



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
ESCOLA DE INFORMÁTICA APLICADA

**Solução de Business Intelligence utilizando a plataforma Microsoft na área da
Segurança Pública**

THIAGO DA SILVA BRITO
RAFAEL DA SILVA OLIVEIRA

Orientador
ASTÉRIO KIYOSHI TANAKA

RIO DE JANEIRO, RJ – BRASIL
NOVEMBRO DE 2017

Catálogo informatizada pelo(a) autor(a)

d862 da Silva Brito, Thiago
Solução de Business Intelligence utilizando a
plataforma Microsoft na área da Segurança Pública /
Thiago da Silva Brito. -- Rio de Janeiro, 2017.
75

Orientador: Astério Kiyoshi Tanaka.
Trabalho de Conclusão de Curso (Graduação) -
Universidade Federal do Estado do Rio de Janeiro,
Graduação em Sistemas de Informação, 2017.

1. Business Intelligence. 2. Analytics
Platforms. 3. Data Warehouse. I. Kiyoshi Tanaka,
Astério, orient. II. Título.

Solução de Business Intelligence utilizando a plataforma Microsoft na área da
Segurança Pública

THIAGO DA SILVA BRITO
RAFAEL DA SILVA OLIVEIRA

Projeto de Graduação apresentado à Escola de
Informática Aplicada da Universidade Federal do
Estado do Rio de Janeiro (UNIRIO) para obtenção do
título de Bacharel em Sistemas de Informação.

Aprovado por:

ASTÉRIO KIYOSHI TANAKA (UNIRIO)

FLAVIA MARIA SANTORO (UNIRIO)

RIO DE JANEIRO, RJ – BRASIL.

NOVEMBRO DE 2017

Agradecimentos

Esse projeto é o resultado de muitas horas de dedicação, e não seria possível a sua realização sem a colaboração das pessoas abaixo, por isso, gostaríamos de prestar os nossos sinceros agradecimentos:

Às nossas famílias, pelo apoio incondicional, em todos os momentos dessa jornada.

Ao professor Asterio Takana, pela sabedoria e orientação dada para o desenvolvimento deste projeto.

Aos docentes do curso de Sistemas de Informação, pela determinação a disseminar o conhecimento a seus alunos.

Aos amigos e amigas, pelos momentos dentro e fora da Universidade, pelos trabalhos em conjunto e pelo apoio prestado durante esses anos.

Por fim, gostaríamos de agradecer a todos que participaram deste trabalho de alguma forma e que nos apoiaram durante esta caminhada.

RESUMO

A área de Business Intelligence e todas as áreas que envolvem análise de dados estão em bastante evidência atualmente. Diversas organizações cada vez mais reconhecem a importância e o valor da informação, e por isso estão cada vez mais se aproximando de uma cultura “data-driven”, orientada a dados, para tomada de decisão. Estas organizações utilizam diversas fontes de dados, de diferentes tipos, enquanto o mercado de tecnologia oferece diversas soluções para cada cenário específico. Este trabalho tem como foco o desenvolvimento de uma arquitetura de Business Intelligence, utilizando a plataforma da empresa Microsoft, uma das mais utilizadas atualmente pela sua habilidade de execução e facilidade de uso. Serão utilizados dados do Instituto de Segurança Pública para demonstração dos indicadores de ocorrências criminais em diversas localidades do Estado do Rio de Janeiro, sendo possível realizar diversas análises sobre o perfil das ocorrências, como localização, tipo de ocorrência registrada e o perfil das vítimas. Através do desenvolvimento destes indicadores, poderemos avaliar os principais componentes da plataforma, assim como realizar considerações de arquitetura e performance, e também uma análise de pontos fortes e fracos, realizando um breve comparativo com as diferentes soluções existentes no mercado, tendências e outras tecnologias existentes.

Palavras-chave: Business Intelligence, Data Warehouse, Analytics

ABSTRACT

The Business Intelligence area and all areas involving data analysis are in evidence today. Several organizations are increasingly recognizing the importance and value of information, and so are increasingly approaching a data-driven culture for decision-making. These organizations use diverse data sources of different types, while the technology market offers several solutions for each specific scenario. This work focuses on the development of a Business Intelligence architecture, using Microsoft's platform, one of the most used today for its ability of execution and ease of use. Data from the Public Security Institute will be used to demonstrate indicators of criminal occurrence in various locations in the State of Rio de Janeiro, and will allow to carry out analysis on the profile of criminal occurrences, such as location, type of occurrence recorded and victims profiling. Through the development of these indicators, we will be able to evaluate the main components of the platform, as well as make architectural and performance considerations, as well as an analysis of strengths and weaknesses, making a brief comparison with the different existing solutions in the market, discuss trends and other existing technologies.

Keywords: Business Intelligence, Data Warehouse, Analytics

Índice

1	Introdução	11
1.1	Motivação	11
1.2	Objetivo	12
1.3	Organização do texto	12
2	Conceito de Business Intelligence	14
2.1	Introdução	14
2.2	Data Warehouse	14
2.3	Componentes de uma solução de Data Warehouse	15
2.4	Modelagem Dimensional	16
2.5	Fatos e Métricas	18
2.6	Dimensões	19
3	Plataforma de BI Microsoft	21
3.1	Introdução	21
3.2	SQL Server	21
3.3	Integration Services	24
3.4	Analysis Services	26
3.5	Excel, Reporting Services e Power BI	28
3.6	Posicionamento da Microsoft em pesquisas de mercado	31
4	Implementando o projeto de BI do ISP	34
4.1	Introdução a análise de segurança pública	34
4.2	Entendendo as ocorrências criminais do Rio de Janeiro	34
4.3	Elaborando os requisitos da solução para o ISP	35
4.4	Elaborando o desenho da solução final	41
4.5	Construindo o Data Warehouse	42
4.6	Construindo o ETL	48
4.7	Construindo o modelo semântico	52
4.8	Construindo os Relatórios e Dashboards no Excel e Power BI	53
4.9	Compartilhando relatórios no Reporting Services	59
4.10	Resultados obtidos	62
5	Conclusão	66
5.1	Considerações finais	66
5.2	Limitações do projeto	68
5.3	Trabalhos futuros	68

Índice de Tabela

Tabela 1 - Comparativo de normalização de banco de dados	17
Tabela 2 - Comparativo de índices	22
Tabela 3 - Especificação das máquinas utilizadas no projeto	42

Índice de Figuras

Figura 1 - Componentes básicos de BI (MICROSOFT, 2017)	16
Figura 2 - Modelo estrela e cubo OLAP (KIMBALL, 2013).....	17
Figura 3 - Exemplo de tabela Fato (KIMBALL, 2013).....	19
Figura 4 - Exemplo de dimensão (KIMBALL, 2013).....	20
Figura 5 - Tabelas Temporais no SQL Server (MICROSOFT, 2017)	23
Figura 6 - Analysis Services e componentes (MICROSOFT, 2017)	27
Figura 7 - Excel, PowerView e PowerPivot (MICROSOFT, 2017)	29
Figura 8 - Reporting Services na Web (MICROSOFT, 2017).....	30
Figura 9 - Visão geral do funcionamento do Power BI (MICROSOFT, 2017)	31
Figura 10 - Quadrante Gartner para ferramentas de BI (GARTNER, 2017)	32
Figura 11 - Quadrante Gartner para Banco de Dados (GARTNER, 2017).....	33
Figura 12 - Quadrante Gartner para ETL (GARTNER, 2017).....	33
Figura 13 - Arquivos de texto do ISP	37
Figura 14 - Esquema Registro de Ocorrências	38
Figura 15 - Violência contra a mulher.....	39
Figura 16 - Perfil vítima de letalidade	40
Figura 17 - Armas apreendidas.....	40
Figura 18 - Plano da solução proposta.....	41
Figura 19 - Solução de banco de dados no Visual Studio	43
Figura 20 - Tipos de componentes de storage do SQL Server	44
Figura 21 - Plano de execução de consulta exibindo as partições acessadas	45
Figura 22 - Script de tabela dimensão no SQL.....	46
Figura 23 - Script de tabela Fato no SQL.....	46
Figura 24 - Objetos criados no Data Warehouse.....	47
Figura 25 - Projeto de ETL no Visual Studio.....	49
Figura 26 - Pacote de ETL "FatoViolenciaMulher"	50
Figura 27 - Fluxo da tarefa "Carrega tabela Fato", com destaque no objeto Origem OLEDB.....	51
Figura 28 - Modo debug do pacote no Data Tools.....	51
Figura 29 - Catálogo de projetos do SSIS no SQL Server	52
Figura 30 - Modelo tabular (visão parcial) no Analysis Services	53
Figura 31 - Importando dados do Analysis Services no Excel.....	54

Figura 32 - Interface dinâmica do Excel utilizando dados do ISP	55
Figura 33 - Conectando no Analysis Services pelo Power BI.....	55
Figura 34 - Prévia dos objetos para importação no Power BI.....	56
Figura 35 - Visão geral da interface gráfica do Power BI.....	56
Figura 36 - Antes e depois de filtrar um elemento selecionando-o no mapa	57
Figura 37 - Publicando no serviço Power BI na nuvem	58
Figura 38 - Interface do serviço do Power BI na nuvem, similar ao Desktop	59
Figura 39 - Janela de configuração do Reporting Services	60
Figura 40 - Página Inicial do Reporting Services.....	61
Figura 41 - Aplicação de Relatórios Móveis (Mobile Reports)	62
Figura 42 - Painel com o resumo temporal de Ocorrências	62
Figura 43 - Proporção de ocorrências por RISP e Categoria.....	63
Figura 44 - Proporção de ocorrências por RISP	64
Figura 45 - Painel exibindo o perfil da vítima de Violência contra Mulher.....	65
Figura 46 - Período do dia com ocorrências de violência contra a mulher	65

Lista de Siglas

BI	- Business Intelligence
OLAP	- On-line Analytical Processing
OLTP	- On-line Transaction Processing
SQL	- Structured Query Language
SSIS	- SQL Server Integration Services
SSRS	- SQL Server Reporting Services
SSAS	- SQL Server Analysis Services
RO	- Registro de Ocorrência
PA	- Predictive Analytics
ETL	- Extract, Transform and Load
DAX	- Data Analysis Expressions
MDX	- Multidimensional Expressions

1 Introdução

1.1 Motivação

O termo Business Intelligence, também conhecido como BI, Inteligência de Negócios ou Inteligência Empresarial, corresponde ao processo de coleta, organização, análise e compartilhamento de dados, e possui como objetivo auxiliar na tomada de decisões estratégicas¹. Diversas organizações têm utilizado BI para observar e melhorar sua produtividade, e a busca por profissionais desta área está em alta. De acordo com um estudo recente do site CareerCast², a profissão Cientista de Dados foi considerada a melhor do ano de 2017 nos Estados Unidos. É necessário que este profissional possua a capacidade de extrair valor da informação, e para isso ele deverá contar com ferramentas que facilitem o seu trabalho, tornando todo o processo menos cansativo. Existem diversas tecnologias disponíveis no mercado, e organizações que possuem uma cultura orientada a dados (*data-driven*) alocam recursos consideráveis para a aquisição dessas tecnologias pois reconhecem a sua importância para o negócio.

Entre os diversos fornecedores do segmento, alguns se destacam como líderes, com maior capacidade de execução ou por serem visionários, ou seja, atendem a maioria dos requisitos necessários dos profissionais ou estão sempre inovando ou fornecendo novas funcionalidades inéditas. Dentre os fornecedores, alguns criam produtos para nichos específicos e outros criam para um uso geral por todas as organizações. A Microsoft, empresa conhecida mundialmente pelo desenvolvimento de softwares como o sistema operacional Windows e também por aplicações para escritórios (Office), obteve recente

1 <http://www.blog.msbrasil.com.br/a-importancia-do-business-intelligence>

2 <https://exame.abril.com.br/carreira/melhor-profissao-do-ano-nos-eua-deve-explodir-no-brasil-entenda/>

destaque com sua ferramenta de Business Intelligence³. A principal motivação do trabalho é avaliar como a ferramenta está aderente ao processo de Business Intelligence e quais são os recursos que justificam o seu destaque, aplicando-a em um cenário real⁴.

1.2 Objetivo

O principal objetivo é a implementação de um projeto de Business Intelligence utilizando dados do Instituto de Segurança Pública do estado do Rio de Janeiro, executando as principais etapas do projeto, incluindo a elaboração de requisitos de negócio, requisitos técnicos e desenvolvimento da solução utilizando a plataforma Microsoft, permitindo a análise dos dados de ocorrências criminais.

Outros objetivos incluem: Apresentação dos conceitos relacionados a Business Intelligence, ressaltando a importância da análise de dados e como as organizações estão utilizando BI para o apoio a decisão; Apresentação dos componentes da solução, metodologias e melhores práticas adotadas por referências na área. Apresentação da plataforma de BI da Microsoft, explorando os seus principais recursos, analisando prós e contras, assim como o seu posicionamento no mercado.

1.3 Organização do texto

O presente trabalho está estruturado em capítulos e, além desta introdução, foi desenvolvido da seguinte forma:

- Capítulo II: Revisão bibliográfica do tema, conceitos de Business Intelligence.
- Capítulo III: Análise da ferramenta de BI Microsoft, outros sistemas e tecnologias disponíveis, considerações técnicas relevantes.
- Capítulo IV: Especificação do projeto utilizando a base de dados do ISP, implementação do modelo de Data Warehouse, implementação dos relatórios, painéis e indicadores.

3 <https://powerbi.microsoft.com/pt-br/blog/gartner-positions-microsoft-as-a-leader-in-bi-and-analytics-platforms-for-ten-consecutive-years/>

4 <http://epoca.globo.com/brasil/noticia/2017/07/crise-da-seguranca-publica-no-rio-de-janeiro.html>

- Capítulo V: Conclusões – Reúne as considerações finais.

2 Conceito de Business Intelligence

2.1 Introdução

Com a explosão da Era da Informação, massivas quantidades de dados são geradas diariamente nas organizações, em seus diversos nichos. A fim de aumentar a competitividade, assim como um melhor entendimento sobre o que essa massa de informação bruta pode proporcionar a uma entidade, a implementação da ideia de Business Intelligence tem o objetivo de capturar, consolidar, organizar, armazenar, distribuir, analisar e providenciar acesso rápido e fácil a esses dados (KIMBALL, 2013). O Business Intelligence analisa dados históricos de uma determinada companhia que são gerados pelo negócio, ou derivada de fontes externas, como condições climáticas ou dados demográficos, para chegar a uma finalidade em particular ou uma linha de negócio (BALLARD, FARREL, GUPTA, MAZUELA, VOHNIK – 2006).

Ainda há o chamado BI2, com a inteligência de negócios, aprofundando em campos nos quais já pisava e entrando em outros que reclamam por suas aplicações. Dados de projetos, de blogs e de redes sociais, e de comportamento das pessoas nos seus mais variados papéis, vêm para se juntar ao mundo do varejo, onde tudo começou. O BI2 também fala de BA-Business Analytics, que mistura o BI em tempo real com mining e análises preditivas. Isso permite a criação de mecanismos inferenciais de negócios com latência quase zero e exige maior conteúdo qualitativo das informações (BARBIERI – 2011).

2.2 Data Warehouse

Enquanto os usuários dos sistemas operacionais fazem as organizações funcionarem, gerando pedidos, atraindo clientes, monitorando status operacionais, entre outras atividades, os usuários de um DW/BI observam a organização funcionando e medem sua performance. Enquanto usuários operacionais lidam com uma transação por vez,

usuários de um DW/BI quase nunca lidam com este cenário (WILEY, 1996). Se tornou claro que um sistema de DW/BI possui diferentes necessidades, clientes, estruturas e ritmos que os sistemas operacionais. Para isso, na década de 80 foi idealizada uma estrutura de dados que suportasse a análise rápida de um grande volume de dados. Com isso, surgiu o conceito de Data Warehouse, traduzido como Armazém de Dados, onde estão presentes os dados operacionais de uma entidade, modelados de uma maneira diferente, a que chamamos de modelagem dimensional. Embora haja quem utilize uma cópia exata dos dados operacionais de uma organização em hardware separado como um Data Warehouse, o seu objetivo principal é resolver as seguintes preocupações dos usuários:

- **Informação acessível facilmente.** O conteúdo de um DW deve ser de fácil entendimento, intuitiva e óbvia para o usuário de negócios
- **Informação consistente.** O conteúdo no DW deve ser crível, sendo montado, organizado e limpo. Consistência também implica em rótulos comuns e definições de termos para toda a organização.
- **Adaptabilidade.** As necessidades, condições e tecnologias estão suscetíveis a mudança. O sistema DW deve estar preparado para tal.

Em termos gerais, um Data Warehouse oferece uma solução de armazenamento de dados centralizado, contendo um grande volume de informações históricas, otimizado para operações de leitura e confiável para a tomada de decisão. Embora tecnologia seja um fator determinante para a construção de um sistema deste tipo, ela é apenas um meio. As necessidades de negócio deverão ser a maior prioridade do projeto, portanto captar as regras de negócio em uma modelagem adequada é a maior responsabilidade de um Data Warehouse.

2.3 Componentes de uma solução de Data Warehouse

Uma solução de Data Warehouse geralmente consiste nos seguintes elementos:

- **Data Sources.** Origem de informações de negócio para o Data Warehouse (DW), incluindo bancos de dados de aplicações transacionais (OLTP) e dados exportados de sistemas proprietários.
- **Processo de ETL.** Um fluxo para acesso aos dados nos Data Sources, modificando-os para conformidade com o modelo de dados do DW, e carregando-os no DW.

- **Áreas de preparação de dados (Data Staging).** Áreas intermediárias para onde os dados são transferidos e armazenados, preparados para importação no DW.
- **Data Warehouse.** Um banco de dados relacional construído para fornecer consultas de dados históricos, com alta performance, para análise.
- **Modelos de dados (Data Models).** Construídos a partir dos dados do DW, modelos de dados serão utilizados por um analista para produzir informações para o negócio.
- **Ferramentas de Visualização.** Estas ferramentas analíticas (OLAP) acessam os modelos de dados e o Data Warehouse, permitindo a construção de relatórios, métricas, painéis e indicadores para o negócio.

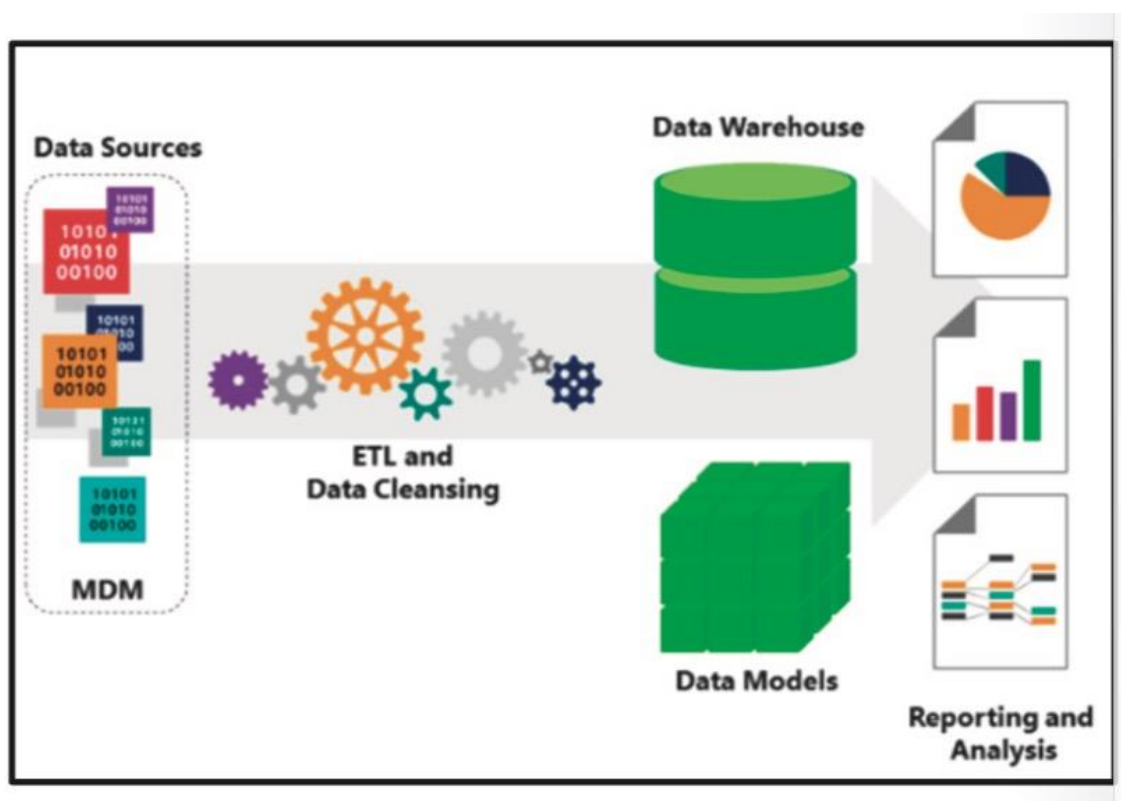


Figura 1 - Componentes básicos de BI (MICROSOFT, 2017)

2.4 Modelagem Dimensional

A modelagem dimensional é amplamente aceita como a técnica preferida para apresentação de dados analíticos porque aborda dois requisitos simultâneos: Entregar dados que são intuitivos aos usuários de negócio, e entregar respostas rápidas às queries (KIMBALL – 2011).

Estruturas de bancos de dados relacionais (SQL) operacionais geralmente utilizam uma modelagem normalizada, na qual podemos classificar em Formas Normais (1FN, 2FN, 3FN, entre outras). Essas formas buscam remover a redundância de dados e garantir que a operação seja otimizada, dividindo os dados entre várias estruturas discretas, as quais chamamos de entidades relacionais. Entretanto, modelos normalizados podem ser muito complexos, aumentando também a complexidade de uma consulta feita por um usuário, podendo impactar negativamente a operação em andamento.

Modelos dimensionais implementados em um sistema gerenciador de banco de dados relacional (SGBD) ou aplicações OLAP (*Online Analytical Processing*) são geralmente classificados como esquemas estrela ou esquemas floco de neve, devido a sua estrutura. A figura 2 a seguir demonstra um esquema estrela e um cubo OLAP.

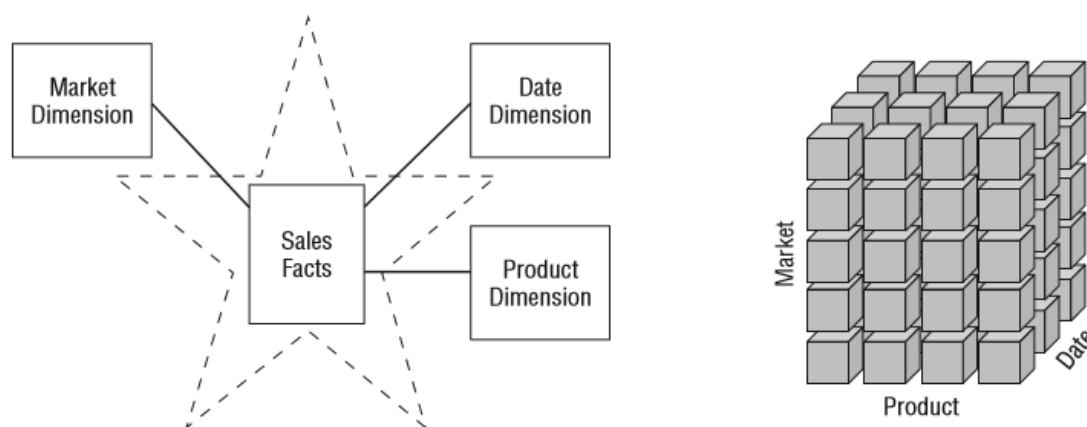


Figura 2 - Modelo estrela e cubo OLAP (KIMBALL, 2013)

O objetivo de ambos os esquemas é a desnormalização, que ocorre quando o modelo apresenta certa redundância de dados visando a performance. Um esquema estrela possui maior nível de desnormalização que um esquema floco de neve. A tabela a seguir compara os diferentes níveis encontrados em soluções de Data Warehouse:

Tabela 1 - Comparativo de normalização de banco de dados

Tipo de Esquema	Estrutura da Tabela	Vantagens	Desvantagens

Completamente normalizado	Contém o ID e Descrição da tabela e a chave FK da tabela pai	Requer pouco armazenamento em disco; não há redundância	Exige muitos joins
Moderadamente desnormalizado	Contém o ID e Descrição da tabela e todas as chaves FKs das tabelas pai de nível acima	Reduz o número de joins	Requer mais armazenamento em disco e controle de redundância
Totalmente desnormalizado	Contém o ID e Descrição da tabela e todas as chaves FKs e descrições das tabelas pai de nível acima	Reduz totalmente o número de joins	Requer muito armazenamento em disco e aumenta a complexidade do controle de redundância

O modelo dimensional e seu nível de desnormalização deve considerar os dois componentes principais: Fatos e Dimensões.

2.5 Fatos e Métricas

A tabela Fato em um modelo dimensional armazena os dados de métricas do negócio. Corresponde ao maior volume de dados de um Data Warehouse, já que este tipo de tabela possui granularidade que corresponde ao fato que está sendo observado. Por exemplo, uma área de negócios gostaria de acompanhar volume de vendas em todas as suas filiais. A figura 3 exemplifica este cenário:

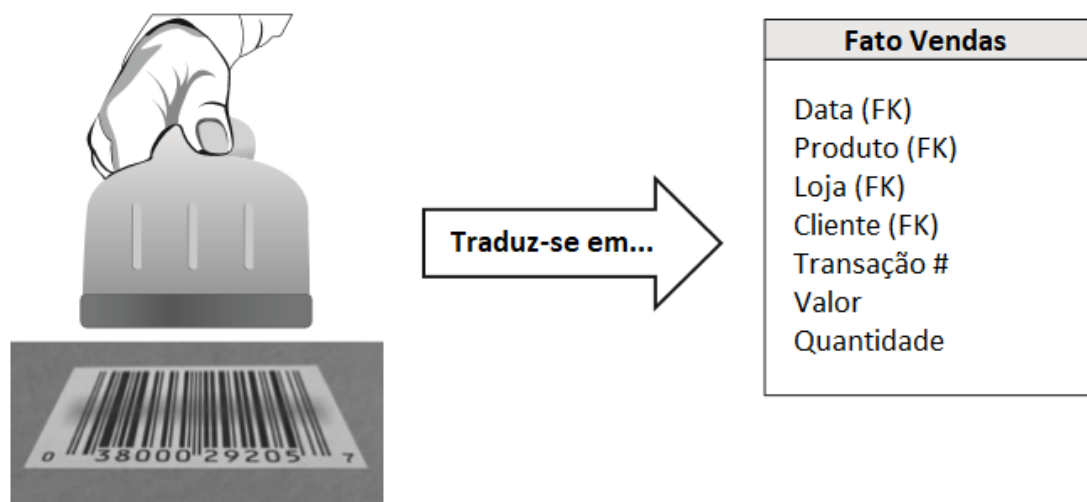


Figura 3 - Exemplo de tabela Fato (KIMBALL, 2013)

Neste caso, para cada evento de Venda teremos um registro na tabela Fato Vendas, sendo as métricas “Valor” e “Quantidade”. Poderíamos realizar algumas agregações, de modo a reduzir o volume de dados armazenado, porém a sua granularidade é um requisito dos usuários de negócios.

O conjunto de tabelas Fato correspondem a cerca de 90% do volume total de um Data Warehouse, e geralmente suas chaves primárias (Primary Key) são compostas de todas as chaves estrangeiras (Foreign Key), ou de atributos que garantem a unicidade do registro (KIMBALL – 2011). Elas tendem a possuir maior quantidade de linhas e menor quantidade de colunas, e podem ser classificadas em vários tipos: Transacionais, Agregadas, Periódicas, Acumulativas, Fato sem fato. Cada tipo possui diferentes níveis de granularidade e periodicidade de carga de dados.

2.6 Dimensões

No exemplo da figura 1.1, temos os atributos “Data”, “Produto”, “Loja” e “Cliente” representados na tabela “Fato Vendas” como chaves estrangeiras (FKs). Apesar da tabela fato possuir apenas um identificador destes atributos, estes são chamados de dimensões, pois descrevem as circunstâncias dos quais ocorreram os fatos.

Tabelas de dimensão contém dados textuais associados a um evento/fato. Eles descrevem o “quem, o que, onde, quando, como e porque” o fato ocorreu (KIMBALL – 2013). A figura 1.2 exemplifica uma tabela de dimensão:

Dimensão Produto
Chave do Produto (PK/SK)
Chave Natural do Produto (NK)
Descrição do Produto
Tipo do Produto
Marca
Peso
Largura
Comprimento
...

Figura 4 - Exemplo de dimensão (KIMBALL, 2013)

Uma dimensão possui certos atributos que facilitam sua manutenção. Uma chave primária própria, chamada de Surrogate Key, geralmente é um número inteiro gerado automaticamente pelo SGBD. Uma chave natural é o atributo que garante a unicidade da dimensão no Data Warehouse. Com isso podemos realizar alterações nos valores das dimensões sem precisar alterar as tabelas Fato que a dimensão se relaciona. Existem diferentes tipos de dimensões que podem ser implementadas em um Data Warehouse:

- **Degenerate Dimension.** Dimensão que foi inserida como coluna em uma Fato, mas não possui uma tabela própria.
- **Slowly changing Dimension.** Dimensão que possui comportamento específico ao ser atualizada. Algumas SCDs podem ser programadas para atualizem o registro, enquanto outras geram um novo, mantendo o histórico de alterações. Todas as dimensões são SCD, exceto a dimensão Tempo, pois este não precisa ser atualizado.
- **Role-Playing Dimension.** É uma dimensão que desempenha múltiplos papéis em uma tabela fato. Tempo pode ser uma dimensão Role Playing, pois uma tabela fato pode possuir diferentes dimensões de tempo.
- **Conformed Dimension.** Ralph Kimball define como uma dimensão que possui o mesmo significado para todas as tabelas fato com que se relaciona.
- **Junk Dimension.** Junção de várias dimensões com cardinalidade pequena em uma única tabela.

Embora as dimensões possam ter muitos atributos, os sistemas de BI deverão ser responsáveis por agregar os dados de acordo com os atributos escolhidos pelo usuário.

3 Plataforma de BI Microsoft

3.1 Introdução

Em um mercado repleto de soluções de Business Intelligence, algumas delas se destacam como líderes de mercado, títulos adquiridos pelas inovações que trazem ou pela capacidade de execução, ou seja, cumprem bem o papel que prometem. Neste capítulo, serão demonstrados os recursos da plataforma Microsoft. A plataforma possui os seguintes componentes principais:

- Banco de dados relacional SQL Server,
- Ferramenta de integração de dados (ETL) Integration Services
- Ferramenta OLAP Analysis Services
- Power BI, Excel e Reporting Services.

3.2 SQL Server

O SQL Server é um sistema gerenciador de bancos de dados relacional (SGBD) desenvolvido pela Microsoft em parceria com a Sybase em 1988. O SQL Server inclui, além dos componentes presentes em bancos de dados relacionais, outros componentes e recursos específicos que podem ser utilizados para implementar elementos de uma arquitetura de Data Warehouse.

Columnstore Index

Um índice é uma estrutura em disco associada a uma tabela ou exibição, que agiliza a recuperação das linhas de uma tabela ou exibição. Um índice contém chaves criadas de uma ou mais colunas da tabela ou exibição. Essas chaves são armazenadas em uma estrutura (árvore B) que habilita o SQL Server a localizar a linha ou as linhas associadas

aos valores de chave de forma rápida e eficaz (MICROSOFT - 2017). Os índices podem ser classificados em dois tipos:

- **Clusterizados.** Os índices clusterizados classificam e armazenam as linhas de dados com base em seus valores de chave.
- **Não clusterizados.** Os índices não clusterizados tem uma estrutura distinta das linhas de dados, contendo os valores de chave do índice, onde cada valor possui um ponteiro para a linha de dados que contém o valor da chave.

A Microsoft introduziu o Columnstore Index (índices baseados em coluna) no SQL Server 2012. Este novo tipo de índice referencia os dados em uma maneira colunar e utiliza compressão até dez vezes maior do que um índice baseado em linha, para redução de espaço em disco utilizado e de operações de I/O necessárias para consultas (queries). Muitos SGBDs voltados para Data Warehouse, como o Amazon Redshift, possuem estruturas de índices semelhantes. A tabela a seguir mostra um comparativo com os índices baseados em linha:

Tabela 2 - Comparativo de índices

Índice baseado em linha	Índice baseado em coluna
Pode ser clusterizado e não-clusterizado	Pode ser clusterizado e não-clusterizado
Melhor performance em operações feitas em linhas, inserções e updates.	Melhor performance em operações feitas em uma tabela, como agregações ou queries analíticas.
Melhor utilização em bancos OLTP	Melhor utilização em Data Warehouses
Todos os dados em uma linha são processados	Apenas as colunas necessárias são processadas

Os índices podem ser combinados para utilização dos benefícios que ambos oferecem, porém isto vem com um custo, neste caso, de espaço utilizado. Quando se está procurando por um dado específico, índices baseados em linha geralmente funcionam melhor. Quando se está processando muitos dados em uma tabela, índices baseados em coluna geralmente são a melhor opção. Por exemplo, em um Data Warehouse utilizando funções de agregação para retornar a soma de todas as vendas em uma região. Portanto, conhecer os objetivos dos usuários do sistema facilita a escolha.

Tabelas Temporais

O recurso de tabelas temporais foi introduzido no SQL Server 2016, que oferece suporte para fornecer informações sobre um registro (linha) em qualquer ponto do tempo, funcionando de maneira similar a um controle de versão. Este recurso é útil especialmente em Data Warehouses que possuem Slowly Changing Dimensions, um tipo de dimensão em que é necessário manter histórico de alterações realizadas. A figura 5 mostra o fluxo dos dados nas tabelas temporais:

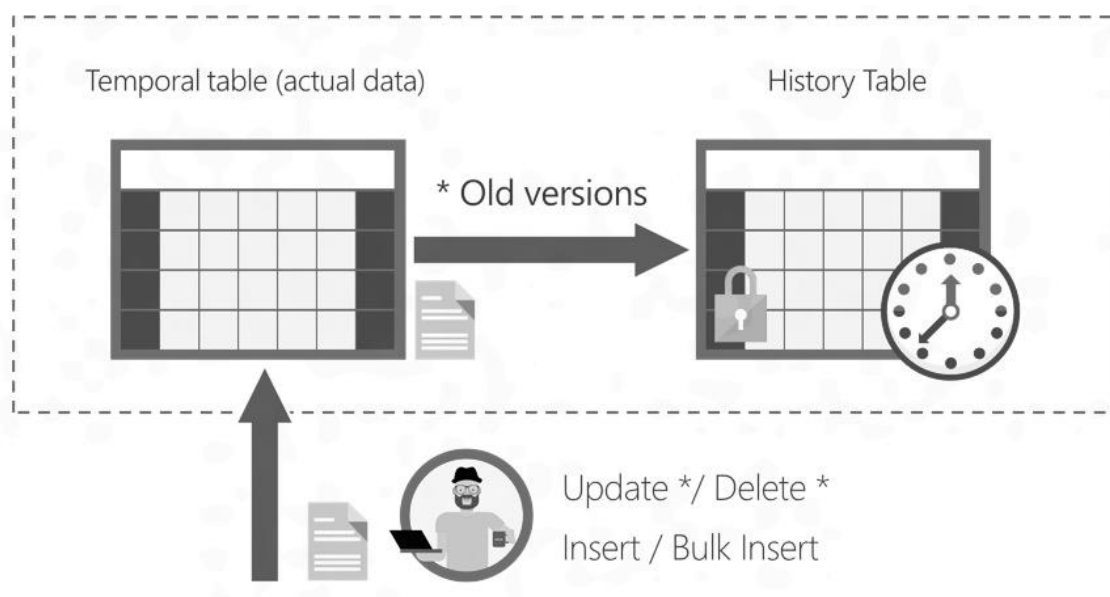


Figura 5 - Tabelas Temporais no SQL Server (MICROSOFT, 2017)

Os dados recentes ficam armazenados na tabela Temporal enquanto os dados que já foram atualizados ficam armazenados em uma tabela de histórico. Sem este recurso teríamos que utilizar outros meios para processar estes dados. Isso é útil, por exemplo, em uma dimensão de Produtos, onde se pode obter o preço unitário de um item ao longo do tempo.

Esquema de Banco de Dados e Filtros Bitmap

Existem diversas metodologias para construção de um esquema de Data Warehouse. O esquema-estrela, um tipo de modelagem de dados dimensional proposta por Kimball, é amplamente utilizada atualmente. Embora O SQL Server possua meios para otimizar consultas realizadas em qualquer tipo de esquema, consultas construídas para Data

Warehouses geralmente processam milhões de linhas. O otimizador de consultas do SQL Server detecta automaticamente esquemas-estrela e cria planos de consulta mais eficientes. Um método utilizado para gerar um plano de execução eficiente é através da Filtragem Bitmap, que utiliza uma representação compacta de um conjunto de valores de uma tabela em uma parte da árvore de operação, para filtragem de linhas em uma segunda tabela em uma outra parte da árvore. Essencialmente, o filtro aplica um “*semi-join*”, ou seja, apenas as linhas na segunda tabela que estão qualificadas para a junção são processadas. Este tipo de operação pode melhorar de maneira significativa a performance de uma consulta que utilizam esquemas-estrela, removendo as linhas de uma tabela Fato que não se qualificam, logo no início do plano de execução da consulta. Sem a otimização do Bitmap, todas as linhas da tabela Fato são processadas através de uma parte da árvore de operação antes do “*join*” com as dimensões.

Outros recursos

O SQL Server conta com outros recursos para bancos de dados de maneira geral, como alta disponibilidade, processamento massivo em paralelo (Parallel Data Warehousing), tabelas em memória, recuperação de desastres, segurança em nível de linha (Row Level Security) entre outros, que também são importantes para manter a saúde do ambiente de Data Warehouse, fornecendo escalabilidade e segurança.

A Microsoft também conta com os serviços em sua nuvem privada, a Azure, e conta com um produto de banco de dados específico para BI, o Azure SQL Data Warehouse, uma aplicação de SQL Server com os recursos pré-configurados para permitir uma implantação mais rápida de projetos de Data Warehouse. O SQL Server também possui integração com as linguagens R e Python, muito utilizada por cientistas de dados, para aplicações preditivas ou machine learning.

3.3 Integration Services

Uma parte significativa de uma criação de um Data Warehouse é a implementação de um ou mais processos de ETL (Extração, Transformação e Carga de dados). A Microsoft introduziu sua ferramenta de ETL, o Integration Services (SSIS), junto ao SQL Server 2005, onde substituiu sua ferramenta anterior, o DTS (Data Transformation Services). O SSIS é uma plataforma para integração de dados e workflow de aplicações.

Possui diversos recursos para integração de dados com Data Warehouses e outras fontes de dados - Bancos de dados OLTP, arquivos, serviços Web, entre outros.

Recursos

O SSIS é uma plataforma incluída com o SQL Server e consiste de um serviço que pode ser instalado à parte, que gerencia a execução de fluxos de ETL. Desenvolvedores podem criar fluxos de carga de dados utilizando o Visual Studio Data Tools, uma IDE, também da Microsoft, voltada para aplicações de Business Intelligence. O principal componente de um projeto do SSIS é chamado de Pacote, e contém uma variedade de elementos que definem um fluxo de trabalho:

- **Conexões.** Uma conexão inclui todas as propriedades para uma fonte de dados. As conexões de um projeto podem ser reaproveitadas em diversos pacotes.
- **Eventos.** Um pacote possui tarefas que podem disparar determinadas ações, de acordo com o seu comportamento. Estas ações são chamadas de eventos.
- **Parâmetros.** Parâmetros permitem a atribuição de valores à propriedades ou variáveis de um pacote, no tempo de execução.
- **Restrições.** As tarefas de um pacote podem ser relacionadas através de restrições de precedência. A restrição permite a execução de tarefas de maneira organizada.
- **Tarefas.** Tarefas são ações executadas em um fluxo de um pacote. Neste caso, temos duas categorias de tarefas: Tarefas de Controle e Tarefas de Fluxo de Dados. As tarefas de controles são responsáveis pela ordem do fluxo como um todo, enquanto as tarefas de Fluxo de Dados são as responsáveis pelo processo de ETL do produto.
- **Variáveis.** As variáveis de um pacote permitem armazenar resultados, tomar decisões ou afetar a configuração do pacote.

A tarefa do tipo “Fluxo de Dados” possui diversas transformações disponíveis para leitura, transformação e gravação de dados. Algumas dessas transformações são:

- **Origem e Destinos.** Este componente permite selecionar as fontes de dados de origem e de destino do fluxo.
- **Conversão de Dados.** Converte o tipo de dados de uma coluna.

- **Merge Join.** Realiza a operação de “*join*” entre duas fontes de dados distintas.
- **Coluna Derivada.** Permite a criação de colunas a partir de expressões ou condições.
- **Slowly Changing Dimension.** Permite a atualização de dimensões com base em colunas-chave da mesma.
- **Agregação.** Realiza operações de agregação, como Soma, Média, Contagem, Máximo e Mínimo de um fluxo de dados.
- **Lookup.** Compara os dados de um fluxo com os dados existentes em outra tabela, para recuperação de valores ou redirecionamento do fluxo.
- **Auditoria.** Cria atributos de controle para facilitar a auditoria dos dados.

Os desenvolvedores também podem criar tarefas de script em C# ou Visual Basic, utilizando o componente Tarefa de Script, permitindo maior customização dos fluxos. Embora seja possível realizar processos de ETL utilizando SQL, Stored Procedures ou Cópias entre bancos de dados, o SSIS permite maior organização, escalabilidade, e oferece uma interface de fácil uso para seus desenvolvedores.

Em geral, o SSIS é utilizado para aplicações de BI, pois bancos de dados de Data Warehouse geralmente possuem janelas para cargas de dados. Porém, ele pode ser utilizado para outros fins, em sistemas que possuem como característica uma carga de dados baseada em lotes. Para dados em tempo real, onde a latência máxima permitida é de menos de um minuto, outra solução deverá ser considerada.

3.4 Analysis Services

O SQL Server Analysis Services (SSAS) é uma plataforma de processamento analítico (OLAP) que fornece dados sobre os quais os relatórios podem ser construídos. É possível utilizá-lo para criar e gerenciar modelos de dados, e agregar dados de múltiplas fontes. Foi lançado no ano 2000 e desde então sofreu diversas modificações. Inicialmente, o Analysis Services suportava apenas modelos Multidimensionais (cubos OLAP). Atualmente, é possível criar dois tipos diferentes de modelos semânticos: Multidimensional ou Tabular (modelagem Relacional). Os modelos semânticos permitem que os usuários analisem os dados sem precisar entender as complexidades de uma camada de banco de dados. Os modelos expõem os dados em um formato que os

usuários podem interagir mais facilmente, acessando-os através de ferramentas de visualização – incluindo o Power BI e outras de diferentes fornecedores – ou através de ferramentas de escritório como o Excel, SharePoint, etc. A figura 6 mostra os componentes do Analysis Services.

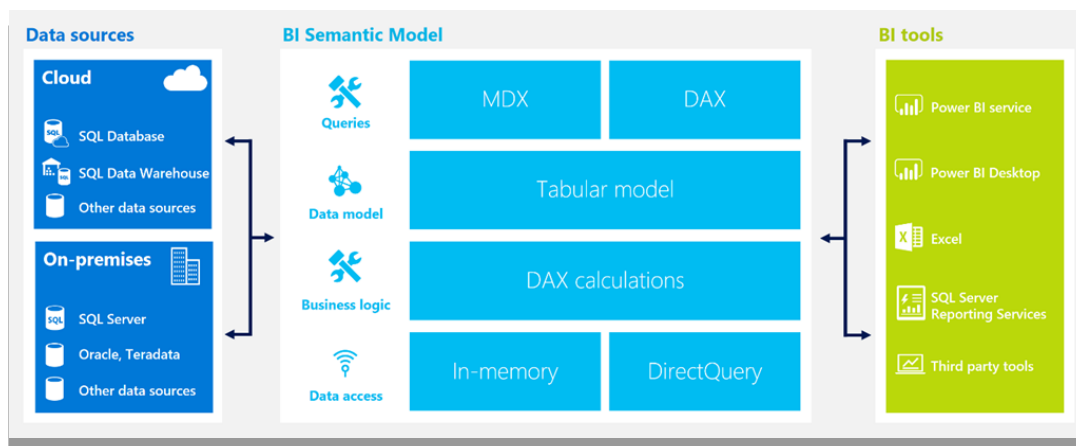


Figura 6 - Analysis Services e componentes (MICROSOFT, 2017)

Modelo Multidimensional

Este tipo de modelo está presente em todas as versões do SSAS. Modelos multidimensionais expõem os dados através de dimensões e cubos. Ele utiliza a linguagem MDX (Multidimensional Expressions) para implementação de lógicas de negócios, e fornece acesso aos dados através de armazenamento relacional (ROLAP) ou multidimensional (MOLAP). O método ROLAP acessa os dados diretamente da fonte de dados, enquanto o método MOLAP armazena cubos pré-processados, para maior performance nas consultas.

Modelo Tabular

O modelo tabular foi introduzido no SSAS 2012. Ele é construído a partir de tabelas e seus relacionamentos, tornando-o mais fácil de trabalhar para profissionais sem experiência com modelos multidimensionais. O modelo tabular utiliza a linguagem DAX (Data Analysis Expressions) para implementação de lógicas de negócio, e fornece acesso aos dados utilizando o mecanismo de processamento em memória, chamado xVelocity, ou o DirectQuery, para acesso direto aos dados do Data Warehouse.

3.5 Excel, Reporting Services e Power BI

O front-end de uma arquitetura de Business Intelligence é o ponto de acesso, a camada da plataforma onde os usuários irão interagir com o Data Warehouse, ou qualquer outra fonte de dados para análise, gerar as métricas e indicadores para a tomada de decisão, e compartilhar a informação a nível organizacional. Em geral, estas ferramentas suportam a criação de relatórios e painéis interativos, com diferentes tipos de visualizações, como KPIs, gráficos, mapas, entre outros

Excel

A Microsoft possui o já bem estabelecido Excel, um dos softwares mais utilizados em ambientes corporativos, e também por diversos profissionais e acadêmicos da área da informação para suportar atividades relacionadas a análise de dados. É possível se conectar ao Analysis Services, SQL Server ou outras fontes, e utilizando os recursos existentes da ferramenta é possível obter insights para o apoio a tomada de decisão. É possível utilizar também os componentes extras do Excel para estender sua capacidade como ferramenta de BI:

- **PowerPivot.** Permite a criação de modelos tabulares que armazenam os dados em planilhas. Utilizando esta abordagem é possível criar cálculos complexos, sem precisar estar conectado a fonte de dados. Utiliza a linguagem DAX, similar ao modelo Tabular do Analysis Services.
- **PowerView.** Ferramenta de visualização que permite a criação de relatórios utilizando diversos tipos de gráficos e tabelas, chamados de visualizações, em que é possível mostrar a mesma informação de diversas maneiras.
- **PowerMap.** Ferramenta de visualização de informações geográficas

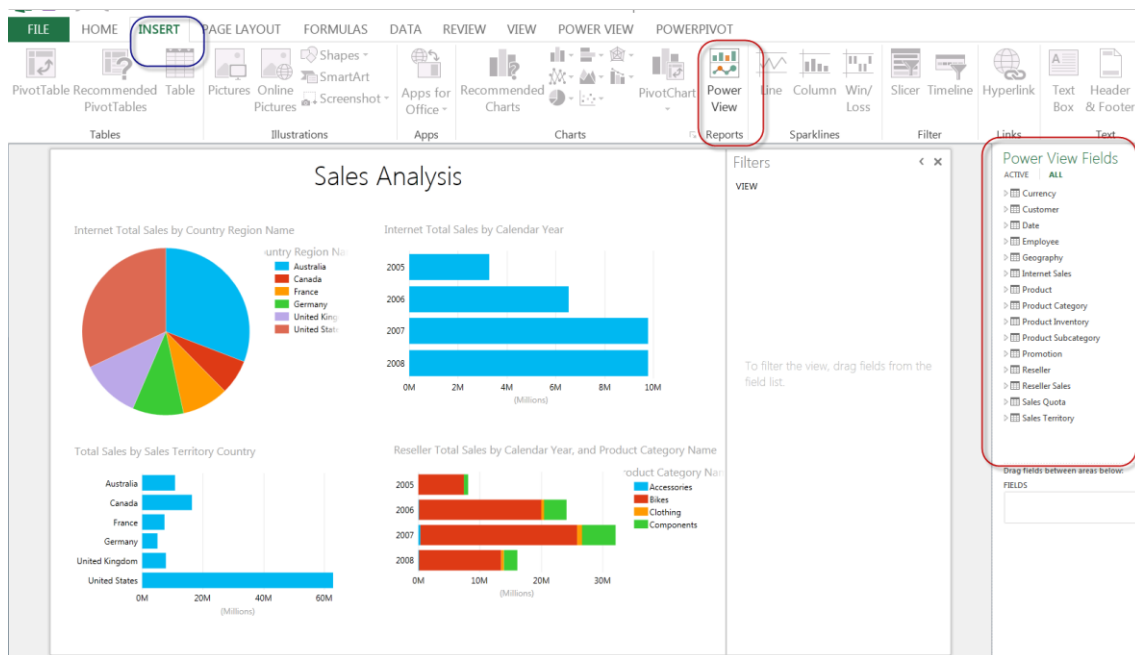


Figura 7 - Excel, PowerView e PowerPivot (MICROSOFT, 2017)

Reporting Services

Embora o Excel possua recursos avançados de visualização e exploração de dados, ainda se torna difícil manter um ambiente de Business Intelligence com relatórios centralizados e distribuídos. Para isso, o SQL Server Reporting Services (SSRS) fornece um ambiente de gerenciamento de ciclo de vida de relatórios. Usuários conseguem publicar e seguir relatórios, que por sua vez podem ser distribuídos em diversas maneiras através do serviço de compartilhamento do SSRS, e administradores podem definir alertas e permissões de visualização e de escrita. Com o Reporting Services também é possível criar relatórios e painéis sem a utilização do Excel. Ele utiliza o Report Designer, parte do Visual Studio Data Tools, ou o Report Builder, que possui uma interface similar à interface do Office, porém não depende do mesmo.

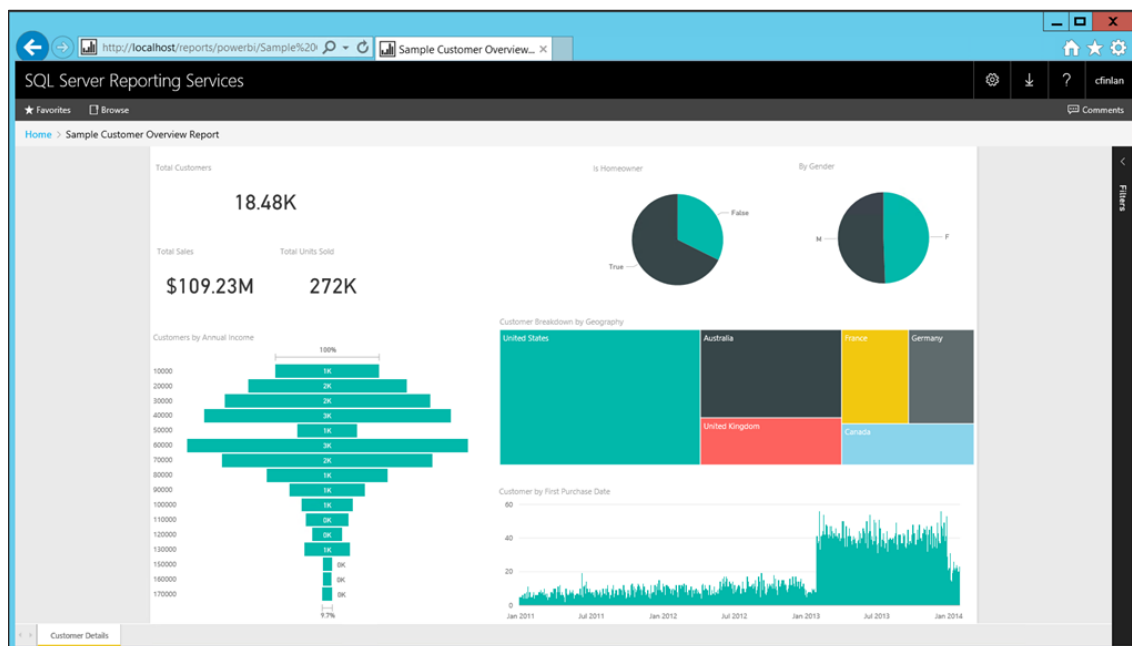


Figura 8 - Reporting Services na Web (MICROSOFT, 2017)

Power BI

O Power BI é um serviço de Business Intelligence que oferece visualizações interativas com capacidades de self-service, ou seja, os usuários finais podem criar os seus relatórios e painéis sem a necessidade de conhecimentos avançados de banco de dados. Originalmente baseado nos add-ins do Excel (Power Query, Power Pivot e Power View), a Microsoft desenvolveu o produto visando oferecer um serviço independente, baseado em nuvem, com mais recursos que as aplicações originais do Excel. Seu lançamento foi em Setembro 2013 com o serviço do Office 365, e depois lançado ao público geral em 2015. De acordo com a especificação do fornecedor, seus principais recursos são:

- **Integração com o Office 365.** O Power BI oferece uma integração nativa com outros aplicativos da plataforma Office 365, como o próprio Excel, SharePoint, Outlook, etc.
- **Aplicação Desktop, Mobile e Web.** Usuários constroem relatórios e painéis no Power BI Desktop, realizando a preparação de dados e utilizando diversos tipos de visualizações nativas, ou customizadas, publicando-os no serviço do Power BI na nuvem, tornando o acesso possível de qualquer navegador ou dispositivo móvel.

- **Suporte a diversas conexões de dados.** Um dos pontos fortes da ferramenta, o Power BI oferece uma vasta gama de conectores de dados: MS SQL Server, Oracle, Amazon Web Services, Azure, Hadoop, Spark, Salesforce, Facebook, Analysis Services, arquivos texto, Excel, entre outros.

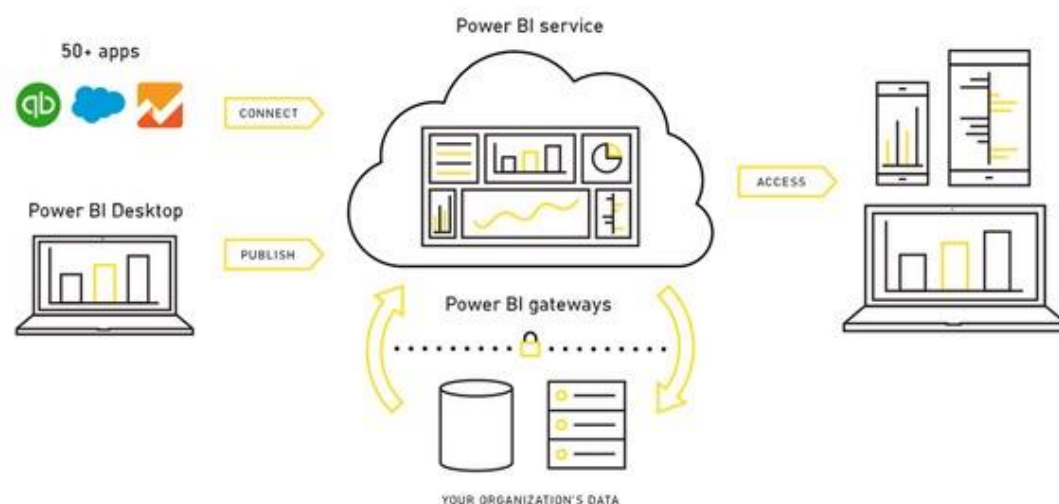


Figura 9 - Visão geral do funcionamento do Power BI (MICROSOFT, 2017)

3.6 Posicionamento da Microsoft em pesquisas de mercado

Para ajudar na escolha de ferramentas, grupos como o Gartner e Forrest realizam análises dos recursos de cada ferramenta e comparando-as e destacando os pontos fortes e fracos. A Microsoft está posicionada como líder há mais de 10 anos consecutivos no segmento de Analytics Platform de acordo com a Gartner, renomada empresa de pesquisa da área de tecnologia. Todos os anos, a Gartner disponibiliza uma avaliação de diversas soluções do mercado, de diferentes categorias (BI, ERPs, CRMs, etc.) classificando-as em: Líderes, Visionários, Concorrentes, e Fornecedores de Nicho. A figura a seguir mostra o Quadrante Mágico para Soluções de BI e Analytics de 2017⁵:

5 <https://www.gartner.com/doc/3611117/magic-quadrant-business-intelligence-analytics>

Figure 1. Magic Quadrant for Business Intelligence and Analytics Platforms



Figura 10 - Quadrante Gartner para ferramentas de BI (GARTNER, 2017)

De acordo com o quadrante, a Microsoft está posicionada como líder, junto com as ferramentas de BI das empresas Tableau e Qlik. Para tal classificação, diversos fatores são julgados, como: Facilidade de uso, recursos de visualização, customização, performance, entre outros. Entretanto, é importante lembrar que algumas ferramentas possuem recursos além da capacidade de visualização e exploração, como integração de dados (ETL), análise estatística avançada, e até banco de dados. Esta categoria do quadrante inclui os fornecedores de ferramentas analíticas e de visualização, excluindo do estudo os recursos de bancos de dados e ETL.

O Gartner também apontou o SQL Server em 2017 como líder em soluções para Gerenciamentos de Dados para Analytics, por melhorias em recursos existentes e também por conta de novos recursos, como Índices baseados em Colunas (Columnstore Index) e Tabelas Temporais, competindo com grandes fornecedores tradicionais de BI como Teradata, IBM, SAP, Oracle, e também a novos competidores como Amazon.

Figure 1. Magic Quadrant for Data Management Solutions for Analytics



Figura 11 - Quadrante Gartner para Banco de Dados (GARTNER, 2017)

O Gartner, que também avalia as ferramentas de integração de dados, colocou a Microsoft no quadrante de Concorrentes no ano de 2017:

Figure 1. Magic Quadrant for Data Integration Tools



Figura 12 - Quadrante Gartner para ETL (GARTNER, 2017)

4 Implementando o projeto de BI do ISP

4.1 Introdução a análise de segurança pública

Para implementar um projeto de Business Intelligence, o grupo pesquisou por uma área de assunto relevante para poder explorar a capacidade das ferramentas da Microsoft. A questão da segurança pública do Estado do Rio de Janeiro é noticiada diariamente através das mídias sociais e outros meios de comunicação, e muito se discute a eficiência dos órgãos públicos para organizar o policiamento ostensivo, que hoje está sob responsabilidade da Polícia Militar, e o policiamento judiciário, responsabilidade da Polícia Civil. Para entender estes fatos, foi necessário utilizar uma base de dados confiável. O Instituto de Segurança Pública (ISP) reúne informações de registros criminais e de atividade policial do estado do Rio de Janeiro. As estatísticas divulgadas são construídas a partir dos Registros de Ocorrência (RO) lavrados nas delegacias de Polícia Civil do Estado do Rio de Janeiro, além de informações complementares. As informações produzidas baseiam-se na data em que foi confeccionado o Registro de Ocorrência e, além disso, informações relevantes como população, divisão territorial da base de segurança, informações institucionais, etc. Todos os dados estão de acordo com a Lei de Acesso à Informação. A motivação de criação desse projeto é demonstrar os recursos da plataforma de BI da Microsoft, aplicando dados reais do ISP.

4.2 Entendendo as ocorrências criminais do Rio de Janeiro

Diariamente, milhares de ocorrências são registradas em delegacias espalhadas pelo Estado, e para monitorar as metas e ações planejadas estas delegacias estão organizadas em regiões administrativas, denominadas RISP (Regiões Integradas de Segurança Pública). Cada RISP contém divisões internas, na intenção de facilitar o compartilhamento de informações e promover a sinergia entre a Polícia Civil e Polícia Militar.

O ISP funciona como o ponto de convergência dessas áreas, onde sua principal responsabilidade é administrar as regiões e seus recursos como um todo. Os dados coletados a partir dos registros de ocorrência das delegacias e informações geográficas contribuem para a análise do histórico de crimes de diversos tipos ocorridos em todo o Estado.

Tipos de ocorrências criminais

Os tipos de ocorrências registradas nas RISP's do Estado são classificados em:

- **Crimes violentos.** Homicídio doloso, lesão corporal seguida de morte, latrocínio (roubo seguido de morte), tentativa de homicídio, lesão corporal dolosa, estupro
- **Crimes de trânsito.** Homicídio culposos, lesão corporal culposa.
- **Crimes contra o patrimônio.** Roubo, furto, sequestro, extorsão, estelionato.
- **Atividade policial.** Apreensão de drogas, apreensão de armas, flagrante, recuperação de veículos
- **Outros registros.** Ameaça, pessoas desaparecidas, policiais mortos em serviço

O aumento do número de homicídios⁶ se dá pelos constantes confrontos entre policiais e facções criminosas, e também a outros fatores menos noticiados, como o feminicídio - morte de mulheres em razão de seu sexo.

4.3 Elaborando os requisitos da solução para o ISP

O projeto surgiu da necessidade de unificar - em uma única solução de bancos de dados - informações de registros de ocorrências do ISP, que atualmente encontra-se de maneira dispersa em seu website. Para a geração de indicadores, é necessário desenhar uma solução para integrar os dados disponibilizados de maneira mais intuitiva, para que possa ser traçado um perfil dos crimes ocorridos no estado do Rio de Janeiro.

O que o sistema de Business Intelligence deve responder

6 <http://agenciabrasil.ebc.com.br/geral/noticia/2017-07/dados-do-isp-confirmam-aumento-da-violencia-no-rio-no-primeiro-semester-de>

O grupo elaborou um questionário, se colocando no lugar de um usuário final e de stakeholders do projeto, e elaborou as perguntas de acordo com os dados disponibilizados e o que poderia ser extraído deles:

- **Quais são as ocorrências mais comuns e onde são registradas?** Responder a esta pergunta permite a mobilização de forças policiais para realizar um patrulhamento ostensivo de curto prazo, dos locais com maior incidência de crimes.
- **Quais regiões e sub-regiões do Estado do Rio de Janeiro possuem uma maior quantidade de registros de ocorrência?** Responder esta pergunta permite um gerenciamento das regiões administrativas do Estado. Em um nível mais gerencial, entender o mapa do crime permite o planejamento de recursos orçamentários a longo prazo.
- **Quando foi registrado uma maior quantidade de ocorrências? Em quais épocas do ano isso se repete?** A sazonalidade das ocorrências registradas pode ser um fator determinante para um melhor remanejamento de forças policiais. Por exemplo, em períodos com maior quantidade de turistas e estrangeiros, em período de férias ou estações do ano específicas, a prática de roubos e furtos pode subir.
- **Qual o tempo médio entre a data de ocorrência do fato e sua comunicação?** Muitos crimes são comunicados dias após a ocorrência do fato. Entender como este tempo influencia no desfecho da ocorrência é necessário para realizar ações de prevenção e conscientização.
- **Qual o perfil das vítimas de Letalidade registrados?** O Rio de Janeiro é um dos Estados brasileiros com o maior índice de homicídios registrados. Entender o perfil das vítimas pode ser útil para prevenir crimes deste tipo.
- **Qual o perfil das vítimas de Violência Contra a Mulher?** Grande parte dos homicídios registrados possuem como vítimas pessoas do sexo feminino. Ainda existem muitos crimes não notificados⁷. Entender o perfil da vítima pode facilitar a prevenção de violência contra a mulher.

7 <https://oglobo.globo.com/rio/estudo-mostra-que-rio-tem-subnotificacao-de-feminicidios-1-22008241>

A figura 8 a seguir demonstra os diversos arquivos encontrados no website do ISP⁸:


































 38SeriePoliciaisMilitaresMortos2017.xls	13/08/2017 18:22	Planilha do Micro...	248 KB
 39SeriePoliciaisCivisMortos2016.xls	13/08/2017 18:23	Planilha do Micro...	240 KB
 42SerieTotaldeRO2017.xls	13/08/2017 18:21	Planilha do Micro...	260 KB
 AispPorRegiao.csv	13/08/2017 18:14	Arquivo CSV	181 KB
 Armas.csv	13/08/2017 18:56	Arquivo CSV	13.643 KB
 Armas2.csv	04/09/2017 23:14	Arquivo CSV	13.333 KB
 BaseDpDicionarioDeVariaveis.xlsx	13/08/2017 18:16	Planilha do Micro...	23 KB
 BaseDPEvolucaoMensualCisp.csv	13/08/2017 18:11	Arquivo CSV	3.985 KB
 BaseUppDicionarioDeVariaveis.xlsx	13/08/2017 18:16	Planilha do Micro...	12 KB
 ChefeDePolicia.csv	07/08/2017 21:41	Arquivo CSV	1 KB
 ComandanteGeral.csv	07/08/2017 21:41	Arquivo CSV	1 KB
 ComandantesBatalhao.csv	07/08/2017 21:41	Arquivo CSV	321 KB
 CorrespondenciaCispAisp.csv	13/08/2017 18:14	Arquivo CSV	409 KB
 CorrespondenciaCispMunicipioCod.csv	13/08/2017 18:15	Arquivo CSV	10 KB
 CorrespondenciaCispMunicipioCodAoLongoDoTe...	13/08/2017 18:14	Arquivo CSV	968 KB
 DelegadosDp.csv	07/08/2017 21:41	Arquivo CSV	831 KB
 DicionarioVariaveisDOMensualEstadoDesde1991.xls	13/08/2017 18:16	Planilha do Micro...	15 KB
 DOMensualEstadoDesde1991.csv	02/09/2017 15:38	Arquivo CSV	59 KB
 DOMensualEstadoDesde1991.xlsx	02/09/2017 15:35	Planilha do Micro...	97 KB
 HierarquiaComando.csv	04/09/2017 23:28	Arquivo CSV	18 KB
 HierarquiaDelegacias.csv	04/09/2017 23:28	Arquivo CSV	18 KB
 Letalidade_PerfilVitima.csv	02/09/2017 13:32	Arquivo CSV	4.336 KB
 Letalidade_PerfilVitima_2014.csv	02/09/2017 13:18	Arquivo CSV	3.572 KB
 Letalidade_PerfilVitima_2015.csv	13/08/2017 18:37	Arquivo CSV	1.737 KB
 Letalidade_PerfilVitima_2016.csv	13/08/2017 18:36	Arquivo CSV	1.670 KB
 PoliciaisMortos1998EmDiante.csv	13/08/2017 18:12	Arquivo CSV	3 KB
 PopulacaoEvolucaoMensualCisp.CSV	30/08/2017 23:54	Arquivo CSV	521 KB
 PopulacaoEvolucaoMensualCisp.xlsx	13/08/2017 18:18	Planilha do Micro...	744 KB
 PopulacaoEvolucaoMensualMunic.csv	13/08/2017 18:19	Arquivo CSV	617 KB
 RelacaodasRISP_AISP.CSV	13/08/2017 17:37	Arquivo CSV	12 KB
 SecretarioDeSeguranca.csv	07/08/2017 21:41	Arquivo CSV	1 KB
 SeriesHistoricasLetalidadeViolenta.xlsx	13/08/2017 18:13	Planilha do Micro...	203 KB
 UppEvolucaoMensualDeTitulos.csv	13/08/2017 18:12	Arquivo CSV	430 KB

Figura 13 - Arquivos de texto do ISP

Foi idealizada uma solução de Data Warehouse que centralizasse as informações vitais do ISP. A missão do projeto é:

- Facilitar o acesso à informação através de interfaces amigáveis, e de fácil entendimento para usuários que não estão familiarizados com tecnologia ou não são proficientes em buscas em banco de dados.
- Centralizar os dados disponíveis em uma única solução
- Permitir a análise descritiva dos dados

Sendo assim, foi elaborado um esquema lógico para integrar informações institucionais (localização das delegacias, áreas sob responsabilidade, etc.) e informações operacionais, ou seja, os registros de ocorrências encontrados nos diferentes arquivos:

8 <http://www.ispdados.rj.gov.br/>

Esquema “Registro de Ocorrências”

O objetivo do esquema é oferecer uma visão agregada das ocorrências registradas. Para isso, o esquema de “Registro de Ocorrência” conta com as dimensões “Unidade de Segurança”, “Data de Ocorrência” e “Tipo de Ocorrência”.

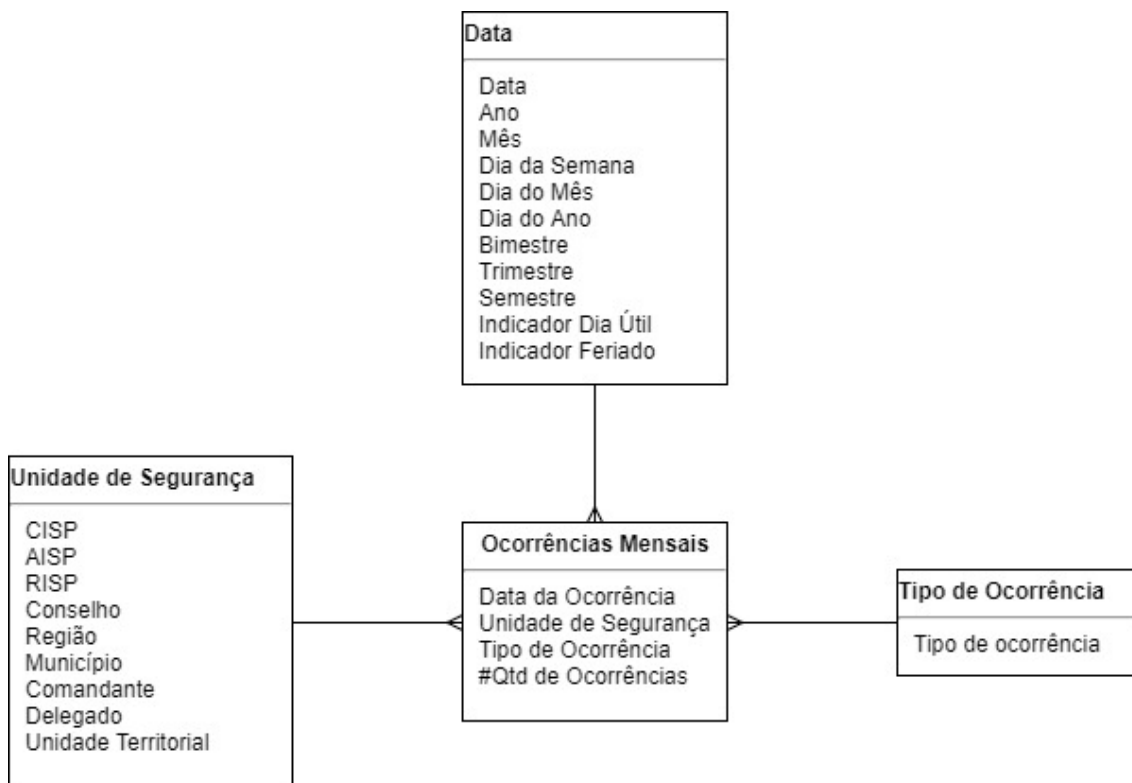


Figura 14 - Esquema Registro de Ocorrências

Esquema “Violência contra a Mulher”

O objetivo do esquema é permitir analisar os casos de violência contra a mulher. Para isso, o esquema conta com dimensões que caracterizam o perfil da vítima, assim como dimensões que caracterizam o local e tempo em que ocorreu o fato. As métricas identificadas são: Quantidade de ocorrências e Quantidade de dias até a comunicação do fato.

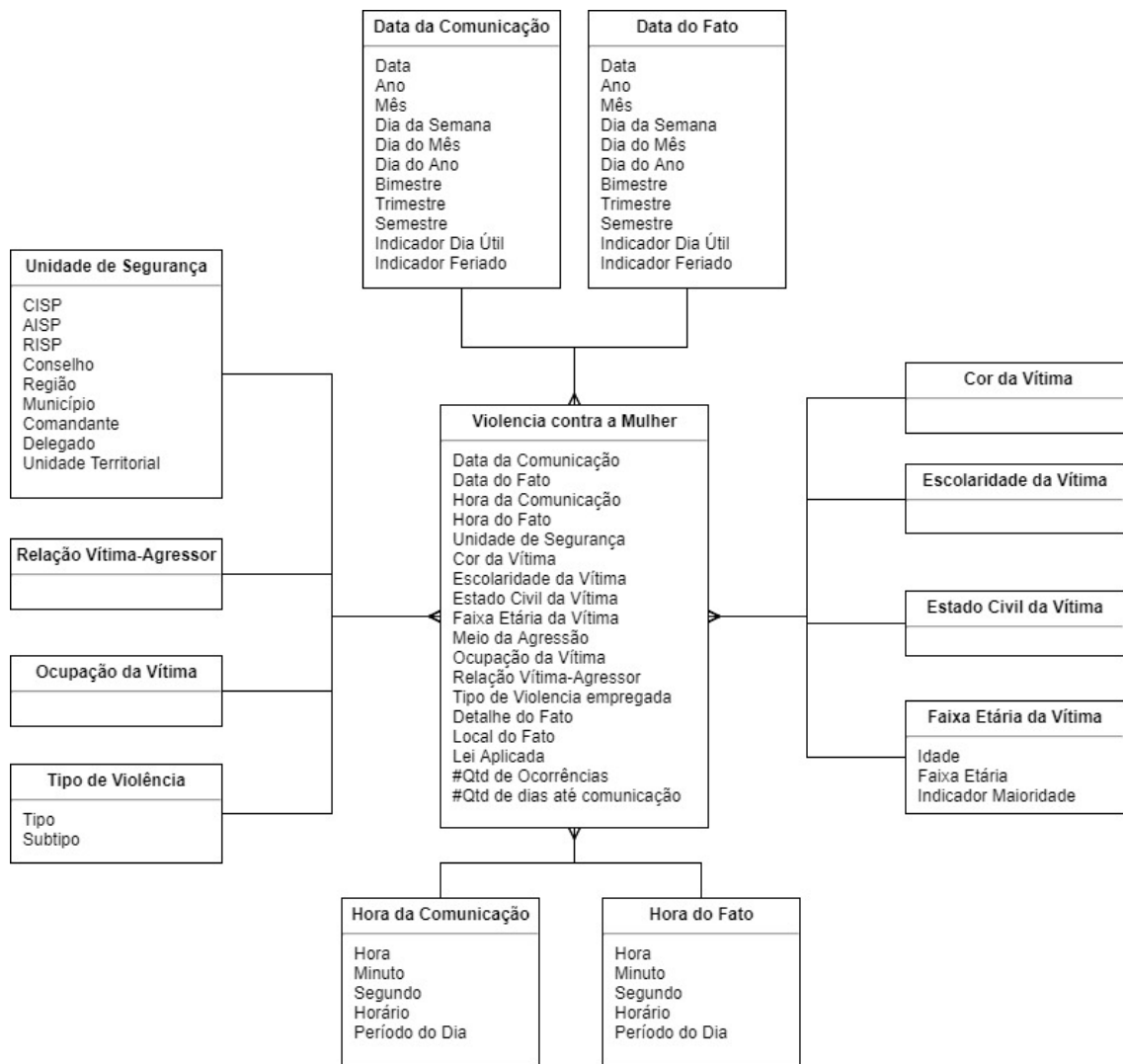


Figura 15 - Violência contra a mulher

Esquema “Letalidade violenta”

O objetivo do esquema é analisar o perfil da vítima de letalidade. O esquema conta com dimensões que caracterizam o perfil da vítima, assim como dimensões que caracterizam o local e tempo em que ocorreu o fato. As métricas identificadas são: Quantidade de ocorrências e Quantidade de dias até a comunicação

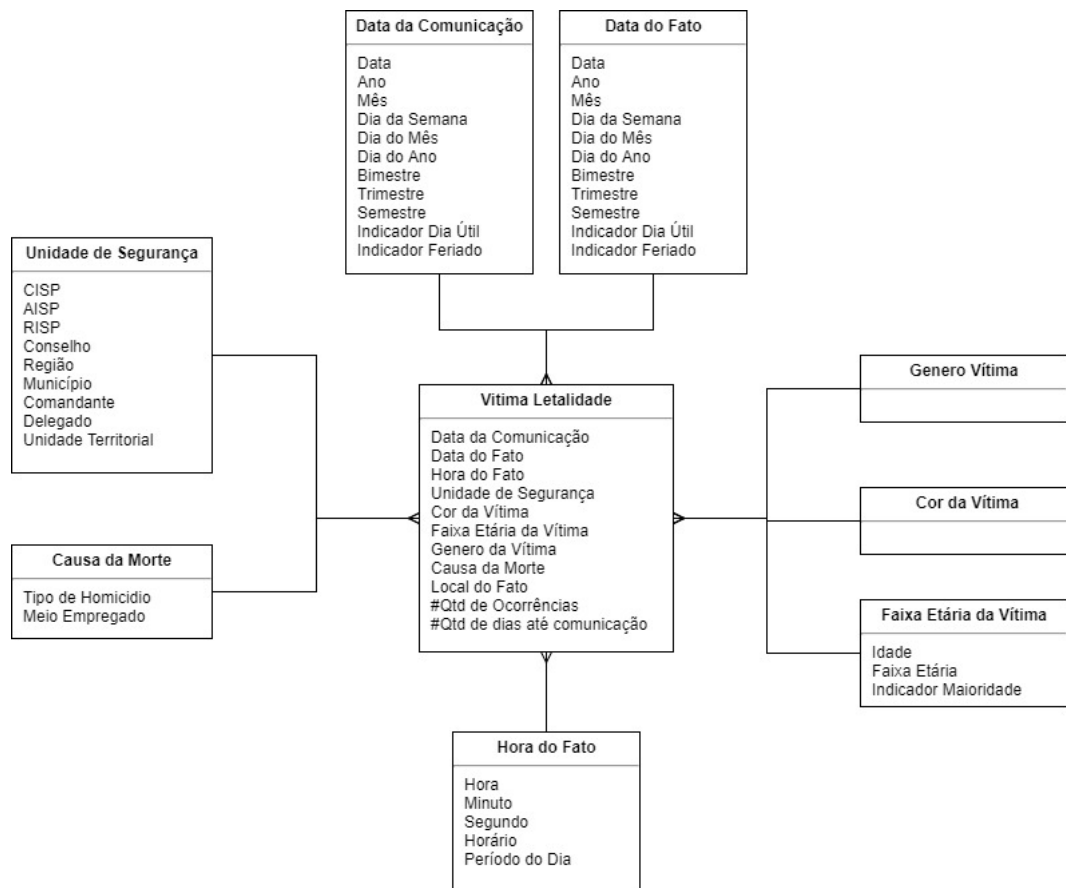


Figura 16 - Perfil vítima de letalidade

Esquema “Armas apreendidas”

Para identificar quais regiões possuem a maior quantidade de armas apreendidas, para ações táticas de combate ao crime.

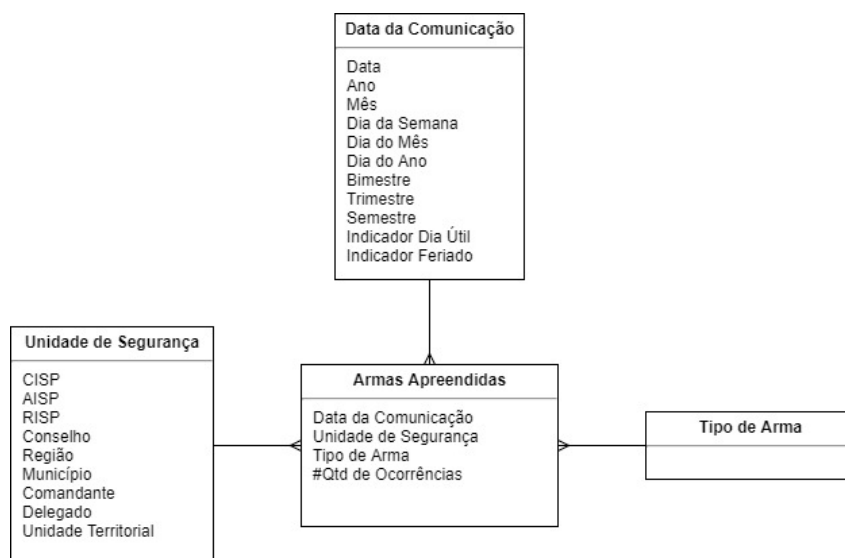


Figura 17 - Armas apreendidas

Outros Esquemas

Existe a possibilidade de desenvolvimento de outros esquemas lógicos ligados a indicadores de crimes. Porém, o ISP não disponibiliza informações suficientes para geração de fatos e dimensões, pois seus registros são limitados até certo intervalo de tempo.

4.4 Elaborando o desenho da solução final

O foco desta etapa foi elaborar uma solução que atenda aos requisitos dos usuários e patrocinadores do projeto, e que também esteja alinhada com o escopo de uma solução de Business Intelligence, contendo seus principais componentes. Para o projeto do ISP, foi utilizada a solução de BI da Microsoft em todas as camadas, permitindo uma integração dos seus componentes. A figura a seguir ilustra o desenho geral da solução proposta:

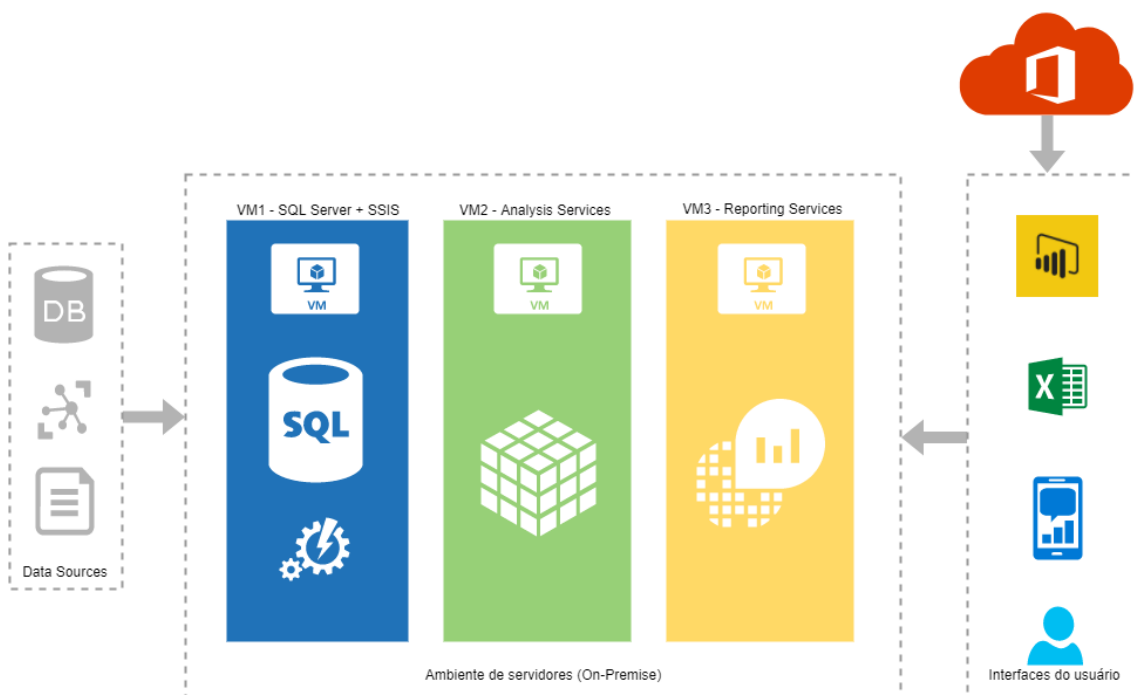


Figura 18 - Plano da solução proposta

Foi utilizado o Microsoft SQL Server 2016 Developer, que conta com os recursos citados no capítulo anterior, instalado em uma máquina virtual Windows Server. As bases de dados do ISP não possuem volume considerável, e com os recursos de Data Warehouse do SQL 2016 obtém-se uma maior compressão dos dados, otimizando os

recursos disponíveis. Foi utilizado o Integration Services, instalado na mesma máquina do SQL Server 2016 e o Analysis Services e Reporting Services, cada um com uma máquina exclusiva. As máquinas foram construídas utilizando template simples, com configurações de hardware mínimas necessárias.

A separação dos componentes é comum em ambientes de larga escala, devido a necessidade de processamento de cada um deles. O SQL Server, por exemplo, consome recursos de armazenamento, enquanto outros como o Analysis Services, consomem mais recursos como memória.

Tabela 3 - Especificação das máquinas utilizadas no projeto

Maquina	CPU	Memória	Espaço	Softwares instalados
BIDW01	4 Processadores	8GB	80GB	MSSQL Server 2016 MSSQL Integration Services
BIAS01	4 Processadores	4GB	40GB	MSSQL Analysis Services
BIRS01	4 Processadores	4GB	40GB	MSSQL Reporting Services
Local (Usuário)	4 Processadores	8GB	256GB	MS Excel, Power BI Desktop, SSRS Report Builder, Mobile Reports, Visual Studio 2017

4.5 Construindo o Data Warehouse

Com o modelo lógico, é possível modelar o Data Warehouse. Foi criada uma instância de banco de dados na máquina BIDW01, e um projeto de banco de dados no Visual Studio Data Tools. É possível manter o versionamento de todos os objetos de um banco de dados: Tabelas, Stored Procedures, Usuários, Permissões, etc. É possível também fazer a comparação do banco de dados que está implantado com o projeto de banco de dados na IDE Visual Studio, identificando e tratando as diferenças que são encontradas. Isto é particularmente útil para manter o versionamento do banco. O projeto de banco DataWarehouse criado para o projeto contém:

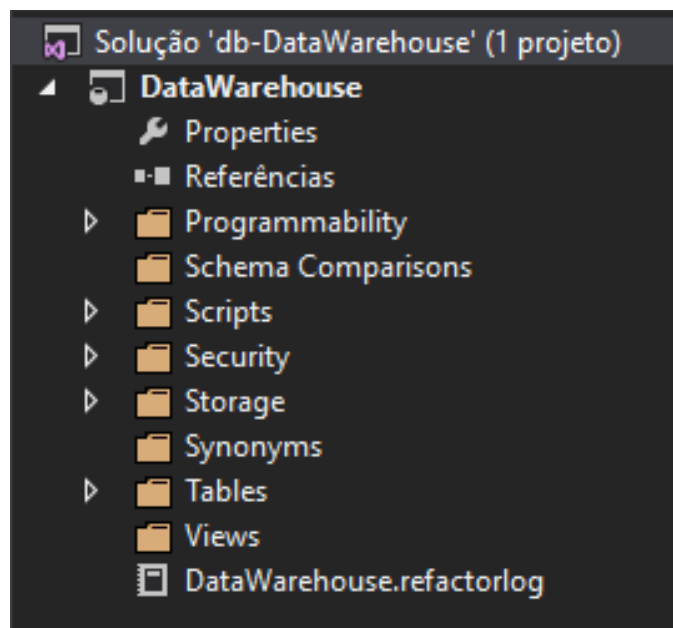


Figura 19 - Solução de banco de dados no Visual Studio

- **Programmability.** Funções, Sequences, Triggers e Stored Procedures.
- **Scripts.** Contém arquivos .sql que podem ser executados antes ou depois de um deploy no banco de dados.
- **Security.** Contém usuários e logins do banco de dados, roles (grupos), permissões e esquemas.
- **Storage.** Contém os grupos de arquivos, esquemas de partição e funções de partição do banco de dados.
- **Synonyms.** Contém os sinônimos do banco de dados, para facilitar a identificação de objetos.
- **Tables.** Contém as tabelas do banco de dados
- **Views.** Contém as visões do banco de dados

O seguinte script foi executado para criar o banco de dados DataWarehouse utilizando compatibilidade com o SQL Server 2016:

```

/***** Object: Database [DataWarehouse] *****/
CREATE DATABASE [DataWarehouse]
    CONTAINMENT = NONE
    ON PRIMARY
ALTER DATABASE [DataWarehouse] SET COMPATIBILITY_LEVEL = 130
GO

```

Storage

Ao criar o banco de dados foi necessário criar os arquivos que irão armazenar os dados e os logs de transação. Para o projeto, foram criados os seguintes grupos de arquivos:

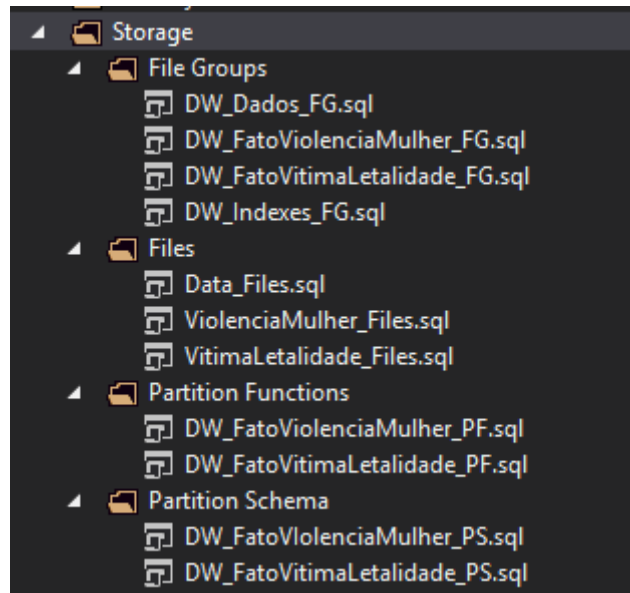


Figura 20 - Tipos de componentes de storage do SQL Server

Os grupos de arquivos (File Groups) podem conter um ou mais arquivos. Isso facilita a manutenção do banco de dados e realiza particionamentos mais eficientes. Por exemplo, a Fato de Violência contra a Mulher contém 20 File Groups, onde cada um mantém uma parte dos dados da tabela. O esquema de partição (Partition Schema) especifica quais File Groups pertencem à tabela. As funções de partição (Partition Functions) dividem as partições de acordo com valores específicos de uma coluna. Isso também se traduz em eficiência nas consultas e melhorias de performance, pois ao filtrar uma consulta pela coluna da função de partição, o SQL Server faz a busca na partição específica onde está o dado. Por exemplo, ao realizar a consulta na tabela fato “ViolenciaMulher”:

```
SELECT *  
FROM Datawarehouse.Fato.ViolenciaMulher  
WHERE DataComunicacao BETWEEN 20160101 AND 20161227
```

Seu particionamento está sendo feito com os seguintes intervalos:

```
CREATE PARTITION FUNCTION [DW_FatoViolenciaMulher_PF](INT)  
AS RANGE RIGHT  
FOR VALUES (
```

```

20150101,
20150701,
20160101,
20160401,
20160701,
20161001,
20170101,
20170201,
20170301,
20170401,
20170501,
20170601,
20170701,
20170801,
20170901,
20171001,
20171101,
20171201,
20180101
);

```

O plano de execução do SQL Server mostra que foram acessadas apenas as partições 4 a 7, pois a coluna DataComunicação é a chave da função de particionamento:

Operação Física	Verificação de Índice Columnstore
Operação Lógica	Clustered Index Scan
Ordenado	True
Paralelo	False
Particionado	True
Partições Reais Acessadas	4..7

Figura 21 - Plano de execução de consulta exibindo as partições acessadas

O particionamento pode deixar as consultas sem filtro mais lentas, mas em um ambiente de Data Warehouse deve-se sempre ter em mente de que as consultas são, em sua maioria, compostas por filtros de data.

Organização das Tabelas

Para qualquer sistema de banco de dados, organização e padronização são extremamente importantes. Em um banco de dados composto por muitos objetos de tabelas, que servem diferentes propósitos, é mais vantajoso organizá-las em esquemas diferentes para fatos e dimensões, ou nomeá-los de acordo com a área de assunto, para melhor entendimento do modelo de dados. Foram criados três esquemas (“schemas” no SQL Server) para o projeto de DataWarehouse:

- **Dimensão.** Schema criado para as tabelas de Dimensão
- **Fato.** Schema criado para as tabelas de Fato

- **Staging.** Schema criado para o processo de ETL

As tabelas devem seguir o padrão de localização em esquemas, e também devem possuir colunas de auditoria, para facilitar a identificação dos processos que realizaram operações de escrita. A tabela [Dimensao].[UnidadeSeguranca], por exemplo, possui a seguinte estrutura:

```
CREATE TABLE [Dimensao].[UnidadeSeguranca] (
    [IdUnidadeSeguranca] INT IDENTITY (1, 1) NOT NULL,
    [IdAltUnidadeSeguranca] VARCHAR(30) NOT NULL,
    [IdCisp] INT NOT NULL,
    [IdAisp] INT NOT NULL,
    [IdRisp] INT NOT NULL,
    [DescCisp] VARCHAR (16) NULL,
    [DescAisp] VARCHAR (16) NULL,
    [DescRisp] VARCHAR (64) NULL,
    [Conselho] VARCHAR (32) NULL,
    [Regiao] VARCHAR (32) NULL,
    [Municipio] VARCHAR (64) NULL,
    [ConselhoComunitarioSeguranca] VARCHAR (64) NULL,
    [Comandante] VARCHAR (64) NULL,
    [Delegado] VARCHAR (64) NULL,
    [UnidadeTerritorial] VARCHAR (256) NULL,
    [ETL_DataInicio] DATETIME2 (3) CONSTRAINT DF_UnidadeSeguranca_DataInicio DEFAULT SYSUTCDATETIME() NOT NULL,
    [ETL_DataFim] DATETIME2 (3) CONSTRAINT DF_UnidadeSeguranca_DataFim DEFAULT '9999-12-31 23:59:59.999' NOT NULL,
    [ETL_IndAtivo] BIT CONSTRAINT DF_UnidadeSeguranca_IndAtivo DEFAULT 1 NOT NULL,
    [ETL_Usuario] VARCHAR (30) NOT NULL,
    [ETL_ExecutionInstanceGUID] UNIQUEIDENTIFIER NULL,
    CONSTRAINT [PK_DimensaoUnidadeSeguranca] PRIMARY KEY CLUSTERED ([IdUnidadeSeguranca] ASC) ON [DW_Dados_FG01]
);
GO
```

Figura 22 - Script de tabela dimensão no SQL

De acordo com Kimball (2013), a chave primária da tabela dimensão deve ser uma Surrogate Key. Isso facilita as operações sobre uma linha da tabela. Também é importante adicionar colunas de controle. No caso do projeto, toda dimensão possui as colunas com o sufixo ETL, que identificam: Data de inserção e modificação do registro, Usuário que realizou a inserção ou modificação e identificador do processo de ETL. O padrão se repete para as tabelas de Fato:

```
CREATE TABLE [Fato].[VitimaLetalidade] (
    [DataComunicacao] INT NOT NULL,
    [IdVitimaLetalidade] BIGINT IDENTITY (1, 1) NOT NULL,
    [DataFato] INT NOT NULL,
    [HoraFato] INT NOT NULL,
    [IdUnidadeSeguranca] INT NOT NULL,
    [IdFaixaEtariaVitima] INT NOT NULL,
    [IdGeneroVitima] INT NOT NULL,
    [IdCorVitima] INT NOT NULL,
    [IdMeioEmpregado] INT NOT NULL,
    [IdTipoHomicidio] INT NOT NULL,
    [LatitudeFato] VARCHAR(20) NULL,
    [LongitudeFato] VARCHAR(20) NULL,
    [QtdDiasAteComunicacao] INT NOT NULL,
    [QtdRegistros] INT NOT NULL,
    [ETL_DataCriacao] DATETIME2(3) CONSTRAINT DF_FatoVitimaLetalidade_DataCriacao DEFAULT SYSUTCDATETIME() NOT NULL,
    [ETL_DataModificacao] DATETIME2(3) NULL,
    [ETL_Usuario] VARCHAR (30) NOT NULL,
    [ETL_ExecutionInstanceGUID] UNIQUEIDENTIFIER NULL,
    CONSTRAINT [PK_FatoVitimaLetalidade] PRIMARY KEY NONCLUSTERED ([DataComunicacao] ASC, [IdVitimaLetalidade] ASC) ON [DW_FatoVitimaLetalidade_PS] ([DataComunicacao]),
    CONSTRAINT [FK_VitimaLetalidade_DataComunicacao] FOREIGN KEY ([DataComunicacao]) REFERENCES [Dimensao].[Data] ([IdData]),
    CONSTRAINT [FK_VitimaLetalidade_DataFato] FOREIGN KEY ([DataFato]) REFERENCES [Dimensao].[Data] ([IdData]),
    CONSTRAINT [FK_VitimaLetalidade_HoraFato] FOREIGN KEY ([HoraFato]) REFERENCES [Dimensao].[Hora] ([IdHora]),
    CONSTRAINT [FK_VitimaLetalidade_UnidadeSeguranca] FOREIGN KEY ([IdUnidadeSeguranca]) REFERENCES [Dimensao].[UnidadeSeguranca]([IdUnidadeSeguranca]),
    CONSTRAINT [FK_VitimaLetalidade_FaixaEtariaVitima] FOREIGN KEY ([IdFaixaEtariaVitima]) REFERENCES [Dimensao].[PerfilFaixaIdade]([IdFaixaIdade]),
    CONSTRAINT [FK_VitimaLetalidade_GeneroVitima] FOREIGN KEY ([IdGeneroVitima]) REFERENCES [Dimensao].[PerfilGenero]([IdGenero]),
    CONSTRAINT [FK_VitimaLetalidade_CorVitima] FOREIGN KEY ([IdCorVitima]) REFERENCES [Dimensao].[PerfilCor]([IdCor])
);
GO

CREATE CLUSTERED COLUMNSTORE INDEX [CSIX_FatoVitimaLetalidade] ON [Fato].[VitimaLetalidade] ON [DW_FatoVitimaLetalidade_PS]([DataComunicacao])
GO
```

Figura 23 - Script de tabela Fato no SQL

De acordo com Kimball (2013), a chave primária de uma tabela fato é composta de todas suas chaves estrangeiras (Foreign Keys), garantindo que uma tupla da fato seja única. Para o projeto do ISP, estamos garantindo que não inserimos dados duplicados através do ETL, e ainda fazemos a junção das Foreign Keys as suas tabelas dimensão. A tabela “VitimaLetalidade” possui um índice baseado em colunas clusterizado, particionado pela coluna DataComunicação. Ao fim da construção das tabelas, temos a seguinte estrutura de dados no banco DataWarehouse:

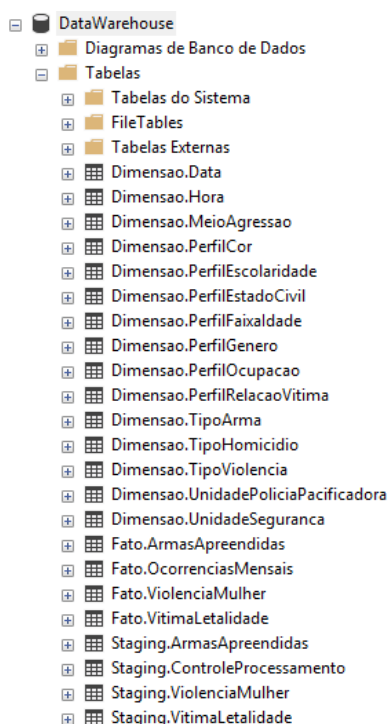


Figura 24 - Objetos criados no Data Warehouse

Regras de Acesso

O ambiente criado para o projeto possui dois usuários: “ETLAppUser” e “BIAppUser”. O primeiro possui permissão de leitura em todos os bancos de dados de origem e permissão de escrita em todo o DataWarehouse, enquanto o ultimo possui permissão de leitura aos esquemas Fato e Dimensão, removendo as tabelas desnecessárias do alcance dos usuários finais.

Considerações

O processo de criação do banco de dados do Data Warehouse se mostrou mais fácil utilizando a IDE Visual Studio Data Tools, disponibilizada pela Microsoft. Ela fornece uma visualização de todos os componentes do banco de dados em uma única interface,

facilitando a identificação de erros na construção. A ferramenta de comparação também se mostrou útil ao identificar diferenças entre o projeto e o banco de dados. O banco de dados foi criado em alguns segundos após a implantação do projeto na instância criada.

4.6 Construindo o ETL

Foi utilizado o SQL Server Integration Services para realizar a transformação e carga de dados dos arquivos disponibilizados pelo ISP para o Data Warehouse. A carga se dá em três passos principais:

- Carga de dados para o esquema “Staging” (Apenas para Fatos)
- Carga de dados para o esquema “Dimensão”
- Carga de dados para o esquema “Fato”

Foi criada uma solução no Visual Studio Data Tools, e um projeto de Integration Services:

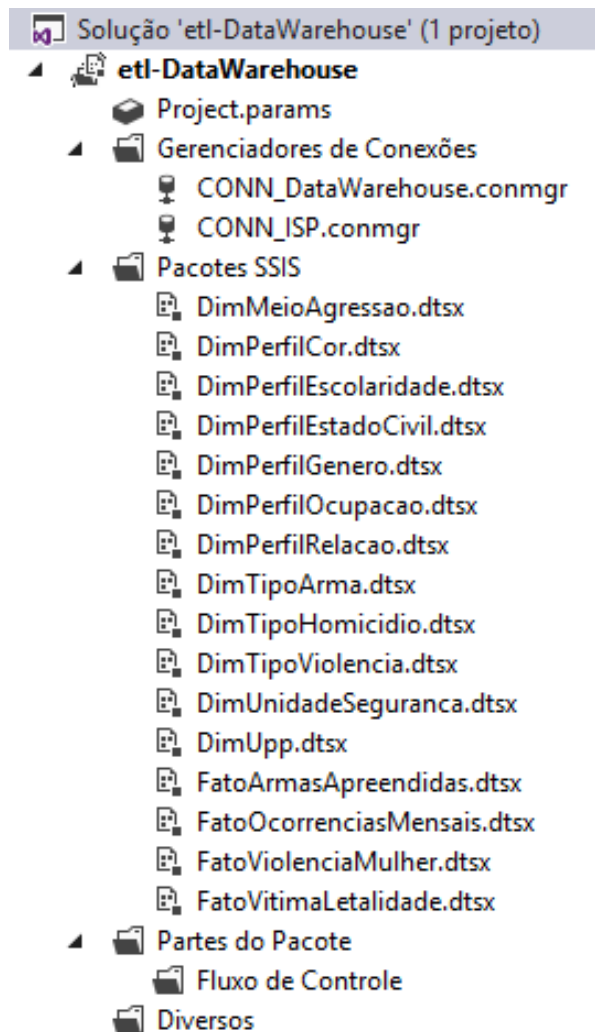


Figura 25 - Projeto de ETL no Visual Studio

No capítulo anterior foram apresentados os componentes do Integration Services. Colocando-os em prática, foram construídos os pacotes de carga das tabelas de dimensão e tabelas de fatos, exceto para as dimensões Data e Hora, onde foram utilizadas Stored Procedures. O motivo para a criação dessas Procedures é que as dimensões citadas não sofrerão alterações com o tempo. A figura a seguir mostra fluxo de controle para o pacote “FatoViolenciaMulher”:

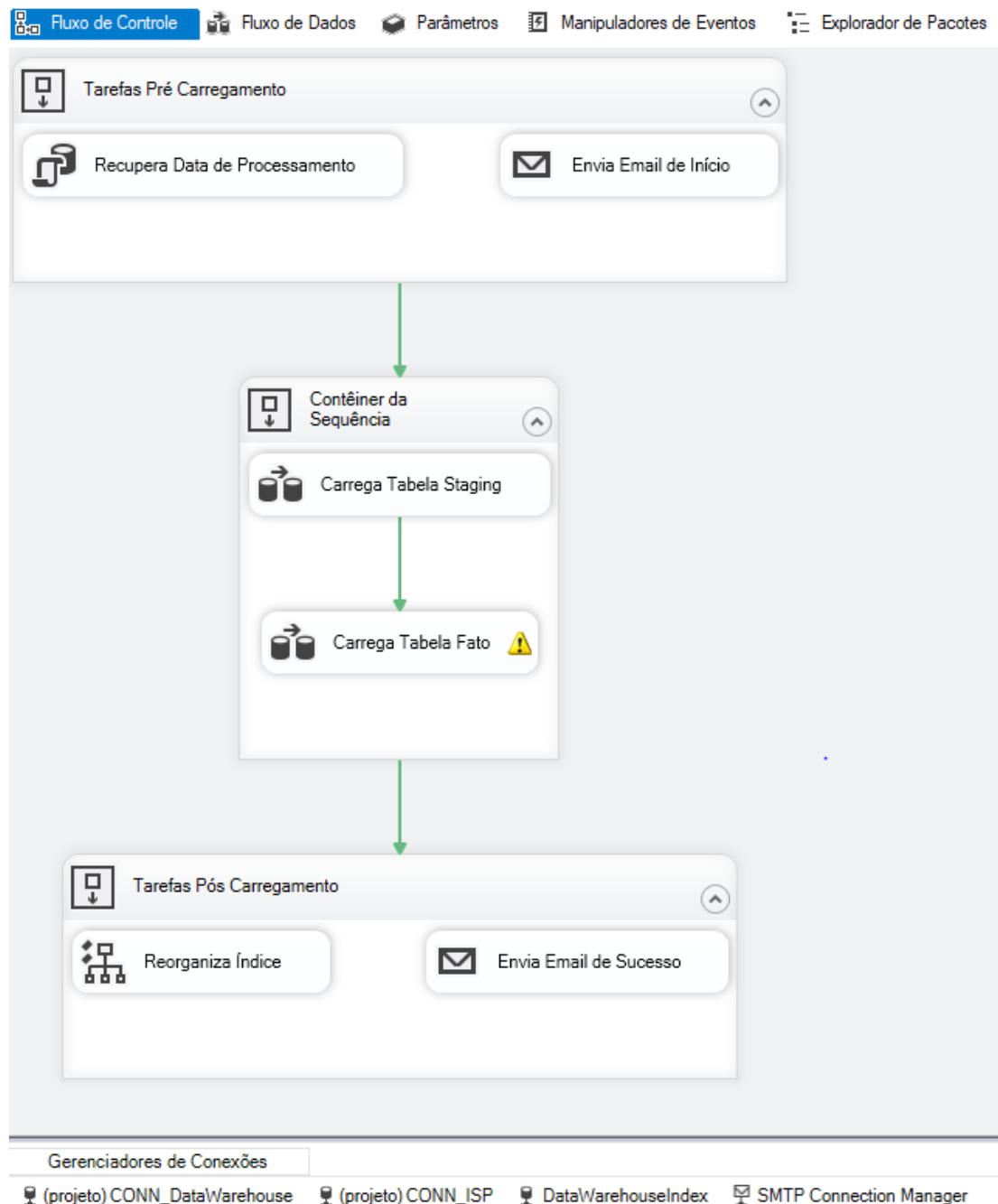


Figura 26 - Pacote de ETL "FatoViolenciaMulher"

Ao visualizar os componentes é possível ter uma ideia das tarefas que o fluxo executa. Isto ocorre porque o Integration Services oferece uma programação “visual” dos fluxos, facilitando o seu entendimento. O pacote “FatoViolenciaMulher” carrega informações de maneira incremental, a partir de uma data especificada via parâmetro. A seguir, os dados brutos do ISP são carregados no esquema “Staging” do Data Warehouse, e em seguida, são transformados no passo “Carrega Tabela Fato”. A figura a seguir mostra o fluxo da tarefa “Carrega Tabela Fato”:

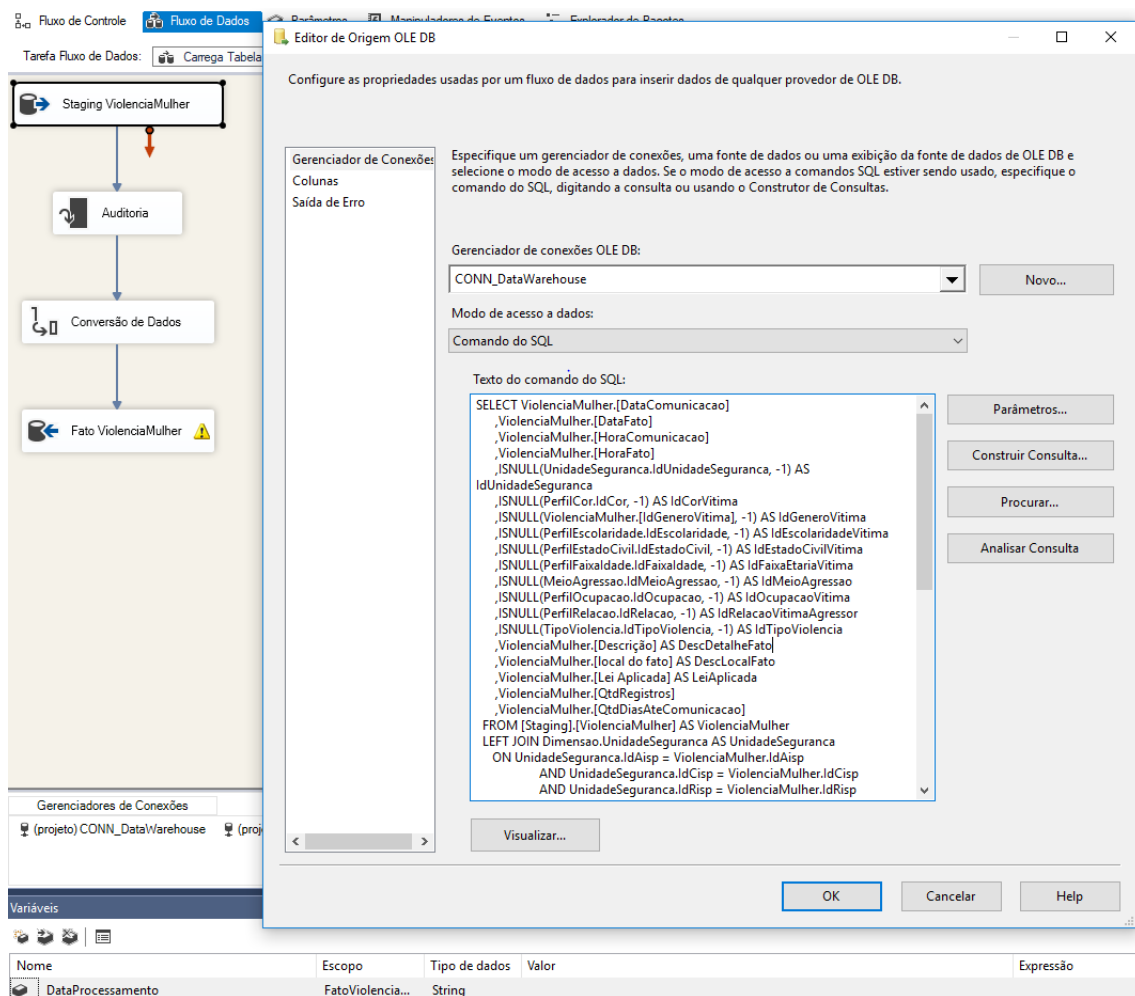


Figura 27 - Fluxo da tarefa "Carrega tabela Fato", com destaque no objeto Origem OLEDB

O componente inicial realiza uma consulta SQL no esquema Staging, e realiza tratamentos de limpeza e de correção dos dados da origem. Em seguida, é feita uma auditoria dos dados, incluindo no fluxo as colunas de controle do ETL. No fim, a tabela Fato é carregada com os dados da consulta. O mesmo padrão de desenvolvimento se repete aos demais pacotes da solução. Os testes e implantação do projeto podem ser feitos via Visual Studio.

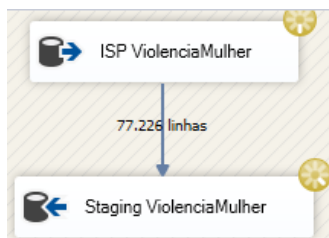


Figura 28 - Modo debug do pacote no Data Tools

O Integration Services possui um banco de dados próprio denominado SSISDB. Este banco de dados possui informações dos pacotes importados para o catálogo do Integration Services, mantendo um histórico das alterações e das execuções sem a necessidade de criação de uma estrutura própria:

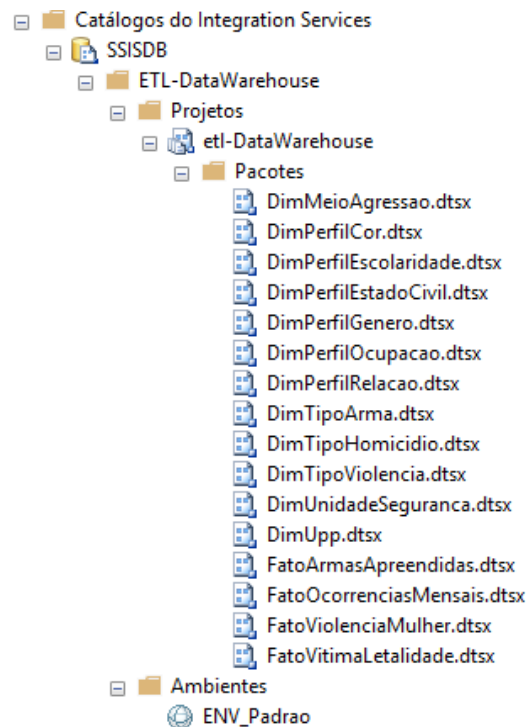


Figura 29 - Catálogo de projetos do SSIS no SQL Server

Considerações

O SQL Server Integration Services apresenta os componentes necessários para o desenvolvimento do projeto. Não é necessário que o desenvolvedor tenha conhecimentos avançados de programação, apesar do SSIS possuir componentes de scripts em C# e VB.

4.7 Construindo o modelo semântico

O SQL Server Analysis Services foi utilizado como uma camada intermediária entre o banco de dados e o usuário final. O Analysis Services foi construído seguindo o modelo Tabular, tecnologia que está ganhando mais espaço entre os utilizadores da plataforma de BI Microsoft. Neste projeto, a instalação foi feita em uma máquina virtual (BIAS01) separada do banco Data Warehouse. A figura a seguir exibe o Modelo Tabular no Analysis Services utilizando as tabelas do Data Warehouse:

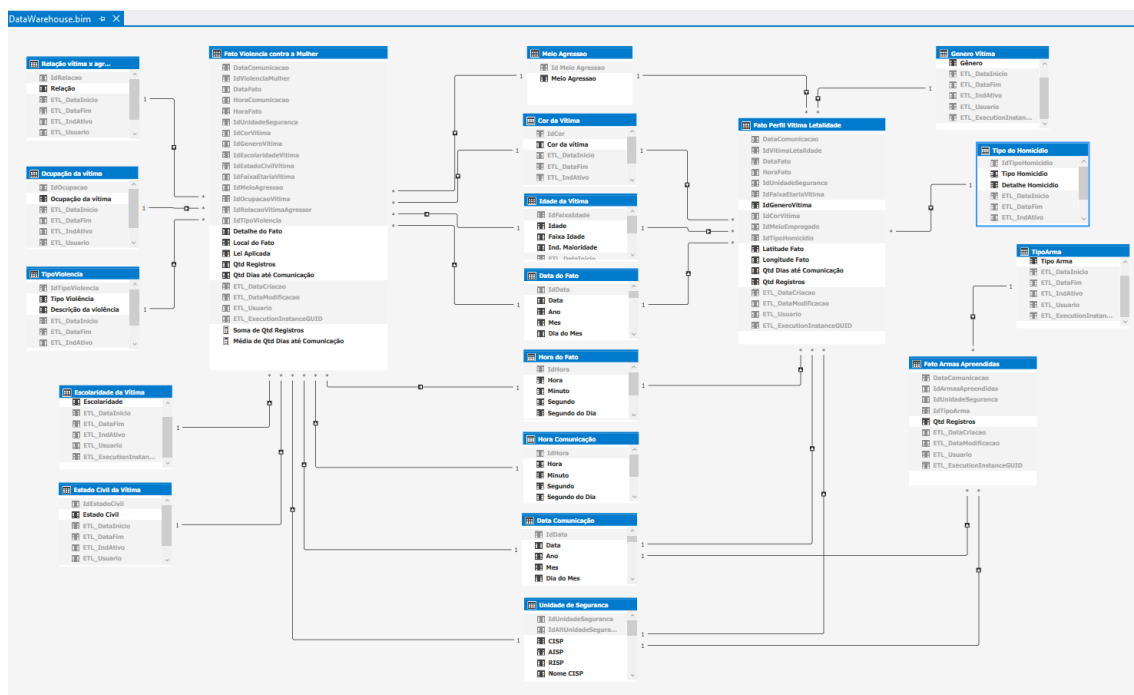


Figura 30 - Modelo tabular (visão parcial) no Analysis Services

No Analysis Services pode-se ocultar as colunas que não são importantes para exibição, como as colunas de controle das tabelas, e as colunas identificadoras. Métricas customizadas podem ser criadas, como somas, médias, valores máximos e mínimos, alterar o nome das colunas para que sejam mais amigáveis para o usuário, e optar por utilizar o armazenamento em memória de tabelas específicas. O modelo Tabular do Analysis Services armazena suas estruturas de tabelas no formato de arquivos JSON.

4.8 Construindo os Relatórios e Dashboards no Excel e Power BI

Para a última camada da aplicação, a camada de visualização, foram utilizadas as ferramentas Excel e Power BI. O Excel está disponível na plataforma Office 365, um produto do tipo “Software como Serviço” (SaaS) que contém aplicações de produtividade. Outros produtos da plataforma incluem Word, PowerPoint, OneDrive, SharePoint, entre outros.

Excel

O Excel é uma ferramenta de produtividade, onde seu principal recurso é a visualização em formato de planilhas. É possível carregar dados de diversas fontes de dados,

incluindo o SQL Server e Analysis Services. Utilizando o menu “Dados”, o modelo DataWarehouse é importado:

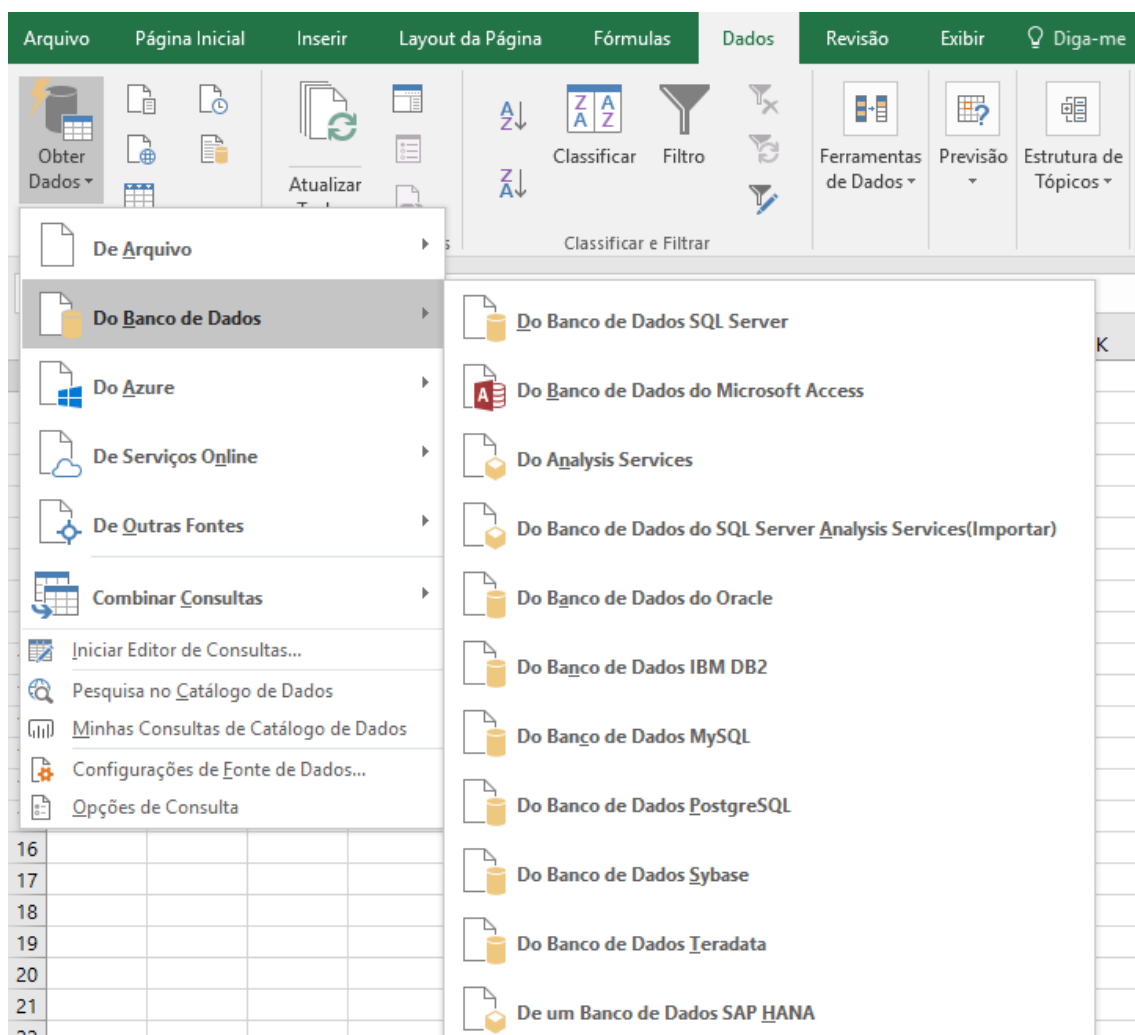


Figura 31 - Importando dados do Analysis Services no Excel

A interface do Excel permite a criação de tabelas dinâmicas ou gráficos. O usuário escolhe os atributos a partir de uma lista, que corresponde aos atributos do modelo do Analysis Services. Ao selecionar os atributos, o Excel realiza uma consulta no Analysis Services, e este último realiza uma consulta no SQL Server. Na figura a seguir, algumas informações sobre o estado civil da vítima e a evolução anual dos registros de violência contra a mulher, agrupado por região do Estado podem ser vistas. Pode-se perceber um decréscimo nas ocorrências registradas no período observado:

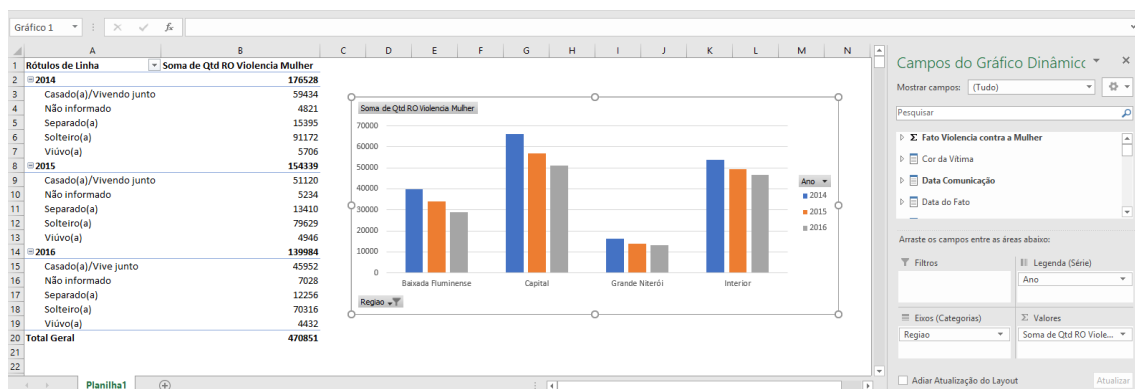


Figura 32 - Interface dinâmica do Excel utilizando dados do ISP

Além das métricas existentes no modelo, também é possível criar fórmulas nas tabelas e gráficos dinâmicos utilizando a linguagem DAX, uma linguagem muito similar as fórmulas do Excel, para maior manipulação dos dados. Com isso, os usuários possuem todos os recursos disponíveis no Excel enquanto exploram os dados dos modelos.

Power BI

Como dito anteriormente, o Power BI é uma ferramenta de Business Analytics, também fornecida pela Microsoft, que possui integração com a plataforma Azure e Office 365. O Power BI foi criado a partir dos componentes Power Pivot e Power View do Excel, portanto, possuem diversas similaridades. É possível realizar o download da versão gratuita do Power BI Desktop no website⁹ da Microsoft. Ao realizar a instalação do Power BI, é habilitada a conexão ao DataWarehouse, como mostra a figura a seguir:

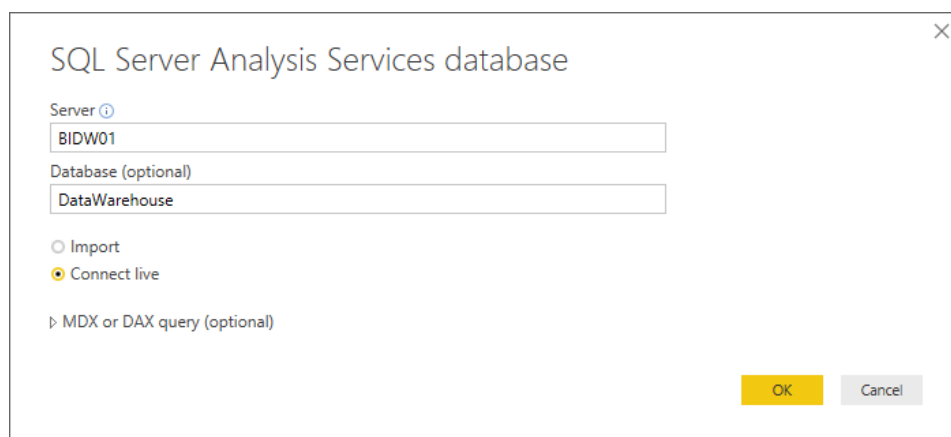


Figura 33 - Conectando no Analysis Services pelo Power BI

⁹ <https://powerbi.microsoft.com/pt-br/downloads/>

O Power BI realiza uma prévia dos objetos que serão acessados na base de dados, com a opção de importação dos dados para um arquivo que pode ser compartilhado entre usuários.



Figura 34 - Prévia dos objetos para importação no Power BI

Ao se conectar na base de dados, o Power BI exibe um menu lateral que é utilizado para seleccionar os objetos do modelo. Ao clicar em qualquer objeto, uma visualização automática é gerada, de acordo com a natureza do atributo escolhido.

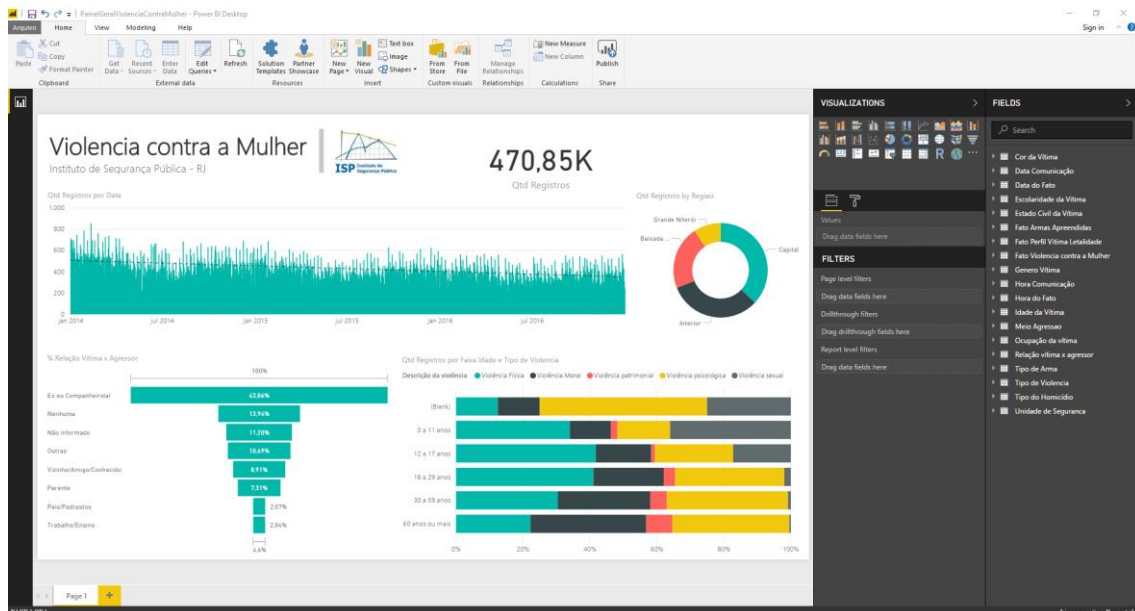


Figura 35 - Visão geral da interface gráfica do Power BI

O design das visualizações pode ser customizado, e além disso, é responsivo às interações. O exemplo a seguir mostra o antes e depois ao selecionar o município do Rio de Janeiro no mapa:

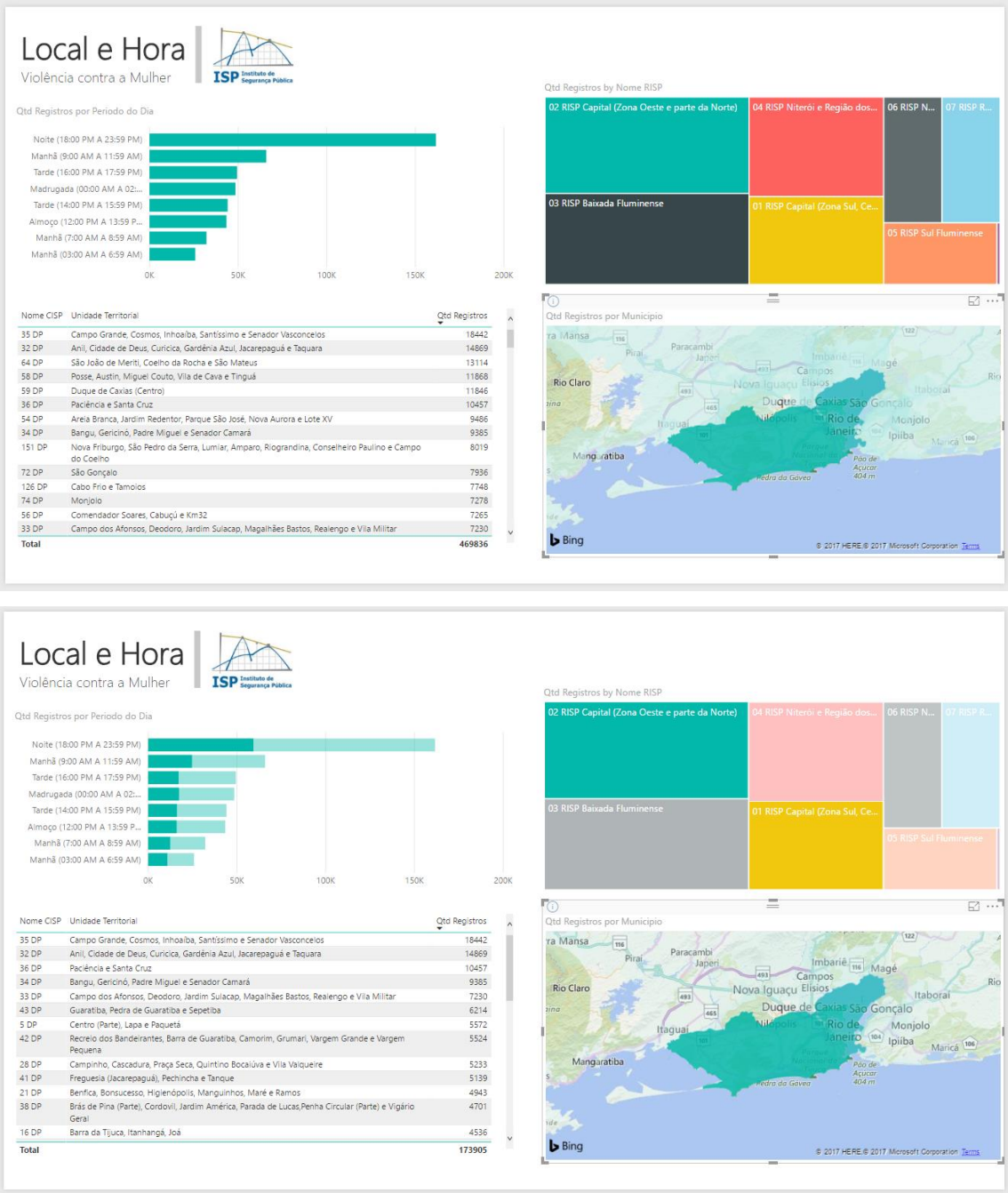


Figura 36 - Antes e depois de filtrar um elemento selecionando-o no mapa

Ao finalizar a construção dos painéis e indicadores, é possível publica-los no serviço do Power BI na nuvem. O serviço é acessado através de uma conta no ambiente do Office 365, e dependendo do plano adquirido, é possível utilizar os recursos mais avançados da

ferramenta, como: Maior tamanho de espaço de armazenamento, atualização dos dados de uma maneira mais frequente, compartilhamento com múltiplos usuários, etc.

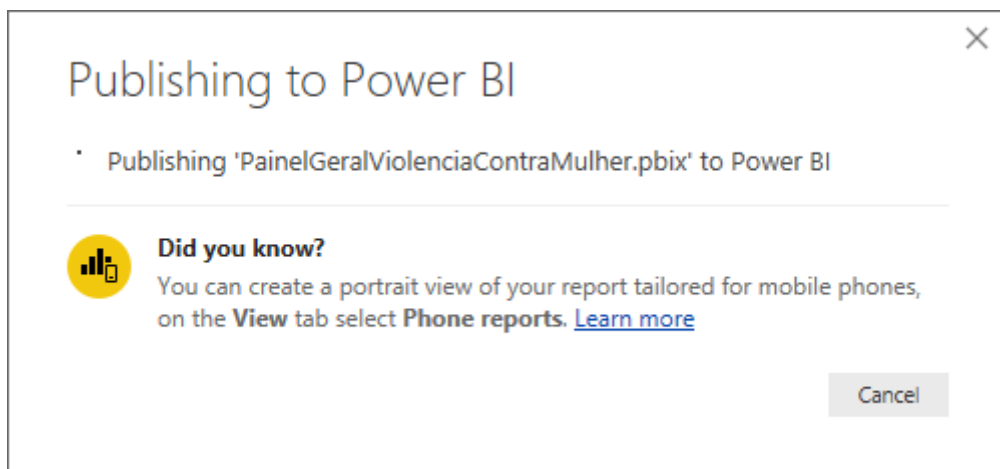


Figura 37 - Publicando no serviço Power BI na nuvem

Para acesso aos dados que utilizam conexão direta no serviço do Power BI, é necessário a instalação do serviço On-Premise Data Gateway¹⁰, que age como uma ponte entre dados locais (que não estão na nuvem) e o Power BI, Power Apps, e outros aplicativos. Com isso, pode-se ter a mesma experiência de navegação do aplicativo Power BI Desktop na web:

10 <https://powerbi.microsoft.com/en-us/gateway/>



Figura 38 - Interface do serviço do Power BI na nuvem, similar ao Desktop

4.9 Compartilhando relatórios no Reporting Services

Algumas organizações possuem restrições quanto o uso de aplicações na nuvem. O SQL Server possui o recurso Reporting Services, que funciona como uma camada de compartilhamento de relatórios. Ele também oferece recursos para criação de relatórios utilizando o Report Builder¹¹ ou Visual Studio. A configuração do Report Services é feita por um utilitário do SQL Server, e utiliza um repositório de banco de dados, similar ao catálogo do Integration Services, para armazenar os objetos salvos.

¹¹ <https://docs.microsoft.com/pt-br/sql/reporting-services/install-windows/install-report-builder>

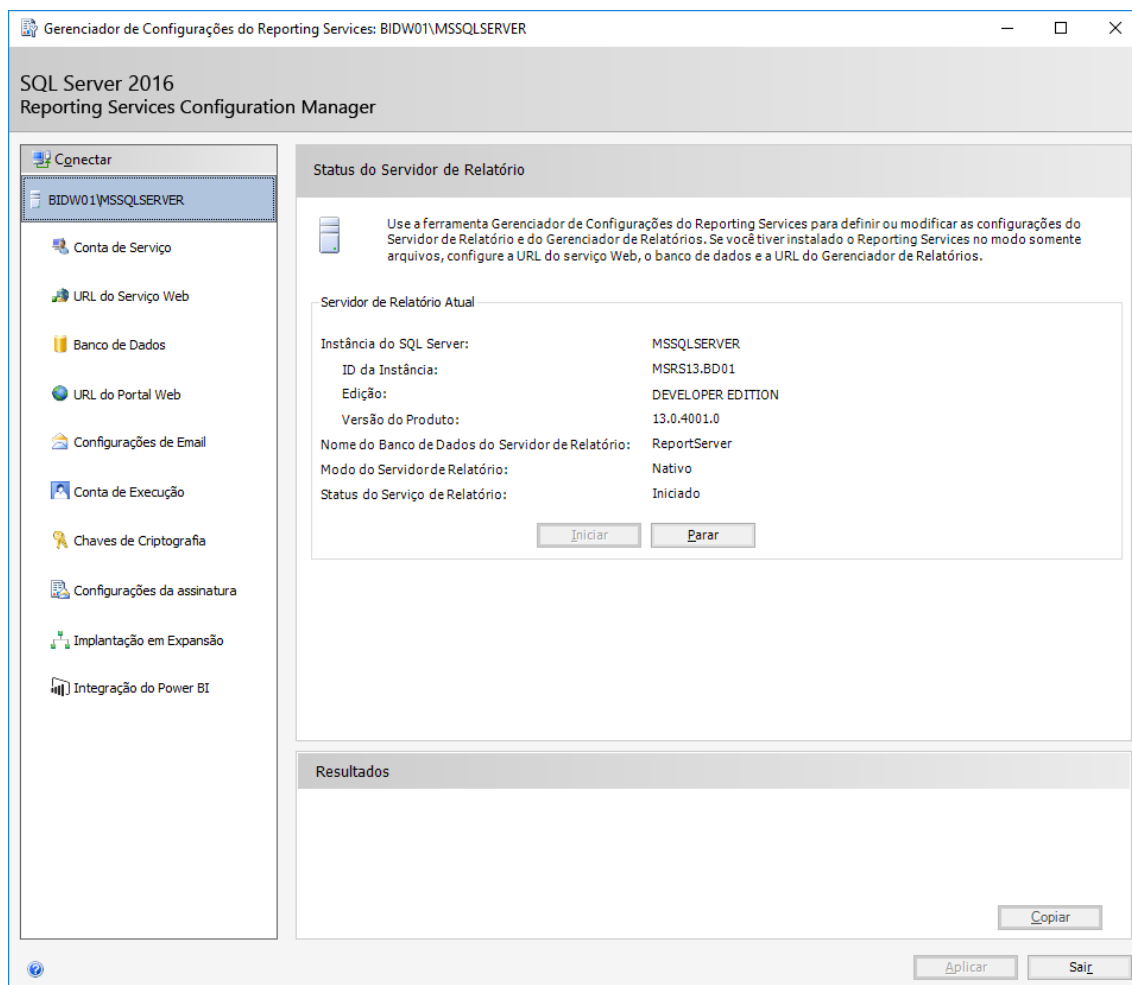


Figura 39 - Janela de configuração do Reporting Services

Ao configurar o Reporting Services, um serviço web está disponível para acesso aos usuários que possuem credenciais válidas. O portal do Reporting Services pode ser customizado com KPIs e relatórios de acesso rápido, e possui uma estrutura de pastas e arquivos.

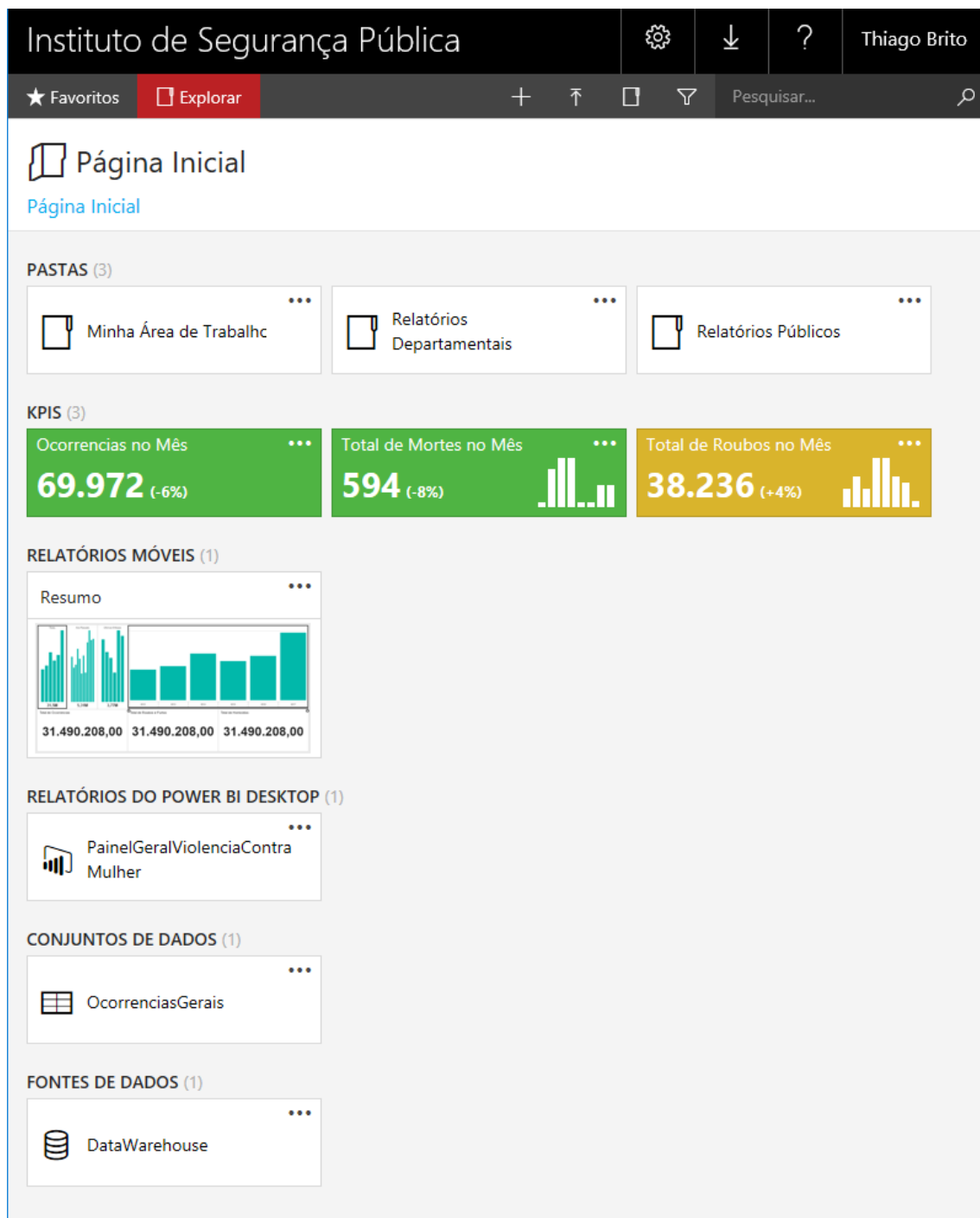


Figura 40 - Página Inicial do Reporting Services

É possível fazer o upload de qualquer tipo de arquivo, porém a intenção é utilizar o SSRS para compartilhar as extensões do Power BI, Excel e Report Builder.

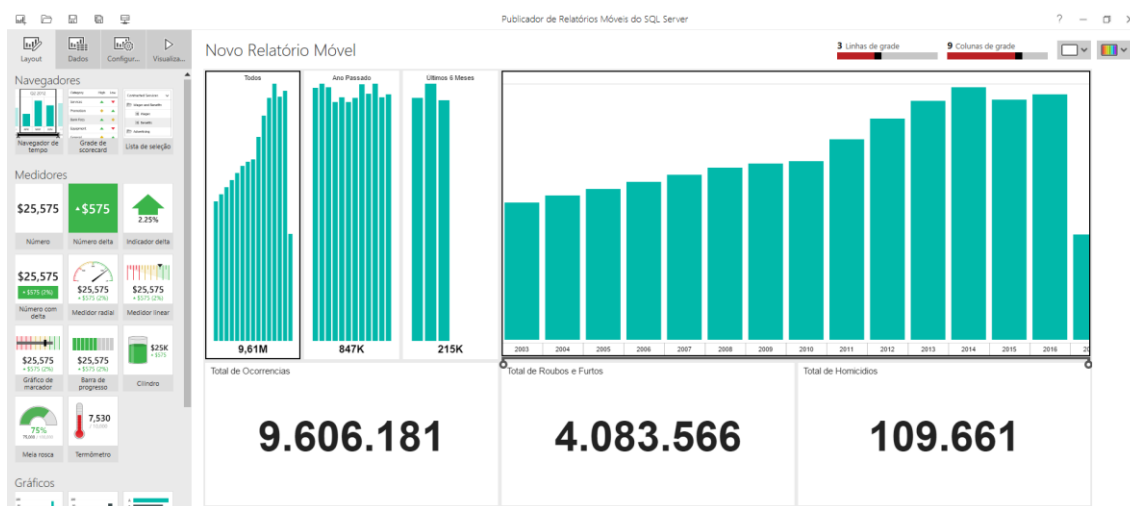


Figura 41 - Aplicação de Relatórios Móveis (Mobile Reports)

4.10 Resultados obtidos

Com o ambiente em funcionamento, foi possível analisar os dados do ISP e tirar conclusões dos dados observados. No âmbito da segurança pública, aspectos como tempo e local são determinantes para a adoção de estratégias que otimizem a operação da Polícia Militar e Polícia Civil. Com o Data Warehouse, as perguntas a seguir são respondidas com maior facilidade:

- **Quais são as ocorrências mais comuns e onde são registradas? Quais regiões e sub-regiões do Estado do Rio de Janeiro possuem uma maior quantidade de registros de ocorrência? Em qual período?**

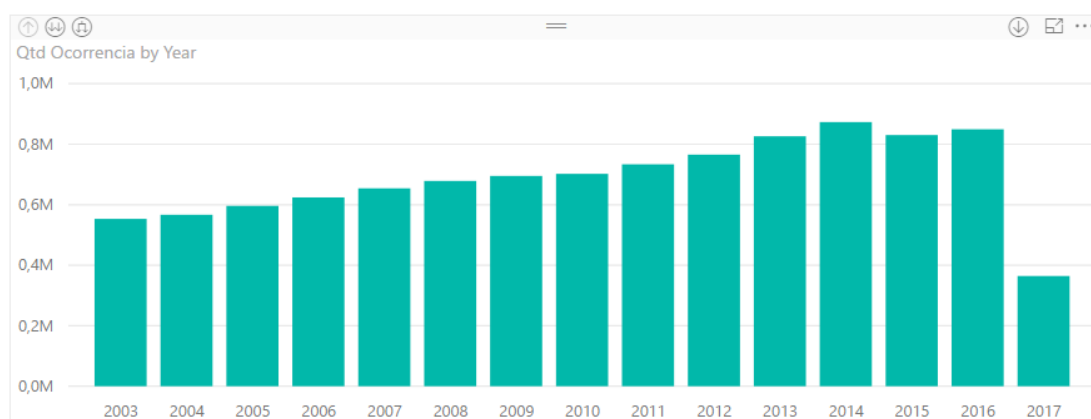


Figura 42 - Painel com o resumo temporal de Ocorrências

As visualizações mostram que o período de 2014 é o que conteve o maior número de registros de ocorrências. Na ferramenta Power BI é possível realizar operações de

“drill” para descobrir períodos mais específicos (Dia, Mês, Ano, Dia da semana, etc.). Foi construído um painel que contém os resumos das ocorrências registradas da tabela fato “Ocorrências Mensais”. Os gráficos mostram que a maioria dos crimes ocorrem na Capital, ou de natureza de Crime contra o patrimônio (roubos, furtos, sequestros). A proporção de crimes violentos também é significativa, visto que a média de mortes para cada 100 mil habitantes no Rio de Janeiro é de 22, segundo dados do próprio ISP. Este valor se aproxima a média de países em guerra, como a Síria, e mostra a situação preocupante que o Estado do Rio de Janeiro vive.

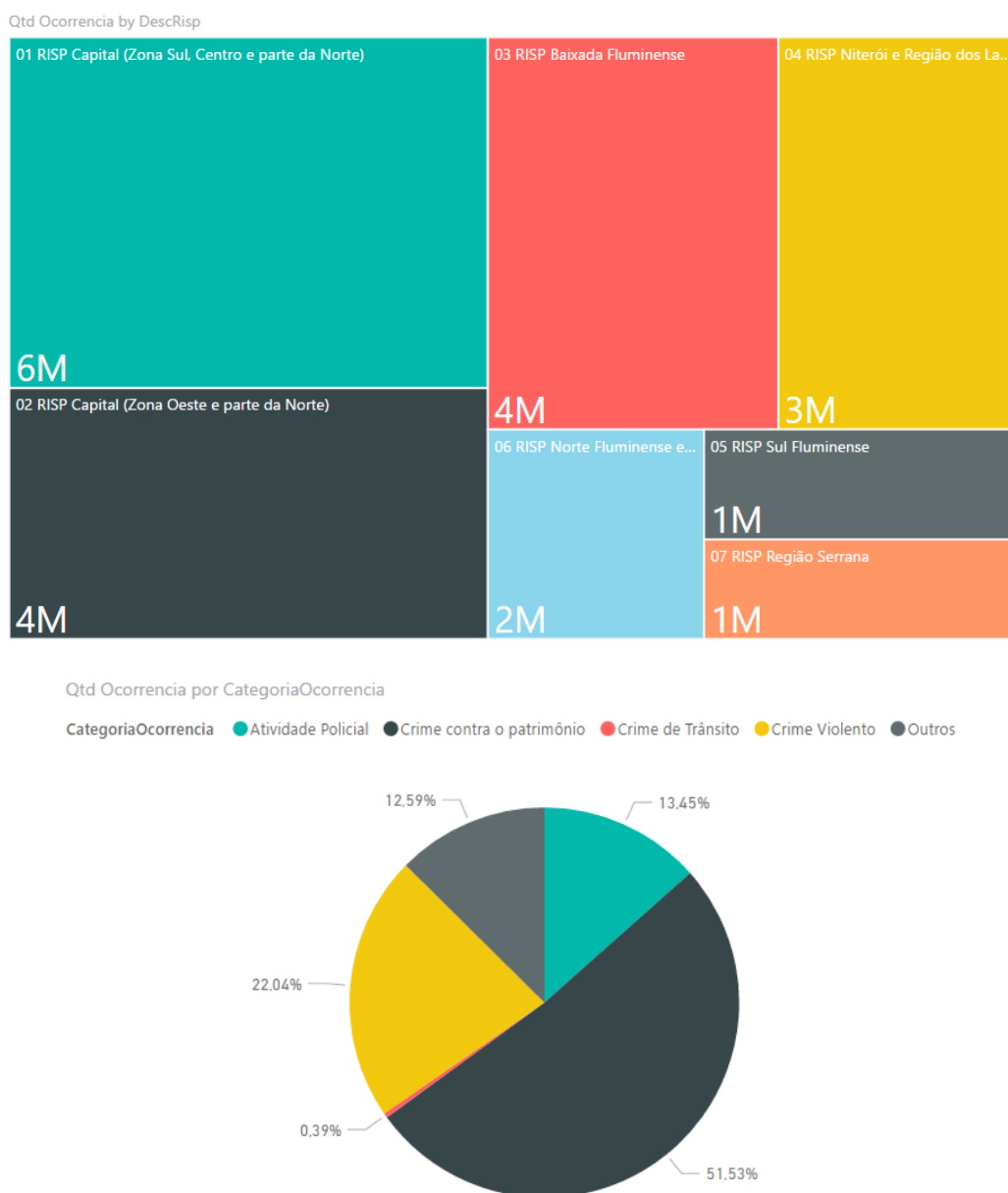


Figura 43 - Proporção de ocorrências por RISP e Categoria

Também podemos observar na figura a seguir que a 1ª RISP (Capital, Zona Sul, Centro e Parte da Norte) concentram o maior número de ocorrências gerais, porém a 3ª RISP concentra uma maior quantidade de crimes violentos. Com isso é possível realizar ações mais assertivas com base nas ocorrências observadas.

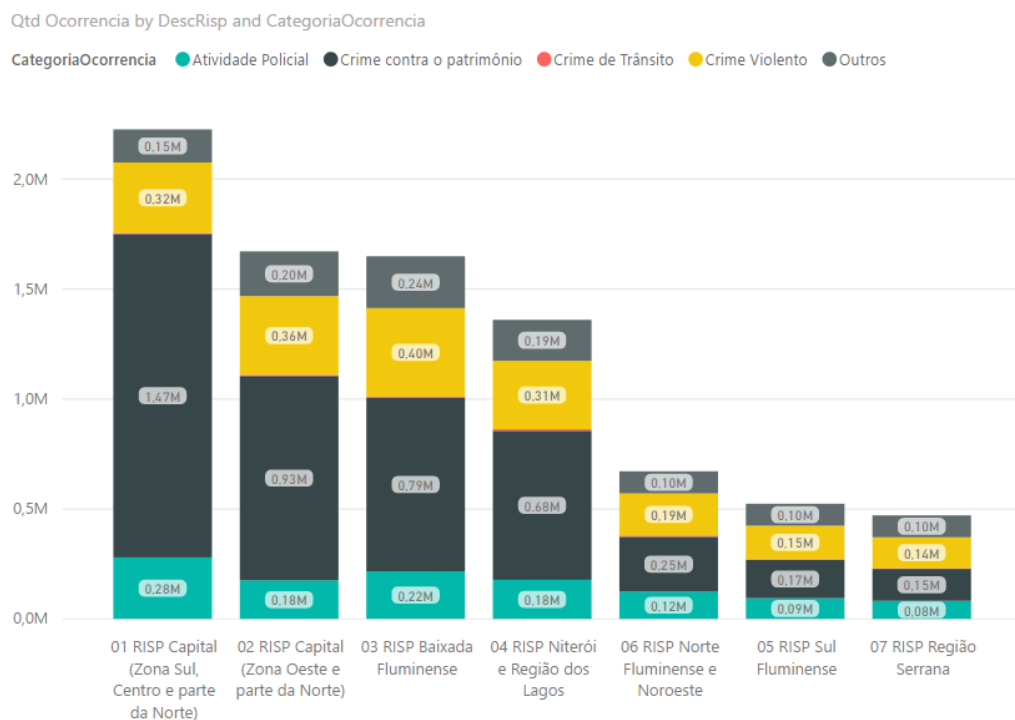


Figura 44 - Proporção de ocorrências por RISP

- **Qual o perfil das vítimas de Violência Contra a Mulher?**

A figura 41 mostra um dos painéis construídos para o fato “Violência contra a mulher”. É possível navegar através das diferentes dimensões construídas para obter dados descritivos. A informação mostra que as vítimas mulheres, em sua maioria de cor branca e solteiras, sofrem mais com a agressão de ex-companheiros. No gráfico de barras do canto inferior direito podemos observar a proporção do tipo de violência utilizada em diferentes faixas etárias.

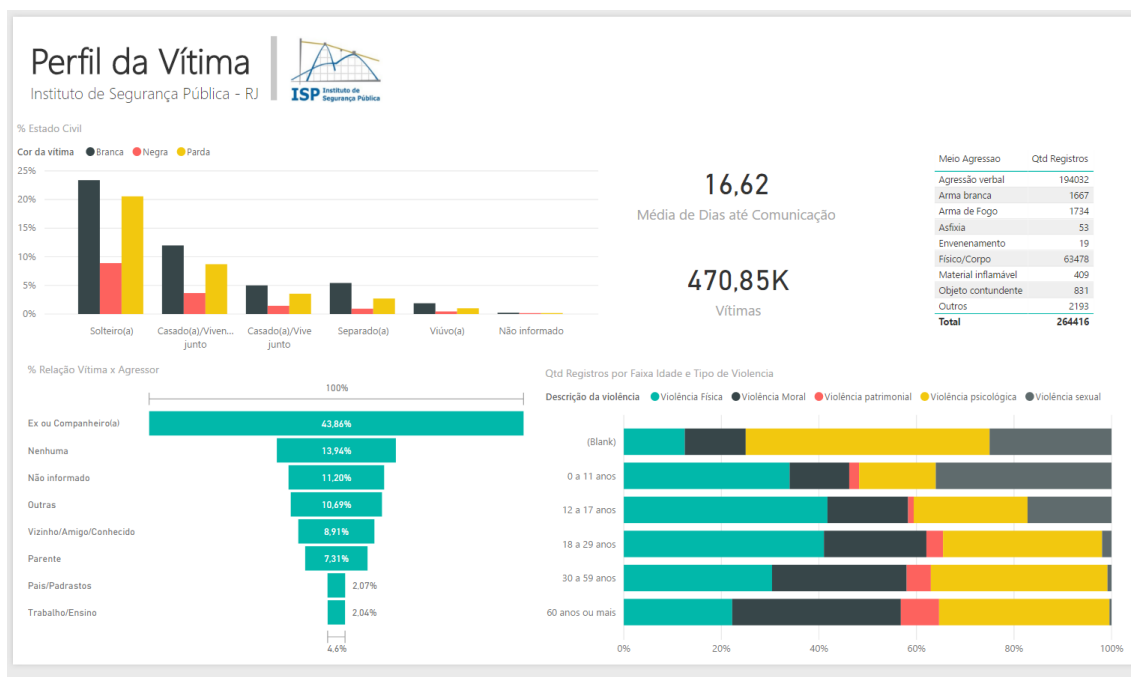


Figura 45 - Painel exibindo o perfil da vítima de Violência contra Mulher

No total, 470 mil vítimas foram registradas desde 2014, um valor que corresponde a quase 5% do total de registros feitos em delegacias. Ao obter estes insights, os analistas de dados podem tirar algumas conclusões, por exemplo, ao observarem que os maiores casos são de violência doméstica, e que a média de dias até a comunicação do crime é de 16 dias, os órgãos públicos podem criar ações para conscientizar as mulheres a alertarem as autoridades mais rapidamente para que este tipo de ocorrência não evolua para um caso de violência letal.

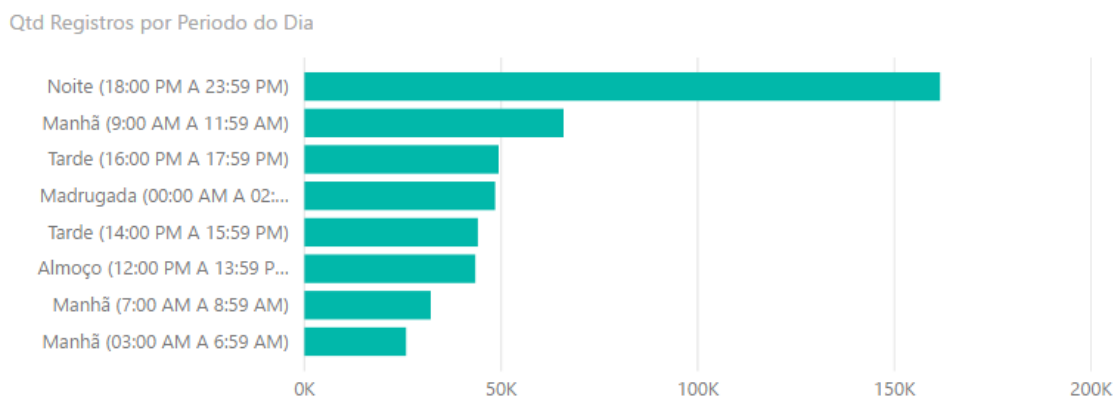


Figura 46 - Período do dia com ocorrências de violência contra a mulher

5 Conclusão

5.1 Considerações finais

O sucesso de um projeto de Business Intelligence depende de diversos fatores: Estratégias bem definidas, engajamento das partes envolvidas, tecnologias utilizadas, entre outros. Este projeto de graduação avaliou os fatores relacionados a tecnologia, e utilizou dados do Instituto de Segurança Pública do Rio de Janeiro para exploração de diversos recursos da plataforma da Microsoft, com a reprodução de um ambiente de Business Intelligence utilizando a versão Developer¹² do SQL Server 2016, voltada para ambientes de laboratório (não produtivos), mas que conta com todos os recursos das versões licenciadas.

O capítulo 2 introduziu conceitos de BI, Data Warehouse, OLAP e ETL, que são importantes para que o leitor entenda o contexto dos capítulos seguintes.

O Capítulo 3 apresentou a plataforma de BI Microsoft, que consiste em uma ampla seleção de ferramentas para diversos fins. Foi apresentado o banco de dados Microsoft SQL Server, e os recursos que o tornam uma solução viável de Data Warehouse, alinhada com os conceitos apresentados previamente. O Integration Services é uma ferramenta voltada para a criação de fluxos de ETL (Extração, transformação e carga de dados) integrado à plataforma, e apresenta recursos que facilitam a transformação de dados de qualquer origem. O Analysis Services é uma aplicação importante para a geração de uma camada semântica, exibida para o usuário final. O usuário pode visualizar o modelo de dados sem precisar de entendimento de banco de dados. As funcionalidades das ferramentas de visualização também foram analisadas, como o

12 <https://www.microsoft.com/en-us/sql-server/developer-get-started/>

Excel, Power BI e Reporting Services. Em um cenário de pequena/média escala, uma ferramenta como o Excel ou Power BI atende a maioria das necessidades de seus usuários, oferecendo maior agilidade na obtenção de insights. Alinhar o uso dessas ferramentas com o Reporting Services ou com o serviço do Power BI na nuvem permite a implantação da solução em um ambiente de grande escala.

O Capítulo 4 é dedicado para a construção de um projeto de BI com os dados obtidos no ISP. Alinhar o armazenamento dos dados com a lógica de negócios pode ser vantajoso para melhorar as consultas, e faz uma grande diferença para o administrador de banco de dados. Porém, para um analista de negócios que não possui experiência com tratamento de dados, o processo de geração de relatórios e indicadores pode se tornar um trabalho cansativo, onde serão necessárias muitas horas para organizar e cruzar informações distintas. É preferível que o arquiteto tenha o conhecimento dos dados que serão armazenados no banco para obter um dimensionamento mais eficiente. Deve-se ressaltar que alguns problemas na usabilidade tornam a experiência frustrante em alguns momentos, causando retrabalho para desenvolvedores menos experientes na ferramenta. Em um primeiro momento, os fluxos para extrair as informações dos arquivos do ISP parecem simples, porém as quantidades de arquivos distintos disponibilizados adicionaram complexidade que resultaram na criação do esquema “Staging”. Já os requisitos, foram elaborados de acordo com as informações disponíveis nas fontes de dados. Utilizando as metodologias apresentadas no capítulo 2 e as tecnologias apresentadas no capítulo 3, foi possível implementar os componentes de uma solução de BI e avaliá-los. Também foi possível responder as perguntas elaboradas na fase inicial do projeto.

Resultados obtidos

Os objetivos chave do projeto foram alcançados:

- Destacar a importância de Business Intelligence para a tomada de decisão
- Analisar e implementar as tecnologias apresentadas
- Obter de insights de ocorrências, letalidade e violência contra a mulher, apreensão de armas e crescimento de população.

5.2 Limitações do projeto

O ambiente limitado aos recursos de hardware não garantiu uma análise mais assertiva sobre a performance da “Engine” do SQL Server, embora seja possível observar ganhos utilizando particionamento e índices baseados em colunas. Algumas inconsistências da base de dados impediram a utilização de recursos mais avançados das ferramentas.

5.3 Trabalhos futuros

Para trabalhos futuros, é possível estender o escopo do projeto utilizando a linguagem R ou Python, integrados ao SQL Server a partir da versão 2017, aplicando análises estatísticas. A realização de paralelos com seus concorrentes de mercado, como a Amazon, Teradata, Oracle, entre outros fornecedores, ou com a crescente demanda de tecnologias que suportem Big Data.

Referências Bibliográficas

Barbieri, C. (2011), “BI2 – Business Intelligence – Modelagem e Qualidade”

Época (2017), “A crise da segurança pública no Rio de Janeiro”
Disponível em <<http://epoca.globo.com/brasil/noticia/2017/07/crise-da-seguranca-publica-no-rio-de-janeiro.html>>

Acessado em 22 de outubro de 2017.

Gartner (2017), “Magic Quadrant for Business Intelligence and Analytics”
Disponível em < <https://www.gartner.com/doc/3611117/magic-quadrant-business-intelligence-analytics>>

Acessado em 30 de outubro de 2017.

Gasparini, C. (2017), “Melhor profissão do ano nos EUA deve explodir no Brasil”.
Disponível em <<https://exame.abril.com.br/carreira/melhor-profissao-do-ano-nos-eua-deve-explodir-no-brasil-entenda/>>

Acessado em 12 de outubro de 2017.

Johal, A., Hordel, R., Allix, G. e Joyner, L., (2014), “Analyzing Data with Power BI”.

KdNuggets (2017), “New Leader, Trends, and Surprises in Analytics, Data Science, Machine Learning Software Poll”.

Disponível em <<https://www.kdnuggets.com/2017/05/poll-analytics-data-science-machine-learning-software-leaders.html>>

Acessado em 30 de outubro de 2017

Kimball, R. e Ross, M. (2013), “The Data Warehouse Toolkit”, 3th edition.

Malcolm, G. e O’Neill, C. (2014), “Implementing a SQL Data Warehouse”.

Microsoft (2017), “Gartner positions Microsoft as a Leader in BI and Analytics”. Disponível em <<https://powerbi.microsoft.com/pt-br/blog/gartner-positions-microsoft-as-a-leader-in-bi-and-analytics-platforms-for-ten-consecutive-years/>>

Acessado em 22 de outubro de 2017.

O Globo (2017), “Estudo mostra que Rio tem subnotificação de feminicídios”. Disponível em <<https://oglobo.globo.com/rio/estudo-mostra-que-rio-tem-subnotificacao-de-feminicidios-1-22008241>>

Acessado em 10 de novembro de 2017

Stephens, M. (2011), “A importância de Business Intelligence”. Disponível em <<http://www.blog.msbrasil.com.br/a-importancia-do-business-intelligence>>

Acessado em 10 de outubro de 2017.