

Revista da Graduação

Vol. 5

No. 1

2012

20

Seção: Faculdade de Informática

UM AMBIENTE DE KDD & BI PARA REDES TELECOM

Abraão Garcia Reckziegel; Adilson Favero

Este trabalho está publicado na Revista da Graduação.

ISSN 1983-1374

<http://revistaseletronicas.pucrs.br/ojs/index.php/graduacao/article/view/11417>

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE INFORMÁTICA
CURSO DE BACHARELADO EM SISTEMAS DE INFORMAÇÃO

ABRAÃO GARCIA RECKZIEGEL
ADILSON FAVERO

UM AMBIENTE DE KDD & BI PARA REDES TELECOM

Porto Alegre

2011

ABRAÃO GARCIA RECKZIEGEL
ADILSON FAVERO

UM AMBIENTE DE KDD & BI PARA REDES TELECOM

Trabalho apresentado como requisito à
colação de grau na Faculdade de Informática
do Programa de Curso de Bacharelado em
Sistemas de Informação da Pontifícia
Universidade Católica do Rio Grande do Sul.

Orientador: Prof. Dr. Duncan Dubugras Ruiz

Porto Alegre

2011

ABRAÃO GARCIA RECKZIEGEL

ADILSON FAVERO

UM AMBIENTE DE KDD & BI PARA REDES TELECOM

Trabalho apresentado como requisito à
colação de grau na Faculdade de Informática
do Programa de Curso de Bacharelado em
Sistemas de Informação da Pontifícia
Universidade Católica do Rio Grande do Sul.

Aprovado em: _____ de _____ de _____

BANCA EXAMINADORA:

Prof. Dr. Duncan Dubugras Ruiz

Prof. Dr. Michael da Costa Mora

Prof. Dr. Eduardo Henrique Pereira de Arruda

Porto Alegre

2011

Dedicamos este trabalho aos nossos pais, que tanto apoiaram e incentivaram o nosso crescimento profissional.

RESUMO

Atualmente as operadoras de telefonia móvel coletam milhares de dados sobre as chamadas executadas pelos seus clientes em suas redes. Estas informações servem como base para o órgão ANATEL aplicar indicadores de desempenho dos serviços prestados por estas operadoras.

O problema enfrentado pelas operadoras é a identificação em tempo hábil dos dados que registram chamadas com problemas, e que por sua vez, degradam os indicadores frente a ANATEL. Análises são necessárias para possibilitar uma atuação que vise à melhoria dos indicadores, porém, são demasiadamente longas, complexas e de grande esforço manual para obtê-las.

A solução mostrada neste documento tem como finalidade construir um ambiente de BI que facilite a tomada de decisão das operadoras, em tempo hábil no crescimento de seus indicadores ANATEL. Com o auxílio de um modelo *Data Warehouse* e técnicas de KDD, a ferramenta apresentará relatórios dinâmicos, de fácil leitura, construção simples e acesso ilimitado aos gestores destas áreas.

Palavras-chave: *Knowledge-Discovery Data. Data Mining, Business Intelligence. Modelo OLAP. Banco de dados. Telecomunicações. Data Warehouse.*

ABSTRACT

Currently, mobile operators collect data from thousands of calls executed by their customers. These data are needed to calculate performance indicators for the services provided by these operators, as ruled by ANATEL.

The problem faced by mobile operators is to identify, as fast as possible, the data from calls with problems, because they can degrade ANATEL quality indicators. Complex data analyses are needed to address this issue. However, they are too long, complex and high demanding.

We introduce a solution the design and implementation of a BI environment that facilitates the decision-making of operators on time, and improving their ANATEL quality indicators. With the support of a data warehouse model and using KDD techniques, the presented tool makes reporting dynamic and easy to read, with simple construction and unlimited access to the managers of different areas.

Keywords: Knowledge-Discovery. Data Mining, Business Intelligence. OLAP Models. Database. Telecommunications. Data Warehouse.

LISTA DE ILUSTRAÇÕES

Figura 1 – Diagrama de rede 2G/3G	14
Figura 2 – Equação do SMP3.....	16
Figura 3 – Equação do SMP5.....	19
Figura 4 – Equação do SMP6.....	21
Figura 5 – Benefícios da utilização de BI	25
Figura 6 – The OLAP Cube	27
Figura 7 – KDD processo de descoberta do conhecimento	28
Figura 8 – Tarefas de Data Mining	29
Figura 9 – Árvore gerada utilizando o algoritmo J48	30
Figura 10 – Modelo OLAP construído	35
Figura 11 – Diagrama de conversão do arquivo para BD	36
Figura 12 – Captura do gatilho criado em FATO	38
Figura 13 – Captura da tabela FATO após inserts	41
Figura 14 – Camadas SQL Server	42
Figura 15 – Propriedades da Job SSIS	42
Figura 16 – Cronograma da Job SSIS	43
Figura 17 – Componentes do pacote SSIS	43
Figura 18 – Diagrama ETL de DIM_OIGEM e DIM_DESTINO	45
Figura 19 – Diagrama ETL de DIM_TEMPO e FATO.....	48
Figura 20 – Data Source View	49
Figura 21 – Hierarquia de DIM_CAUSE.....	49
Figura 22 – Hierarquia de DIM_COBERTURA	50
Figura 23 – Hierarquia de DIM_DESTINO	50
Figura 24 – Hierarquia de DIM_ORIGEM	51
Figura 25 – Hierarquia de DIM_STATUS.....	51
Figura 26 – Hierarquia de DIM_TEMPO	51
Figura 27 – Printscreen de navegação 1	52
Figura 28 – Printscreen de navegação 2	52
Figura 29 – Propriedades das Medidas.....	54
Figura 30 – Propriedades dos KPIs.....	56
Figura 31 – Deploy dos KPIs	56
Figura 32 – Gerenciador de Soluções SSRS.....	57
Figura 33 – Relatório total indicadores	58
Figura 34 – Relatório SMP3.....	59
Figura 35 – Relatório SMP5.....	60
Figura 36 – Relatório SMP6.....	60
Figura 37 – Chamadas, Causas e Tarifas.....	61
Figura 38 – Top 10 Antenas sem completamente	62
Figura 39 – Interface Web dos documentos publicados.....	62
Figura 40 – Relatório do SSRS no SharePoint Server	63

Figura 41 – Relatórios dinâmicos no SharePoint.....	63
Figura 42 – Árvore IVR_NOK	65
Figura 43 – Árvore SAC_NOK.....	65
Figura 44 – Árvore CALLS_NOT_OK.....	66

LISTA DE TABELAS

Tabela 1 – Campos fornecidos no arquivo do CDR-View	15
Tabela 2 – Indicadores ANATEL	15
Tabela 3 – Regras para Numerador / Denominador SMP3	17
Tabela 4 – Regras para Numerador / Denominador SMP5	19
Tabela 5 – Regras para Numerador / Denominador SMP6	22
Tabela 6 - Requisitos de Hardware	23
Tabela 7 – Requisitos de Software.....	24
Tabela 8 – Mapeamento para carga de DIM_ORIGEM	44
Tabela 9 – Mapeamento para carga de DIM_DESTINO	44
Tabela 10 – Mapeamento para carga de DIM_CAUSE.....	45
Tabela 11 – Mapeamento para carga de DIM_STATUS.....	46
Tabela 12 – Mapeamento para carga de DIM_COBERTURA	46
Tabela 13 – Mapeamento para carga de DIM_TEMPO	47
Tabela 14 – Mapeamento para carga da tabela FATO	47

LISTA DE SIGLAS

ANATEL – Agência Nacional de Telecomunicações
2G – Segunda geração de tecnologia móvel pessoal
3G – Terceira geração de tecnologia móvel pessoal
BI – Business Intelligence
BSC – Base Station Controller
BTS – Base Transceiver Station
CDR View – Call Data Record View
DW – Data Warehouse
DM – Data Mining
ETL – Extract, Transform, Load
FK – Foreign Key
GSM – Global System for Mobile
KDD – Knowledge-Discovery in Database
KPI – Key Performance Indicators
TELECOM – Telecomunicações
MS – Microsoft
MSC – Mobile Service Center
OLAP – Online Analytical Processing
PK – Primary Key
PMM – Período de Medições Mensuráveis
PUCRS – Pontifícia Universidade Católica do Rio Grande do Sul
SLA – Service Level Agreement
SMP – Serviço Móvel Pessoal
SS7 – Signaling System #7
SSAS – SQL Server Analysis Services
SSIS – SQL Server Integration Services
SSRS – SQL Server Report Services
TI – Tecnologia da Informação

SUMÁRIO

1 INTRODUÇÃO	12
2 OBJETIVOS	13
2.1 CENÁRIO	13
2.2 CARACTERIZAÇÃO DO PROBLEMA	15
2.2.1 O indicador SMP3	16
2.2.2 Contabilização do SMP3.....	17
2.2.3 O indicador SMP5	19
2.2.4 Contabilização do SMP5.....	19
2.2.5 O indicador SMP6	21
2.2.6 Contabilização do SMP6.....	21
2.3 REQUISITOS DE UMA SOLUÇÃO.....	22
2.3.1 Recursos de Hardware	22
2.3.2 Recursos de Software.....	23
3 FUNDAMENTAÇÃO TEÓRICA	25
3.1 BUSINESS INTELLIGENCE	25
3.2 DATA WAREHOUSE	26
3.3 MODELO OLAP.....	26
3.3.1 Navegabilidade do Cubo.....	27
3.4 KNOWLEDGE - DISCOVERY IN DATABASE.....	28
3.4.1 Data Mining	29
3.4.2 BI & KDD aplicados em Telecom	31
3.5 FERRAMENTAS PARA BI	31
3.5.1 Microsoft Excel	31
3.5.2 Microsoft SQL Server Management Studio.....	32
3.5.3 Microsoft SQL Server Integration Services	32
3.5.4 Microsoft SQL Server Analysis Services	32
3.5.5 Microsoft SQL Server Reporting Services	32
3.5.6 Microsoft Office SharePoint Server	33

4 UM AMBIENTE DE KDD & BI PARA REDES TELECOM	34
4.1 QUESTÕES DE NEGÓCIO.....	34
4.2 CONSTRUÇÃO DO MODELO.....	34
4.3 DESCRIÇÃO DAS TABELAS	36
4.4 APLICAÇÃO DAS REGRAS	38
4.5 O PROCESSO DE ETL	41
4.5.1 População via SSIS	42
4.5.2 ETL de DIM_ORIGEM e DIM_DESTINO	44
4.5.3 ETL de DIM_CAUSE, DIM_STATUS e DIM_COBERTURA	45
4.5.4 ETL de DIM_TEMPO e FATO	46
4.6 A GERAÇÃO DO CUBO.....	48
4.7 A IMPLEMENTAÇÃO DOS KPIs	52
5 EXPERIMENTAÇÃO DA SOLUÇÃO	57
5.1 RELATÓRIOS.....	57
5.1.1 Totalização dos Indicadores	58
5.1.2 Relatório SMP3	58
5.1.3 Relatório SMP5	59
5.1.4 Relatório SMP6	60
5.1.5 Chamadas, Causas e Tarifas	61
5.1.6 Cidades e Antenas com falhas	61
5.2 PUBLICAÇÃO	62
5.3 MINERAÇÃO	64
5.4 DEPOIMENTOS	66
6 CONSIDERAÇÕES FINAIS	68
6.1 LIÇÕES APRENDIDAS	68
6.2 TRABALHOS FUTUROS	68
REFERÊNCIAS.....	70
GLOSSÁRIO	71

1 INTRODUÇÃO

Segundo a ANATEL (18/11/2010, <http://www.anatel.gov.br/>): “O Brasil atingiu a marca de mais de um aparelho celular por habitante, superando países como Estados Unidos, Japão e França”. A prioridade constante das empresas que oferecem serviços de telecomunicações no país é garantir a qualidade da rede e atender as solicitações dos clientes.

Uma determinada operadora de telefonia móvel, proprietária de uma ferramenta que registra todas as chamadas ocorridas em sua rede, busca identificar em sua base de dados informações que auxiliem na tomada de decisões frente aos seus indicadores de desempenho. Devido à grandiosidade dos registros de chamadas, este trabalho torna-se demorado, complexo e pouco difundido sem o auxílio de ferramentas de BI.

A solução proposta neste documento é modelar um *data warehouse* que receba os dados de registros das chamadas na operadora, automatizando o complexo processo de análise dos registros e gerando de forma dinâmica e automática relatórios que facilitem o acompanhamento e tomada de decisão dos gestores envolvidos.

Este documento visa abranger todos os passos que foram necessários até que o produto final fosse implantado no ambiente do cliente, contemplando assim um embasamento teórico que descreve a solução abordada, uma seção explicativa sobre a tecnologia utilizada, os objetivos atingidos pelo projeto e todo o fluxo de desenvolvimento, bem como o resultado alcançado e algumas considerações finais.

2 OBJETIVOS

O presente trabalho tem como objetivo auxiliar os gestores e analistas da empresa Alcatel-Lucent na tomada de decisões destinadas à qualidade dos indicadores ANATEL, por intermédio de *Business Intelligence* e *Knowledge-Discovery in database*, para isto, implantaremos uma solução global que permita fornecer relatórios de desempenho de rede, navegabilidade nas bases de dados de chamadas e acompanhamento dos indicadores SMP3, SMP5 e SMP6.

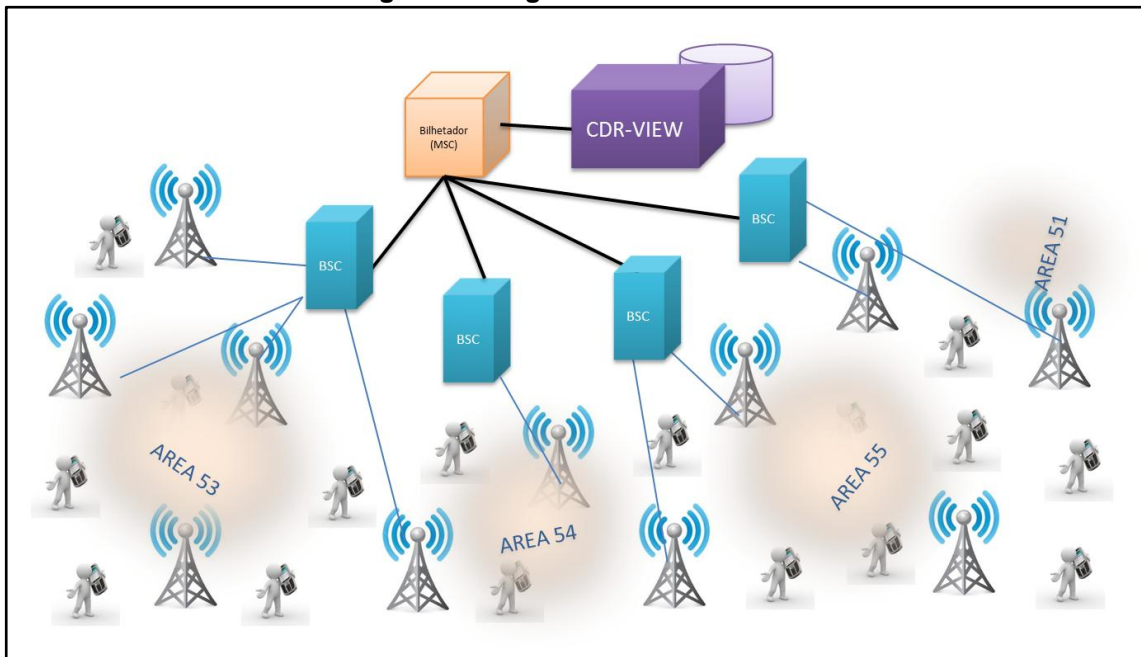
Atualmente os analistas e gestores procuram executar qualquer tipo de ação que possa vir ajudar nos indicadores, ainda que não seja comprovado que esta ação foi realmente eficaz. Prévias dos indicadores são realizadas em planilhas do MS Excel, levando de quatro a cinco dias de trabalho dos colaboradores para encontrar possíveis falhas na rede. Como prática, os dados são extraídos uma semana antes da data de medições acordada junto a ANATEL, isto significa que, caso encontrado algum ofensor nos indicadores, restarão apenas dois ou três dias para atuação na falha. Também devemos levar em consideração que no período de uma semana o cenário coletado possa ter agravado ou normalizado, perdendo assim a referência.

Nossa finalidade é acompanhar estes ofensores no tempo estabelecido pela plataforma CDR-View, ou seja, com um dia de atraso, nossa solução deverá consolidar dados de cerca de três milhões de chamadas, nos PMMs, e gerar relatórios sobre motivos de desconexão, áreas de cobertura em municípios que se encontram congestionadas, chamadas que demoram até o completamento e crescimento de tráfego.

2.1 CENÁRIO

Atualmente as redes que proporcionam serviços de segunda e terceira geração (2G/3G), utilizam-se do padrão aberto GSM, desenvolvido pelo grupo 3GPP na década de 80 e adotado no Brasil há 10 anos. A estrutura física que hoje geralmente é aplicada a companhias telefônicas no Brasil esta formada basicamente em cinco nodos que se comunicam via protocolo SS7 conforme a figura a seguir:

Figura 1 - Diagrama de rede 2G/3G



Fonte: Alcatel-Lucent.

Com base na figura acima, veremos que cada elemento de rede tem uma função específica:

- **Terminais de acesso (celulares):** Constituem-se nos aparelhos clientes, de uso proprietário, que determinam as solicitações de acesso a serviços quando registradas na rede.

- **BTS:** As antenas apresentadas na figura representam as estações rádio base (ERB) ou BTS que fornecem cobertura e comunicação do aparelho cliente com a BSC.

- **BSC:** Realizam a função de controle e multiplexação das informações provindas das BTSs. Nelas também são armazenados dados intermediários sobre chamadas e status dos terminais de acesso.

- **MSS:** Realiza a função de principal servidor, análise de dígitos e gateway para os outros pontos de interconexão (outras operadoras e serviços). Nesta central também são gerados os CDRs e registros de status das chamadas.

- **CDR-View:** Plataforma cliente que aglomera as informações de todas as chamadas (CDR) e solicitações registradas na rede em um período de 24 (vinte e quatro) horas. Nesta, são derivadas informações sobre serviços de tarifação, cobertura e desconexões na rede. A plataforma opera sob um sistema *client*, acessado por uma interface *Java applet*. Diariamente a plataforma armazena

informações com cerca de três milhões de chamadas, isto equivale, em média, 6GB de informações que serviu de base para este trabalho.

A plataforma permite que façamos a importação destas informações no formato CSV (Arquivo de Colunas Separadas por Vírgulas), que servirá como entrada para a solução implementada. Estes arquivos, chamados tecnicamente de “Bilhetes” constituem uma tabela única composta pelos seguintes campos:

Tabela 1 – Campos fornecidos no arquivo do CDR-View

Nome da Coluna	Descrição Técnica	Exemplo de Dado
SimCard	Código de criptografia do assinante.	"724314121370115"
Origem	Número do celular de quem origina.	"5185852962"
OpEntrada	Operadora que esta originando a ligação.	"BRT CELULAR - RS"
TipoA	Categoria do cliente que origina a ligação.	"Pré-pago"
DataHora	Data e hora do início da ligação.	"21/07/2011 09:00:02"
PMM	Período em que a ligação contabiliza para ANATEL.	"PMM1"
Destino	Número discado pelo originador da chamada.	"145134801000"
OpSaida	Operadora proprietária do número de destino.	"GVT - RS"
IndPortB	Indica se o destino é portado para outra operadora.	"Não Portado"
TAT	Tempo até o Atendimento da chamada.	"44"
TET	Tempo de Estabelecimento (até chamar).	"0"
Xnumber	Código que define o serviço de caixa postal.	"--"
CallRel	Código que define a desconexão da chamada.	"0000030A"
STCellA	BTS (antena) que o Originador se encontra ao disar.	"TPS411B"
BSC	BSC onde a BTS se encontra.	"BSC04PAE"
Bilhetador	MSC onde a BSC se encontra.	"MSS77PAE"
Área	Código de área onde o Originador se encontra.	"51"
RecType	Sigla que define o status contabilizado da chamada.	"UCA"

Fonte: CDR-View.

2.2 CARACTERIZAÇÃO DO PROBLEMA

Como métricas que regularizam o desempenho das redes móveis de cada operadora frente aos clientes, a ANATEL, cultiva os indicadores dispostos na tabela a seguir:

Tabela 2 – Indicadores ANATEL

Indicador	Descrição
SMP1	Taxa de reclamações.
SMP2	Taxa de reclamações de congestionamento de voz a cada 1000 (mil) clientes em operação.
SMP3	Taxa de chamadas completadas para os centros de atendimento.
SMP4	Taxa de atendimento pela telefonista / atendente do Call Center.

SMP5	Taxa de chamadas originadas completadas.
SMP6	Taxa de Estabelecimento de chamadas.
SMP7	Taxa de queda de ligação.
SMP8	Taxa de resposta ao usuário.
SMP9	Taxa de resposta aos pedidos de informação.
SMP10	Taxa de atendimento pessoal ao usuário.
SMP11	Taxa de atendimentos relativos a contas por 1000 (mil) contas emitidas.
SMP12	Taxa de recuperação de defeitos.

Fonte: PGMQ 2010.

Os indicadores supramencionados são aplicados às diversas áreas das operadoras telecom, eles seguem períodos de coletas estatísticas denominados PMMs que estão distribuídos em três fases do dia:

- **PMM1:** Coleta de dados de chamadas das 09h às 11h;
- **PMM2:** Coleta de dados de chamadas das 14h às 16h;
- **PMM3:** Coleta de dados de chamadas das 20h às 22h.

O ponto focal deste trabalho é auxiliar nos indicadores relacionados à otimização e desempenho de rede. Sendo assim, especificaremos a seguir as regras utilizadas na contabilização dos indicadores SMP3, SMP5 e SMP6.

2.2.1 O indicador SMP3

Segundo a ANATEL (PGMQ SMP Consulta Pública nº 27/2010, p. 22): “As chamadas originadas na rede da prestadora e destinadas ao seu Centro de Atendimento devem ser completadas, em cada PMM, no mínimo, em 98% (noventa e oito por cento) dos casos”. Vejamos a partir da equação de contabilização do indicador exposta abaixo:

Figura 2 - Equação do SMP3

$$SMP3 = \frac{A}{B} \times 100$$

Fonte: PGMQ 2010.

A - número total de chamadas originadas na rede da prestadora para os Centros de Atendimento, atendidas por Sistemas de Autoatendimento ou pelas telefonistas/atendentes, em cada PMM, no mês;

B - número total de tentativas de originar chamadas na rede da prestadora para os Centros de Atendimento, contadas a partir da alocação do canal de voz, em cada PMM, no mês.

2.2.2 Contabilização do SMP3

Para mensurar o numerador e denominador, criam-se filtros nos arquivos CSV, dando-lhes um nome de regra. Tais regras se apresentam da seguinte forma:

Tabela 3 – Regras para Numerador / Denominador SMP3

	Descrição	Regras (filtros)
SMP3	Numerador Número de chamadas originadas completadas para o Centro de Atendimento.	SAC_OK_MOC+SAC_OK_UCA + IVR_OK_MOC + IVR_OK_UCA
	Denominador Número de tentativas de originar chamadas para o Centro de Atendimento, a partir da alocação do canal de voz.	SAC_OK_MOC+SAC_OK_UCA+SAC_NOK + IVR_OK_MOC + IVR_OK_UCA + IVR_NOK

Fonte: Alcatel-Lucent.

Onde reconhecemos os seguintes filtros para construção das regras:

- **SAC_OK_MOC:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “MOC”, SimCard com início “72431” ou “72416”, Destino igual a 'B1414', '1053', 'B144'; 'B800'; 'B803', 'B893' ou 'B894' e TAT menor que 302 segundos;
- **SAC_OK_UCA:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, SimCard com início “72431” ou “72416”, Destino igual a 'B1414', '1053', 'B144'; 'B800'; 'B803', 'B893' ou 'B894', Xnumber diferente de 5503140*; 5501646100100; 55??100100; 100100; 55100100; 0??88022057, CallRel igual a 00000319; 00000304; 00000306; 00000010; 00000040; 00000315; 00000300; 00000301; 00000307; 00000314; 00000817; 00000706; 00000C01; 00000A0D; 00000012; 00000D07; 00000000; 00000308; 00000819; 00000312; 00000212; 00000B17; 00000C00; 00000C02; 00000818; 00000305; 00000816; 00000318; 00000815; 00000206; 00000310; 00000006; 00000016, e STCellA não nulo;

- **IVR_OK_MOC:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “MOC”, SimCard com início “72431” ou “72416”, Destino igual a ‘B111’, ‘B222’, ‘B801’, ‘B804’ ou ‘B805’ e TAT menor que 302 segundos;
- **IVR_OK_UCA:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, SimCard com início “72431” ou “72416”, Destino igual a ‘B111’, ‘B222’, ‘B801’, ‘B804’ ou ‘B805’, CallRel igual a 00000319; 00000304; 00000306; 00000010; 00000040; 00000315; 00000300; 00000301; 00000307; 00000314; 00000817; 00000706; 00000C01; 00000A0D; 00000012; 00000D07; 00000000; 00000308; 00000819; 00000312; 00000212; 00000B17; 00000C00; 00000C02; 00000818; 00000305; 00000816; 00000318; 00000815; 00000206; 00000310; 00000006; 00000016, e STCellA não nulo;
- **SAC_NOK:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, SimCard com início “72431” ou “72416”, Destino igual a ‘B1414’, ‘1053’, ‘B144’, ‘B800’, ‘B803’, ‘B893’ ou ‘B894’, Xnumber diferente de 5503140*; 5501646100100; 55??100100; 100100; 55100100; 0??88022057, CallRel diferente de 00000319; 00000304; 00000306; 00000010; 00000040; 00000315; 00000300; 00000301; 00000307; 00000314; 00000817; 00000706; 00000C01; 00000A0D; 00000012; 00000D07; 00000000; 00000308; 00000819; 00000312; 00000212; 00000B17; 00000C00; 00000C02; 00000818; 00000305; 00000816; 00000318; 00000815; 00000206; 00000310; 00000006; 00000016, e STCellA não nulo;
- **IVR_NOK:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, SimCard com início “72431” ou “72416”, Destino igual a ‘B111’, ‘B222’, ‘B801’, ‘B804’ ou ‘B805’, CallRel diferente de 00000319; 00000304; 00000306; 00000010; 00000040; 00000315; 00000300; 00000301; 00000307; 00000314; 00000817; 00000706; 00000C01; 00000A0D; 00000012; 00000D07; 00000000; 00000308; 00000819; 00000312; 00000212; 00000B17; 00000C00; 00000C02; 00000818; 00000305; 00000816; 00000318; 00000815; 00000206; 00000310; 00000006; 00000016, e STCellA não nulo;

Ao final da contagem das tuplas que pertencem a cada regra, a fórmula abaixo é aplicada para obtenção do valor final do indicador:

$$SMP3 = 100 * (SAC_OK_MOC + SAC_OK_UCA + IVR_OK_MOC + IVR_OK_UCA) / (SAC_OK_MOC + SAC_OK_UCA + SAC_NOK + IVR_OK_MOC + IVR_OK_UCA + IVR_NOK)$$

2.2.3 O indicador SMP5

Segundo a ANATEL (PGMQ SMP Consulta Pública nº 27/2010, p. 22): “As tentativas de obtenção do canal de sinalização e controle devem permitir o estabelecimento de uma chamada, em cada PMM, no mínimo em 67% (sessenta e sete por cento) dos casos”. Logo, isto se transcreve na seguinte equação:

Figura 3 - Equação do SMP5

$$SMP5 = \frac{A}{B} \times 100$$

Fonte: PGMQ 2010.

A - total de acessos ao canal de sinalização e controle para estabelecer uma chamada, em cada PMM, no mês; *B* - total de tentativas de acessar o canal de sinalização e controle, em cada PMM, no mês;

2.2.4 Contabilização do SMP5

Assim como no indicador SMP3, este indicador apresenta regras com filtros dos dados apresentadas assim:

Tabela 4 – Regras para Numerador / Denominador SMP5

	Descrição	Regras (filtros)
SMP5	Numerador Número de chamadas originadas completadas.	CALLS_OK+AN_NOK+OK_801 +OK_804+VM_NOK
	Denominador Número de tentativas de originar chamadas, a partir da alocação do canal de voz.	CALLS_OK+CALLS_NOT_OK +AN_NOK + VM_NOK

Fonte: Alcatel-Lucent.

Onde reconhecemos os seguintes filtros para construção das regras:

- **CALLS_OK**: Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “MOC” e TAT menor que 302 segundos;

- **CALLS_NOT_OK:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, Xnumber diferente de ‘5503140*’; 5501646100100; 55??100100; 100100; 55100100; 0??88022057, CallRel diferente de ‘00000319’; ‘00000304’; ‘00000306’; ‘00000010’; ‘00000040’; ‘00000315’; ‘00000300’; ‘00000301’; ‘00000307’; ‘00000314’; ‘00000817’; ‘00000706’; ‘00000C01’; ‘00000A0D’; ‘00000012’; ‘00000D07’; ‘00000000’; ‘00000308’; ‘00000819’; ‘00000312’; ‘00000212’; ‘00000B17’; ‘00000C00’; ‘00000C02’; ‘00000818’; ‘00000305’; ‘00000816’; ‘00000318’; ‘00000815’; ‘00000206’; ‘00000310’; ‘00000006’; ‘00000016’ e STCellA diferente de nulo;
- **AN_NOK:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, Xnumber diferente de ‘5503140*’; 5501646100100; 55??100100; 100100; 55100100; 0??88022057, CallRel igual a ‘00000319’; ‘00000304’; ‘00000306’; ‘00000010’; ‘00000040’; ‘00000315’; ‘00000300’; ‘00000301’; ‘00000307’; ‘00000314’; ‘00000817’; ‘00000706’; ‘00000C01’; ‘00000A0D’; ‘00000012’; ‘00000D07’; ‘00000000’; ‘00000308’; ‘00000819’; ‘00000312’; ‘00000212’; ‘00000B17’; ‘00000C00’; ‘00000C02’; ‘00000818’; ‘00000305’; ‘00000816’; ‘00000318’; ‘00000815’; ‘00000206’; ‘00000310’; ‘00000006’; ‘00000016’ e STCellA diferente de nulo;
- **OK_801:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, Destino igual a ‘B111’ ou ‘B801’ e CallRel igual a ‘0000030B’; ‘00000015’; ‘00000B32’; ‘00000318’;
- **OK_804:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, Destino igual a ‘B222’, ‘B804’ ou ‘B805’ e CallRel igual a ‘0000030A’, ‘0000030B’, ‘00000015’, ‘00000B32’ ou ‘00000318’;
- **VM_NOK:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, Xnumber igual a ‘5503140*’; 5501646100100; 55??100100; 100100; 55100100 ou 0??88022057;
- **DSC_A_5_SEG:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “UCA”, Xnumber diferente de ‘5503140*’; 5501646100100; 55??100100; 100100; 55100100 ou 0??88022057 e CallRel igual a ‘0000030A’, ‘0000030B’, ‘00000202’, ‘00000007’ ou ‘00000013’, TAT menor que 5 segundos e STCellA diferente de nulo;

- **ABAND_A_5_IVR:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a "UCA", Destino igual a "B111", "B222", "B801", "B804" ou "B805", CallRel igual a '0000030A', '0000030B', '00000202', '00000006', '00000007' ou '00000013' TAT menor que 5 segundos e STCellA diferente de nulo;

Ao final da contagem das tuplas que pertencem a cada regra, a fórmula abaixo é aplicada para obtenção do valor final do indicador:

$$SMP5 = 100 * (CALLS_OK + AN_NOK + OK_801 + OK_804 + VM_NOK) / (CALLS_OK + CALLS_NOT_OK + AN_NOK + VM_NOK)$$

2.2.5 O indicador SMP6

Segundo a ANATEL (PGMQ SMP Consulta Pública nº 27/2010, p. 22): "Todas as tentativas de envio de mensagens devem resultar em entrega ao usuário final em até 60 (sessenta) segundos no mínimo em 95% (noventa e cinco por cento) dos casos". Onde transcrevemos sua fórmula na figura abaixo:

Figura 4 - Equação do SMP6

$$SMP6 = \frac{A}{B} \times 100$$

Fonte: PGMQ 2010.

A - Número total de mensagens entregues ao usuário em até 60 (sessenta) segundos, no mês;

B - Número total de tentativas de envio de mensagens na rede da prestadora, no mês.

2.2.6 Contabilização do SMP6

Na mesma linha, o numerador e denominador deste indicador se dividem nas seguintes regras:

Tabela 5 – Regras para Numerador / Denominador SMP6

	Descrição	Regras (filtros)
SMP6	Numerador Número de tentativas de originar chamadas estabelecidas em até 10 segundos, contados a partir da alocação do canal de voz.	CALLS_OK_10
	Denominador Número total de tentativas de originar chamadas.	CALLS_OK_SMP6

Fonte: Alcatel-Lucent.

- **CALLS_OK_10:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “MOC” ou “MTC”, com TET menor que 11 segundos e TAT menor que 302 segundos;
- **CALLS_OK_SMP6:** Regra que contabiliza a quantidade de chamadas que respeitam o filtro de RecType igual a “MOC” ou “MTC” e TAT menor que 302 segundos.

Ao final da contagem das tuplas que pertencem a cada regra, a fórmula abaixo é aplicada para obtenção do valor final do indicador:

$$SMP6_10 = 100 * (CALLS_OK_10) / (CALLS_OK_SMP6)$$

É importante evidenciar que esta série de filtros apresentados anteriormente é o reflexo da grandiosidade de trabalho que os analistas e gestores sofrem ao efetuar uma tomada de decisão em relação aos indicadores ou simplesmente na confecção de um documento final.

2.3 REQUISITOS DE UMA SOLUÇÃO

A seguir mostramos os recursos de *hardware* e *software* necessário durante a implementação desta solução. Informamos que todos os recursos descritos neste trabalho foram concebidos pelo convênio acadêmico MSDN Microsoft e PUCRS.

2.3.1 Recursos de *Hardware*

A tabela abaixo mostra os requisitos do sistema do SQL Server 2008 R2 Enterprise (64 bits) x64:

Tabela 6 - Requisitos de Hardware

Componente	Requisito
Processador	Tipo de processador: <ul style="list-style-type: none"> · Mínimo: AMD Opteron, AMD Athlon 64, Intel Xeon com suporte Intel EM64T, Intel Pentium IV com suporte EM64T Velocidade do processador: <ul style="list-style-type: none"> · Mínimo: 1,4 GHz · Recomendável: 2,0 GHz ou mais rápido
Sistema operacional	Windows Server 2003 SP2 64 bits x64 Datacenter Windows Server 2003 SP2 64 bits x64 Enterprise Windows Server 2003 SP2 64 bits x64 Standard Windows Server 2003 R2 SP2 64 bits x64 Datacenter Windows Server 2003 R2 SP2 64 bits x64 Enterprise Windows Server 2003 R2 SP2 64 bits x64 Standard Windows Server 2008 SP2 64 bits x64 Datacenter Windows Server 2008 SP2 64 bits x64 Datacenter sem Hyper-V Windows Server 2008 SP2 64 bits x64 Enterprise, Windows Server 2008 SP2 64 bits x64 Enterprise sem Hyper-V Windows Server 2008 SP2 64 bits x64 Standard Windows Server 2008 SP2 64 bits x64 Standard sem Hyper-V Windows Server 2008 SP2 64 bits x64 Web Windows 2008 R2 64 bits x64 Datacenter Windows 2008 R2 64 bits x64 Enterprise Windows 2008 R2 64 bits x64 Standard Windows 2008 R2 64 bits x64 Web Windows Server 2008 R2 x64 para Windows Essential Server Solutions
Memória	RAM: <ul style="list-style-type: none"> · Mínimo: 1 GB · Recomendável: 4 GB ou mais · 2 TB (O SQL Server Enterprise Edition dá suporte a um máximo de 2 TB de RAM ou o máximo do sistema operacional, sendo considerado o menor valor).

Fonte: www.microsoft.com

2.3.2 Recursos de *Software*

Em um sistema operacional Windows x64 bits, os seguintes requisitos aplicam-se a todas as instalações do SQL Server 2008 R2:

Tabela 7 – Requisitos de Software

Componente	Requisito
Framework2	A Instalação do SQL Server instala os seguintes componentes de software requeridos pelo produto: <ul style="list-style-type: none"> · .NET Framework 3.5 SP11 · SQL Server Native Client · Arquivos de suporte à Instalação do SQL Server
Software2	A Instalação do SQL Server requer o Microsoft Windows Installer 4.5 ou uma versão mais recente; Depois de instalar os componentes necessários, a Instalação do SQL Server verifica se o computador em que o SQL Server 2008 R2 será instalado também atende a todos os outros requisitos para uma instalação bem-sucedida.
Software de rede	Os requisitos de software de rede para as versões de 64 bits do SQL Server 2008 R2 são iguais aos requisitos para as versões de 32 bits. Os sistemas operacionais com suporte têm software de rede interno. As instâncias padrão e nomeadas autônomas dão suporte aos seguintes protocolos de rede: <ul style="list-style-type: none"> · Memória compartilhada · Pipes nomeados · TCP/IP
Virtualização	O SQL Server 2008 R2 tem suporte em ambientes de máquina virtual em execução na função Hyper-V nas edições Windows Server 2008 SP2 Standard, Enterprise e Datacenter. A máquina virtual deve executar um sistema operacional com suporte para a edição específica do SQL Server 2008 R2 listada posteriormente neste tópico. Além dos recursos exigidos pela partição pai, cada máquina virtual (partição filho) deve ser fornecida com recursos de processador, memória e recursos de disco suficientes para sua instância do SQL Server 2008 R2. Observações: <ul style="list-style-type: none"> · É recomendável que o SQL Server 2008 R2 seja encerrado antes de desligar a máquina virtual.
Software de Internet	O Microsoft Internet Explorer 6 SP1 ou uma versão posterior é necessário para todas as instalações do SQL Server 2008 R2. O Internet Explorer 6 SP1 ou posterior é necessário para o Microsoft Management Console (MMC), SQL Server Management Studio, Business Intelligence Development Studio, o componente Designer de Relatórios do Reporting Services e a ajuda em HTML.
Disco rígido	Os requisitos de espaço em disco variam com os componentes do SQL Server 2008 R2 instalados.
Unidade	É necessária uma unidade de CD ou DVD, conforme apropriado, para a instalação a partir de disco.
Vídeo	As ferramentas gráficas do SQL Server 2008 R2 exigem resolução Super VGA ou superior: resolução mínima de 800 x 600 pixels.
Outros dispositivos	Dispositivo apontador: um mouse Microsoft ou um dispositivo apontador compatível é necessário.

Fonte: www.microsoft.com

3 FUNDAMENTAÇÃO TEÓRICA

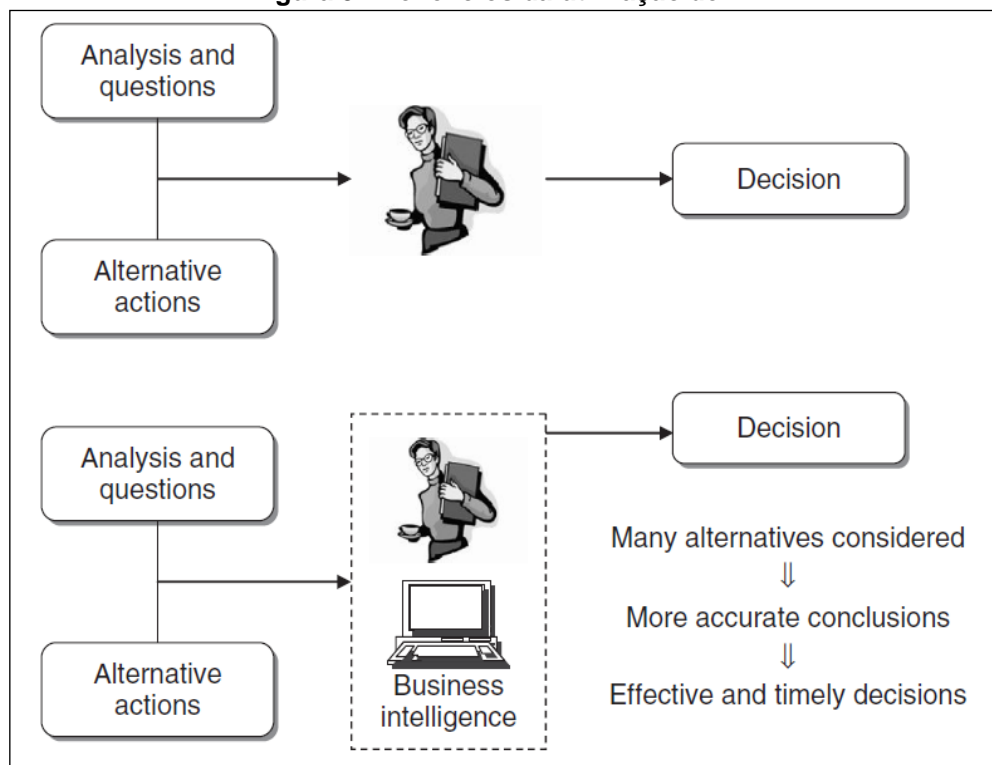
Neste capítulo vamos apresentar um estudo sobre os conceitos de *Business Intelligence*, *Data Warehouse*, Modelos OLAP e *Knowledge-Discovery in database*.

3.1 BUSINESS INTELLIGENCE

Business Intelligence ou inteligência empresarial significa obter as informações primárias da empresa, ou seja, cruzar as informações para que a melhor decisão possa ser tomada. Esse conceito, criado na década de 80, tem por finalidade apoiar a tomada de decisão a nível estratégico.

Na figura a seguir, é possível ilustrar um modelo onde o *workflow* de BI é aplicado, diferenciando-o de uma metodologia básica e revelando seus fundamentais aperfeiçoamentos:

Figura 5 - Benefícios da utilização de BI



Fonte: Business Intelligence: Data Mining and Optimization for Decision Making, p.22.

3.2 DATA WAREHOUSE

Data Warehouse, ou armazém de dados, pode ser definido como um sistema de armazenamento de dados históricos, de forma consolidada. Geralmente é utilizado em grandes corporações e coletando dados de sistemas transacionais diversos. Sua vantagem é ter um desenho de base de dados que favoreça suas consultas, relatórios, a obtenção de informações e uma visão única dos dados corporativos. J. Zubcoff e J. Trujillo destacam: (Data & Knowledge Engineering, 2007, p. 63) “*A Data Warehouse stores data obtained from operational systems and is organized and managed to support reporting and analytical needs*”. Segundo R. Kimball (The Data Warehousing Toolkit, 1996, p. 32) “*The data warehouse must have the right data in it to support the decision making*”.

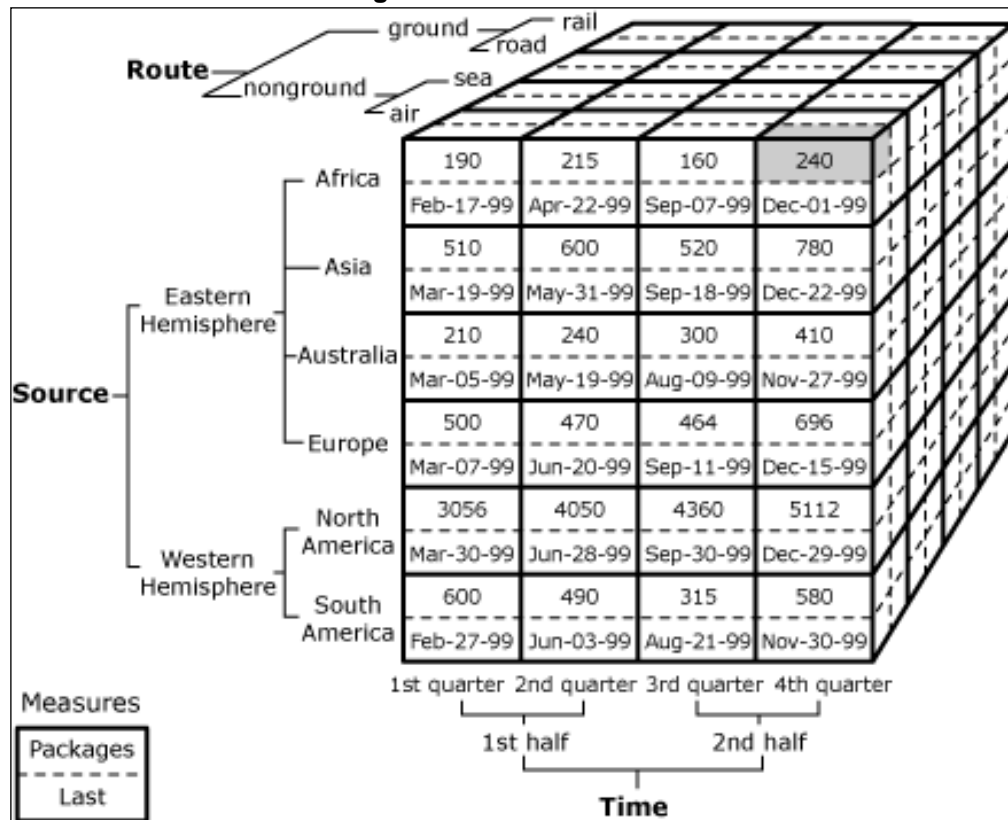
3.3 MODELO OLAP

Segundo Pang-Ning Tan (Introduction to Data Mining, 2007, p. 131): “*OLAP systems also have a strong focus on the interactive analysis of data and typically provide extensive capabilities for visualizing the data and generating statistics.*” Ou seja, este modelo de sistema é utilizado para prover visualizações de informações cruzadas sobre diferentes aspectos em uma base de dados.

A informação é conceitualmente organizada em cubos que armazenam valores quantitativos. Estes valores estão identificados por categorias denominadas dimensões que formam a estrutura de um cubo. Essas dimensões caracterizam-se por ser uma visão do negócio. Este modelo multidimensional simplifica o processo de formular relatórios ou análises comparativas.

Dentro de cada dimensão os dados podem ser organizados em diferentes níveis de detalhe. Por exemplo, dentro da dimensão tempo, pode-se ter uma hierarquia representando os níveis anos, meses, e dias. Nota-se que um usuário poderá realizar operações de *drill up* ou *drill down* entre os níveis para visualizar informação com maior ou menor nível de detalhe. Um exemplo da forma em que as informações são armazenadas em dimensões que constituem um cubo pode ser observado nesta figura:

Figura 6 - The OLAP Cube



Fonte: biresort.net

3.3.1 Navegabilidade do Cubo

A navegabilidade em um cubo OLAP pode ser considerada por Inmon (1997) e Singh (2001) nas seguintes perspectivas:

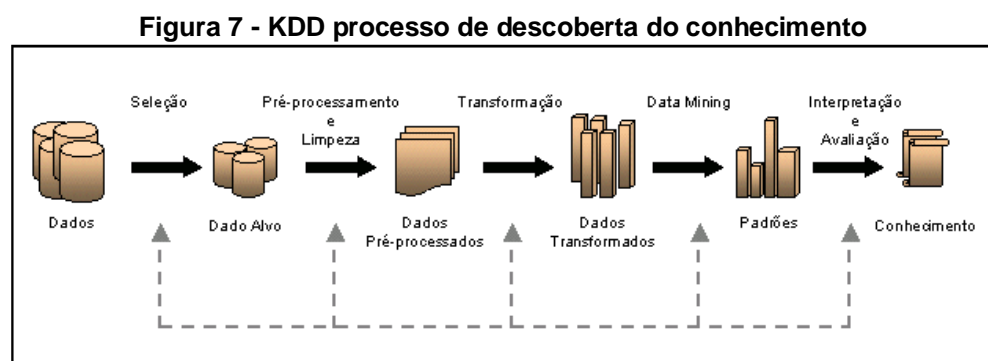
- **Drill-down:** Aumento do nível de detalhe da informação e consequente diminuição do nível de granularidade através da navegação pelos níveis hierárquicos definidos no conjunto de dados que está sendo visualizado;
- **Drill-up:** Diminuição no nível de detalhe da informação e consequente aumento do nível de granularidade através da navegação pelos níveis hierárquicos definidos no conjunto de dados que está sendo visualizado. É o processo inverso ao *drill-down*;
- **Drill-across:** Se refere às navegações realizadas nas dimensões, onde as consultas passam de um nível para o outro sem passar pelos níveis intermediários;
- **Drill-through:** É a funcionalidade utilizada quando se necessita analisar uma informação em um nível de detalhe menor do que o apresentado pelas

tabelas fato, ou seja, é a operação que busca a informação além do nível de granularidade existente na estrutura dimensional apresentada no cubo;

- **Pivoting:** É a alternância entre linhas e colunas na consulta apresentada, a fim de obter uma nova visão das informações;
- **Slice & dice:** Permite a rotação em qualquer sentido dos lados do cubo de dados que está sendo analisado, possibilitando a combinação de qualquer dimensão para a obtenção das informações necessárias.

3.4 KNOWLEDGE - DISCOVERY IN DATABASE

O KDD surgiu na década de 80, com o crescimento dos bancos de dados e a oportunidade de extrair conhecimento dos mesmos. Uma apropriada conceitualização para ele segundo FAYYAD, et al. (1996) é a seguinte: “O processo, não trivial, de extração de informações implícitas, previamente desconhecidas e potencialmente úteis, a partir dos dados armazenados em um banco de dados”. Fazendo-se uma análise desta citação, não trivial significa tornar clara a existência de alguma inferência. Previamente desconhecidos significa que haverá informação implícita deixando claro que esta informação deve trazer consigo algum benefício, o qual deverá possibilitar, ao não especialista no assunto observar relações de interesse. O processo de KDD passa por etapas de seleção, pré-processamento / limpeza, transformação, mineração de dados (*data mining*) e interpretação a nível tático e estratégico (BI), conforme ilustra a figura abaixo:



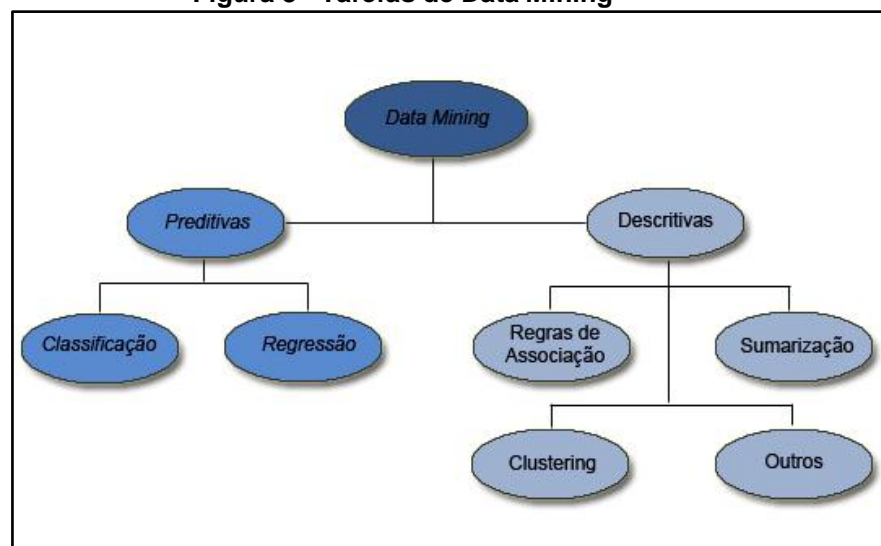
Fonte: PRASS, Fernando Sarturi (p.14, 2004).

3.4.1 Data Mining

Historicamente o processo de *Data Mining* surgiu junto ao KDD, que Pang-Ning Tan e outros destacam: (Introduction to Data Mining, 2007, p. 2-3) “*Data mining is the process of automatically discovering useful information in large data repositories.*” Ou seja, o processo de *data mining* esta inserido ao KDD e seu objetivo é utilizar-se de ferramentas e algoritmos para evidenciar padrões entre dados e informações automaticamente.

As etapas que vigoram o DM consistem em uma adequação dos dados que visa à normalização, preenchimento de informações faltantes ou redundâncias, a construção de repositórios *Data Marts* ou *Data Warehouse* com estas informações, estas, por sua vez, serão analisadas e refinadas por intermédio de ferramentas que possibilitarão uma dinâmica interativa até que os padrões apareçam. Como esta etapa é um processo iterativo, pode ser necessário que seja executada diversas vezes para ajustar o conjunto de parâmetros visando à obtenção de resultados mais adequados aos objetivos estabelecidos. A escolha da tarefa é feita de acordo com os objetivos pré-estabelecidos. As tarefas possíveis de um algoritmo de *Data Mining* podem ser agrupadas em atividades preditivas ou descritivas, consoante figura a seguir:

Figura 8 - Tarefas de Data Mining



Fonte: Adaptação de Rezende, 2003. P.318

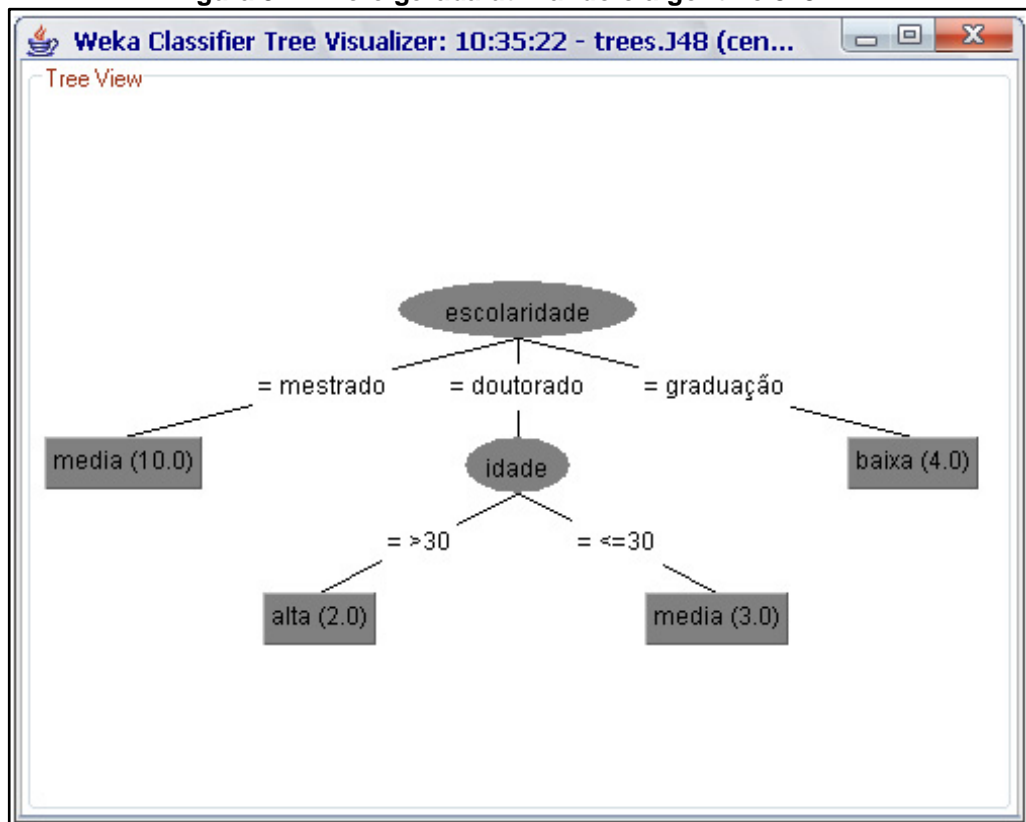
Atividades de predição consistem na generalização de exemplos ou experiências anteriores com respostas conhecidas usadas para classificar novos objetos. É definido um conjunto de padrões que possam ser usados para classificar

novos objetos. Esses padrões são construídos com base na análise prévia de um conjunto de dados de amostragem, que possuem objetos corretamente classificados.

A classificação pode ser entendida como a busca por uma integração que permita associar corretamente cada registro X , de um banco de dados, a um único rótulo categórico Y (objeto de saída), que corresponde à classe. Assim, uma vez sendo definida, essa função pode ser utilizada para prever a classe de novos registros, ou seja, o modelo é capaz de prever em qual classe se enquadram novos exemplos.

Um exemplo de classificação através do DM é apresentado na figura a seguir, onde um determinado banco de dados contém informações sobre escolaridade, idade e renda de pessoas. A questão de negócio é a classificação da renda em relação à escolaridade e idade.

Figura 9 - Árvore gerada utilizando o algoritmo J48



Fonte: danuzafaria.pbworks.com

A estrutura possui as seguintes características:

- Cada nó interno é um teste em um atributo preditivo;
- Uma ramificação partindo de um nó interno representa um resultado para o teste (por exemplo, Escolaridade = "Doutorado");

- Uma folha da árvore representa um rótulo de classe (por exemplo, Renda = "baixa", Renda = "media" ou Renda= "alta");
- Em cada nó da árvore, um atributo deve ser escolhido para dividir as observações do conjunto de treinamento em classes, na medida do possível.
- Uma nova observação é classificada seguindo um caminho na árvore, da raiz até a folha.

Não existe uma técnica que resolva todos os problemas de *data mining*. Diferentes técnicas servem para diferentes propósitos, cada uma oferecendo vantagens e desvantagens. A escolha da técnica está fortemente relacionada com o tipo de conhecimento que se deseja extrair ou com o tipo de dado no qual ela será aplicada.

3.4.2 BI & KDD aplicados em Telecom

Diferente dos exemplos clássicos do uso de BI & KDD em bases de dados de apoio a vendas e lucros, as redes de telecomunicações propõem um cenário rico do ponto de vista operacional, ou seja, através das práticas de KDD e DM aqui descritas podemos deduzir se determinado número de assinantes tem dificuldades para fazer chamadas em um cenário composto de área de cobertura, tarifa e hora do dia, atuando na falha antes mesmo da abertura da reclamação pelo cliente. Por sua vez, este fator está ligado diretamente aos conceitos de BI.

3.5 FERRAMENTAS PARA BI

Nas próximas seções, apresentamos as ferramentas utilizadas na implementação da solução, bem como suas principais funcionalidades.

3.5.1 Microsoft Excel

O Excel é um programa que nos permite criar planilhas inteligentes. Ele reconhece as informações digitadas e com elas realiza cálculos, verificações, comparações, gráficos e tabelas.

3.5.2 Microsoft SQL Server Management Studio

Considerado a base de uma implantação de Banco de Dados da Microsoft. Com o auxílio do SQL Server Integration Services (SSIS), podemos estabelecer a integração das fontes de dados providas da plataforma CDR View e fazer o agendamento de atualizações automáticas no banco de dados do SQL Server. Ele também cria e gerencia o modelo OLAP do data warehouse, sobre o qual construirá o restante da solução de BI.

3.5.3 Microsoft SQL Server Integration Services

Plataforma para criar integração de dados em nível corporativo e soluções de transformações de dados. Permite copiar ou baixar arquivos, atualizar *data warehouse*, fazendo a limpeza e mineração de dados e gerenciar objetos e dados do SQL Server.

3.5.4 Microsoft SQL Server Analysis Services

O SSAS é a ferramenta de OLAP que permite decompor, investigar, fazer uma busca detalhada e analisar os dados de negócios. Os cubos OLAP são construídos sobre o *data warehouse* do SQL Server e o SSAS nos permite realizar a navegabilidade destes cubos. A ferramenta nos permite, ainda, criar KPIs, que neste caso, serão os indicadores ANATEL.

3.5.5 Microsoft SQL Server Reporting Services

O SSRS oferece relatórios formatados com gráficos e medidores de desempenho. Os relatórios podem ser implantados como um serviço de tempo de execução no portal de usuários do SharePoint, ou podem ser agendados e automaticamente distribuídos aos usuários.

3.5.6 Microsoft Office SharePoint Server

Consiste em um portal *web* personalizável onde podemos compartilhar documentos e implantar painéis, *scorecards* e relatórios de modo que o acesso dos usuários seja simplificado e seguro. Por intermédio deste, o usuário final realizará consultas de mineração de dados, módulos de relatório e os valores dos indicadores ANATEL, acessando módulos diferentes sem ter que fazer *login* e *logout* em diversas aplicações.

4 UM AMBIENTE DE KDD & BI PARA REDES TELECOM

4.1 QUESTÕES DE NEGÓCIO

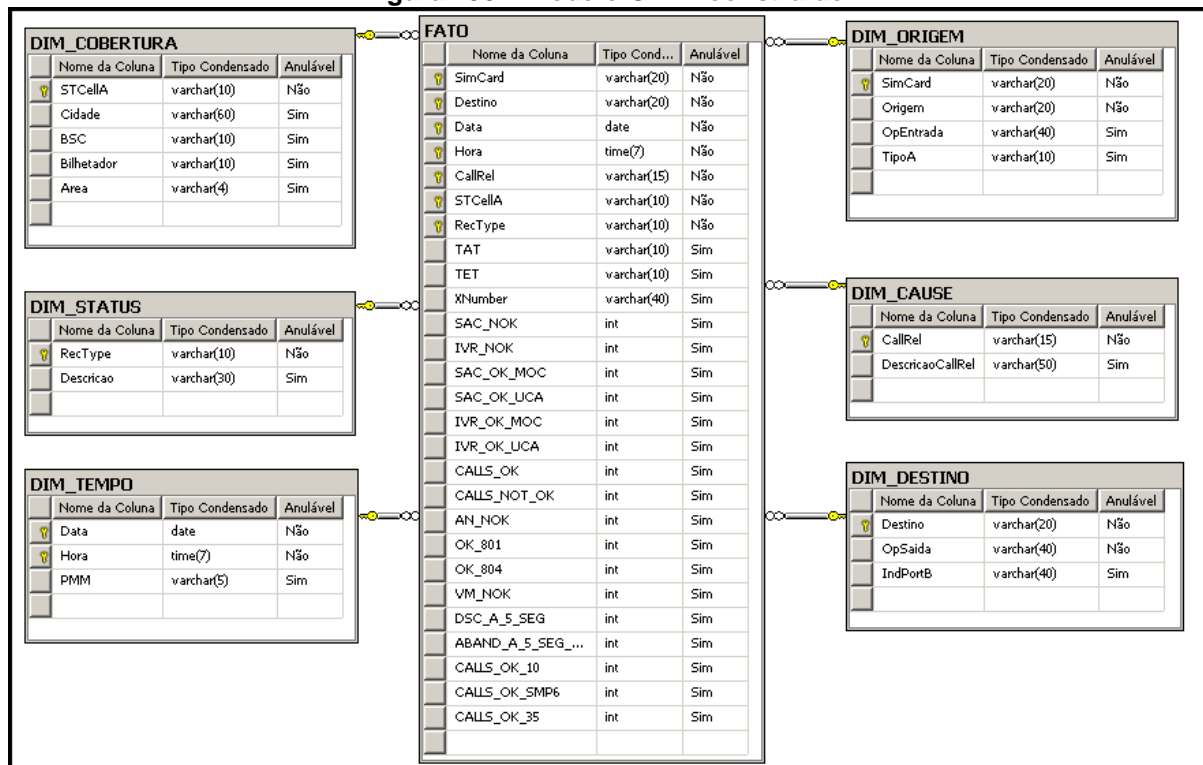
Com base nos cenários e problemas enfrentados pela equipe da Alcatel-Lucent, as questões de negócio abaixo são evidenciadas para o surgimento do modelo OLAP:

- **Valores do SMP3:** É necessário saber qual o valor de porcentagem deste indicador, incluindo a quantidade de chamadas em cada regra que constitui este indicador;
- **Valores do SMP5:** É necessário saber qual o valor de porcentagem deste indicador, incluindo a quantidade de chamadas em cada regra que constitui este indicador;
- **Valores do SMP6:** É necessário saber qual o valor de porcentagem deste indicador, incluindo a quantidade de chamadas em cada regra que constitui este indicador;
- **Chamadas, tarifas e causas:** Para controle dos índices de tarifação, é necessário saber em quais os cenários as chamadas estão sendo desconectadas e tarifando de forma errônea ou não;
- **Antenas sem completamento:** Para garantir a integridade da rede de cobertura e perceber problemas antes mesmo da abertura de reclamações pelos clientes, é necessário saber quais as antenas e cidades que possuem um alto índice de desconexão por falha.

4.2 CONSTRUÇÃO DO MODELO

Considerando-se as regras apresentadas nos indicadores e as dimensões construídas em torno do conhecimento técnico dos analistas da Alcatel-Lucent, o modelo OLAP de estrela foi o que melhor, expressa uma chamada real como FATO e propicia um ambiente claro para a resolução das questões de negócio, ao qual podemos visualizar na figura a seguir:

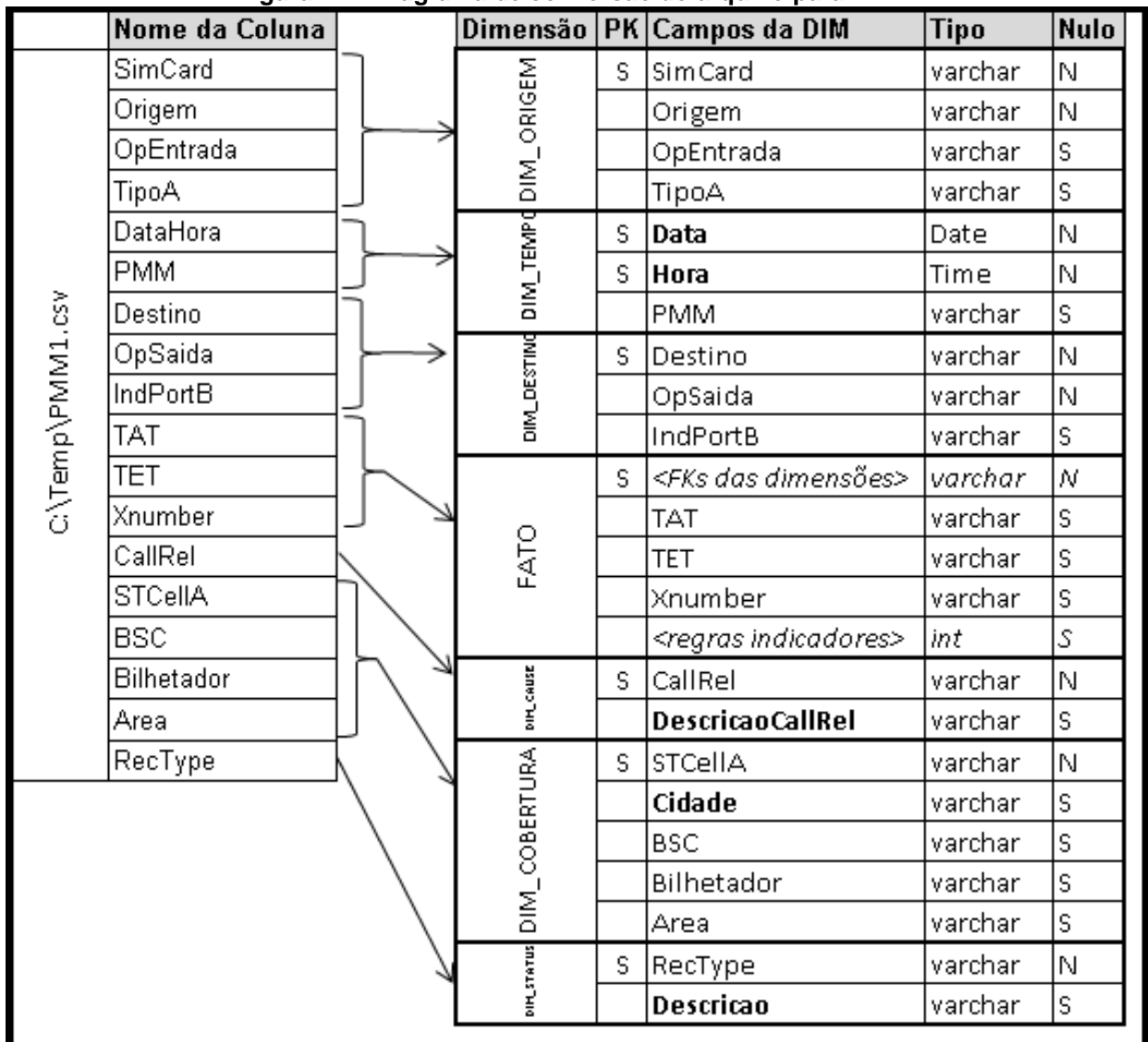
Figura 100 – Modelo OLAP construído



Fonte: Abraão Reckziegel e Adilson Favero.

Observa-se que as regras responsáveis pelos filtros foram atribuídas à tabela FATO, visando o cálculo dos indicadores. Também foram criados campos de descrição para RecType e CallRel, Cidade para STCellA e corrigido o formato de DataHora para Data e Hora separadamente, onde, no momento dos relatórios de análise técnica, aumentará a granularidade da pesquisa. Na figura a seguir, mostramos o mapeamento das colunas do arquivo CSV para o formato *data warehouse*:

Figura 11 – Diagrama de conversão do arquivo para BD



Fonte: Abraão Reckziegel e Adilson Favero.

4.3 DESCRIÇÃO DAS TABELAS

- **DIM_ORIGEM:** Esta tabela guarda o dimensionamento dos clientes que utilizam a rede da operadora, ela é composta pelos atributos SimCard, Origem, OpEntrada e TipoA. Esta é uma tabela de constante aumento, pois a cada dia novos clientes ingressam na operadora, consideramos, inclusive, clientes de outras operadoras em *roaming* na rede desta;
- **DIM_DESTINO:** Esta tabela guarda o dimensionamento dos destinos aos quais os clientes discam, ela é composta pelos atributos Destino, OpSaida e IndPortB. Poderá haver qualquer destino discado pelos clientes, tanto erro de discagem como números de outras operadoras Telecom fixa ou móveis;

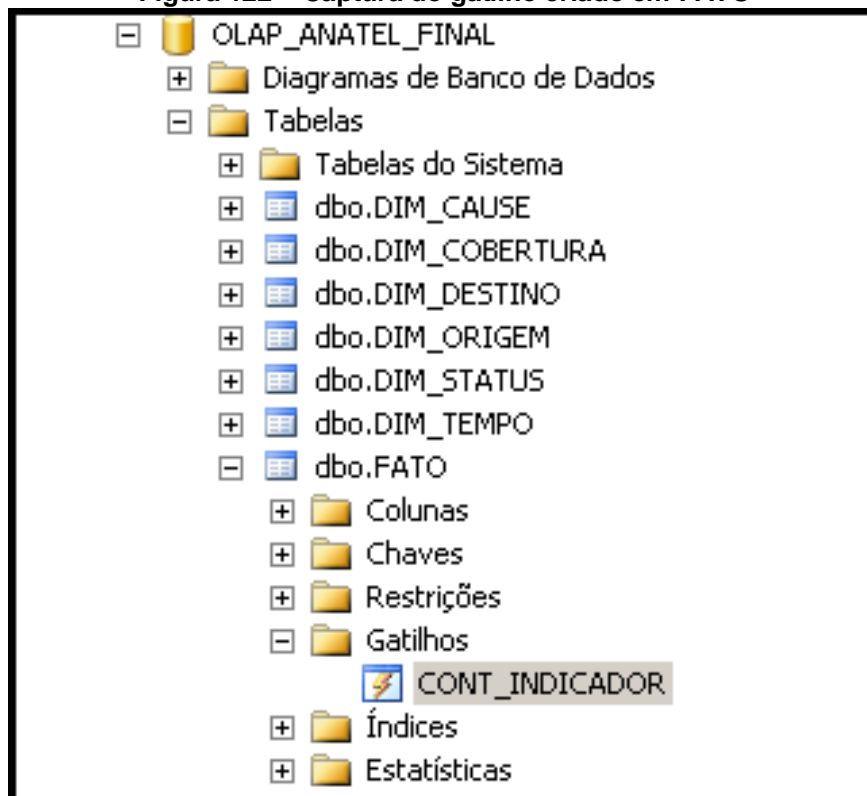
- **DIM_COBERTURA:** Esta tabela guarda o dimensionamento da cobertura de rede móvel da operadora, sendo composta pelos atributos STCellA, Cidade, BSC, Bilhetador e Área. Ela hierarquiza todo o plano de cobertura da operadora, desde uma específica antena que fornece cobertura a um bairro, passando pelas centrais de controle de uma cidade até o domínio de um código de área local. Esta dimensão também tende a aumentar conforme os projetos de expansão da operadora Telecom;
- **DIM_CAUSE:** Esta tabela guarda informações técnicas sobre o motivo da desconexão da chamada. Ela é composta pelos atributos CallRel e DescricaoCallRel. Uma chamada pode ter mais de 800 (oitocentos) tipos de desconexões diferentes, desde uma desconexão normal, quando o usuário finaliza a chamada em seu aparelho, até por falta de crédito, baixo nível de sinal de cobertura ou congestionamento para um destino. O atributo DescricaoCallRel foi adicionado, pois as informações de CallRel são geradas em código hexadecimal no CDR-View, dificultando a leitura durante o processo de análise;
- **DIM_STATUS:** Esta tabela guarda o dimensionamento do tipo de tarifação inserida no contexto da chamada. Ela é composta pelos atributos RecType e DescricaoRecType, este último foi adicionado, pois as informações de RecType são apresentadas em siglas, podendo haver 30 (trinta) casos diferentes de tarifação. Este campo também é base nos filtros de todos indicadores ANATEL;
- **DIM_TEMPO:** Esta tabela guarda o dimensionamento das chamadas, apontando o momento exato de início de cada uma delas. Ela é composta pelos atributos Data, Hora e PMM, pelo campo PMM se consegue facilmente extrair os valores dos indicadores dentro de cada período especificado pela ANATEL;
- **FATO:** Esta tabela guarda o registro de cada ligação executada na rede. Com a união das chaves das dimensões surgirá o registro de uma origem, recebendo cobertura em uma antena da rede, ligando em um determinado horário para um destino, que por sua vez houve atendimento e tarifação ou não. Além dos atributos SimCard, Destino, Data, Hora, CallRel, STCellA e RecType, provindos das outras tabelas, contamos com o atributo TAT, que especifica o tempo de duração da chamada, o atributo TET, que especifica o tempo levado até o tom de discagem e o atributo Xnumber, que mostrará a presença do serviço de Caixa Postal (*Voice Mail*) durante a chamada. Outro importante ponto de modelagem desta tabela foi a aquisição dos

campos responsáveis pelas regras dos indicadores: SAC_OK_MOC, SAC_OK_UCA, IVR_OK_MOC, IVR_OK_UCA, SAC_NOK, IVR_NOK, CALLS_OK, AN_NOK, OK_801, OK_804, VM_NOK, CALLS_NOT_OK, CALLS_OK_10, CALLS_OK_SMP6. Estes terão a função de registrar se a chamada em questão (tupla da FATO) contabiliza ou não em uma das regras, conforme será descrito na próxima seção.

4.4 APLICAÇÃO DAS REGRAS

Para que o modelo reconheça os filtros das regras, leia cada registro de chamada e incremente a regra adequada, tudo de forma automática, foi criado um gatilho (*trigger*) na tabela FATO, que realiza a leitura de cada nova inserção na tabela, aplicando os filtros necessários e, caso atenda, contabiliza cada regra especificada pela ANATEL:

Figura 122 – Captura do gatilho criado em FATO



Fonte: Abraão Reckziegel e Adilson Favero.

Inicialmente o gatilho conta com a notação “INSTEAD OF INSERT” que nos permite manipular os dados antes da inserção e realizar as comparações que, caso verdadeiras, incrementam as respectivas regras. Um resumo comentado do que foi desenvolvido será apresentado a seguir:


```
CREATE TRIGGER [dbo].[CONT_INDICADOR]
  ON [dbo].[FATO]
  INSTEAD OF INSERT
AS
BEGIN
  DECLARE
    [...] --Declaração das variáveis do gatilho

  SELECT
    [...] --Seleção das colunas inseridas em FATO
  FROM inserted I
```

```

IF( -- exemplo de filtro para SAC_NOK

    (@SimCard LIKE ('72431%') OR @SimCard LIKE ('72416%'))
    AND @RecType IN ('UCA')
    AND @Destino IN ('B1414', '1053', 'B144', 'B800',...
    AND @CallRel NOT IN ('00000C01','00000C02', ...
    AND (@XNumber NOT IN ('5503140','5501646100100', ...
    AND (@STCellA NOT IN ('','--') OR ...

)SET @SAC_NOK = 1

IF(
    [...] --Comparações da regra IVR_NOK
)SET @IVR_NOK = 1

IF(
    [...] --Comparações da regra SAC_OK_MOC
)SET @SAC_OK_MOC = 1

IF(
    [...] --Comparações da regra SAC_OK_UCA
)SET @SAC_OK_UCA = 1

[...] --Comparações das demais regras

--Após cada regra receber o valor 0 ou 1, o insert final dos
dados é realizado:

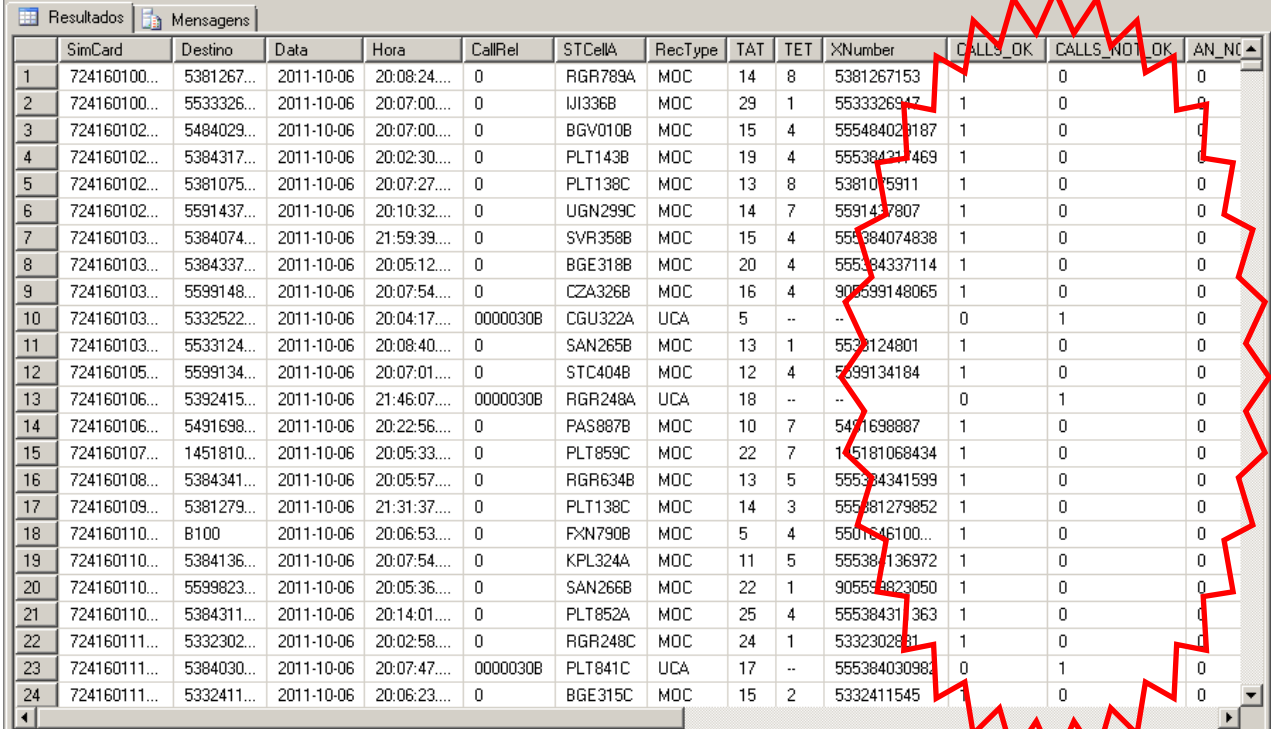
    INSERT INTO FATO VALUES (... ,
                                @SAC_NOK,
                                @IVR_NOK,
                                @SAC_OK_MOC,
                                @SAC_OK_UCA,
                                @IVR_OK_MOC,
                                @IVR_OK_UCA,
                                @CALLS_OK,
                                @CALLS_NOT_OK,
                                @AN_NOK,
                                @OK_801,
                                @OK_804,
                                @VM_NOK,
                                @DSC_A_5_SEG,
                                @ABAND_A_5_SEG_IVR,
                                @CALLS_OK_10,
                                @CALLS_OK_SMP6)

END

```

Após o processo de população da tabela FATO uma imagem de como se apresentam as contabilizações das regras após a ação do gatilho pode ser observada assim:

Figura 13 – Captura da tabela FATO após inserts



	SimCard	Destino	Data	Hora	CallRel	STCellA	RecType	TAT	TET	XNumber	CALLS_OK	CALLS_NOT_OK	AN_NC
1	724160100...	5381267...	2011-10-06	20:08:24...	0	RGR789A	MOC	14	8	5381267153	0	0	0
2	724160100...	5533326...	2011-10-06	20:07:00...	0	IJI336B	MOC	29	1	5533326947	1	0	0
3	724160102...	5484029...	2011-10-06	20:07:00...	0	BGV010B	MOC	15	4	555484029187	1	0	0
4	724160102...	5384317...	2011-10-06	20:02:30...	0	PLT143B	MOC	19	4	555384317469	1	0	0
5	724160102...	5381075...	2011-10-06	20:07:27...	0	PLT138C	MOC	13	8	5381075911	1	0	0
6	724160102...	5591437...	2011-10-06	20:10:32...	0	UGN299C	MOC	14	7	5591437807	1	0	0
7	724160103...	5384074...	2011-10-06	21:59:39...	0	SVR358B	MOC	15	4	555384074838	1	0	0
8	724160103...	5384337...	2011-10-06	20:05:12...	0	BGE318B	MOC	20	4	55534337114	1	0	0
9	724160103...	5599148...	2011-10-06	20:07:54...	0	CZA326B	MOC	16	4	905599148065	1	0	0
10	724160103...	5332522...	2011-10-06	20:04:17...	0000030B	CGU322A	UCA	5	--	--	0	1	0
11	724160103...	5533124...	2011-10-06	20:08:40...	0	SAN265B	MOC	13	1	5533124801	1	0	0
12	724160105...	5599134...	2011-10-06	20:07:01...	0	STC404B	MOC	12	4	5599134184	1	0	0
13	724160106...	5392415...	2011-10-06	21:46:07...	0000030B	RGR248A	UCA	18	--	--	0	1	0
14	724160106...	5491698...	2011-10-06	20:22:56...	0	PAS887B	MOC	10	7	5491698887	1	0	0
15	724160107...	1451810...	2011-10-06	20:05:33...	0	PLT859C	MOC	22	7	145181068434	1	0	0
16	724160108...	5384341...	2011-10-06	20:05:57...	0	RGR634B	MOC	13	5	55534341599	1	0	0
17	724160109...	5381279...	2011-10-06	21:31:37...	0	PLT138C	MOC	14	3	555381279852	1	0	0
18	724160110...	B100	2011-10-06	20:06:53...	0	FXN790B	MOC	5	4	5507646100...	1	0	0
19	724160110...	5384136...	2011-10-06	20:07:54...	0	KPL324A	MOC	11	5	555384136972	1	0	0
20	724160110...	5599823...	2011-10-06	20:05:36...	0	SAN266B	MOC	22	1	90559823050	1	0	0
21	724160110...	5384311...	2011-10-06	20:14:01...	0	PLT852A	MOC	25	4	555384311363	1	0	0
22	724160111...	5332302...	2011-10-06	20:02:58...	0	RGR248C	MOC	24	1	5332302881	1	0	0
23	724160111...	5384030...	2011-10-06	20:07:47...	0000030B	PLT841C	UCA	17	--	555384030988	0	1	0
24	724160111...	5332411...	2011-10-06	20:06:23...	0	BGE315C	MOC	15	2	5332411545	0	0	0

Fonte: Abraão Reckziegel e Adilson Favero.

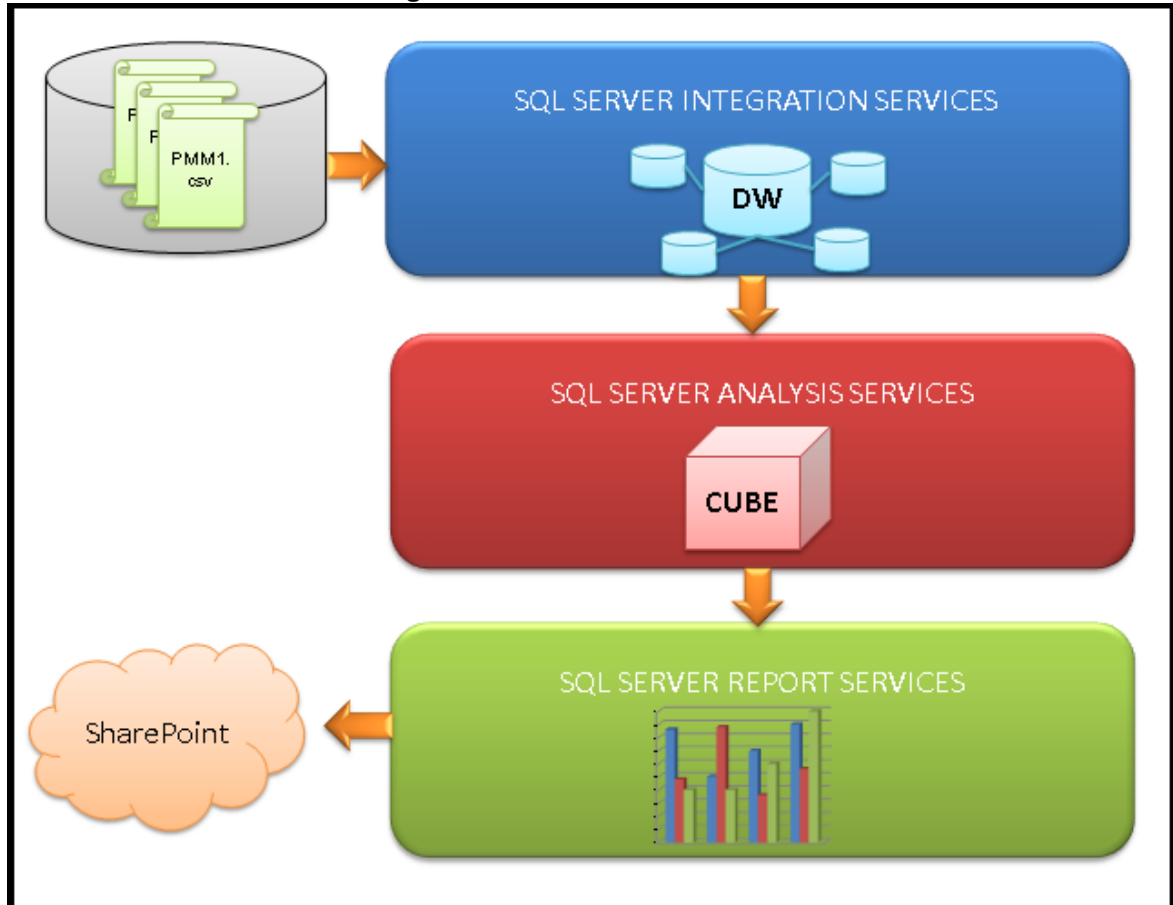
4.5 O PROCESSO DE ETL

Com a necessidade de automatização do processo de inserção dos dados no *data warehouse*, implementamos uma cadeia de processos na ferramenta SQL Server Integration Services, esta cadeia permitirá a coleta do arquivo, limpeza dos dados repetidos, inserção / validação no banco, processamento do cubo e exclusão dos arquivos inúteis.

Conforme descrito anteriormente, a ferramenta CDR-View exporta diariamente (as 00h01min), três arquivos no formato CSV para o caminho “C:\Temp” de um servidor remoto, estes arquivos contém dados de chamadas registradas nos três PMMs ANATEL do dia anterior, tendo eles os nomes de “PMM1.csv”, “PMM2.csv” e “PMM3.csv”. Após a realização dos processos do SSIS, iremos projetar os relatórios de análise via SSAS e SSRS. Ao final, os relatórios são disponibilizados no MS

SharePoint a todos os envolvidos. Referido processo, baseado nos conceitos de ETL, pode ser expresso de forma macro na imagem a seguir:

Figura 14 – Camadas SQL Server



Fonte: Abraão Reckziegel e Adilson Favero.

4.5.1 População via SSIS

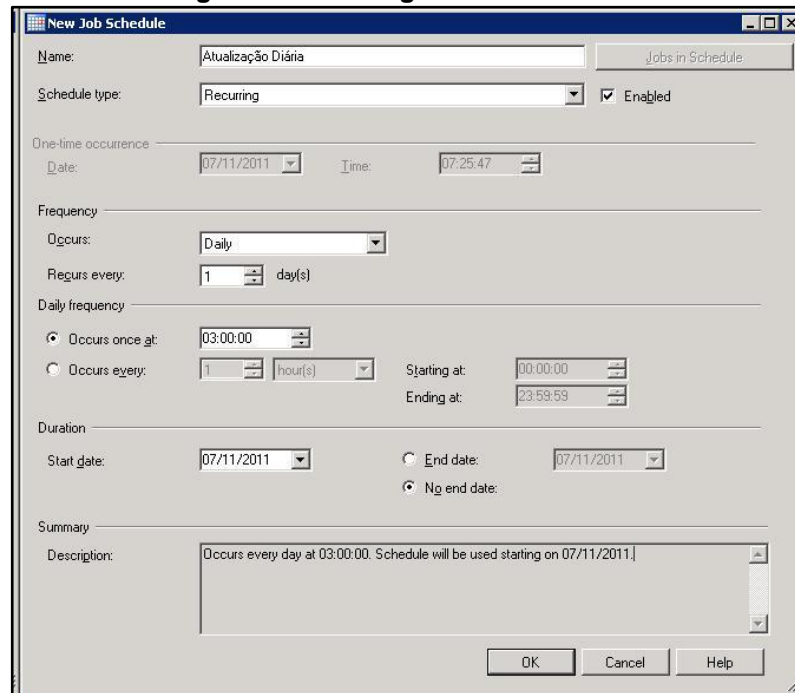
No servidor remoto, onde se encontram os arquivos, implementamos uma *Job* a ser executada todos os dias às 03h, que tem a função de realizar o *start* do pacote SSIS. Vejamos a imagem a seguir:

Figura 15 – Propriedades da Job SSIS



Fonte: Abraão Reckziegel e Adilson Favero.

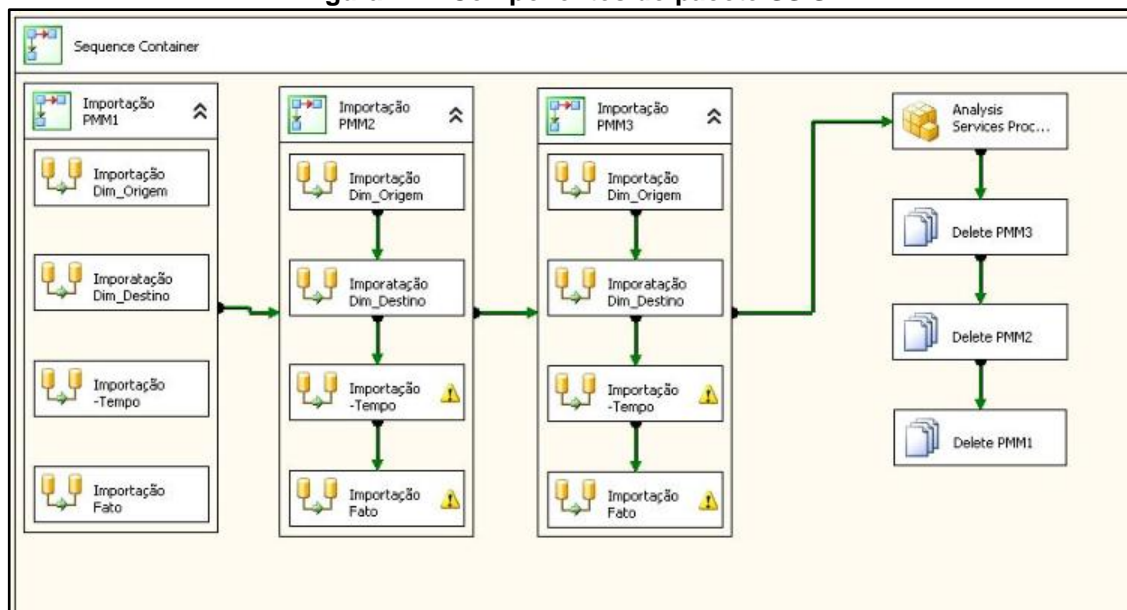
Figura 16 – Cronograma da Job SSIS



Fonte: Abraão Reckziegel e Adilson Favero.

O pacote executado na *Job* importa linearmente os arquivos “PMM1.csv”, “PMM2.csv” e “PMM3.csv”, nesta ordem, realizando os *inserts* nas dimensões, processando o cubo e excluindo os arquivos não mais necessários. A figura a seguir apresenta a tela de desenvolvimento do SSIS onde o pacote é visto de forma geral:

Figura 17 – Componentes do pacote SSIS



Fonte: Abraão Reckziegel e Adilson Favero.

Desmembrando o processo acima exposto, seguem as transformações aplicadas na importação, correção e inserção dos dados no banco, bem como mapas de ETL desenvolvidos nas cadeias de processos:

4.5.2 ETL de DIM_ORIGEM e DIM_DESTINO

Na tabela dimensional DIM_ORIGEM, é selecionado o atributo SimCard para servir como chave primária já que cada cliente da operadora possui o seu. Em DIM_DESTINO o próprio número discado corresponde a uma chave primária, pois cada cliente de outra operadora é reconhecido assim.

Tabela 8 – Mapeamento para carga de DIM_ORIGEM

DE: PMM[...].csv		PARA: DIM_ORIGEM		
Coluna	Tipo	Atributo	Tipo	PK
SimCard	Texto	SimCard	varchar(20)	X
Origem	Texto	Origem	varchar(20)	
OpEntrada	Texto	OpEntrada	varchar(40)	
TipoA	Texto	TipoA	varchar(20)	

Fonte: Abraão Reckziegel e Adilson Favero.

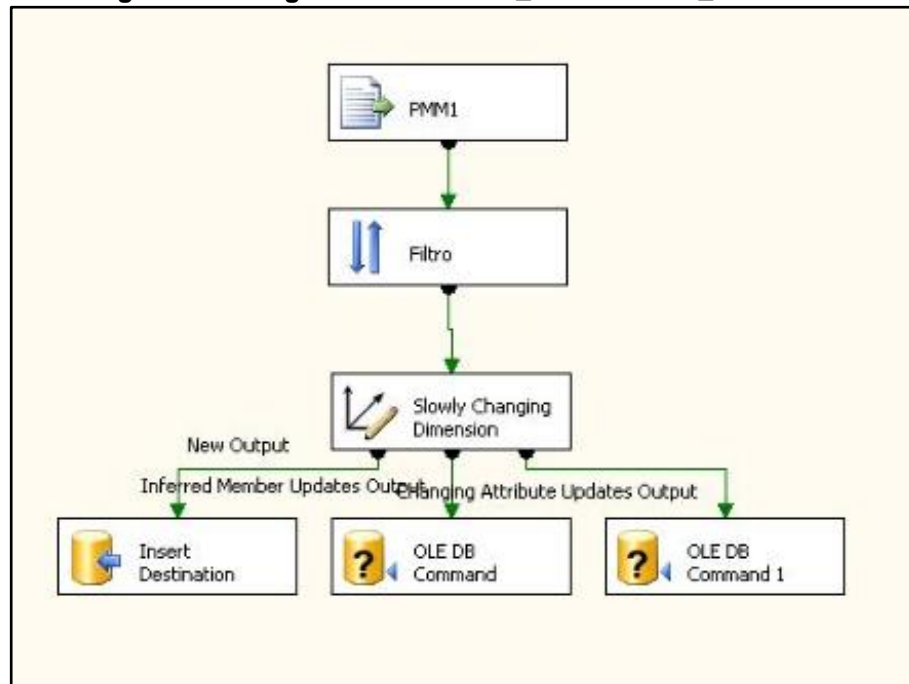
Tabela 9 – Mapeamento para carga de DIM_DESTINO

DE: PMM[...].csv		PARA: DIM_DESTINO		
Coluna	Tipo	Atributo	Tipo	PK
Destino	Texto	Destino	varchar(40)	X
OpSaida	Texto	OpSaida	varchar(40)	
IndPortB	Texto	IndPortB	varchar(40)	

Fonte: Abraão Reckziegel e Adilson Favero.

Criamos um mapa de ETL que faz a leitura dos dados do arquivo PMM1.csv, PMM2.csv e PMM3.csv, realiza as conversões necessárias descritas nas tabelas acima, complementa os *merges* para que não haja chave primária repetida e então carrega as informações no *data warehouse*, conforme diagrama na figura abaixo:

Figura 18 – Diagrama ETL de DIM_OIGEM e DIM_DESTINO



Fonte: Abraão Reckziegel e Adilson Favero.

4.5.3 ETL de DIM_CAUSE, DIM_STATUS e DIM_COBERTURA

Na tabela dimensional DIM_CAUSE, selecionamos o atributo CallRel para servir como chave primária já que cada desconexão possui seu código hexadecimal de identificação, com a ajuda de uma planilha fornecida pelos analistas da Alcatel-Lucent, que descreve cada código hexadecimal, realizamos de uma única vez a inserção de dados nesta dimensão, contemplando assim todos os possíveis cenários de desconexão de chamadas. As conversões de tipos e chaves podem ser verificadas da seguinte maneira:

Tabela 10 – Mapeamento para carga de DIM_CAUSE

DE: PMM[...].csv		PARA: DIM_CAUSE		
Coluna	Tipo	Atributo	Tipo	PK
CallRel	texto	CallRel	varchar(15)	X
		*DescricaoCallRel	varchar(15)	

Fonte: Abraão Reckziegel e Adilson Favero.

Para a tabela DIM_STATUS, realizamos o mesmo que em DIM_CAUSE, ou seja, selecionamos o atributo RecType para servir como chave primária já que cada tarifa aplicada possui sua sigla. Com informações complementares fornecidas pela

equipe da Alcatel-Lucent, criamos o campo Descrição, que contempla informações adicionais sobre RecType. Para melhor entendimento anexamos a tabela abaixo:

Tabela 11 – Mapeamento para carga de DIM_STATUS

DE: PMM[...].csv		PARA: DIM_STATUS		
Coluna	Tipo	Atributo	Tipo	PK
RecType	texto	RecType	varchar(20)	X
		*Descricao	varchar(30)	

Fonte: Abraão Reckziegel e Adilson Favero.

Para a tabela DIM_COBERTURA, realizamos o mesmo que em DIM_CAUSE E DIM_STATUS, ou seja, selecionamos o atributo STCellA para servir como chave primária já que cada antena no Brasil possui sua sigla única. Com informações complementares fornecidas pela equipe da Alcatel-Lucent, criamos o campo Cidade, que contempla informações adicionais sobre a cidade em que a antena esta instalada. Para melhor entendimento anexamos a tabela abaixo:

Tabela 12 – Mapeamento para carga de DIM_COBERTURA

DE: PMM[...].csv		PARA: DIM_COBERTURA		
Coluna	Tipo	Atributo	Tipo	PK
STCellA	texto	STCellA	varchar(10)	X
		*Cidade	varchar(60)	
BSC	texto	BSC	varchar(10)	
Bilhetador	texto	Bilhetador	varchar(10)	
Area	texto	Area	varchar(4)	

Fonte: Abraão Reckziegel e Adilson Favero.

4.5.4 ETL de DIM_TEMPO e FATO

Para a tabela DIM_TEMPO realizamos uma conversão do formato de DataHora (dia/mês/ano hora:minuto:segundo) em Data (dia/mês/ano) e Hora (hora:minuto:segundo), ou seja, ocorre a divisão da coluna DataHora do arquivo em dois campos distintos para que os relatórios possam contemplar um maior nível de detalhamento. A união de Data e Hora para chaves primárias é essencial no contexto da FATO pois uma origem discando para o mesmo destino, pode apresentar a mesma cobertura, desconexão e tarifa, mas não pode ser feita no mesmo dia, hora, minuto e segundo. A tabela abaixo exemplifica as conversões realizadas e tipos dos atributos utilizados:

Tabela 13 – Mapeamento para carga de DIM_TEMPO

DE: PMM[...].csv		PARA: DIM_TEMPO		
Coluna	Tipo	Atributo	Tipo	PK
DataHora	data	*Data	Date	X
		*Hora	Time(7)	X
PMM	texto	PMM	varchar(4)	

Fonte: Abraão Reckziegel e Adilson Favero.

Na tabela FATO assumimos os atributos chave das dimensões descritas nas tabelas anteriores. Também inserimos as colunas TAT, TET e Xnumber provindas do arquivo bruto (CSV). Campos que contemplam as regras ANATEL foram adicionados nesta tabela que, com o auxílio do gatilho, serão atualizados após a execução deste processo. A tabela abaixo exemplifica o mapeamento adotado:

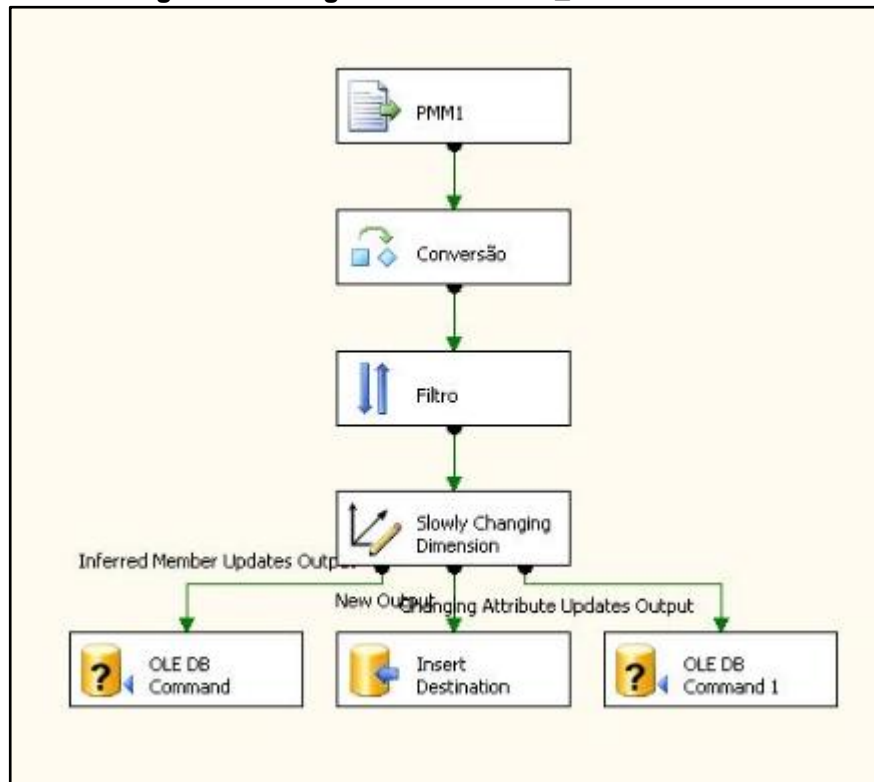
Tabela 14 – Mapeamento para carga da tabela FATO

DE: PMM[...].csv		PARA: FATO		
Coluna	Tipo	Atributo	Tipo	PK / FK
SimCard	texto	SimCard	varchar(20)	X
Destino	texto	Destino	varchar(40)	X
DataHora	texto	*Data	Date	X
		*Hora	Time(7)	X
CallRel	texto	CallRel	varchar(15)	X
STCellA	texto	STCellA	varchar(10)	X
RecType	texto	RecType	varchar(20)	X
TAT	texto	TAT	varchar(10)	
TET	texto	TET	varchar(10)	
Xnumber	texto	Xnumber	varchar(40)	
		*SAC_NOK	int	
		*IVR_NOK	int	
		*SAC_OK_MOC	int	
		*SAC_OK_UCA	int	
		*[Demais regras ...]	int	

Fonte: Abraão Reckziegel e Adilson Favero.

Para as tabelas DIM_TEMPO e FATO também criamos um mapa de ETL que faz a leitura dos arquivos brutos (CSV), realiza as conversões necessárias descritas nas tabelas anteriores, realiza os merges para que não haja chave primária repetida e então carrega as informações no *data warehouse*, conforme diagrama desta figura:

Figura 19 – Diagrama ETL de DIM_TEMPO e FATO

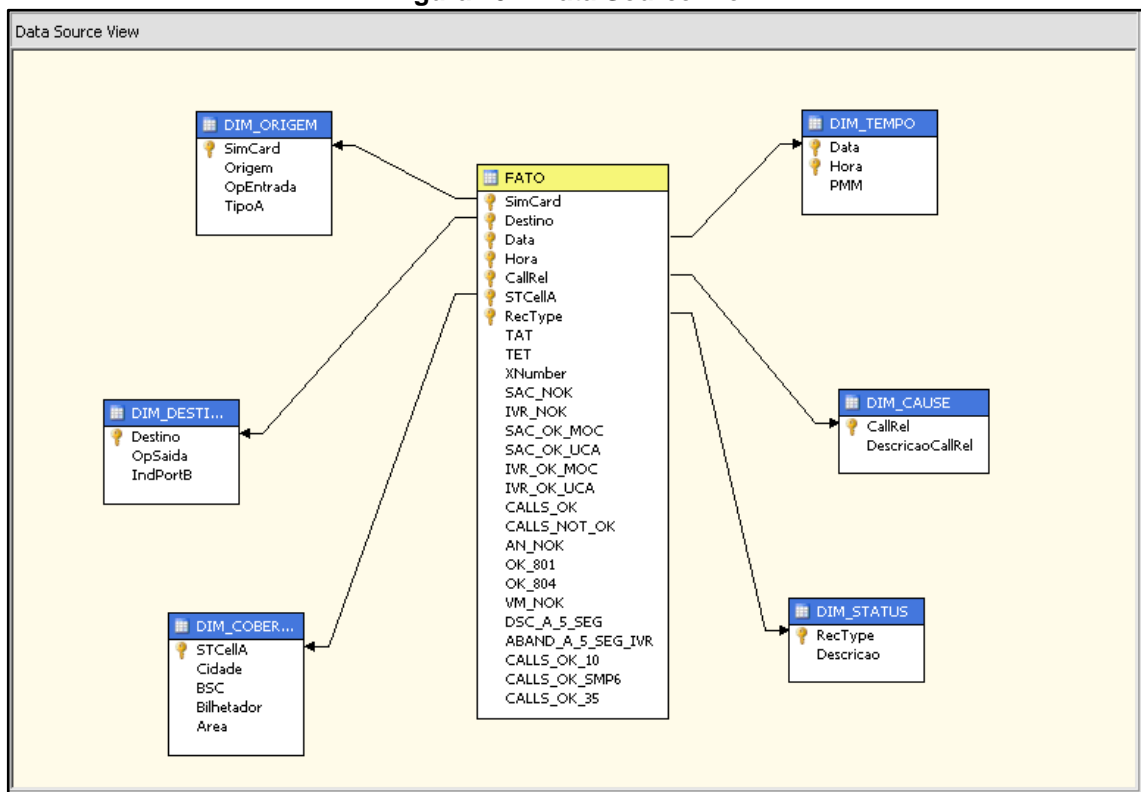


Fonte: Abraão Reckziegel e Adilson Favero.

4.6 A GERAÇÃO DO CUBO

Com o modelo OLAP implementado e o *data warehouse* abastecido de forma periódica (SSIS), partimos para o processamento do cubo na ferramenta SQL Server Analysis Services (SSAS). Ao utilizarmos nosso *data warehouse* como nova fonte de dados no SSAS, obtemos o seguinte diagrama relacional, que servirá como base para o processamento do cubo almejado:

Figura 20 – Data Source View



Fonte: Abraão Reckziegel e Adilson Favero.

Para a aplicação das hierarquias que visam aumentar a eficiência na navegabilidade do cubo criando relações entre os atributos das dimensões, seguimos as normas citadas a seguir:

- **Hierarquia de DIM_CAUSE:** Selecionamos o atributo chave CallRel como principal item, em seguida sua descrição, conforme figura abaixo:

Figura 21 – Hierarquia de DIM_CAUSE



Fonte: Abraão Reckziegel e Adilson Favero.

- **Hierarquia de DIM_COBERTURA:** Selecionamos o atributo Bilhetador, pois corresponde a central MSC, que por sua vez pode conter mais de um código de área em seu domínio, onde sucessivamente recebe dezenas de BSCs, responsáveis por centenas de Cidades que estão cobertas por milhares de antenas (STCellA). Nesta linha, a figura abaixo representa o mapa hierárquico da DIM_COBERTURA:

Figura 22 – Hierarquia de DIM_COBERTURA



Fonte: Abraão Reckziegel e Adilson Favero.

- **Hierarquia de DIM_DESTINO:** Para esta dimensão, verificamos primeiramente se o destino é portado (IndPortB), sendo ou não, partimos para dezenas de operadoras de destino (OpSaida), podendo uma única operadora ser proprietária de centenas de clientes (Destino). Nesta linha, a figura abaixo representa a hierarquia concebida:

Figura 23 – Hierarquia de DIM_DESTINO



Fonte: Abraão Reckziegel e Adilson Favero.

- **Hierarquia de DIM_ORIGEM:** Nesta dimensão, inserimos como topo do nível hierárquico o atributo TipoA (“Pré-pago” ou “Pós-pago”), que abre o tipo de plano correspondente ao cliente na rede. Nos próximos níveis inserimos o atributo OpEntrada, que varia entre clientes da própria operadora e outras operadoras em *roaming* na rede desta, seguido dos milhares de números de origem e seus SimCards. Nesta linha de raciocínio, a figura abaixo representa a hierarquia concebida:

Figura 24 – Hierarquia de DIM_ORIGEM



Fonte: Abraão Reckziegel e Adilson Favero.

- **Hierarquia de DIM_STATUS:** Selecionamos o atributo chave RecType como principal item, em seguida sua descrição, conforme a figura abaixo:

Figura 25 – Hierarquia de DIM_STATUS



Fonte: Abraão Reckziegel e Adilson Favero.

- **Hierarquia de DIM_TEMPO:** Inserimos o atributo PMM no topo da hierarquia, pois se apresentam apenas três (PMM1, PMM2 e PMM3), seguido da data e as horas que compõem o dia, conforme a figura abaixo:

Figura 26 – Hierarquia de DIM_TEMPO



Fonte: Abraão Reckziegel e Adilson Favero.

Com a fonte de dados, dimensões e hierarquias definidas, é realizado o processamento do cubo na ferramenta SSAS. As figuras abaixo mostram relatórios *Ad-hoc* extraídos durante o processo de navegação:

Figura 27 – Printscreen de navegação 1

Dimension	Hierarchy	Operator	Filter Expression
Drop Filter Fields Here			
		Descricao Call Rel	Call Rel
		00000B14	Grand Total
		Total	Total
Cidade	ST Cell A	FATO Count	FATO Count
ALVORADA	AVA000B	1	1
	AVA001A	1	1
	AVA003C	2	2
	AVA877C	1	1
	Total	5	5
CACHOEIRINHA	CHN014C	1	1
	CHN833B	1	1
	Total	2	2
Campos Borges	CBG978A	1	1
	Total	1	1
CANOAS	CAN026C	1	1
	CAN030C	1	1
	CAN071C	1	1
	Total	3	3
CERRO GRANDE DO SUL	CGZ902B	1	1
	Total	1	1
Condor	CWR981B	1	1
	Total	1	1
ERECHIM	ERE330C	1	1
	Total	1	1
Glorinha	GLI904C	1	1
	ROD483B	1	1
	Total	2	2
GRAVATAÍ	GTI094C	1	1
	GTI098B	1	1
	GTI931B	1	1
	Total	3	3
PORTO ALEGRE	PAE043B	1	1
	PAE063C	1	1
	PAE088A	1	1

Fonte: Abraão Reckziegel e Adilson Favero.

Figura 28 – Printscreen de navegação 2

Dimension	Hierarchy	Operator	Filter Expression
Drop Filter Fields Here			
		PMM	
		PMM1	PMM2
		PMM3	Grand Total
		CALLS OK	CALLS NOT OK
		CALLS OK	CALLS NOT OK
		CALLS OK	CALLS NOT OK
		CALLS OK	CALLS NOT OK
Cidade		CALLS OK	CALLS NOT OK
Água Santa		51	96
AGUDO		72	38
AJURICABA		75	33
ALECRIM		48	23
ALEGRETE		2807	1391
Almirante Tamandaré do Sul		134	70
Alpestre		19	5
ALVORADA		13441	6591
Ametista do Sul		17	6
ANDRÉ DA ROCHA		188	77
ANTA GORDA		85	18
ANTONIO PRADO		108	197
ARAMBARÉ		141	63
Aranicá		121	73
ARROIO DO MEIO		327	129
ARROIO DO SAL		239	109
ARROIO DO TIGRE		49	26
ARROIO DOS RATOS		315	164
ARROIO GRANDE		4616	1983
ARROIO TEIXEIRA		42	31
ARVOREZINHA		80	42
ATLÂNTIDA		940	477
ATLÂNTIDA SUL		103	46
AUGUSTO PESTANA		64	22
Aurea		40	101
BAGÉ		4812	2630
BALNEÁRIO PINHAL		794	376
Barão		31	8
Barão de Cotegipe		48	62
Barra do Quarai		34	17
Barra do Ribeiro		257	124

Fonte: Abraão Reckziegel e Adilson Favero.

4.7 A IMPLEMENTAÇÃO DOS KPIS

Próximo ao objetivo central deste projeto, após a conclusão do processamento do cubo, ainda no SSAS, criamos grupos de medidas relacionados

diretamente aos indicadores. Nestes grupos de medidas aplicamos as equações de percentuais descritas pela ANATEL, unindo as regras inseridas na tabela FATO. Cada grupo criado será apresentado a seguir:

- **Grupo de medidas SMP3:** Definimos como um percentual associado à tabela FATO o grupo de medidas “[SMP3]”. A expressão de cálculo deste grupo, desenvolvida em linha de comando no SSAS, pode ser observada abaixo:

```
( [Measures] . [SAC_OK_MOC] + [Measures] . [SAC_OK_UCA] +
[Measures] . [IVR_OK_MOC] + [Measures] . [IVR_OK_UCA] ) /
( [Measures] . [SAC_OK_MOC] + [Measures] . [SAC_OK_UCA] +
[Measures] . [SAC_NOK] + [Measures] . [IVR_OK_MOC] +
[Measures] . [IVR_OK_UCA] + [Measures] . [IVR_NOK] )
```

- **Grupo de medidas SMP5:** Definimos como um percentual associado à tabela FATO o grupo de medidas “[SMP5]”. Segue abaixo a expressão desenvolvida em linha de comando no SSAS:

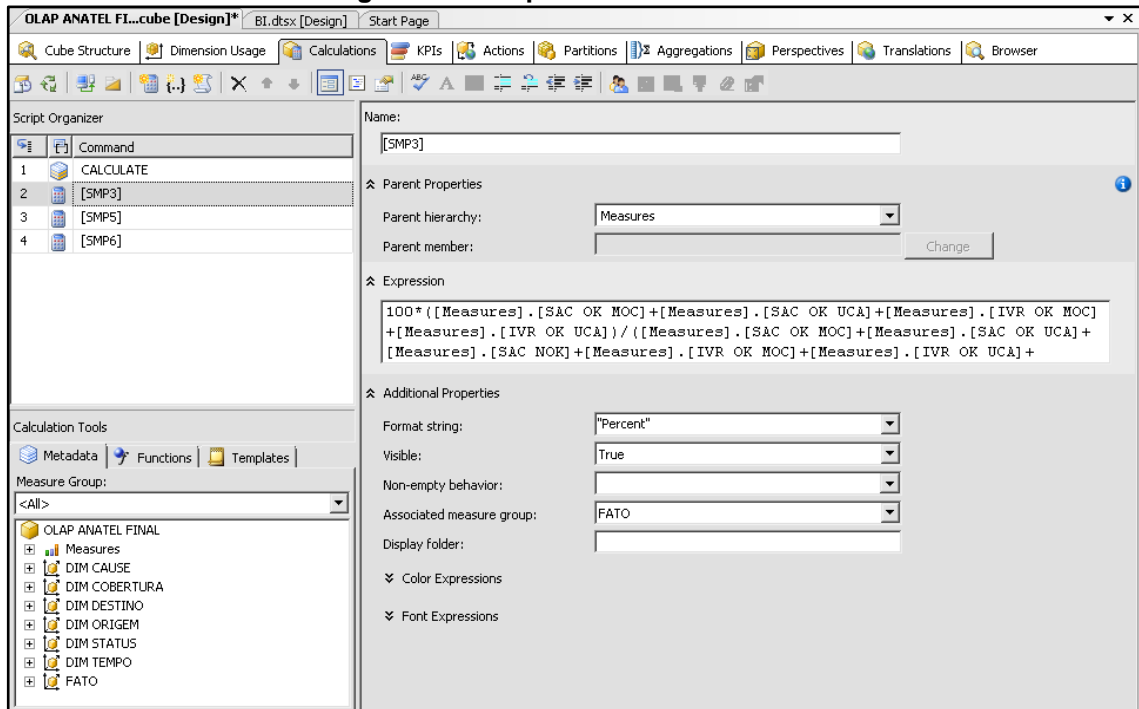
```
( [Measures] . [CALLS_OK] + [Measures] . [AN_NOK] +
[Measures] . [OK_801] + [Measures] . [OK_804] +
[Measures] . [VM_NOK] ) /
( [Measures] . [CALLS_OK] + [Measures] . [CALLS_NOT_OK] +
[Measures] . [AN_NOK] + [Measures] . [VM_NOK] )
```

- **Grupo de medidas SMP6:** Assim como nos anteriores, definimos como um percentual associado o “[SMP6]”. Segue abaixo a expressão desenvolvida em linha de comando no SSAS:

```
( [Measures] . [CALLS_OK_10] ) /
( [Measures] . [CALLS_OK_SMP6] )
```

Para um melhor entendimento, a figura abaixo mostra as propriedades do grupo de medidas SMP3, que se assemelha aos demais grupos de medidas criados, mostrando como funciona o processo na ferramenta SSAS:

Figura 29 – Propriedades das Medidas



Fonte: Abraão Reckziegel e Adilson Favero.

Após a criação dos grupos de medidas, parte-se de fato para a criação dos KPIs. Ainda na ferramenta SSAS, definimos por linha de código as normas que regem o comportamento dos mesmos, indicando os casos de desempenho regular, satisfatório ou não.

Conforme descrito anteriormente, os valores de meta para os indicadores SMP3, SMP5 e SMP6, são 98% (noventa e oito por cento), 67% (sessenta e sete por cento) e 95% (noventa e cinco por cento), nesta ordem. Para que os valores sejam calculados e apresentados de forma amigável a cada processamento dos dados no cubo, definimos uma meta (*Goal*) e uma expressão de status (*Indicator*) para cada criado, conforme será apresentado a seguir:

KPI SMP3 - Expressão de meta:

[Measures].[SMP3] >= .98

Expressão de status deste indicador:

```

Case
  When KPIVALUE("SMP3") >= .98
    Then 1
  When .70 > KpiValue("SMP3") < .98
    Then 0
  Else -1
End

```

- **KPI SMP5** - Expressão de meta:

```
[Measures].[SMP5] >= .67
```

Expressão de status deste indicador:

```

Case
  When KPIVALUE("SMP5") >= .67
    Then 1
  When .40 > KpiValue("SMP5") < .67
    Then 0
  Else -1
End

```

- **KPI SMP6** - Expressão de meta:

```
[Measures].[SMP56] >= .95
```

Expressão de status deste indicador:

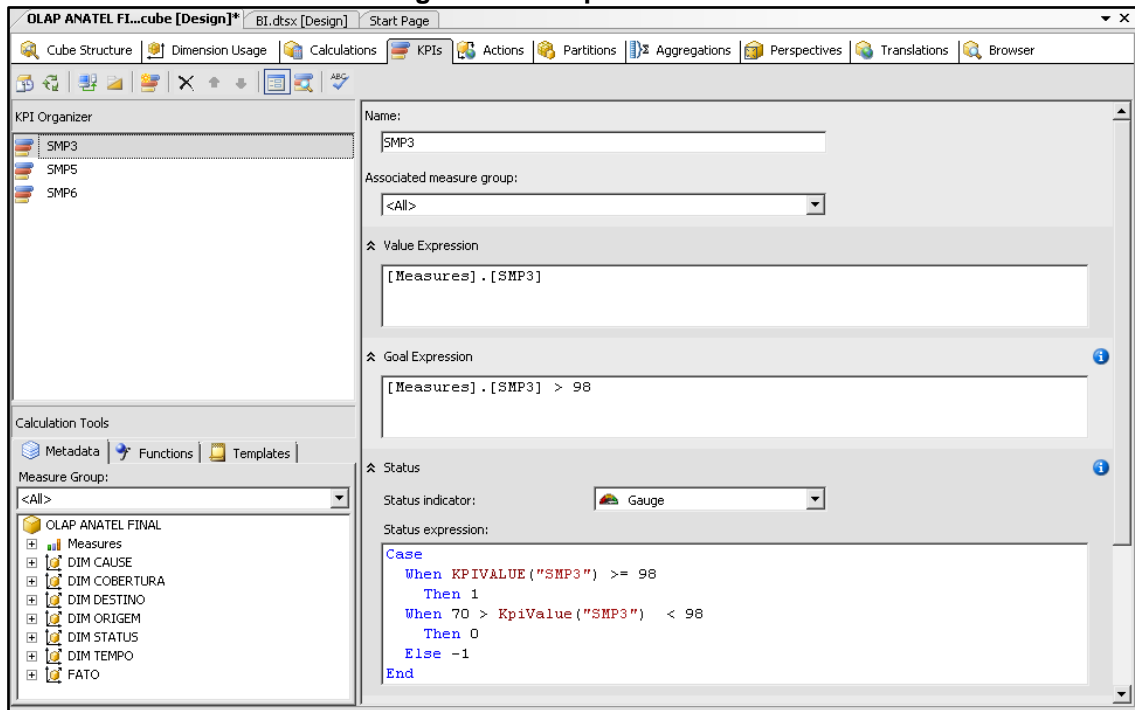
```

Case
  When KPIVALUE("SMP6") >= .95
    Then 1
  When .70 > KpiValue("SMP6") < .95
    Then 0
  Else -1
End

```

As figuras abaixo mostram como os indicadores são criados dentro da ferramenta SSAS, mostrando suas propriedades e por fim o frame de em execução, apresentando ao gestor o andamento dos KPIs:

Figura 30 – Propriedades dos KPIs



Fonte: Abraão Reckziegel e Adilson Favero.

Figura 31 – Deploy dos KPIs

Dimension	Hierarchy	Operator	Filter Expression	Value	Goal	Status	Trend	Weight
SMP3				9358.57%	False			
SMP5				6906.10%	True			
SMP6				9945.04%	True			

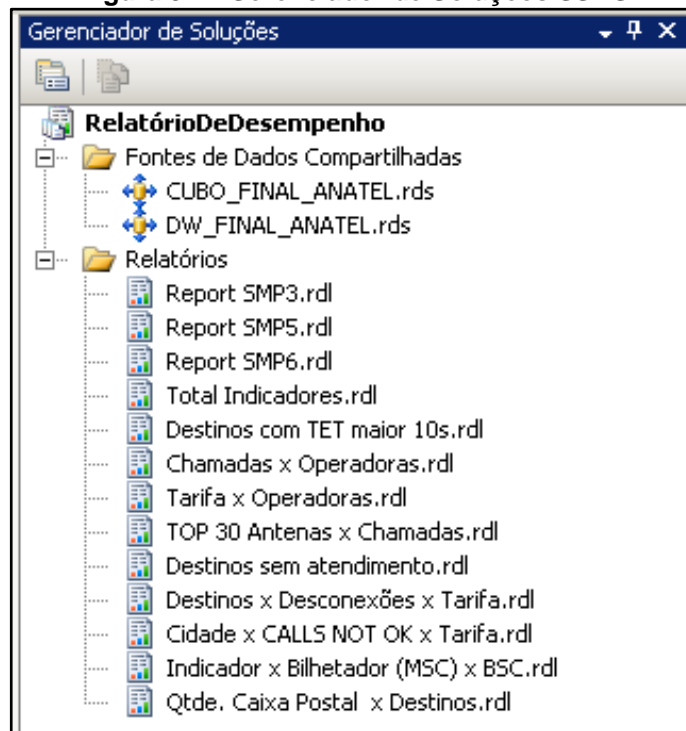
Fonte: Abraão Reckziegel e Adilson Favero.

5 EXPERIMENTAÇÃO DA SOLUÇÃO

5.1 RELATÓRIOS

Após a consolidação do *data warehouse* no SSIS e validação dos KPIs de desempenho no SSAS, é chegado o momento de construção dos relatórios que serão apresentados aos *stakeholders*. Utilizando a ferramenta SQL Server Report Services (SSRS), adicionamos como fonte de dados o cubo projetado no SSAS (CUBO_FINAL_ANATEL), e para consultas mais elaboradas, construídas diretamente por T/SQL (por exemplo: TOP 10, 50, 100), importamos diretamente o *data warehouse* (DW_FINAL_ANATEL), Como podemos observar nesta figura:

Figura 32 – Gerenciador de Soluções SSRS



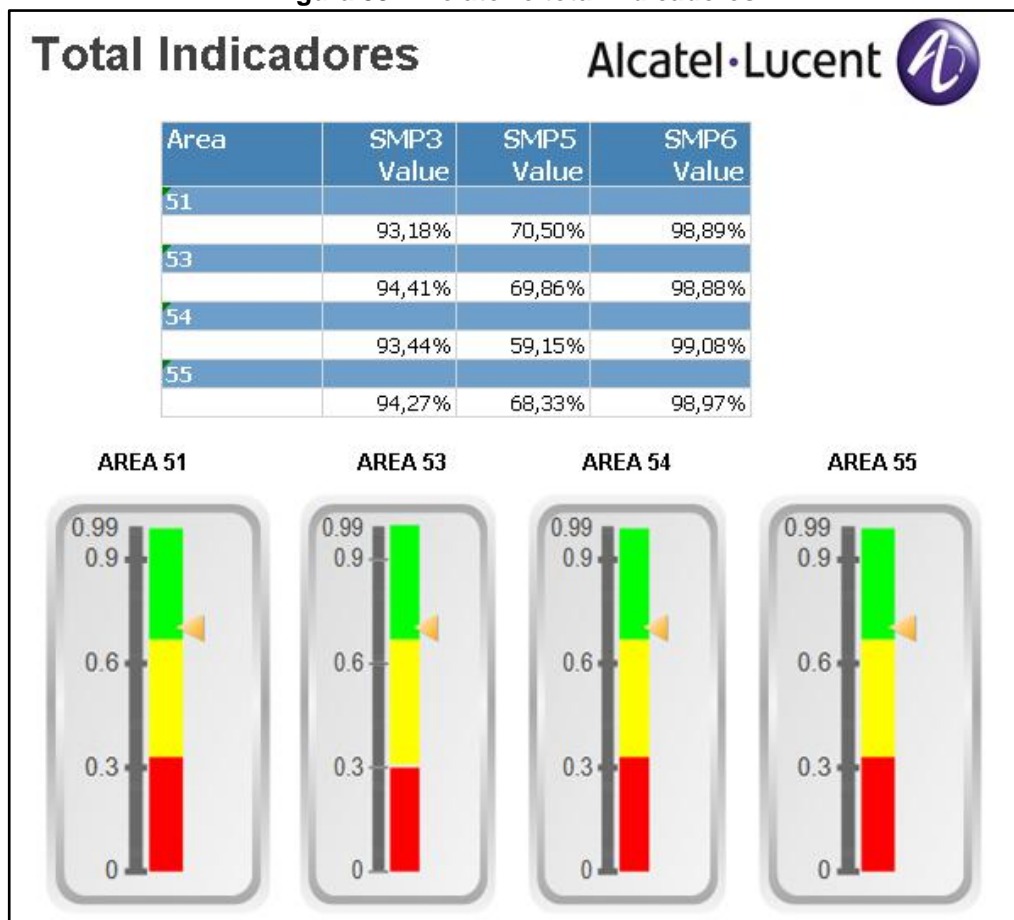
Fonte: Abraão Reckziegel e Adilson Favero.

Nas próximas seções apresentamos a confecção dos relatórios em função dos interesses da equipe da Alcatel-Lucent, bem como a visualização final destes.

5.1.1 Totalização dos Indicadores

Neste relatório selecionamos o atributo Área da dimensão cobertura e os KPIs *Values* de indicadores, criados anteriormente no SSAS. O modelo apresenta cada área de cobertura do estado do Rio Grande do Sul e os valores calculados. Inserimos marcadores de três cores, configurados para apontar a média entre os indicadores de cada área e também o logo da empresa para fins de personalização:

Figura 33 – Relatório total indicadores

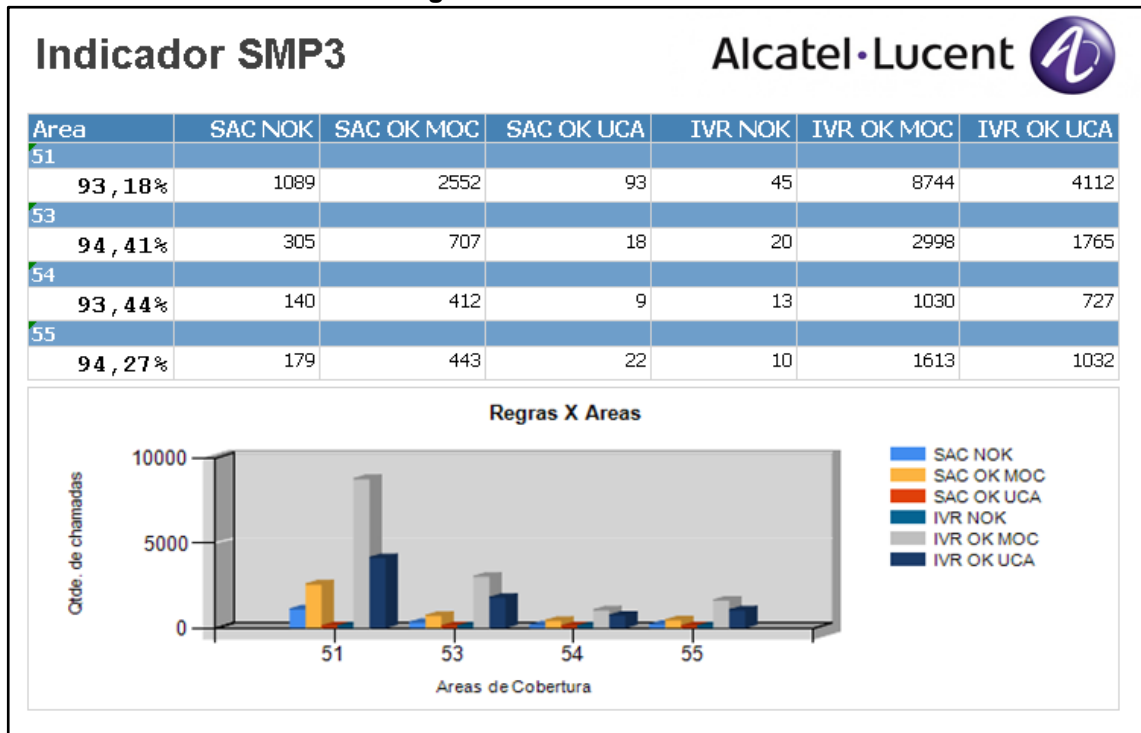


Fonte: Abraão Reckziegel e Adilson Favero.

5.1.2 Relatório SMP3

Para este relatório apresentamos os valores do indicador SMP3 por área de cobertura do Estado do Rio Grande do Sul e a quantidade de registros encontrados por regra que constitui o indicador. Para aumentar a dinâmica, um gráfico de colunas foi inserido apresentando as áreas em relação às regras ANATEL:

Figura 34 – Relatório SMP3

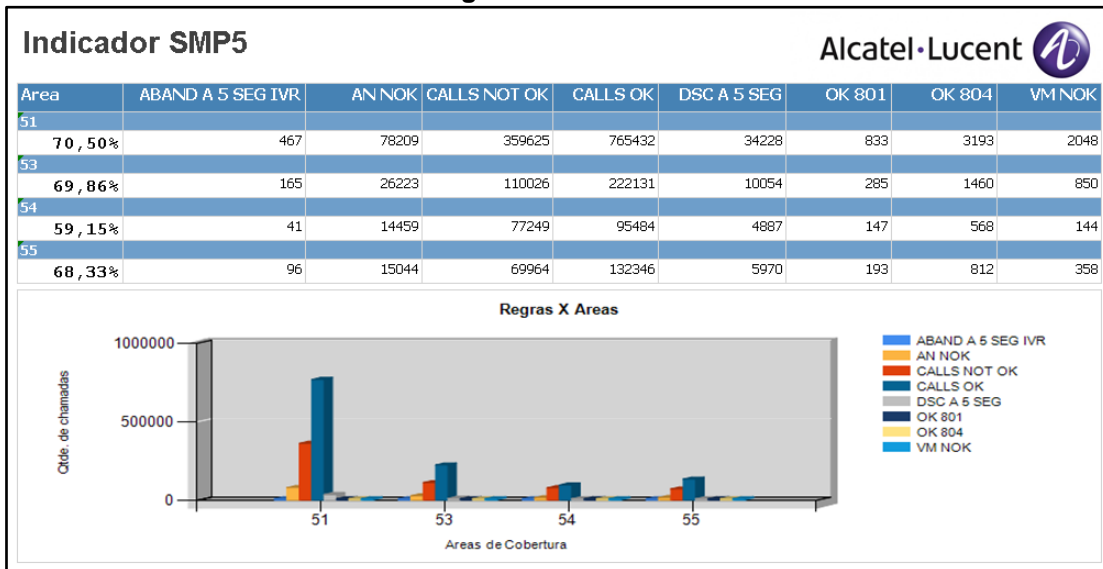


Fonte: Abraão Reckziegel e Adilson Favero.

5.1.3 Relatório SMP5

Neste relatório adotamos a mesma ideia, apresentamos cada área de cobertura do Estado do Rio Grande do Sul e a quantidade de registros encontrados por regra que constitui o indicador. Percebe-se que, por se tratar de um indicador que possui mais regras de contabilização que os demais, o trabalho apresentado neste relatório tem um ganho de tempo incrível relacionado ao trabalho executado anteriormente. Para aumentar a dinâmica, um gráfico de colunas foi inserido apresentando das áreas em relação às regras, conforme mostra a imagem a seguir:

Figura 35 – Relatório SMP5

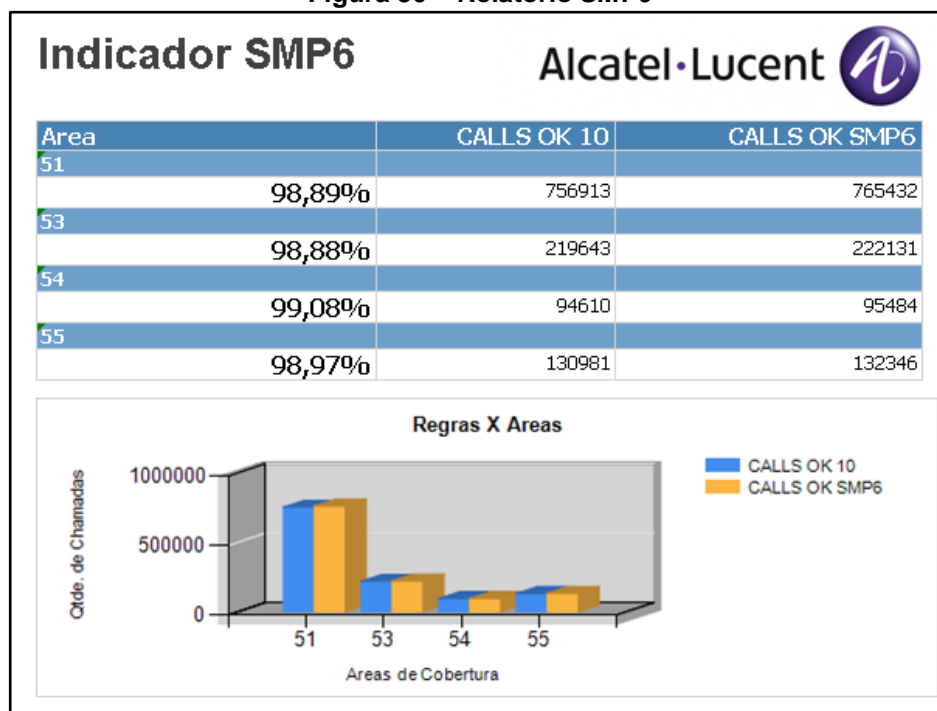


Fonte: Abraão Reckziegel e Adilson Favero.

5.1.4 Relatório SMP6

Neste relatório adotamos a mesma ideia, apresentamos cada área de cobertura do Estado do Rio Grande do Sul e a quantidade de registros encontrados cruzando as regras que constitui o indicador SMP6:

Figura 36 – Relatório SMP6




Fonte: Abraão Reckziegel e Adilson Favero.

5.1.5 Chamadas, Causas e Tarifas

Neste relatório cruzamos a quantidade de registros da tabela FATO agrupadas por Causas de desconexão e Tarifas aplicadas. Este relatório é de grande importância nos casos em que há uma desconexão por falha da rede e mesmo assim o cliente foi cobrado. Também pode se ter o controle de chamadas que deveriam estar sendo tarifadas na rede:

Figura 37 – Chamadas, Causas e Tarifas

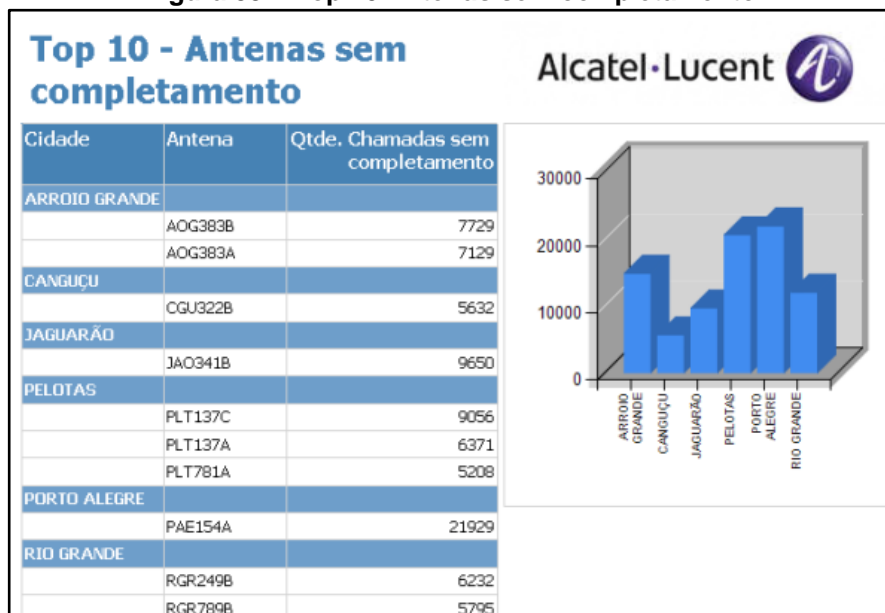
Chamadas, Causas e Tarifas		Alcatel-Lucent 
Descricao da Desconexão	Qtde. Chamadas	Tarifa atribuída
---	7 MOC	
	2751 UCA	
ABSENT SUBSCRIBER	48 MOC	
	48 UCA	
ADDRESS COMPLETE MESSAGE TIME-OUT	2462 UCA	
ADDRESS INCOMPLETE	16 MOC	
	890 UCA	
B ANSWER TIME OUT	11353 MOC	
	330 UCA	
B SUBSCRIBER BUSY	61 MOC	
	134858 UCA	

Fonte: Abraão Reckziegel & Adilson Favero

5.1.6 Cidades e Antenas com falhas

Este relatório tem por objetivo identificar um TOP 10 de antenas que fornecem cobertura na rede e não completam chamadas. Fator decisivo na atuação de reparos antes mesmo do surgimento de reclamações dos clientes ou acompanhamento do crescimento da população de clientes por cidades. Para execução deste relatório cruzamos os registros da tabela FATO, de valores de RecType igual a “UCA”, com os valores de STCellA para cada Cidade, conforme mostra a figura a seguir:

Figura 38 – Top 10 Antenas sem completamento

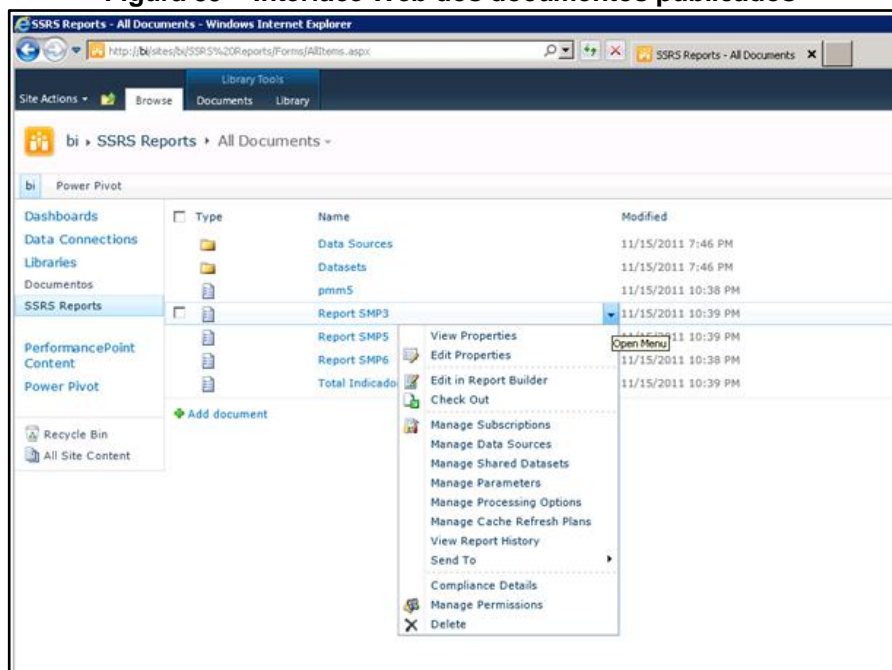


Fonte: Abraão Reckziegel e Adilson Favero

5.2 PUBLICAÇÃO

Após a implementação dos relatórios no SSRS, ligamos a exportação dos mesmos ao domínio local da empresa por intermédio da ferramenta MS SharePoint Server, o que nos permite criar em um ambiente *web* um repositório de informações apresentáveis a todos envolvidos no projeto. A imagem a seguir exibe a forma como os dados são acessados via navegador:

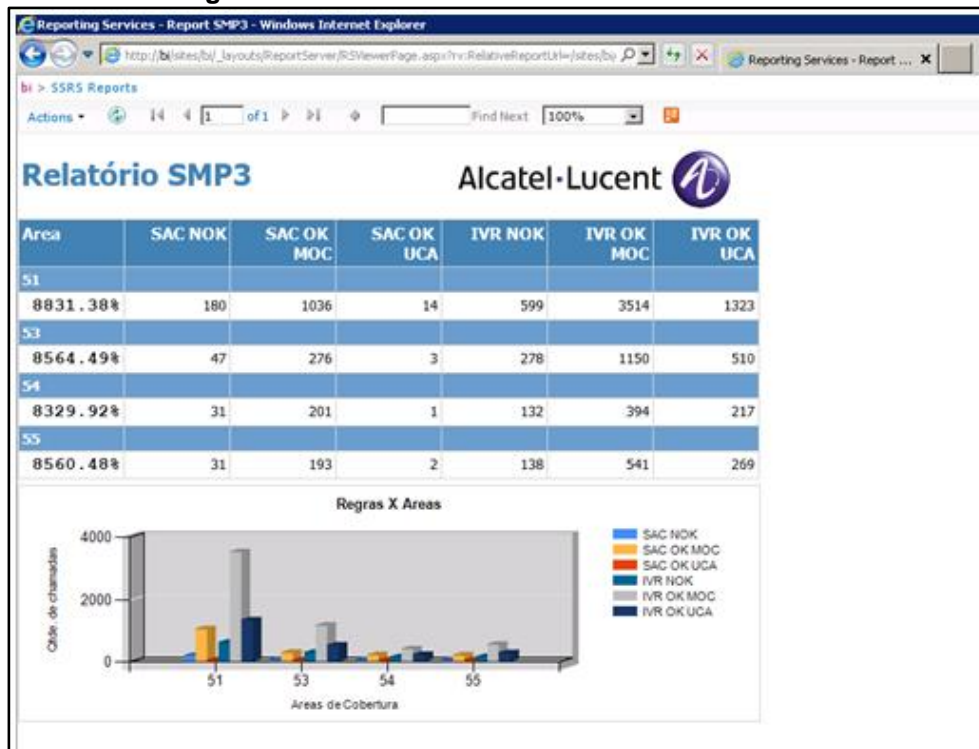
Figura 39 – Interface Web dos documentos publicados



Fonte: Abraão Reckziegel e Adilson Favero

A figura a seguir mostra os relatórios construídos no SSRS, sendo acessados no domínio registrado pela ferramenta SharePoint:

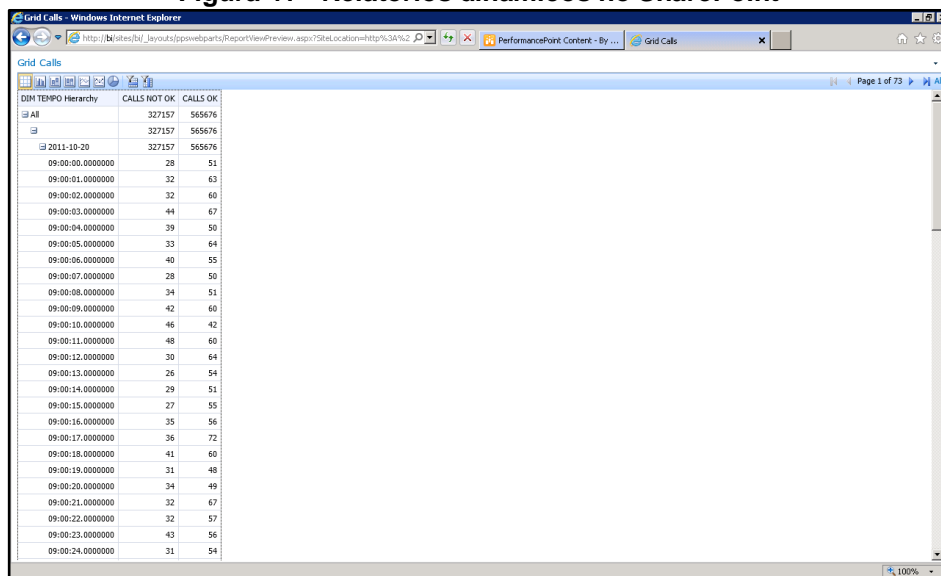
Figura 400 - Relatório do SSRS no SharePoint Server



Fonte: Abraão Reckziegel e Adilson Favero

Através do conteúdo publicado ainda é possível acessar a galeria de documentos e consultar dados dinamicamente, conforme mostra a figura abaixo:

Figura 41 - Relatórios dinâmicos no SharePoint



Fonte: Abraão Reckziegel e Adilson Favero

5.3 MINERAÇÃO

Para extrair informações implícitas dos registros de chamadas, que de alguma forma não poderiam ser evidenciadas nos relatórios propostos, partimos para a aplicação das técnicas de data mining. Considerando que as regras SAC_NOK, IVR_NOK e CALLS_NOT_OK, contabilizam apenas como denominadores dos indicadores SMP3 e SMP5 foram escolhidos estes como atributos alvos de nossos experimentos. O exemplo abaixo pode ilustrar melhor esta escolha, evidenciando na própria fórmula a simplificação algébrica das regras:

$$\begin{aligned}
 \text{SMP3} &= 100 * \\
 & \frac{(\text{SAC_OK_MOC} + \text{SAC_OK_UCA} + \text{IVR_OK_MOC} + \text{IVR_OK_UCA})}{(\text{SAC_OK_MOC} + \text{SAC_OK_UCA} + \text{IVR_OK_MOC} + \text{IVR_OK_UCA} + \text{SAC_NOK} + \text{IVR_NOK})} \\
 \\
 \text{SMP5} &= 100 * \\
 & \frac{(\text{CALLS_OK} + \text{AN_NOK} + \text{OK_801} + \text{OK_804} + \text{VM_NOK})}{(\text{CALLS_OK} + \text{AN_NOK} + \text{VM_NOK} + \text{CALLS_NOT_OK})}
 \end{aligned}$$

Considerando os valores de SAC_NOK, IVR_NOK e CALLS_NOT_OK com valores variando entre 0 e 1, buscamos inferir relações entre os destinos discados pelos assinantes, a desconexão ocorrida durante a chamada, o tempo de estabelecimento da chamada ou qualquer outra relação entre os atributos da tabela FATO que modele o cenário ofensor.

A ferramenta WEKA (Waikato Environment for Knowledge Analysis – 3.7.1), criada pela universidade de Waikato, na Nova Zelândia, foi escolhida para realizarmos testes nos exemplos coletados, possibilitando a execução de diferentes algoritmos até o encontro de um modelo apropriado de cenário ofensor das regras anteriormente citadas.

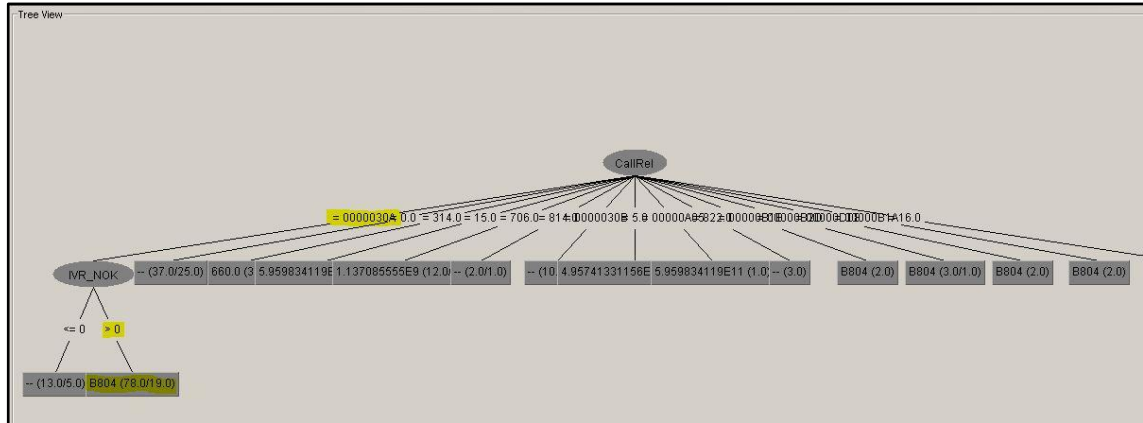
Para haver uma maior precisão, foi um conjunto de testes para cada regra, contendo metade de tuplas com status de regra igual a 1 (um) e metade de tuplas com regra igual a 0 (zero).

Prospecção de IVR_NOK

No conjunto de teste que pertence à regra IVR_NOK, percebemos com o auxílio do algoritmo J48, que quando o assinante disca para o destino *804 (B804 –

Consulta ao saldo), ocorre uma desconexão 0000030A (CLEAR/A ONHOOK DURING SET-UP PHASE – *DESCONECTADAS PELO ASSINANTE ANTES DE CHAMAR*), isto é previsto em 85% dos casos em que o assinante disca *804, pois, conforme figura abaixo, o algoritmo classificou 78 (setenta e oito) tuplas corretamente e falhou em 19 (dezenove):

Figura 42 - Árvore IVR_NOK

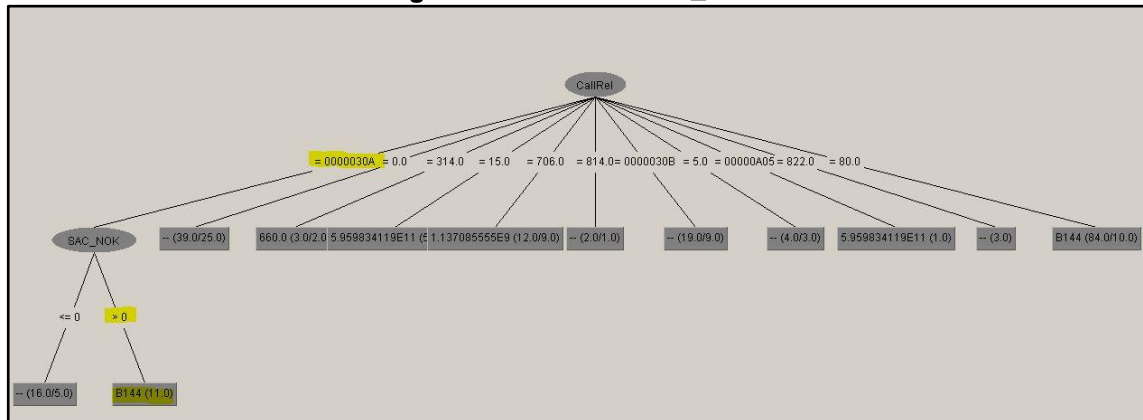


Fonte: Abraão Reckziegel e Adilson Favero

Prospecção de SAC_NOK

No conjunto de teste que pertence à regra SAC_NOK, percebemos com o auxílio do algoritmo J48, que quando o assinante disca para o destino *144 (B144 – Call Center), ocorre uma desconexão 0000030A (CLEAR/A ONHOOK DURING SET-UP PHASE – *DESCONECTADAS PELO ASSINANTE ANTES DE CHAMAR*), em um cenário muito parecido ao da regra IVR_NOK. Isto é previsto em 90% dos casos em que o assinante disca *144 conforme mostra a figura a seguir:

Figura 433 - Árvore SAC_NOK

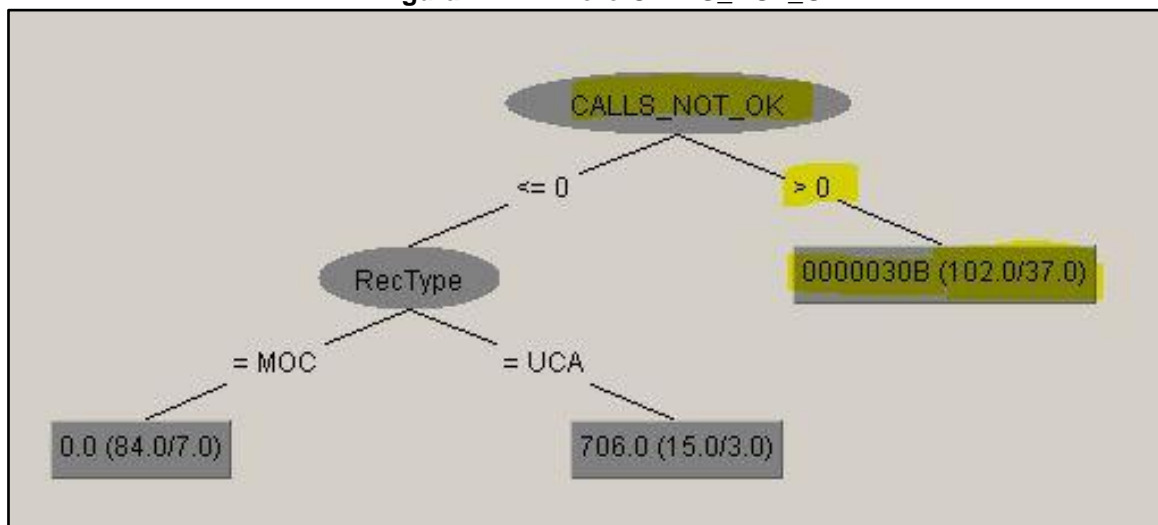


Fonte: Abraão Reckziegel e Adilson Favero

Prospecção de CALLS_NOT_OK

No conjunto de teste que pertence à regra CALLS_NOT_OK, percebemos também com o auxílio do algoritmo J48, que em 73% dos casos, estão ligados diretamente ao CallRel “030B” (CLEAR/A ONHOOK DURING 6 S – Desconexão aos seis segundos pelo usuário), onde na figura abaixo podemos ver 102 (cento e dois) acertos e 37 (trinta e sete) erros encontrados durante a classificação do algoritmo:

Figura 444 - Árvore CALLS_NOT_OK



Fonte: Abraão Reckziegel e Adilson Favero

Os resultados obtidos através dos algoritmos foram extremamente importantes para a simulação dos cenários em que ocorreram falhas, pois verificamos que as chamadas que degradam o indicador são aquelas em que o cliente liga, chama uma vez e desconecta a ligação (conhecido como “toque”). Logo, novas linhas de raciocínio foram tomadas para mudar o comportamento e interação entre os clientes.

5.4 DEPOIMENTOS

Segundo o colaborador Lucas Ferraz, analista do projeto de indicadores da Alcatel-Lucent: “Esta solução, nunca cogitada anteriormente, apresenta ferramentas de uso admirável em tamanha facilidade. Anterior a esta, o trabalho era executado em cinco dias, sem pelo menos guardarmos uma referência histórica dos indicadores. O nível de satisfação do cliente aumentou muito, pois agora podemos ser mais proativos, ao invés de atuarmos paliativamente, em um ou dois dias, ficaram mais

claros nossos objetivos. Também percebemos o quanto temos que evoluir no SMP3, junto aos pontos que devemos manter nos indicadores SMP5 e SMP6”.

6 CONSIDERAÇÕES FINAIS

6.1 LIÇÕES APRENDIDAS

Para os autores deste trabalho, junto ao imenso conhecimento adquirido, fica evidente a capacidade dinâmica que os modelos de *data warehouse* têm para lidar com um grande volume de dados e diferentes fontes. A solução implantada permitiu uma efetiva produção e controle dos indicadores da ANATEL, e aplicando conceitos de BI, ilustra que os relatórios criados nela auxiliam na tomada de decisão empresarial.

A exploração por Mineração de Dados identificou as situações mais frequentes, causadoras de problemas. Isto nos permite garantir certo nível de excelência nos serviços prestados, pois agora os colaboradores ganharam tempo para atuar diretamente na correção das falhas, pulando o antigo processo de identificação que era complexo, tardio e duvidoso.

O proveito de podermos explorar diferentes ferramentas que se integram é vasto, pois construímos uma única solução de boa escalabilidade e que atende, segundo o cliente, todas suas questões de negócio. Outro ponto relevante é a quantidade de colaboradores envolvidos, pois com a solução o processo, como um todo, pode ser limitado a um único membro, permitindo que hajam mais recursos focados na correção das falhas e, conseqüentemente, um melhor desempenho nos indicadores ANATEL.

6.2 TRABALHOS FUTUROS

Realizando-se uma análise na escalabilidade deste trabalho, para que a solução possa ser aplicada aos demais estados do Brasil e, além disso, contemple um período maior de dados, concluímos que se basta expandir os recursos de *hardware*, que até o momento foram utilizados, pois conforme a operadora Telecom o requisito do formato dos dados fornecidos pela plataforma CDR-View é o mesmo para todo o território nacional. Logo, afirmamos que a flexibilidade do modelo OLAP criado permite que este redimensionamento trabalhe desde que haja recursos de armazenamento e processamento paralelos a demanda.

Os testes aferidos com o auxílio de mineração foram inicialmente bem sucedidos, pois projetaram a realidade dos clientes com problemas, mas ainda

assim mereciam ser melhores explorados. Com o auxílio de variadas ferramentas de DM, poderíamos inferir cenários de falhas na rede que não estão relacionados com as regras de contabilização dos indicadores e ainda inseri-los em relatórios da *web*.

Observando os KPIs construídos sobre os indicadores SMP3, SMP5 e SMP6, percebemos que, em outrora, este experimento pode ser expandido às demais bases de dados pertencentes à operadora Telecom, apresentando também valores dos indicadores SMP1, SMP2, SMP4, SMP7 e assim por diante. Para que isto ocorra, uma nova análise sobre as regras de contabilização e fonte de dados destes indicadores deve ser realizada, mas desta forma, tanto no frame de apresentação dos *KPIs* quanto na interface *web*, poderíamos ter uma única área que permitiria o controle de todos indicadores ANATEL, em todo o Brasil.

REFERÊNCIAS

FAYYAD, Usama. **From Data Mining To Knowledge Discovery: an overview**. England: 1996, p. 34.

GONÇALVES, Marcio. **Extração de dados para Data Warehouse**. Rio de Janeiro: Axcel Books, 2003.

H. Inmon, Willian. **Building the Data Warehouse**, Fourth Edition. EUA:2005.

JACOBSON, Reed; MISNER, Stacia; HITACHI Consulting. **Microsoft SQL Server 2005 Analysis Services: Passo a Passo**. Porto Alegre: Bookman, 2007.

KIMBALL, Ralph. **The Data Warehouse Toolkit**. EUA, New York: 2002, p. 436. Disponível em: Biblioteca Central Irmão José Otão, PUCRS - 2011.

MICROSOFT, Corporate; **Software Evaluation Guide of BI solutions**. EUA: Solver, 2008.

Plano Geral de Metas de Qualidade para o Serviço Móvel Pessoal – PGMQ-SMP. Disponível em: <<http://www.anatel.gov.br>> Acesso em: 28 de março de 2011.

TAN, Pang-Ning. **Introduction To Data Mining**. Boston: 2006, p. 796. Disponível em: Biblioteca Central Irmão José Otão, PUCRS - 2011.

VERCELIS, Carlo. **Business Intelligence: Data Mining and Optimization for Decision Making**. United Kingdom: 2009, p.436.

GLOSSÁRIO

Ad-hoc - Expressão latina cuja tradução literal é "para isto" ou "para esta finalidade".

Design - Denomina-se Design qualquer processo técnico e criativo relacionado à configuração, concepção, elaboração e especificação de um artefato. Esse processo normalmente é orientado por uma intenção ou objetivo, ou para a solução de um problema.

Gigabyte - Unidade de medida múltipla do byte. Um gigabyte corresponde a 1073741824 bytes.

Lookup - É a transformação utilizada nos processos de ETL para buscar valores de outra tabela ou view do banco de dados.

Metadados - São dados sobre outros dados. Um item de um metadado pode dizer do que se trata aquele dado, geralmente uma informação inteligível por um computador. Os metadados facilitam o entendimento dos relacionamentos e a utilidade das informações dos dados.

Query - Abstração lógica de uma consulta em bancos de dados.

Ranking - Processo de posicionamento de itens de estatísticas individuais, de grupos ou comerciais, na escala ordinal de números, em relação a outros.

Software - Conjunto de instruções armazenadas em disco(s) ou em chips internos do computador que determinam os programas básicos, utilitários ou aplicativos, que ele tem para serem usados.

Trigger - Recurso de programação executado sempre que o evento associado ocorrer. Trigger é um tipo especial de procedimento armazenado, que é executado sempre que há uma tentativa de modificar os dados de uma tabela que é protegida por ele.

Update - Comando SQL que realiza a mudança de valores de dados em uma ou mais linhas de uma tabela existente.