



Ricardo Henrique Marques Pinto

Datamart Projeto Banda Larga nas Escolas

Rio de Janeiro – RJ

2021

RICARDO HENRIQUE MARQUES PINTO

DATAMART PROJETO BANDA LARGA NAS ESCOLAS
DASHBOARD COM O OBJETIVO DE ANALISAR A SITUAÇÃO DAS
ESCOLAS PUBLICAS E ESTADUAIS EM RELAÇÃO A CONEXÃO
COM A INTERNET

Projeto de pesquisa apresentado ao curso de Small Data Business Intelligence (BI Microsoft), do Instituto Infnet, a ser utilizado como trabalho de conclusão de bloco (PB)

Orientador: prof. Landry Duailibe Salles Filho

Rio de Janeiro – RJ
2021

RESUMO

O projeto tem como objetivo utilizar para análise a base de dados PBLE, que se trata de um projeto de instalação de internet banda larga em escolas públicas distribuídas pelo Brasil. De forma que tenha uma visão detalhada sobre esse projeto, com isso será desenvolvido um banco de análise que possa fornecer melhor desempenho para possíveis consultas.

Através de um Dashboard será utilizado indicadores sobre o projeto PBLE implementado pelo governo para que possa ter uma análise sobre esses dados concreta e objetiva.

Palavras chave: Tecnologia da Informação, Business Intelligence, Modelo Dimensional, Modelo Estrela, ETL, Integration Services, Analysis Services, Dashboard, Power BI, Censo Escolar.

SUMÁRIO

1. INTRODUÇÃO	7
1.1 FONTE DE DADOS CENSO ESCOLAR	7
1.1 FASES DO PROJETO	7
2. ANÁLISE INICIAL DOS DADOS	9
2.1 ORIGEM DOS DADOS	9
2.2 ANÁLISE DOS DADOS	10
3. MODELO DIMENSIONAL	11
3.1 MODELO LÓGICO DIMENSIONAL	11
3.2 MODELO FÍSICO DIMENSIONAL	12
3.3 DICIONÁRIO DE DADOS	13
4. CARGA DOS DADOS (ETL)	15
4.1 ARTEFATOS SCRUM	15
4.2 DOCUMENTAÇÃO ETL	17
4.3 PROJETO ETL	22
5. BANCO DE DADOS DE ANÁLISE MULTIDIMENSIONAL	23
5.1 Projeto do Banco Multidimensional	23
5.1.1 Dimensões	23
5.1.2 Grupos de Medidas	27
5.2 Bus Matriz	27
5.3 Projeto SSAS no Visual Studio	27
5.3.1 Dimensão Prestadora	28
5.3.2 Dimensão Escola	29
5.3.3 Dimensão Município	30
5.3.4 Dimensão Data	---
5.3.5 Cubo Censo Escolar	30
6. BANCO DE DADOS DE ANÁLISE TABULAR	32
6.1 Projeto do Banco Tabular	32
6.1.1 Dimensões	32
6.1.2 Grupos de Medidas	35
6.3 Projeto SSAS no Visual Studio	35
6.3.1 Dimensão Prestadora	36
6.3.2 Dimensão Escola	36
6.3.3 Dimensão Município	37
6.3.4 Dimensão Data	---
6.3.5 Medidas	37
6.3.6 Relacionamentos	38
7. DASHBOARD	39
7.1 Mockup	39

7.2 Dashboard Desenvolvido com Reporting Services	Erro! Indicador não definido.
7.3 Dashboard Desenvolvido com Power BI	40
8. CONCLUSÃO	41

1. INTRODUÇÃO

1.1 FONTE DE DADOS PBLE

O Programa Banda Larga nas Escolas (PBLE) foi lançado em 4 de abril de 2008 pelo governo federal, por meio do Decreto 6424, que altera o Plano Geral de Metas para a Universalização do Serviço Telefônico Fixo Comutado Prestado no Regime Público (PGMU).

- Municipais
- Estaduais

O programa prevê o atendimento de todas as escolas públicas urbanas de nível fundamental e médio, participantes dos programas E-Tec Brasil, além de instituições públicas de apoio à formação de professores: Polos Universidade Aberta do Brasil, Núcleo de Tecnologia Estadual (NTE) e Núcleo de Tecnologia Municipal (NTM).

Finalidade

O Programa **Banda Larga nas Escolas** tem como **objetivo** conectar todas as **escolas** públicas à internet, rede mundial de computadores, por meio de tecnologias que propiciem qualidade, velocidade e serviços para incrementar o ensino público no País.

Periodicidade

O projeto PBLE vem sendo feito desde o ano de 2008 em todas as escolas públicas que se adequem aos requisitos estipulados e assim tendo a implementação desse projeto nas escolas.

1.1 FASES DO PROJETO

a) Análise Inicial dos Dados

Os dados do projeto PBLE serão importados a partir do arquivo CSV de mesmo nome para tabelas no Management Studio, com objetivo de explorar e compreender os dados.

b) Modelo Dimensional

Será desenvolvido um modelo estrela (Star Schema) com sua fato e dimensões, assim como seus aspectos físicos, específicos para o SQL Server.

c) Carga dos Dados (ETL)

A carga dos dados para a fato e as dimensões será feita a partir de um projeto de carga no Integration Services da Microsoft que vem junto do SQL Server.

d) Banco de Dados de Análise

Utilização do Microsoft Analysis Services para desenvolvimento de um banco multidimensional, com informações do PROJETO PBLE, utilizando Pivot Table.

e) Dashboard

Desenvolvimento de Dashboard utilizando o Microsoft Power BI contendo os principais indicadores, com o objetivo de avaliar a distribuição média de velocidade de conexão instalada por escola, município e região..

2. ANÁLISE INICIAL DOS DADOS

2.1 ORIGEM DOS DADOS

Os dados do Projeto Banda Larga estão disponíveis para acesso através do site dados.gov.br linkado logo abaixo.

<https://dados.gov.br/dataset/pble>

Contendo junto de si, um glossário simples, com o significado de cada atributo presente no conjunto de dados.

GLOSSÁRIO DE TERMOS

Prestadora: Empresa de telecomunicações incumbida em fornecer conexão à internet para a escola.

Número INEP: Código utilizado pelo INEP para cadastramentos das escolas. Cada escola possui seu número específico.

Município/UF: Município e Unidade da Federação de localização da escola.

Situação: Indica a condição da conexão à internet da escola.

Tipo de Obrigação: É o próprio Programa de Banda larga nas Escolas Públicas Urbanas – PBLE.

Velocidade de Acesso Instalada: Velocidade da conexão do acesso à internet.

Tecnologia: Indicação da tecnologia de telecomunicações utilizada para o fornecimento da conexão à internet para a escola.

Nome da Escola: Nome pelo qual a escola é conhecida.

Tipo da Escola: Indica responsabilidade de gestão da escola.

Data de Ativação: Data em que a conexão à internet foi efetivada.

Contendo dados sobre a implementação do projeto desde o ano de 2008.

Foi feita uma seleção de quais atributos de colunas seriam utilizados para o projeto, sendo descartados os atributos que não possuíam relevância e apresentavam falta de qualidade nos dados.

2.2 ANÁLISE DOS DADOS

Os dados do projeto PBLE foram importados através de um arquivo csv para dentro do SQL server, sem nenhuma transformação para uma análise rápida dos dados.

Foram encontrados alguns dados duplicados na coluna Inep.

```
SELECT [Nº INEP] , count(*) as QtdLinhas  
FROM [Stage].[PBLE]  
GROUP BY [Nº INEP]  
HAVING count(*) > 1
```

O campo NºINEP recebeu tratamento nas suas linhas para que pudesse ser utilizado como PK da dimensão Escola.

3. MODELO DIMENSIONAL

3.1 MODELO LÓGICO DIMENSIONAL

Modelo Estrela Projeto Banda Larga.



3.2 MODELO FÍSICO DIMENSIONAL

Devido a pouca quantidade de dados, não foi utilizado nenhum recurso de particionamento e compressão das dimensões **Dim_Escola**, **Dim_Municipio**, **Dim_Prestadora**, **Dim_Data**.

Contudo se a base fosse composta de uma enorme quantidade de dados, esses mesmo divididos por anos distintos, seria uma boa opção a utilização de recursos SGBD para a criação de uma partição desses dados por ano, e uma compressão das dimensões.

3.3 DICIONÁRIO DE DADOS

DimAno

Contem o registro de cada ano mês e dia do projeto PBLE.

DimMunicipio

Contém o nome do município e estado.

Coluna: Municipio_ID

Tipo de dado: Int, Pk Surrogate Key

Significado: Chave sequencial criada durante a carga.

Coluna: Municipio_Name

Tipo de dado: varchar (250)

Significado: Nome do Município

Coluna: UF_Name

Tipo de dado: varchar(5)

Significado: Nome do Estado onde está localizado o município

DimEscola

Contém a lista de escolas e o seu tipo.

Coluna: Inep_ID

Tipo de dado: int primary key

Significado: Chave primaria criada a partir da coluna N°INEP.

Coluna: Nome_Escola

Tipo de dado: varchar(250)

Significado: Nome completo da Escola

Coluna: Tipo_Escola

Tipo de dado: varchar(20)

Significado: Informativo se a escola é Municipal ou Estadual

DimPrestadora

Prestadora que instalou o serviço em determinada escola.

Coluna: Prestadora_ID

Tipo de dado: int pk surrogate key

Significado: Chave sequencial criada durante a carga

Coluna: Prestadora

Tipo de dado: varchar(20)

Significado: Nome da Prestadora.

4. CARGA DOS DADOS (ETL)

4.1 ARTEFATOS SCRUM

Backlog Produto

1) Desenvolver Datamart e Banco de Análise PBLE

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

2) Desenvolver Dashboard

“Como gestor público, preciso de um dashboard que possua indicadores sobre os dados do PBLE que possam me mostrar a distribuição e eficiência da velocidade da conexão instalada nessas escolas”.

Backlog Sprint 1

1.1) Carga Dimensão Escola

A dimensão escola receberá a carga a partir de uma tabela Stage com os registros já corrigidos e alterados para que os dados sejam validados pelo Product Owner.

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

”.

1.2) Carga Dimensão Município

A dimensão município receberá carga a partir de uma tabela Stage PBLE com os registros já corrigidos e alterados para que os dados sejam validados pelo Product Owner.

Complemento de História:

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

1.3) Carga Dimensão Ano

A dimensão Ano receberá carga a partir de uma tabela Stage PBLE com os registros já corrigidos e alterados para que os dados sejam validados pelo Product Owner.

Complemento de História:

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

1.4) Carga Dimensão Prestadora

A Dimensão Prestadora receberá carga a partir de uma tabela Stage PBLE com os registros já corrigidos e alterados para que os dados sejam validados pelo Product Owner.

Complemento de História:

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

Backlog Sprint 2

1.5) Carga Fato Internet

A fato Internet receberá dados a partir da tabela Stage PBLE e assim possa ser validado pelo Product Owner.

Complemento de História:

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

1.6) Desenvolver Banco de Análise Multidimensional

O banco de análise Multidimensional receberá dados das dimensões e da tabela fatos para que possa ser validado os dados carregados em um cubo Internet PBLE.

Complemento de História:

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

1.7) Desenvolver Banco de Análise Tabular

O Banco de análise Tabular receberá dados das dimensões e da tabela fatos para que possa ser validado os dados carregados em um banco tabular Internet PBLE.

Complemento de História:

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

Backlog Sprint 4

1.8) Desenvolver Dashboard

“Como gestor público, preciso analisar os dados do Projeto Banda Larga nas Escolas acessando o banco de análise, com o objetivo de avaliar, a qualidade de conexão disponível nas escolas públicas que participaram do projeto”.

4.2 DOCUMENTAÇÃO ETL

No projeto PBLE, apesar de conter dados durante todo ano, houve uma facilidade no processo de ETL pois os mesmos derivam da mesma carga, sem alteração nos dados, assim também não teve a necessidade de fazer um SCD tipo 2.

DimEscola

- Origem: Stage.PBLE

Coluna Origem: NºINEP

Coluna Destino: Inep_ID

Tipo de dado: int

Transformação: Primary Key

Coluna Origem: Nome da Escola

Coluna Destino: nome_escola

Tipo de dado: varchar(250)

Transformação:

Coluna Origem: Tipo Escola

Coluna Destino: tipo_escola

Tipo de dado: varchar(20)

Transformação:

DimPrestadora

- Origem: Stage.PBLE

Coluna Origem: Prestadora
Coluna Destino: prestadora_id
Tipo de dado: int
Transformação: identity surrogate key

Coluna Origem: Prestadora
Coluna Destino: prestadora
Tipo de dado: varchar(20)
Transformação:

DimMunicipio

- Origem: Stage.PBLE

Coluna Origem: Municipio
Coluna Destino: municipio_id
Tipo de dado: int
Transformação: identity surrogate key

Coluna Origem: Municipio
Coluna Destino: Municipio_nome
Tipo de dado: varchar(250)
Transformação:

Coluna Origem: UF
Coluna Destino: UF_nome
Tipo de dado: varchar(5)
Transformação:

FatoInternetPBLE

- Origem: Stage.PBLE

Coluna Origem: Municipio
Coluna Destino: municipio_id
Tipo de dado: int
Transformação: identity surrogate key

Coluna Origem: Prestadora
Coluna Destino: Prestadora_id
Tipo de dado: int
Transformação: identity surrogate key

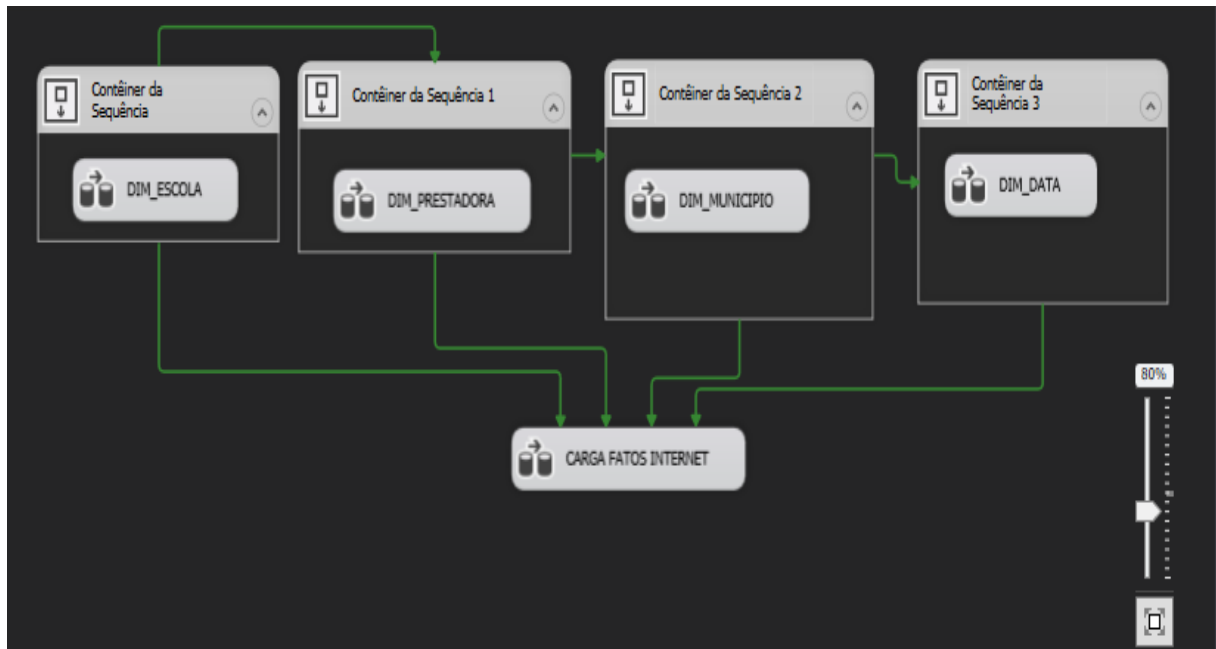
Coluna Origem: N^oINEP
Coluna Destino: Inep_id
Tipo de dado: int
Transformação: primary key

Coluna Origem: Data de Ativação
Coluna Destino: data_id
Tipo de dado: int
Transformação: primary key

4.3 PROJETO ETL

Foi feito um insert nas dimensões a partir da tabela Stage.PBLE que já estava transformada e corrigida.

O Projeto de Carga consistiu em apenas um pacote SSIS onde foi alimentado as dimensões a partir da tabela Stage.PBLE e por fim fazendo uso das dimensões para carregar a fatos com um select join.



5. BANCO DE DADOS DE ANÁLISE MULTIDIMENSIONAL

5.1 Projeto do Banco Multidimensional

5.1.1 Dimensões

Dim Prestadora

Tabela origem dbo.dim_prestadora

Coluna Origem: prestadora_ID

Atributo: Codigo prestadora sk

Pasta: chave

Coluna Origem: prestadora

Atributo: prestadora

Pasta: Característica

Dim Escola

- Tabela Origem: dbo.Dim_Escola

Coluna Origem: Inep_ID

Atributo: Codigo Inep pk

Pasta: chave

Coluna Origem: Nome_Escola

Atributo: Nome_Escola

Pasta: Característica

Coluna Origem: Tipo_Escola

Atributo: Tipo_Escola

Pasta: Característica

Hierarquia: Escola

Tipo Escola

Nome Escola

Hierarquia: Município
Tipo Escola
Nome Escola
Municipio_ID

Dim_Municipio

- Tabela Origem: dbo.Dim_Municipio

Coluna Origem: Municipio_ID

Atributo: Municipio_id pk
Pasta: chave

Coluna Origem: Municipio_name
Atributo: Municipio_name
Pasta: Característica

Coluna Origem: UF_name
Atributo: UF_name
Pasta: Característica

Hierarquia: Município
Uf_Name
Municipio_Name
Inep_ID

Dim_Data

- Tabela Origem: dbo.Dim_Data

Coluna Origem: Data_ID**Atributo:** Data_id pk**Pasta:** chave**Coluna Origem:** Dia**Atributo:** Dia**Pasta:** Característica**Coluna Origem:** Mês**Atributo:** Mês**Pasta:** Característica**Coluna Origem:** Mês_Nome**Atributo:** Mês_Nome**Pasta:** Característica**Coluna Origem:** Ano**Atributo:** Ano**Pasta:** Característica**Coluna Origem:** Data_Completa**Atributo:** Data_Completa**Pasta:** Característica

Hierarquia: Data
Ano
Mês
Dia

5.1.2 Grupos de Medidas

Fato_Internet

- Tabela Origem: dbo.Fato_Internet

Coluna Origem	Medida	Formato	Agregação
Velocidade Acesso	Velocidade Acesso	#, #0	Sum

5.2 Bus Matriz

	Dim Escola	Dim Prestadora	Dim Municipio	Dim Data
Fato Internet	•	•	•	•

5.3 Projeto SSAS no Visual Studio

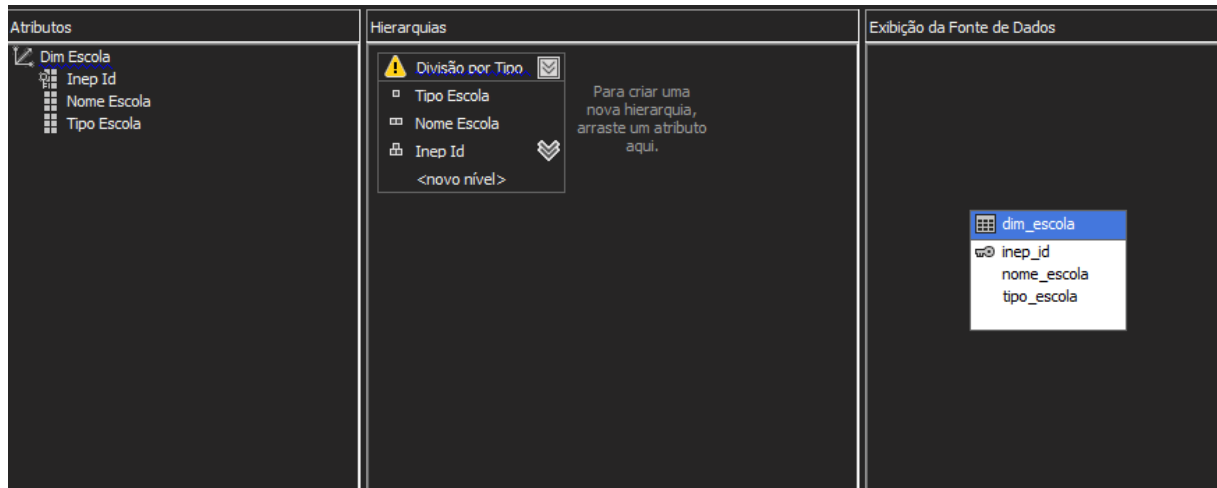
A partir dos dois modelos estrela, foi desenvolvido projeto de um banco Analysis Services Multidimensional, contendo quadro dimensões e dois grupos de medidas.

A dimensão Prestadora possui apenas um atributo Prestadora, com com o nome das empresas que participaram do projeto banda larga.



5.3.1 Dimensão Escola

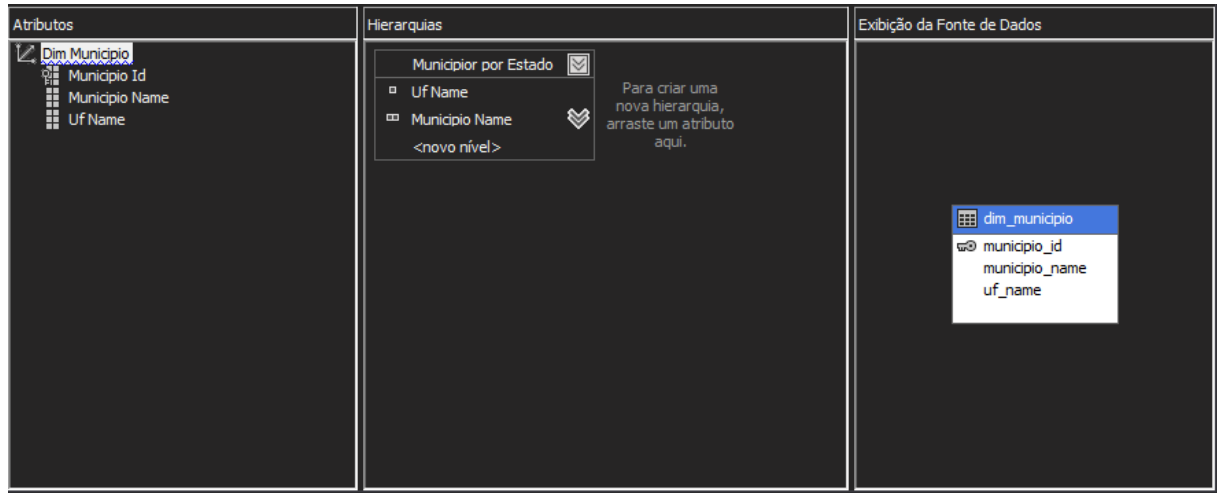
Na dimensão Escola foi criada uma hierarquia, organizados em pastas.



5.3.2 Dimensão Município

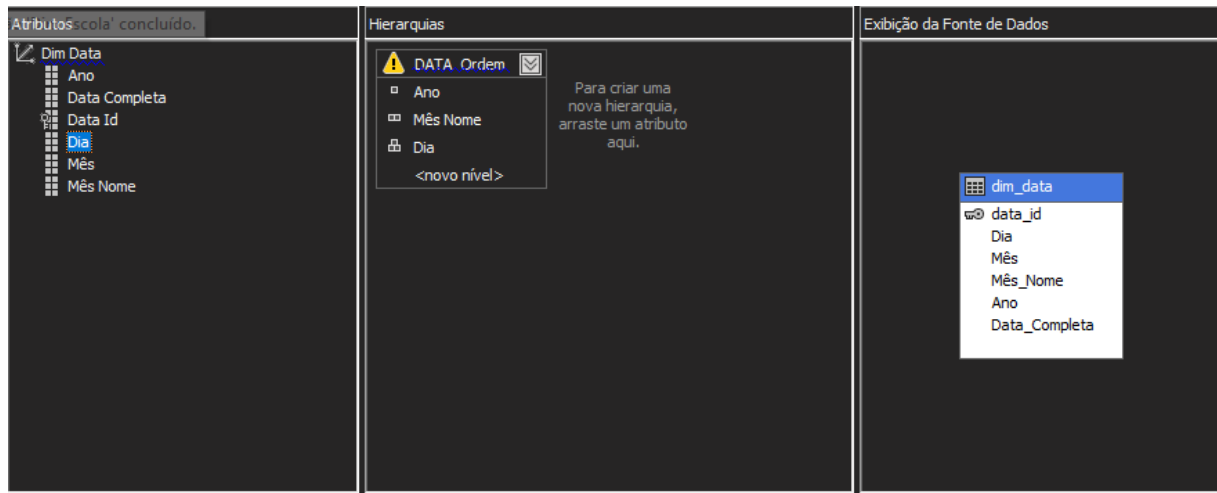
Na dimensão Município, também foi criada apenas uma hierarquia. Organizados em pastas.

Utilizado o KeyDuplicate Error como: Ignore error, pois existem cidades de mesmo nome, porem estados diferentes.



5.3.3 Dimensão Data

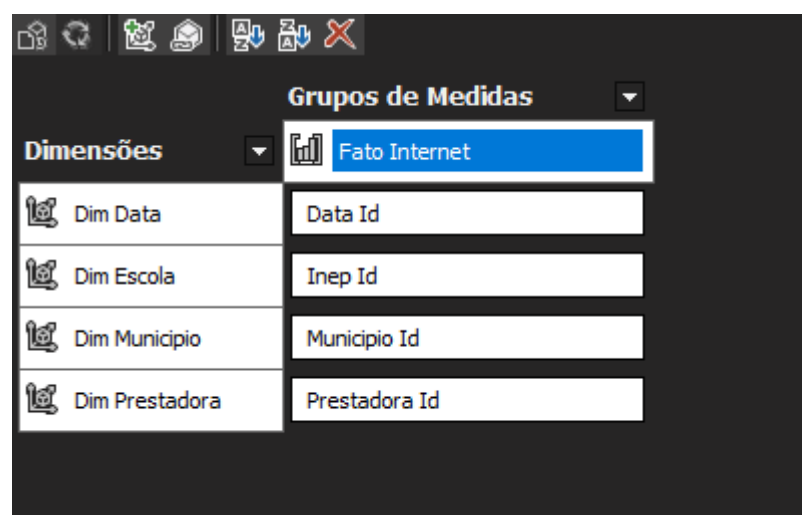
Para **Dim Data** foi criada uma hierarquia a distribuição por ano mês e dia.



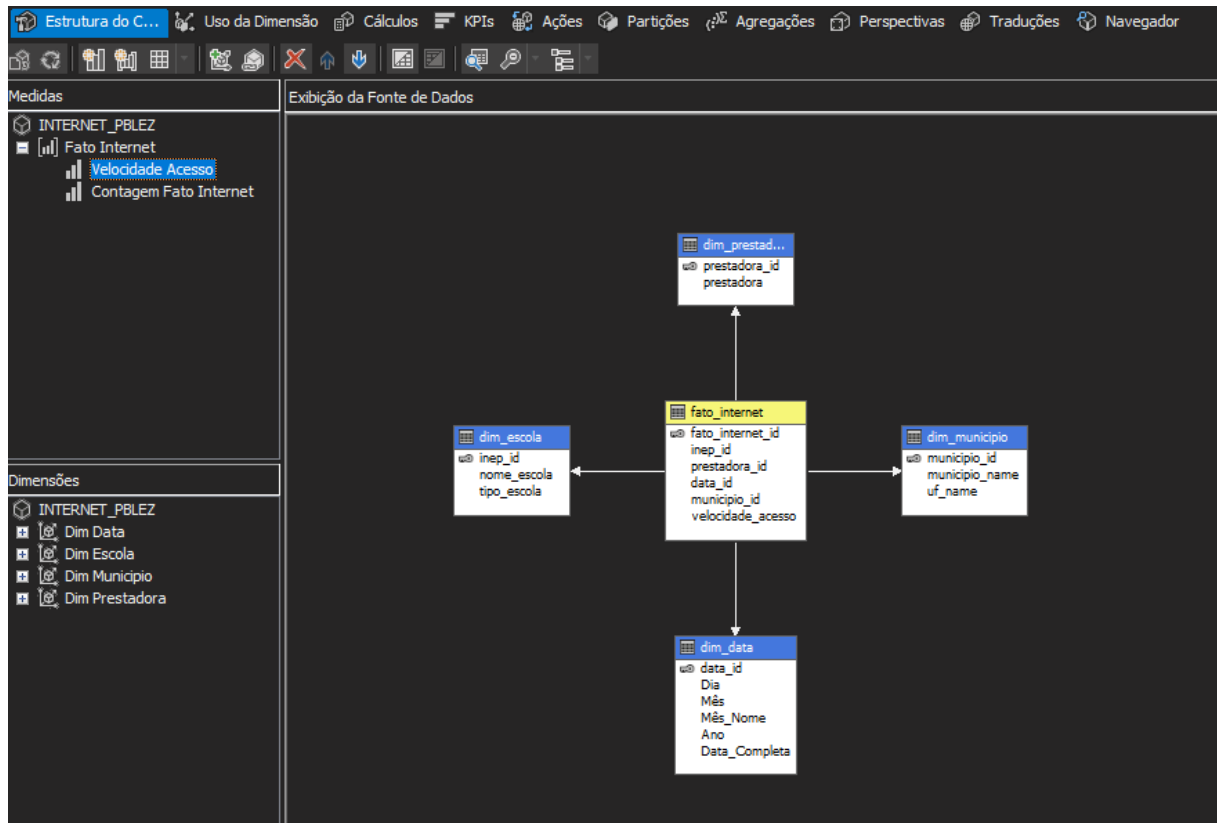
5.3.4 Cubo Internet PBLE

O Cubo **Internet PBLE** tem apenas um grupo de medidas, obtendo sua métrica a partir da Fato_Internet

Relacionamento entre Dimensões e Grupos de medidas:



Estrutura do Cubo Internet PBLE:



6. BANCO DE DADOS DE ANÁLISE TABULAR

6.1 Projeto do Banco Tabular

6.1.1 Dimensões

Dim Prestadora

-Tabela Origem: dbo.Dim_Prestadora

Coluna Origem: prestadora_ID

Atributo: Codigo prestadora sk

Pasta: chave

Coluna Origem: prestadora

Atributo: prestadora

Pasta: Característica

Dim Escola

- Tabela Origem: dbo.Dim_Escola

Coluna Origem: Inep_ID

Atributo: Codigo Inep pk

Pasta: chave

Coluna Origem: Nome_Escola

Atributo: Nome_Escola

Pasta: Característica

Coluna Origem: Tipo_Escola

Atributo: Tipo_Escola

Pasta: Característica

Hierarquia: Escola

Tipo Escola

Nome Escola

Hierarquia: Município
Tipo Escola
Nome Escola
Municipio_ID

Dim_Municipio

- Tabela Origem: dbo.Dim_Municipio

Coluna Origem: Municipio_ID

Atributo: Municipio_id pk

Pasta: chave

Coluna Origem: Municipio_name

Atributo: Municipio_name

Pasta: Característica

Coluna Origem: UF_name

Atributo: UF_name

Pasta: Característica

Hierarquia: Município
Uf_Name
Municipio_Name
Inep_ID

Dim_Data

- Tabela Origem: dbo.Dim_Data

Coluna Origem: Data_ID

Atributo: Data_id pk

Pasta: chave

Coluna Origem: Dia

Atributo: Dia

Pasta: Característica

Coluna Origem: Mês

Atributo: Mês

Pasta: Característica

Coluna Origem: Mês_Nome

Atributo: Mês_Nome

Pasta: Característica

Coluna Origem: Ano

Atributo: Ano

Pasta: Característica

Coluna Origem: Data_Completa

Atributo: Data_Completa

Pasta: Característica

Hierarquia: Data
Ano
Mês
Dia

6.1.2 Grupos de Medidas

FatoInternet

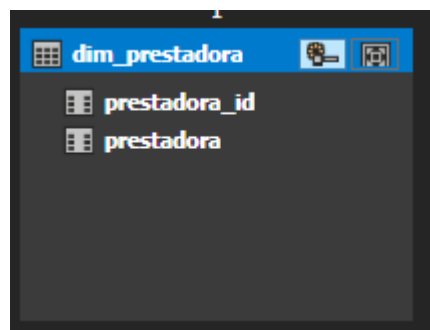
- Tabela Origem: dbo.FatoInternet

Coluna Origem	Medida	Formato	Agregação
Velocidade de Acesso	Velocidade de Acesso	#,#0	Sum

6.2 Projeto SSAS no Visual Studio

A partir do modelo estrela, foi desenvolvido um projeto de Analysis Services Tabular, contendo quatro dimensões e um grupo de medidas.

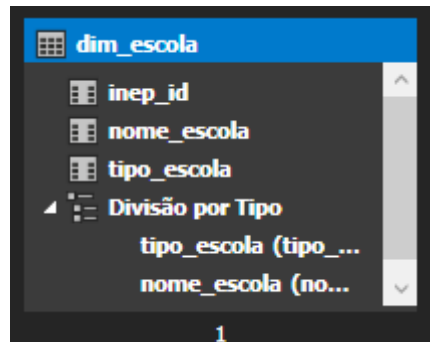
A dimensão Prestadora possui apenas um atributo Prestadora, com cinco membros representando as cinco empresas que instalaram internet, no Projeto Banda Larga.



6.3.1 Dimensão Escola

Na dimensão Escola foi criada uma hierarquia, organizados em pastas.

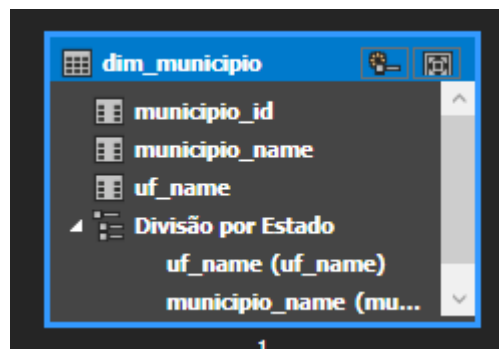
Os atributos foram organizados em pastas, utilizando a propriedade **Display Folder**.



6.3.2 Dimensão Município

Na dimensão Município foi criada uma hierarquia, organizados em pastas.

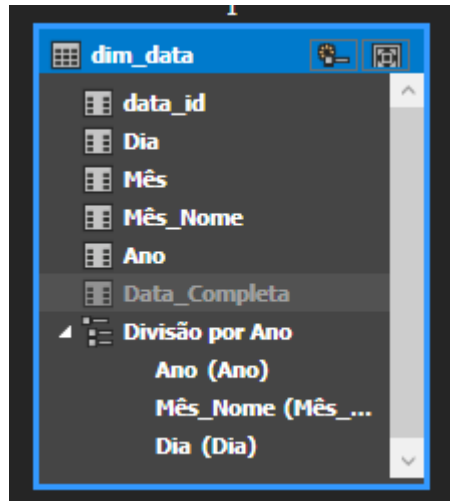
Os atributos foram organizados em pastas, utilizando a propriedade **Display Folder**.



6.3.3 Dimensão Data

Na dimensão Data foi criada uma hierarquia, organizados em pastas.

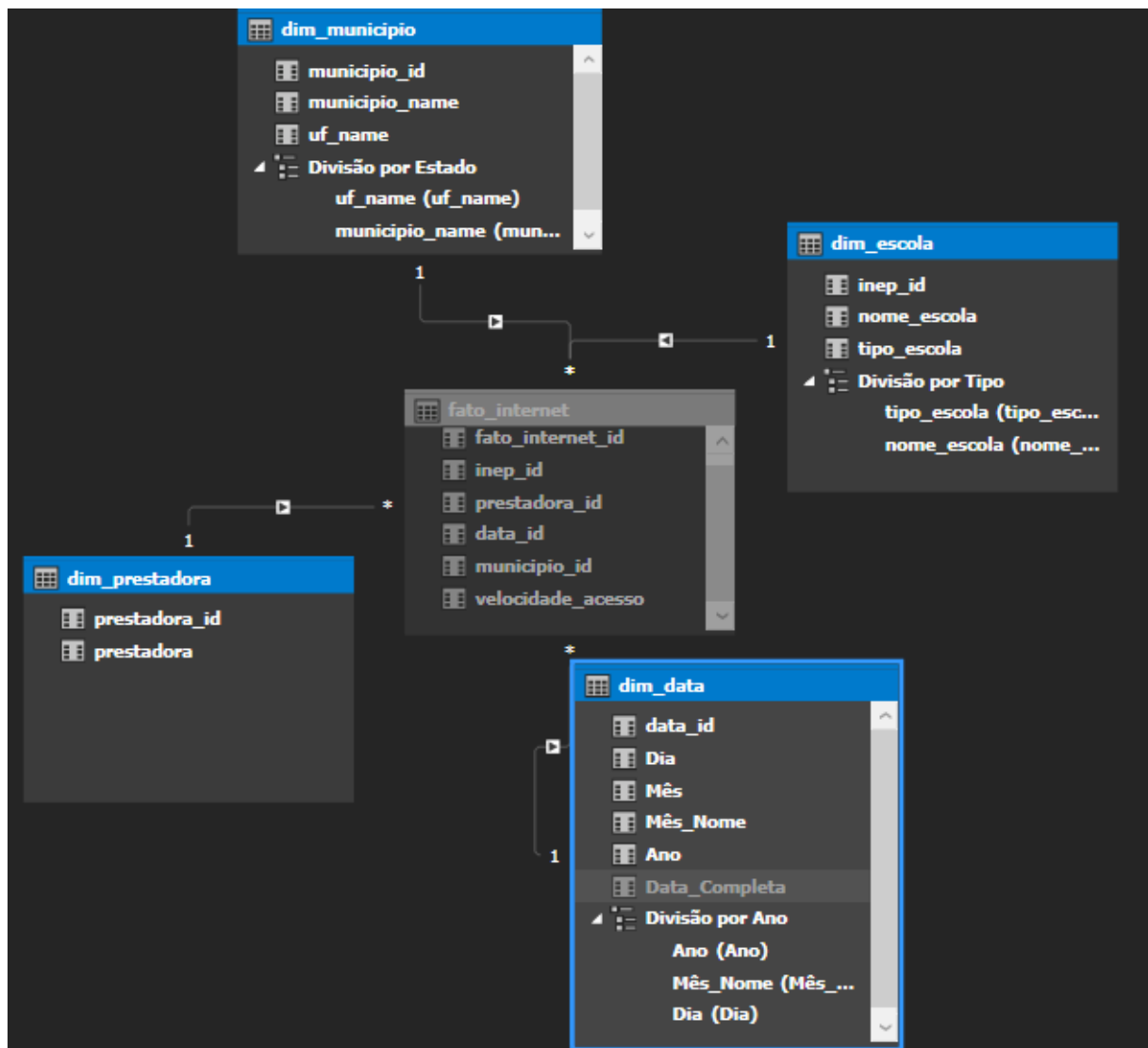
Os atributos foram organizados em pastas, utilizando a propriedade **Display Folder**.



6.3.4 Medidas

O Banco Tabular **Internet PBLE** possui uma medida, proveniente da tabela Fatos_Internet

6.3.4 Relacionamentos



7. DASHBOARD

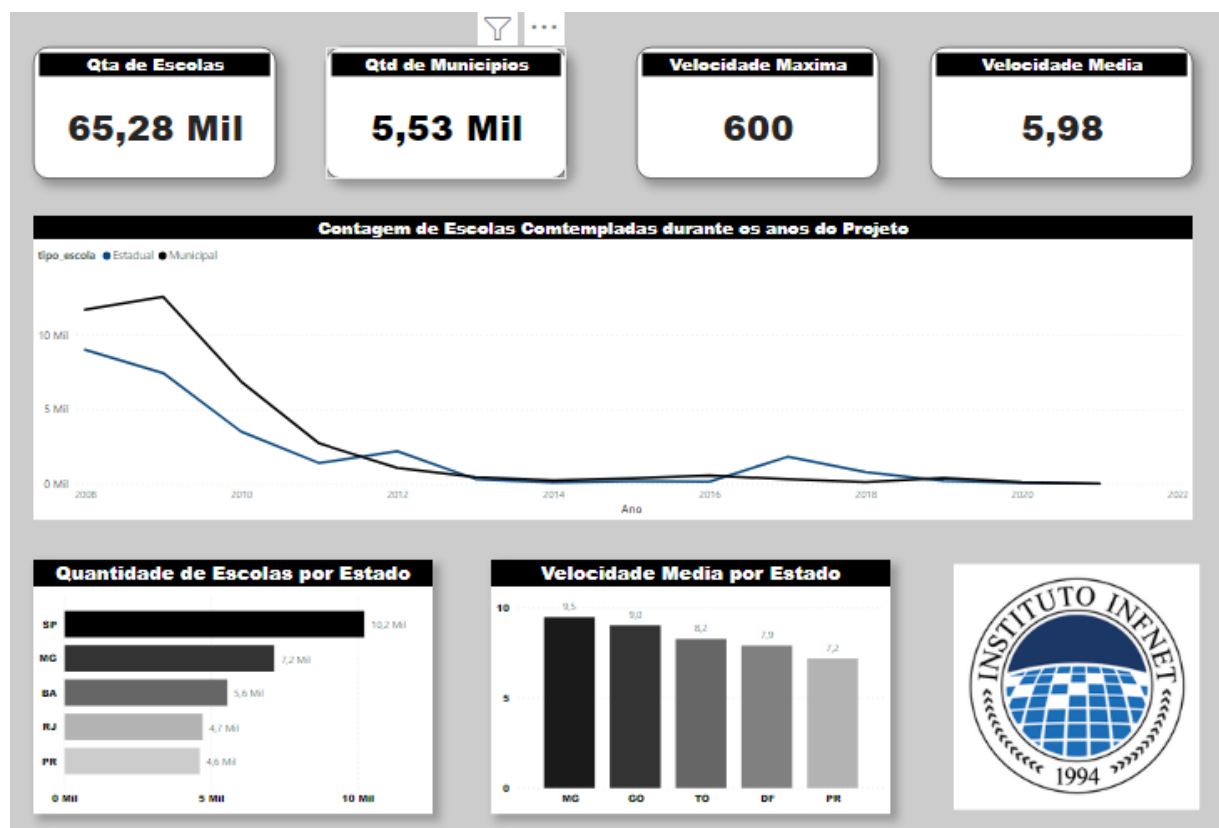
7.1 Mockup

O Mockup do Dashboard foi feito através do power Bi, na sua parte superior ele apresenta cards com a quantidade de escolas, municípios contemplados, a maior velocidade instalada, e a média de velocidade instalada durante todos os anos.

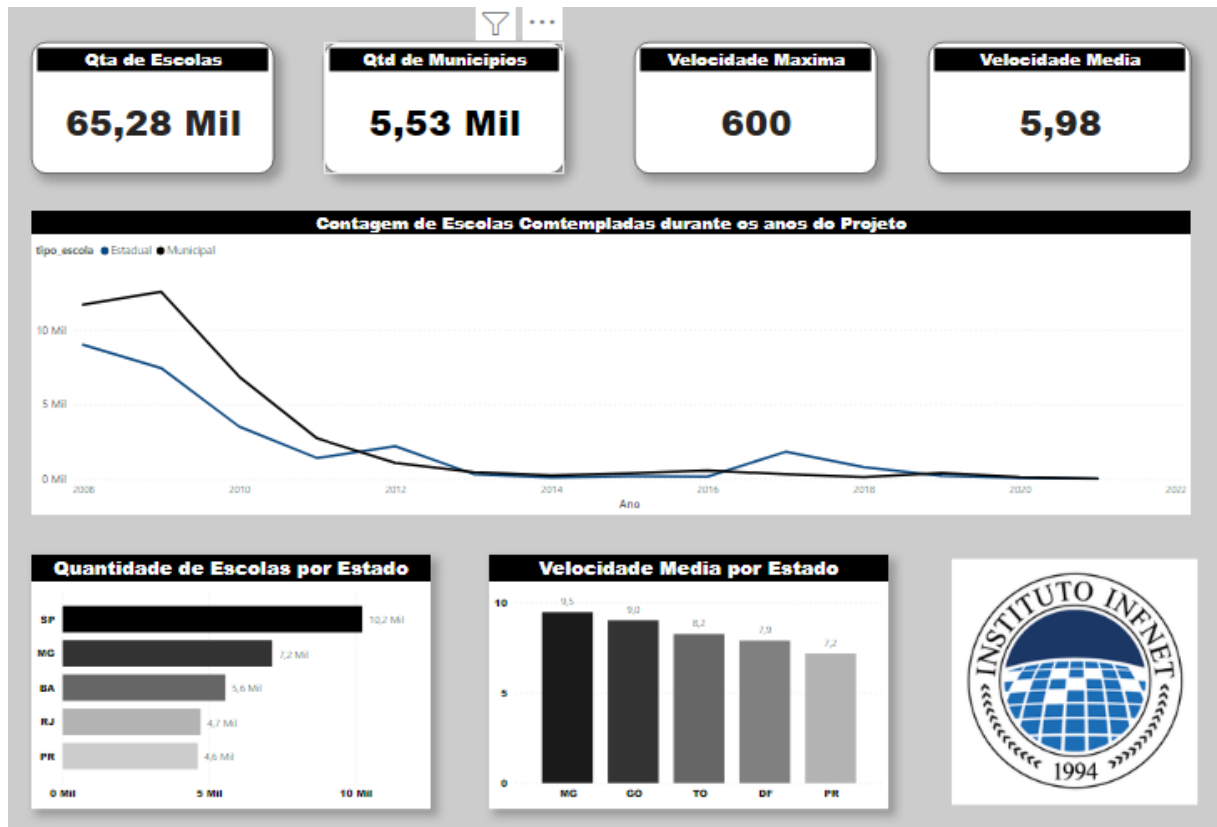
Na região do centro do Dashboard apresenta um gráfico de linhas que apresenta a queda de escolas contempladas com o projeto durante todos os anos.

Na parte inferior apresenta-se através de dois gráficos de barras, a distribuição por estado com maior número de escolas, e os 5 principais estados com a velocidade média mais alta instalada.

O Dashboard foi desenvolvido para se saber a distribuição media de conexão instalada nas escolas do brasil.



7.2 Dashboard Desenvolvido com Power BI



8. CONCLUSÃO

O projeto Banda Larga nas escolas criou uma base de dados contendo indicadores de internet das escolas no Brasil, esses dados podem servir de múltiplas análises para o desenvolvimento de melhor qualidade e avanço tecnológico nas escolas brasileiras.

Entregando um ótimo desempenho e facilitando a busca de indicadores a partir dos dados trabalhados durante o processo.

Assim o Dashboard pode ser um ótimo recurso estratégico para verificar a qualidade de conexão presente nas escolas públicas, e analisar possíveis ações futuras para melhorar a qualidade nas escolas brasileiras.