresa.

## *1.3 Descrição*

O desenvolvimento da monografia se faz em algumas partes.

* Utilização do modelo transacional existente
* Elaboração do modelo multidimensional.
* Desenvolvimento do DW (Data Warehouse).
* Tratamento e transformação dos dados (ETL).
* Criação do DM (Data Mart).
* Análise de resultados (Dashboard).

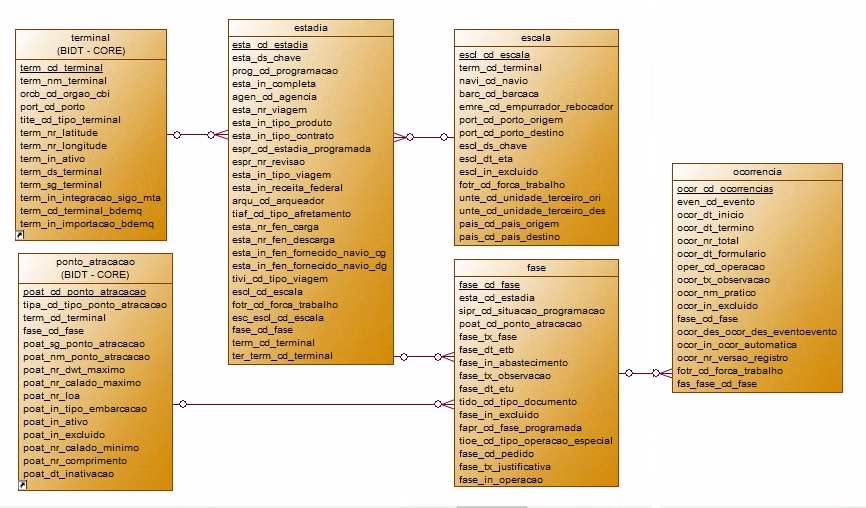
# 2. Conceitos e modelo de dados

Podemos considerar nosso projeto como um Data Mart, visto que tem um uso específico para uma área de negócio.

Ao contrário da implementação de um Datawarehouse empresarial que pode se estender por vários meses ou até anos, um Data Mart é geralmente implementado dentro de alguns meses, fornecendo suporte rápido. Isso se deve ao tamanho menor do Data Mart (menos de 100 GB) e à extração de dados de um número menor de origens.

## *2.1 Modelo de dados transacional*

A **Figura 2.1** apresenta o diagrama entidade relacionamento da base transacional.com as principais tabelas utilizadas como fonte de dados. O objetivo principal desta base de dados é controlar todas as operações realizadas durante a permanência de uma embarcação no berço, existindo para isso, seis tabelas principais. A escala começa quando a embarcação chega ao porto e pode conter várias estadias e fases, uma vez que a embarcação pode atracar e desatracar várias vezes. Observa-se que a embarcação pode ser de três tipos: navio, barcaça e empurrador- rebocador, ligados à escala. A escala sempre se refere a um terminal que possui vários pontos de atracação (berços) nos quais as embarcações realizam suas operações. As ocorrências determinam que tipo de evento ocorre em um período de tempo. No projeto em questão, trabalharemos somente com os eventos de atracação e desatracação para poder determinar o tempo de ocupação do berço num dado mês.

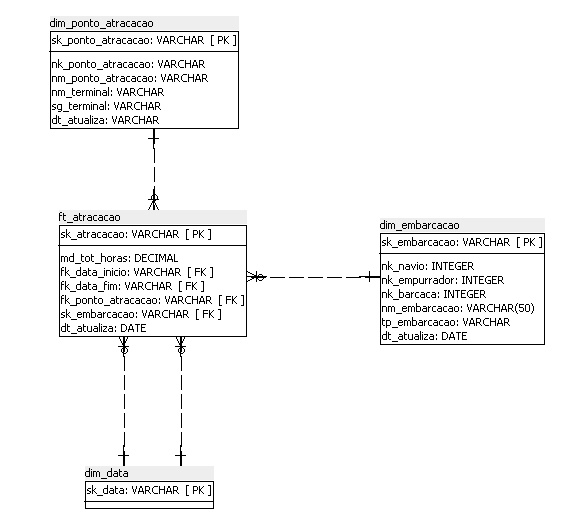


*Fig.2.1 Modelo de dados Transacional*

## *2.2 Modelo de dados dimensional*

Modelo multidimensional é uma técnica de modelagem conceitual de negócios, que facilita a investigação, o resumo e a organização de dados para a análise de negócios. Formado por relações entre dimensões e fatos.

O modelo multidimensional relaciona [tabelas de fatos](https://pt.wikipedia.org/wiki/Tabela_de_factos) com tabelas de dimensões em um [banco de dados](https://pt.wikipedia.org/wiki/Banco_de_dados) do tipo [OLAP](https://pt.wikipedia.org/wiki/OLAP), possibilitando inúmeras combinações.

Sua modelagem é realizada de forma a ganhar performance nas consultas, possuindo regras diferentes das relacionais em sua modelagem. O modelo multidimensional visa somente consultas analíticas. Neste projeto será utilizado o modelo Star Schema (modelo estrela), que consiste em uma tabela fato e várias dimensões. Na Figura 2.2, pode ser observado o modelo estrela que foi estruturado para receber os dados consolidados de tempo de atracação das embarcações.

*Figura 2.2- Modelo de dados Dimensional*

O modelo consiste em três tabelas de dimensões e uma tabela fato.

* DIM\_DATA: Como o tempo de ocupação do berço precisa ser analisado por mês, a dimensão de data possibilita a extração dos dados por mês, ano e agregações tais como trimestre.
* DIM\_EMBARCAÇÃO: Esta dimensão agrupa os três tipos de embarcação existentes, classificando-as em navios, barcaças ou empurrador-rebocador.
* DIM\_PONTO\_ATRACAÇÃO: Nesta dimensão temos os terminais e seus respectivos pontos de atracação, permitindo uma visualização final de forma hierarquizada nesses dois níveis.
* FT\_ATRACAÇÃO: O tempo de atracação de cada embarcação nos pontos de atracação é registrado na tabela fato, relacionado a uma data início de atração e fim de desatracação (período de permanência no berço)

# 

# 3. Tratamento e carga dos dados

Para realizar toda a parte de transformação de dados e carga dos dados no banco de dados dimensional, foi utilizado o software Pentaho Data Integration. As figuras abaixo foram extraídas dessa ferramenta.

O Pentaho Data Integration é o componente da suíte Pentaho usado para criar processos de extração, transformação e carga (do inglês Extraction, Transformation and Loading, ETL) que alimentam o banco de dados.

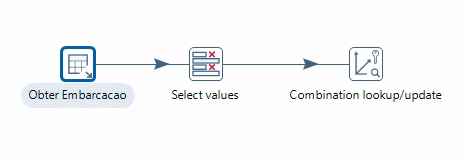
## *3.1 Modelo transacional*

Os dados foram obtidos diretamente do banco de dados transacional utilizado pelo sistema de movimentação e estadia de navios. Visando a preservação de dados confidenciais, foi utilizada a técnica de mascaramento de dados que cria uma versão semelhante aos dados originais, mas sem revelar a verdadeira informação. Esse tratamento foi feito no Power BI, mas poderia ter sido feito na etapa de carga dos dados também.

## 3.2 Carga modelo dimensional

A carga da dimensão Embarcação consistiu na junção das tabelas de navio, barcaça e empurrador\_rebocador com objetivo de consolidar as embarcações na dimensão.

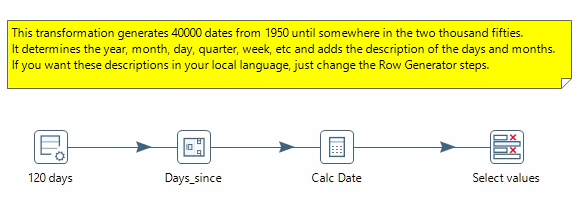
O step *“Combination lookup/update”,* usado na carga de todas as dimensões, facilita a criação de dimensões gerando as surrogate Keys de forma automatizada.

**

*Figura 3.2.1 Carga da dimensão dim\_embarcação*

A carga da dimensão Ponto de Atracação foi feita de forma análoga, buscando junto com cada ponto de embarcação o terminal ao qual ele está associado. Sendo assim, será possível extrair os dados por terminal e por ponto de atracação, de forma hierarquizada.

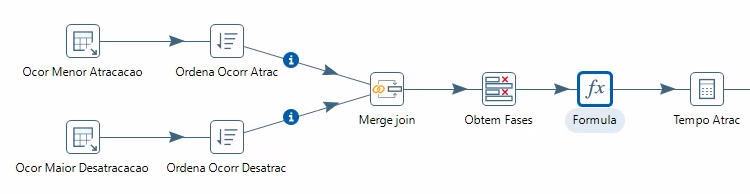
A carga da dimensão Data foi feita a partir de um sample disponibilizado pela Pentaho chamado General - Populate date dimension.ktr



*Figura 3.2.2 Carga da dimensão dim\_data*

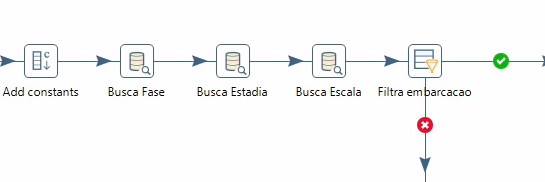
Após a carga das dimensões do modelo, a tabela fato pode ser carregada. A complexidade dessa carga foi bem maior devido aos tratamentos necessários para buscar os tempos de atracação por mês.

A busca pelo tempo de atracação consiste na obtenção da menor data do evento de atracação e maior data do tempo de desatracação, por mês. Considerando que a desatracação pode ocorrer num mês subsequente e a atracação num mês antecessor, foi necessário preencher no step *Formula* com o primeiro dia do mês e o último dia do mês, conforme o caso. Dessa forma, pode-se calcular o tempo de atracação no step *Calculator*



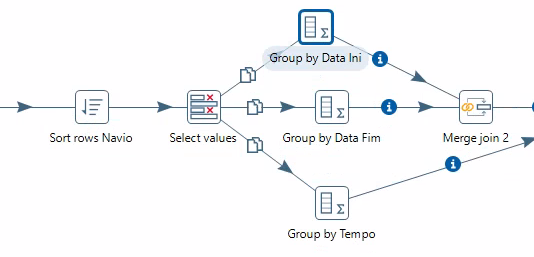
*Figura 3.2.3 Tratamento dos eventos e tempo de atracação*

Em seguida, com as ocorrências já filtradas, obtém-se a fase, estadia e escala associadas para que se possa chegar nas informações que serão associadas às tabelas de dimensões geradas no passo anterior.



*Figura 3.2.4 Busca das fases, estadia e escalada ligados às ocorrências filtradas*

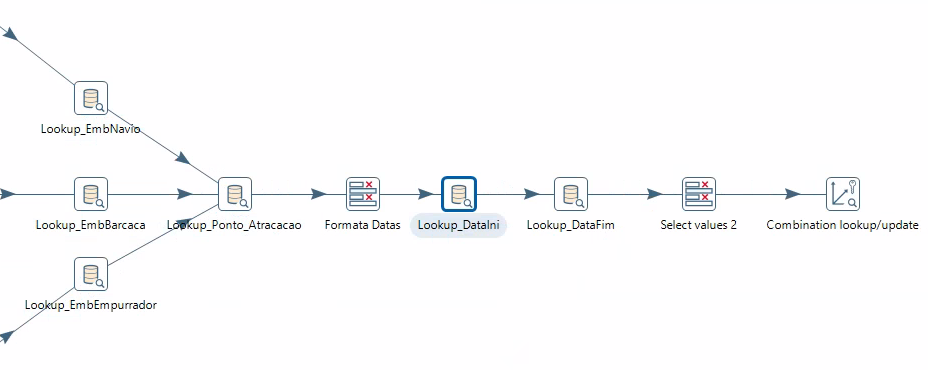
Sabendo que a mesma embarcação pode atracar e desatracar mais de uma vez no mesmo mês, foi necessário fazer um agrupamento por mês/ano, ponto de atracação e embarcação para que os tempos fossem somados e a menor data de atracação e maior data de desatracação fosse considerada. Esse mesmo tratamento foi feito para barcaça e empurrador rebocador.



*Figura 3.2.5 Soma os tempos agrupados por mês/ano, Ponto de Atracação e Embarcação*

Por fim, foram utilizados os steps *Database lookup* para obtenção das surrogate Keys de cada dimensão, com base nas chaves primárias originais (natural Keys), para inserção na tabela fato ft\_atracação.

Optou-se por duas associações entre a tabela ft\_atracação e a tabela dim\_data, uma para a data de início da atracação e outra para a data fim da desatracação.

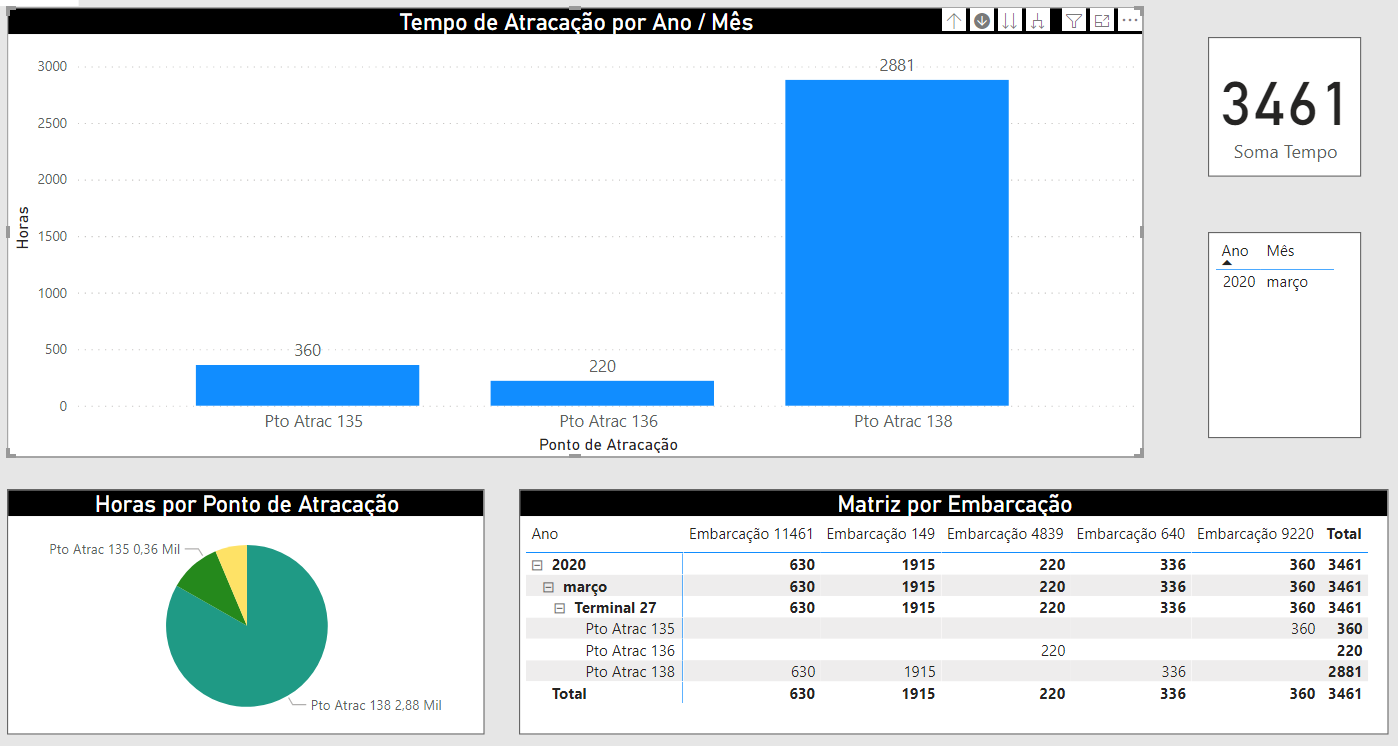


*Figura 3.2.6 Carga da tabela ft\_atracação com base nas dimensões associadas*

# 4. Resultados

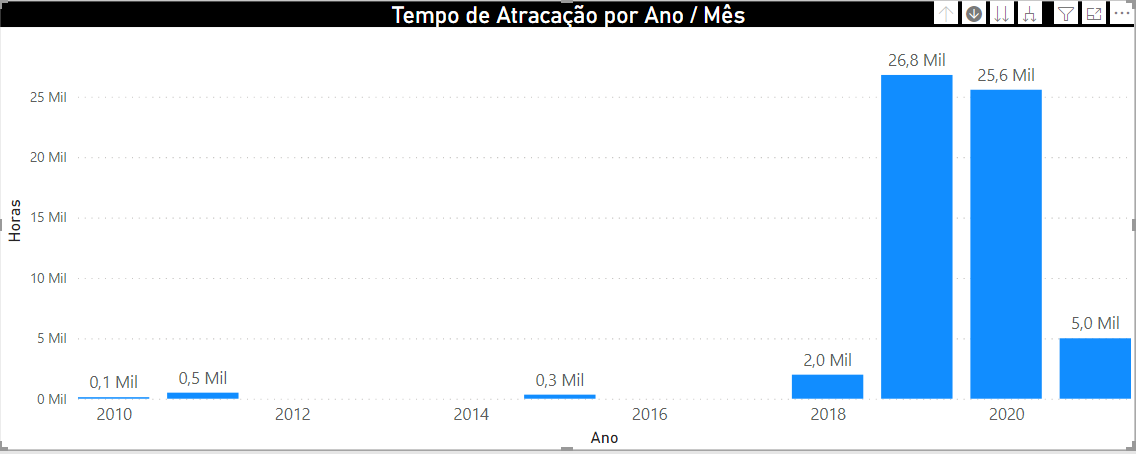
Para possibilitar uma análise mês a mês do tempo de ocupação dos berços, foi criado um dashboard no Power BI contendo os seguintes elementos visuais, explicados separadamente em seguida.

O objetivo do gráfico principal foi possibilitar uma navegação hierarquizada e os demais elementos complementam com visões definidas para Ponto de Atracação e Embarcação.



*Figura 4.1 Dashboard com a visões do tempo de atracação*

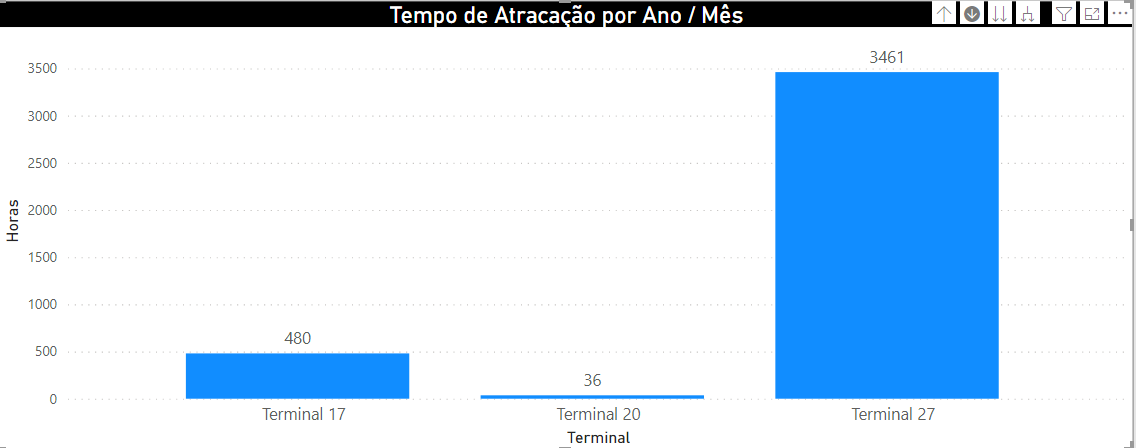
Um gráfico de colunas clusterizado que permite a consulta de forma hierarquizada podendo ser escolhido o ano, mês, terminal, ponto de atracação e embarcação, conforme figuras abaixo.



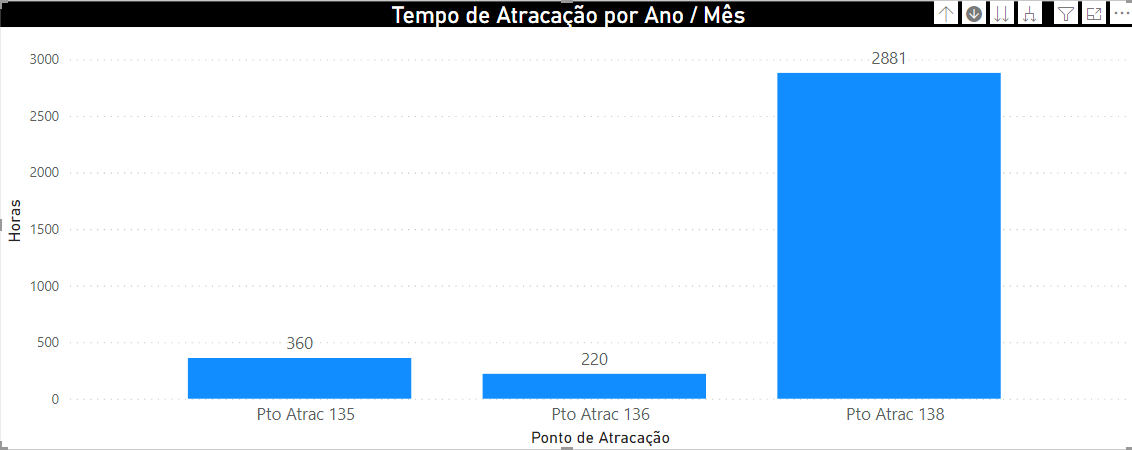
*Figura 4.2 Visão do tempo de atracação por ano*



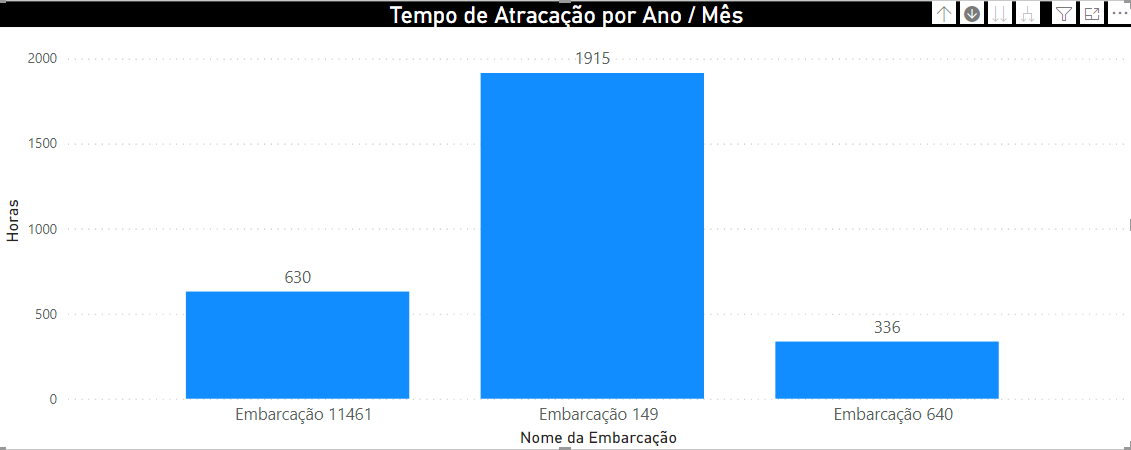
*Figura 4.3 Visão do tempo de atracação por mês*

**

*Figura 4.4 Visão do tempo de atracação por Terminal*

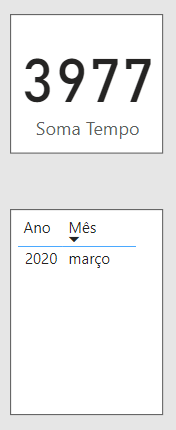
**

*Figura 4.5 Visão do tempo de atracação por Ponto de Atracação*

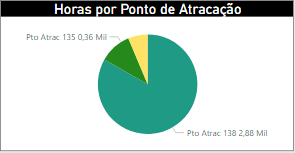


*Figura 4.6 Visão do tempo de atracação por Embarcação*

Ao lado do gráfico principal, a soma total do tempo e o ano e mês escolhidos na navegação permanecem em destaque.

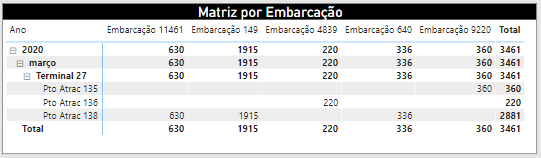


Para fornecer uma visão somente do Ponto de Atracação ao longo da navegação pela hierarquia, o gráfico de pizza abaixo foi elaborado.



*Figura 4.2 Visão Permanente do tempo de atracação por Ponto de Atracação*

Ao lado do gráfico de pizza, possibilitando uma análise mais detalhada das embarcações, foi elaborada uma matriz por Embarcação, apresentando os tempos distribuídos por Embarcação.



*Figura 4.2 Visão Permanente do tempo de atracação por Ponto de Atracação*

# 6. Conclusões

Neste projeto, a lógica implementada no processo de transformação do ETL possibilitou um tratamento mais fácil das informações armazenada no banco transacional, que tem o foco operacional. Os dados foram consolidados conforme os requisitos definidos para a análise consolidada dos tempos de atracação.

Com o uso da modelagem dimensional, o resultado final do tempo de atracação foi gerado na tabela fato de maneira a facilitar a análise dos dados pela ferramenta do PowerBI.

## 6.1. Trabalhos futuros

Esse projeto nasceu como um “*Data Mart independente”*, o que significa que foi construído sem um Datawarehouse preexistente. Esse tipo de projeto visa atender a objetivos de curto prazo mas, ao longo do tempo, podem se tornar difíceis de gerenciar.

Portanto, é importante considerar a possibilidade de usar a abordagem chamada de “[*Método de design dimensional de Kimball*](https://www.mssqltips.com/sqlservertip/2976/comparing-data-warehouse-design-methodologies-for-microsoft-sql-server/)”, na qual os data marts são consolidados posteriormente (bottom-up).

A grande vantagem dessa abordagem é o tempo reduzido na entrega de valor para o cliente. No entanto, é preciso manter a visão geral do Data Warehouse de forma que se evite a criação de dimensões redundantes ou se deixe de criar alguma dimensão. Apesar do foco ser um processo de negócio específico, é importante manter a visão do todo para que essa abordagem propicie bons resultados na evolução do projeto.

# Referências Bibliográficas

[*https://pt.wikipedia.org/wiki/Modelo\_multidimensional*](https://pt.wikipedia.org/wiki/Modelo_multidimensional)

<https://www.astera.com/pt/type/blog/data-mart-vs-data-warehouse/>

https://www.mssqltips.com/sqlservertip/2976/comparing-data-warehouse-design-methodologies-for-microsoft-sql-server/