

Mapeamento da Simetria de Tráfego em Internet Exchanges Points (IXP)

Sandro Lucas Azevedo Ferreira¹
Orientador: César Augusto Hass Loureiro¹

¹Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - Campus Porto Alegre, Curso Tecnologia em Sistemas para Internet
Rua Cel. Vicente, 281, Porto Alegre - RS

osandrolucas@gmail.com, cesar.loureiro@poa.ifrs.edu.br

Resumo: O estudo realizado em Bertholdo, et. al (2021) pelo grupo de pesquisa *Design and Analysis of Communication Systems* (DACs) da Universidade de Twente, localizada em Enschede/Holanda, identificou a presença de assimetria na troca de tráfego realizada em diferentes *Internet Exchanges Points* (IXPs) do mundo, que são infraestruturas consideradas essenciais de entrega de conteúdo na Internet.

A partir desta constatação, o presente artigo objetiva apresentar uma solução baseada em *Business Intelligence* (BI) capaz de processar, transformar e analisar dados de diferentes IXPs. Os dados utilizados foram coletados através de um método baseado em sondagem ativa *anycast* que foi desenvolvido pelo grupo de pesquisa DACs e como resultados foram apresentados, por meio de um *dashboard*, os diferentes tipos de assimetria de tráfego dos Sistemas Autônomos (AS) em cada um dos cinco IXPs estudados.

A partir dos resultados obtidos, foi possível observar que o percentual de simetria é 7,5% maior nas redes /24 em comparação às redes /23, o que demonstra que redes mais específicas tendem a ter um tráfego mais simétrico. Também foi possível observar que aproximadamente 20% do tráfego da internet não é simétrico, o que pode ser um ponto de melhoria para os provedores de conteúdo.

Palavras-chave: Internet Exchanges Points (IXP); Ponto de Troca de Tráfego (PTT); Business Intelligence (BI); Autonomous System (AS).

1 Introdução

Internet Exchanges Points (IXPs) são elementos essenciais na Internet. Eles desempenham um papel importante fornecendo infraestrutura para troca de tráfego, atraindo Provedores de Serviços de Internet (ISPs), Redes de Distribuição de Conteúdo (CDNs) e Provedores de Nuvem, levando diferentes opções para as operadoras de redes se conectarem. (BERTHOLDO, et al, 2021)

O benefício mais importante dos IXPs é não cobrar dos participantes pelo volume de tráfego utilizado, permitindo a possibilidade de um rápido crescimento. Um exemplo recente é o estudo realizado em CANDELA, et. al (2020) onde alguns IXPs relataram crescimento de até 60% durante o período da pandemia de Covid-19.

Além disso, outro aspecto importante para os participantes ao se conectarem a um IXP é que eles se conectam diretamente com centenas de outros sistemas autônomos (ASes). Essa conexão direta pode reduzir os preços de conexão e melhorar a latência de acesso a conteúdo.

De acordo com Bertholdo, et. al (2021), em seu artigo *Forecasting the Impact of IXP Outages Using Anycast*, foi constatado que apesar da conexão com o IXP ser muito vantajosa, ocorrem problemas de assimetria no tráfego, isto é, nem todo o conteúdo solicitado por um participante via IXP é realmente respondido pela mesma conexão ao IXP. Para exemplificar, suponhamos um cenário onde um usuário da Internet está situado em Porto Alegre e requisita conteúdo a uma plataforma de *streaming* de vídeo, na qual seu provedor de acesso encaminha ao IXP de Porto Alegre (RS), a plataforma por sua vez consegue identificar a localidade deste usuário, porém, ao invés de entregar este conteúdo através do IXP de Porto Alegre, opta por entregar através de sua conexão de trânsito na Internet, caracterizando a assimetria e consequentemente o aumento na latência na entrega do conteúdo, conforme ilustrado na Figura 1.

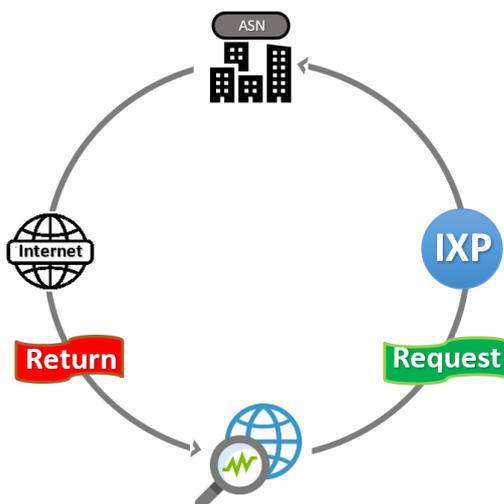


Figura 1. Assimetria de Tráfego

Fonte: Elaboração pelo Autor

Neste contexto, este trabalho utiliza os dados coletados pelo grupo de pesquisa da universidade de Twente, responsáveis pelo artigo supracitado, e apresenta o desenvolvimento de uma solução, utilizando ferramentas e conceitos de *Business Intelligence (BI)*, para medir o nível de simetria de tráfego em cinco IXP mundiais.

Para tanto, este trabalho pretende atingir os seguintes objetivos específicos:

- Criar uma arquitetura capaz de realizar o processo de extração, transformação e carregamento (ETL) para cada amostra de dados gerada.
- Definir os indicadores que devem ser apresentados.
- Construir um *dashboard* para apresentação dos dados analisados.
- Disponibilizar toda a infraestrutura construída para que o grupo de pesquisa possa reutilizá-la em futuros estudos.

Este artigo está organizado em 8 seções, na seção 2, abordaremos a fundamentação teórica e revisões bibliográficas, seguido dos trabalhos relacionados. Na seção 4 é

apresentada a metodologia para o desenvolvimento da pesquisa. Na seção 5 temos a explicação sobre a origem e captura dos dados, na seção 6 são apresentadas as ferramentas utilizadas e o processo para a criação da solução, na seção 7 temos os resultados preliminares e por fim, as considerações finais.

2 Fundamentação Teórica

Nessa seção abordaremos sobre o funcionamento da Internet, sobre os ASes e a sua relação com os IXPs, bem como a definição do Processo ETL (*Extraction, Transformation and Loading*), os conceitos de *Business Intelligence* e *Dashboard*.

2.1 A INTERNET, ASes, e IXPs

A Internet pode ser vista como uma rede de redes interligadas, que utiliza como principal protocolo de comunicação o IP (*Internet Protocol ou TCP/IP*), assim permitindo que todos os dispositivos que estiverem conectados a ela consigam interagir entre si e que possam funcionar sobre qualquer tecnologia de telecomunicações. (NICBR, 2014)

Os endereçamentos de IP são atribuídos às instituições chamadas de Sistemas Autônomos (AS, do inglês *Autonomous System*), que são redes com independência técnica e administrativa, que colaboram entre si e seguem padrões tecnológicos em comum para poderem interoperar e formar a Internet. Cada AS é identificada na Internet através de um número chamado *Autonomous System Number (ASN)*¹. Existem diversos tipos de ASes, tais como os que oferecem serviços ou conteúdos na rede e são conhecidos como provedores de serviço ou provedores de conteúdo, os que oferecem acesso a rede, conhecidos como provedores de acesso, e até mesmo os provedores destes provedores, conhecidos como provedores de trânsito. Todas as aplicações, sites e serviços que conhecemos, como por exemplo Google, Twitter, Netflix, Facebook, Youtube, entre outros, são exemplos de ASes, assim como os usuários da Internet, que utilizam endereçamento IP de um provedor de acesso, que também é um AS, ou seja, todos os endereços IPs pertencem a um AS.

A interconexão entre esses ASes forma o *backbone* da Internet. O funcionamento é tecnicamente simples, pois resumidamente acontecem envios de pacotes de um lado para o outro, sem diferenciação ou qualquer tipo de verificação sobre seu conteúdo, sendo que este tipo de serviço normalmente é oferecido por provedores de trânsito. A internet foi criada neste formato para que possa ser escalável, ou seja, para permitir que pessoas ou empresas que estejam conectadas à ela possam criar novas aplicações e serviços sem precisar realizar qualquer ajuste em seu núcleo, ou negociar com as outras milhares de pessoas e empresas que já estavam conectadas. (NICBR, 2014)

Para que os ASes conheçam os endereços IPs das redes de outros ASes, utiliza-se um protocolo chamado BGP (*Border Gateway Protocol*). O BGP é um protocolo de roteamento que tem como principal objetivo aprender e ensinar caminhos (rotas). Através deste protocolo, o AS que anteriormente detinha somente as informações da sua rede interna, passa a trocar informações com os ASes vizinhos, e através dessa troca cada AS constrói o seu mapa, conhecido como Tabela BGP. A partir da Tabela BGP cada roteador da Internet tem acesso aos melhores caminhos para cada um dos destinos na rede. (SILVA, 2013)

¹ Disponível em: <https://www.upx.com/post/asn-crescimento-isp>

Diferentes ASes podem se interconectar por meio de enlaces privados entre si, dois a dois, o que possui um custo elevado e por conta disso hoje em dia cada vez mais ASes usam um tipo de infraestrutura compartilhada na rede, conhecido como IXP (*Internet Exchange Point*). Essas infraestruturas conectam fisicamente diversos ASes, o que acaba sendo um formato vantajoso porque com um único enlace físico, um AS pode se interligar com alguns ou todos os outros participantes do IXP.

Na Internet não há uma autoridade central que determine como as redes devem se interligar ou qual o tipo de acordo deve ser feito, ou seja, cada uma delas tem a liberdade e autoridade para negociar livremente com os seus pares a fim de melhorar sua qualidade de conexão para atingir o objetivo de seu negócio, seja ele acesso ou distribuição de conteúdo (NICBR, 2014). Através dessa liberdade e autoridade, os ASes utilizam de vários tipos de relacionamentos entre si, dentro ou fora dos IXPs. Por exemplo, um AS pode permitir que todos os pacotes vindos de outro AS utilizem a sua infraestrutura para chegar na Internet, sendo esse tipo de relacionamento conhecido como Trânsito, porque a rede permite a passagem dos pacotes, levando-os para outras redes, e nesse cenário normalmente se cria uma relação comercial entre os participantes. Também existem casos onde dois ou mais ASes podem permitir o acesso aos clientes e serviços uns dos outros, mutuamente, mas sem levar os pacotes para a Internet, e esse tipo de relacionamento é conhecido como Troca de Tráfego, onde normalmente acontece uma relação de colaboração por meio do IXP. (FERREIRA, 2020)

2.2 SISTEMA DE *BUSINESS INTELLIGENCE*

Business Intelligence (BI) é um termo genérico que combina arquitetura, banco de dados, ferramentas analíticas, aplicações e metodologias. O maior objetivo do BI é oferecer um acesso interativo aos dados, permitindo a manipulação e análise deles. O processo de BI é baseado na transformação e organização dos dados para gerar informação, e auxiliar na tomada de decisão. (SHARDA et. al, 2016)

SHERMAN (2014) explica que a diferença entre dados e informação é que dados são crus, randômicos e desorganizados. Informação é o dado de forma organizada, estruturada e processada, que pode ser utilizado para adquirir conhecimento. Existem diversas formas para transformar o dado em informação, e no presente estudo utilizaremos o processo de ETL para processamento dos dados e um *Dashboard* para a apresentação dos resultados obtidos:

- ETL é um processo que consiste em três etapas: extração, transformação e carregamento. O processo de ETL é tudo que fica entre os sistemas de fonte operacional e a área de apresentação que normalmente é um *dashboard* ou um relatório. (KIMBALL; ROSS, 2013)
- Em termos de definição, "*Dashboard* é uma apresentação visual do dado usado para monitorar condições e/ou facilitar o entendimento" (WEXLER et. al, 2017). Com diversas ferramentas e indicadores, o *dashboard* permite ao usuário construir suas próprias questões e gerar relatórios estratégicos, para facilitar a visualização e entendimento dos dados.

Através dos conceitos apresentados, será explicado na seção cinco a forma de captura das estatísticas de tráfego de determinados IXPs para a geração de um *dashboard* que possibilite a análise dos dados.

3 Trabalhos Relacionados

De forma resumida, os IXPs são infraestruturas físicas com o papel de interconectar os ASes. Existem alguns estudos que dissertam sobre o importante papel que os IXPs desempenham quando falamos sobre a conectividade com a Internet. A conectividade oferecida pelos IXPs permite reduzir a atenuação causada pela distância, redução de custos e maior facilidade de troca de tráfego entre os seus membros. (WERMANN, 2016). Desde suas implantações iniciais, os IXPs encurtaram os caminhos, fornecendo melhorias técnicas e econômicas. Hoje em dia, eles também atraem ASes não locais que possuem interesse em se conectarem a redes específicas presentes em um IXP, elevando o escopo do IXP a um nível de cobertura nacional ou global. (NOMIKOS et. al, 2018)

Já existe um consenso sobre os IXPs serem infraestruturas importantes que servirão de apoio para novas tecnologias (por exemplo, IoT e 5G) (HOESCHELE et. al, 2021). Ao mesmo tempo, ainda acontecem debates para determinar se eles devem ser classificados como “infraestruturas críticas” e conseqüentemente sofrerem implicações governamentais. Para a APNIC (2021), eles possuem um papel importante na topologia da internet mas não são classificados como críticos, entretanto podem vir a sofrer atribuições de políticas específicas.

Um estudo focado na anatomia das infraestruturas de IXPs públicos, com olhar específico para a internet do Brasil, é encontrado em Brito et al. (2015). Este trabalho trouxe como contribuição a primeira análise do ecossistema de IXPs em operação no Brasil, compilando informações sobre amostras das tabelas BGPs, que compreendem desde a caracterização dos tipos de membros desses ambientes até a construção de grafos de conectividade em nível de Sistema Autônomo (AS).

A maioria dos estudos tentou entender os IXPs utilizando dados públicos de diferentes fontes, tais como os bancos de dados Euro-IX², PeeringDB³, PCH⁴, dados de traceroute, LookingGlass IXP e coletores de rota, como RIS⁵ e Route Views⁶. Mais recentemente, em 2012, foi criada uma nova plataforma chamada IX-F⁷ (*Internet eXchange Federation*) que tem como objetivo principal construir uma comunidade IXP global e ajudar no desenvolvimento dessas infraestruturas em todo o mundo, porém sua adoção ainda é limitada.

Esses trabalhos relacionados servem de embasamento e complemento para o proposto nesse artigo, pois até o momento não foram encontrados trabalhos sobre a análise de assimetria na troca de tráfego em IXPs.

4 Metodologia

Para realizar o desenvolvimento deste trabalho, foi utilizado o método incremental, focado no envio periódico de demandas. Este modelo combina elementos do Modelo em Cascata aplicados de maneira iterativa, ou seja, o progresso acontece através de sucessivos refinamentos, melhorados a cada iteração. (PRESSMAN, 2005).

² Disponível em: <https://ixpdb.euro-ix.net/en/ixpdb/ixps/>

³ Disponível em: <https://www.peeringdb.com>

⁴ Disponível em: <https://www.pch.net/ixp/dir>

⁵ Disponível em: <https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris>

⁶ Disponível em: <http://routeviews.org/>

⁷ Disponível em: <http://www.ix-f.net>

Os dados brutos são enviados pelo DACS, no qual a cada demanda realizada, disponibiliza mais dados para refinar os resultados apresentados no *dashboard* e incluir novos indicadores.

A Figura 2 ilustra o ciclo de desenvolvimento utilizado no desenvolvimento da solução.

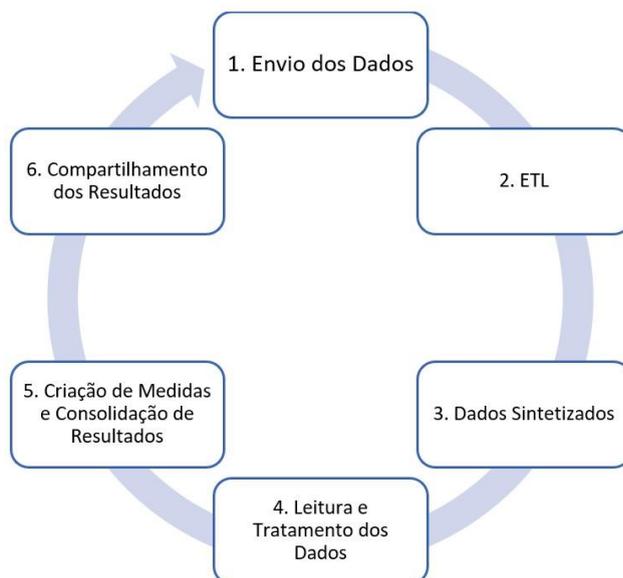


Figura 2. Ciclo de Desenvolvimento

Fonte: Elaboração pelo Autor

A seguir está o detalhamento dos passos listados na Figura 2.

Envio dos Dados: Este é o momento no qual o grupo de pesquisa DACS envia os arquivos com os dados brutos, contendo a lista de ASes conectados, os dados de troca de pacotes e a lista BGP para cada um dos IXP analisados.

ETL: Neste momento, os dados citados no passo anterior passam pela etapa de ETL, onde ocorrem alguns ajustes como a criação de novas variáveis, e armazenamento em um banco de dados.

Dados Sintetizados: Subsequente ao passo 2, são gerados os arquivos com os organizados e sintetizados, agrupados a nível de AS.

Leitura e Tratamento dos Dados: Neste momento os dados sintetizados são carregados no software Power BI e passam pela última etapa de transformação antes da exibição dos resultados, onde acontece a criação de campos baseados nas regras de negócio definidas, a alguns testes de conformidade dos dados.

Criação de Medidas e Consolidação de Resultados: Este é o momento no qual o *dashboard* uma vez já construído, passa a ser atualizado a partir dos dados gerados no passo 4 e a informação é apresentada por meio de gráficos e tabelas.

Compartilhamento dos Resultados: Este é o encerramento e reinício do ciclo, onde os resultados são compartilhados com o DACS e ocorrem as definições dos próximos passos a serem seguidos.

O ciclo supracitado iniciou em outubro de 2021 e desde então ocorrem reuniões duas vezes por semana. Estima-se que até o término do trabalho tenham ocorrido 42 ciclos.

5 Origem dos dados

Conforme detalhado na seção Plano de Medição do artigo Bertholdo, et. al (2021), para a coleta dos dados foram implantados sites *anycast* em cada um dos cinco IXPs analisados e outro site *anycast* em um provedor de trânsito de Internet. O prefixo da rede *anycast* é propagado para os IXPs e para o provedor de trânsito por BGP, conforme mostrado na Figura 3. A política de roteamento fornece preferência ao recebimento de tráfego no IXP sobre o provedor de trânsito (*Drain site*), assim como o prefixo mais específico /24 tem preferência sobre o /23.

O sistema de medição ativo (*Pinger*) também é um site *anycast* gerando pacotes ICMP para cada rede /24 mapeada no estudo de W. B. de Vries, et. al (2017). Se o pacote ICMP enviado pelo *Pinger* utilizar a rota via IXP, a resposta ao ICMP será enviada pelo site *anycast* de dentro do IXP (IXP-1), caso contrário a resposta será enviada pelo *Drain Site*. Coletando a informação de onde a requisição foi respondida, tem-se a cobertura de redes atendidas por cada IXP.

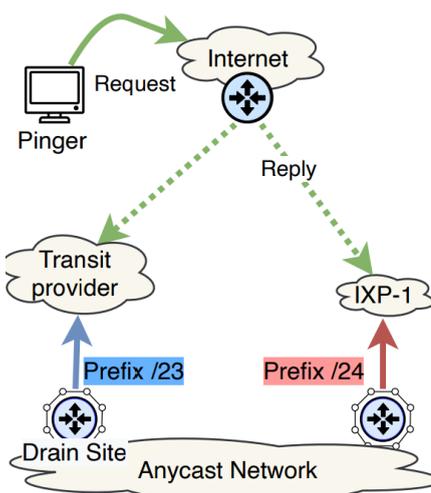


Figura 3. Método de Sondagem Ativa com Sites *Anycast*

Fonte: Bertholdo, et. al (2021)

Esse método foi utilizado para avaliar os IXPs em termos de cobertura e representatividade regional e também para mostrar como o tráfego deve fluir se porventura um IXP fique indisponível, pois, essa informação deve ajudar as operadoras a se prepararem para lidar com estes possíveis eventos. É importante salientar que outras ferramentas já conhecidas no mercado também foram consideradas para realizar essa captura dos dados, tais como Sflow⁸ e Netflow⁹ que apesar de possuírem dados bastante completos, deixaram de ser opções devido ao difícil acesso aos dados, pois necessitam da cooperação dos IXPs e além disso não poderia ser aplicada em uma visão global por conta de leis de privacidade. (BERTHOLDO, et al, 2021)

⁸ Disponível em: <https://sflow.org>

⁹ Disponível em: https://www.cisco.com/c/pt_br/tech/quality-of-service-qos/netflow/index.html

O método sondagem ativa com sites *anycast* foi implementado em oito dos dez maiores IXPs do mundo, a exceção foi o IX.br/RS, de Porto Alegre (RS), que mesmo não possuindo uma grande expressividade de troca de tráfego, colaborou com o projeto e foi utilizado para fins de validação. O Quadro 1 apresenta a lista dos cinco IXPs analisados no presente estudo, com informações complementares como seu website e a sua velocidade de troca de tráfego por segundo.

Quadro 1: IXPs contemplados neste estudo

Nome do IXP	Tráfego	Endereço
IX.br/SP	15 Tbps	www.ix.br
AMS-IX	11 Tbps	ams-ix.net
LINX	7 Tbps	www.linx.net
SIX	2 Tbps	seattleix.nrt
IX.br/RS	0,5 Tbps	www.ix.br

Fonte: PeeringDB (Maio de 2022)

6 Desenvolvimento da Solução

Para o desenvolvimento da solução foram utilizadas diversas ferramentas, a seguir estão elencadas as que desempenharam os papéis mais importantes:

- **Pentaho Data Integration**

O Pentaho Data Integration é um componente da suíte Pentaho usado para criar processos de extração, transformação e carga (do inglês *Extraction, Transformation and Loading*, ETL) que alimentam os bancos de dados (COSTA, 2017). Em nosso trabalho, foi utilizado o SGBD relacional MySQL.¹⁰

- **Microsoft Power BI**

O Microsoft Power BI é uma coleção de serviços de softwares, aplicativos e conectores que trabalham juntos para transformar suas fontes de dados não relacionadas em informações coerentes, visualmente envolventes e interativas.¹¹

6.1 EXTRAÇÃO, TRANSFORMAÇÃO E CARREGAMENTO DOS DADOS

Para a prova de conceito da solução desenvolvida no presente artigo, o grupo de pesquisa da universidade de twente disponibilizou uma série de arquivos, coletados no período de julho a outubro de 2021, sendo:

- 60 arquivos com dados de troca de pacotes, coletados nos dias 20/09/2021 e 05/10/2021, em cinco diferentes IXPs, totalizando 18,4 GB e 42 milhões de registros.
- 02 arquivos chamados de *hitlist*, coletados nos dias 30/07/2021 e 03/08/2021, onde estão discriminadas as melhores rotas para cada um dos 7,8 milhões de IPs mapeados que realizou conexão através do IXP, podendo ser constatado se existe uma rota específica ou somente rota padrão.

¹⁰ Disponível em: <https://downloads.mysql.com/docs/refman-4.1-pt.pdf>

¹¹ Disponível em: <https://docs.microsoft.com/pt-br/power-bi/fundamentals/power-bi-overview>

Além dos arquivos disponibilizados pelo DACS, também foi necessário realizar o levantamento da relação das listas de ASes conectados em cada um dos IXPs estudados. As coletas foram realizadas em Fevereiro de 2022 através do website de cada IXP, e foram obtidos os resultados apresentados no Quadro 2.

Quadro 2: ASes Participantes em cada IXP

Nome do IXP	Ases Conectados	Endereço
IX.br/SP	2384	https://ix.br/particip/sp
AMS-IX	804	https://www.ams-ix.net/ams/connected-networks
LINX	733	https://portal.linx.net/members/members-ip-asn
SIX	337	https://www.seattleix.net/participants/
IX.br/RS	302	https://ix.br/particip/rs

Fonte: Elaboração pelo Autor

Após este levantamento, ocorre o processo de ETL, que lê o conjunto de dados que foi previamente organizado em pastas agrupadas por data de amostra a fim de facilitar a leitura.

Como existem diferentes tipos de arquivos, com diferentes estruturas, foram criados diferentes processos de leitura dos dados, cada um com uma particularidade para atender todas as possíveis estruturas de arquivos. Nestes processos, os dados passam por algumas etapas de transformações, onde são realizados ajustes e enriquecimento da base, como por exemplo definições de tipos de dados, criação de novos campos e criação de colunas auxiliares para posterior armazenamento no banco de dados. Subsequente a este processo, são gerados os arquivos com os dados tratados e sintetizados, cada um com a sua determinada regra de negócio, conforme ilustrado na Figura 4.

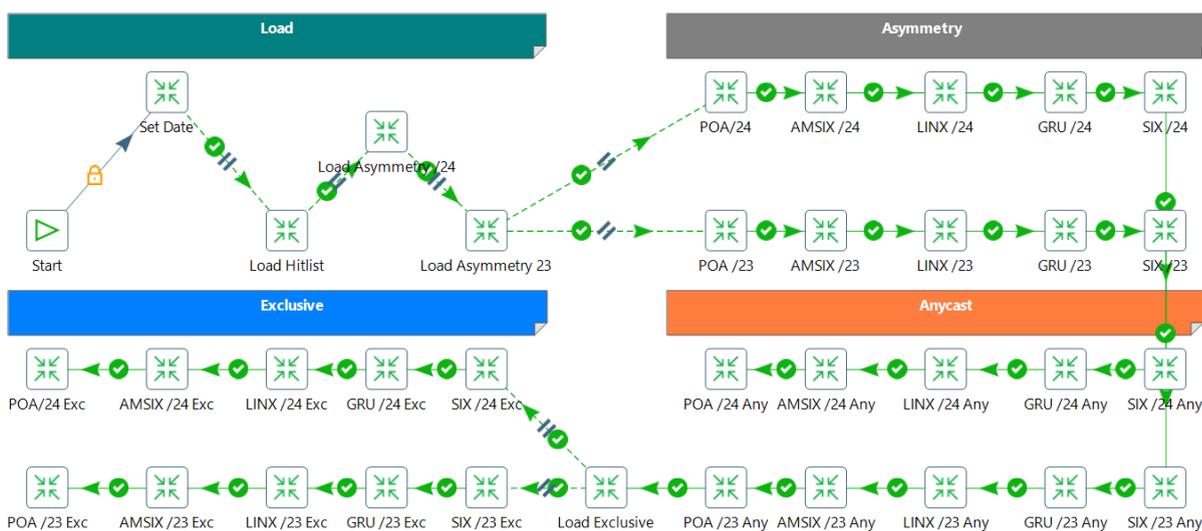


Figura 4. Processo de ETL com Pentaho Data Integration

Fonte: Elaboração pelo Autor

Essa arquitetura foi desenvolvida para que o processo de leitura, transformação e armazenamento de dados possa ser realizado sob demanda, garantindo que o período de análise ou a quantidade de IXPs analisados possam ser expandidos.

6.2 CLASSIFICAÇÃO DOS RESULTADOS

Para a verificação da Simetria dos ASes e de suas redes foram criadas três classificações:

- **Simétrico:** São determinados como sendo do tipo Simétrico todas as redes em que os pacotes foram coletados pelo site *anycast* que está dentro do IXP e que o AS possui uma rota específica na sua tabela BGP para a rede em questão, isto é, o tráfego entrou e saiu do IXP por uma única rota.
- **Assimétrico de Entrada:** São determinados como sendo do tipo assimétrico de entrada todos os pacotes em que o AS recebeu o tráfego a partir da Internet mas respondeu pelo IXP.
- **Assimétricos de Saída:** São determinados como sendo do tipo assimétrico de saída todos os pacotes em que o AS recebeu o tráfego a partir do IXP porém respondeu pela Internet.

7 Resultados

Os resultados são apresentados por meio de um *Dashboard*, permitindo a interação dos usuários com os dados, por meio de filtros e a criação de novas visões de forma ágil, graças ao conceito de *Self Service BI*¹² utilizado na ferramenta Microsoft Power BI que visa dar autonomia para o usuário trabalhar com os dados.

Na página inicial do *Dashboard* é exibida uma tabela que foi intitulada como “*result table*” onde constam os resultados obtidos a partir de cada amostra analisada e seu principal objetivo é servir de insumo para análises posteriores realizadas pelo DACS.

Para análise dos resultados, foram criadas interfaces que possuem uma gama de filtros, permitindo que o usuário possa visualizar os resultados nos três métodos que estão sendo analisados, *Asymmetry*, *Anycast* e *Exclusive*. As análises de *Anycast* e *Exclusive* não serão abordadas no presente artigo.

Na Figura 5 temos a representação da segunda página do *dashboard*, onde os dados são apresentados por meio de tabelas, agrupados por data da amostra, IXP, sub-rede e pelas categorias previamente explicadas na seção 6.2. Na parte inferior estão os resultados que dizem respeito ao comportamento dos ASes, e na parte superior está uma visão a nível das redes destes ASes.

¹² Afinal, o que é self service BI? <https://farolbi.com.br/afinal-o-que-e-self-service-bi/>. Acesso em 27/12/2021

Resumo por Categoria			
% de Redes por Tipo			
Data Amostra	% Redes Simétricas	% Assimetria de Entrada	% Assimetria de Saída
<input type="checkbox"/> 05/05			
<input type="checkbox"/> br-gru			
/23	84,74%	0,82%	14,44%
/24	86,20%	1,71%	12,10%
<input type="checkbox"/> us-sea			
/23	90,24%	7,02%	2,74%
/24	90,47%	8,19%	1,34%

% de ASes por Tipo				
Data Amostra	Assimétrico - Só Entrada	Assimétrico - Só Saída	Sem Redes na Amostra	Simétrico
<input type="checkbox"/> 05/05				
<input type="checkbox"/> br-gru				
/23	0,87%	9,51%	12,79%	76,82%
/24	1,20%	6,28%	12,84%	79,67%
<input type="checkbox"/> us-sea				
/23	10,26%	11,28%	8,72%	69,74%
/24	9,95%	5,76%	8,90%	75,39%

Figura 5. Resumo por Categoria - Método *Asymmetry*

Fonte: Elaboração pelo Autor

Os resultados de forma mais simplificada são mostrados na terceira página do *dashboard* por meio de um gráfico de linhas agrupadas pela data da amostra, Sub-Rede e IXP (Figura 6) tem por objetivo exibir o percentual de simetria das redes de cada um dos IXPs.

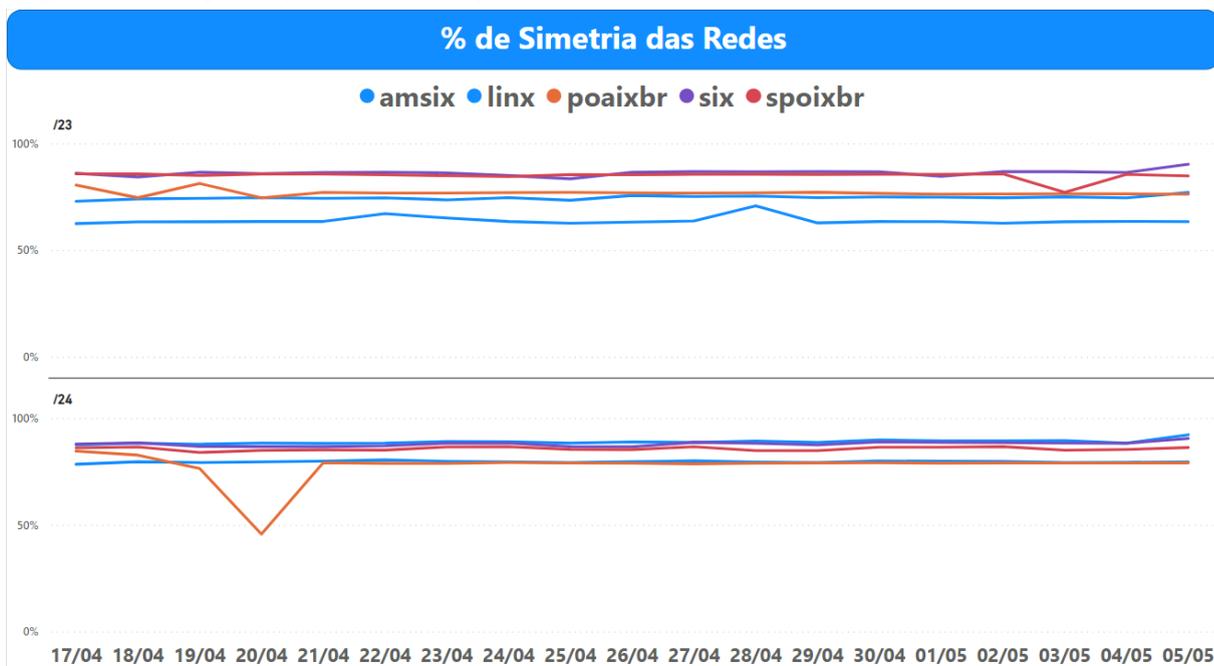


Figura 6. % de Redes Simétricas por IXP

Fonte: Elaboração pelo Autor

Na quarta página do *dashboard* é possível observar o total de ASes e de suas redes em cada uma das classificações, de forma gráfica e analítica (Figura 7), possibilitando acompanhar como cada AS se comportou ao longo das amostras estudadas.



Figura 7. Classificação dos ASes

Fonte: Elaboração pelo Autor

Com base nos dados fornecidos, ainda seria possível criar outros gráficos e tabelas, em níveis ainda mais detalhados, porém essas visualizações não foram apontadas como necessárias pelo DACS.

8 Considerações Finais

O desenvolvimento do ETL, a definição dos indicadores e o desenvolvimento do *Dashboard* proposto neste trabalho estão estruturados e em pleno funcionamento. Após várias reformulações na ferramenta, foi possível chegar em uma solução que executa o ciclo citado na seção 4, fornecendo os resultados para três diferentes classificações de simetria de tráfego, olhando os dados dos cinco IXPs analisados.

Durante o processo de desenvolvimento foi possível identificar anomalias nos dados de alguns IXPs e após investigação por meio do *dashboard* e do bancos de dados que foram criados neste estudo, foi possível elencar os pontos anômalos e direcionar ao DACS para que melhorias fossem aplicadas na captura dos dados, o que contribuiu para a evolução do método de sondagem ativa *anycast*.

Em relação aos resultados obtidos, foi possível observar conforme apresentado na Figura 8, que o percentual de simetria é 7,5% maior nas redes /24 em comparação às redes /23, isso é, redes mais específicas tendem a ter tráfego mais simétrico. Os resultados

demonstram que aproximadamente 20% do tráfego da Internet não é simétrico, podendo ser um ponto de melhoria para os provedores de conteúdo, que entregam o conteúdo a seus usuários por caminhos diferentes do solicitado.

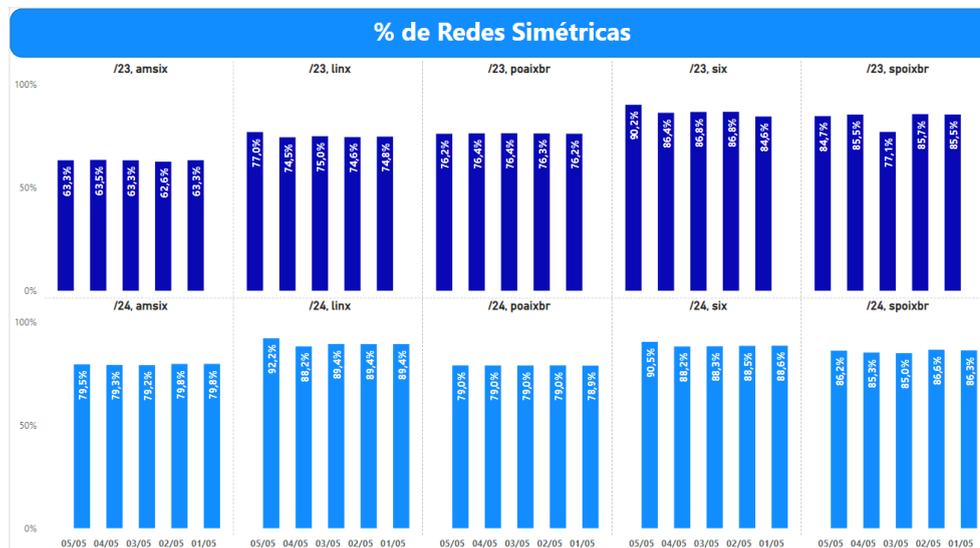


Figura 8. Simetria de Redes /23 e /24

Fonte: Elaboração pelo Autor

Para geração dos resultados apresentados até o momento, foram processados aproximadamente 250 Gigabytes com mais de 1,3 bilhões de registros. No estudo anterior, o grupo DACS processava os resultados um dia de coleta em aproximadamente 24 horas, e agora com essa implementação, foi possível mapear todo o processo de uma só vez, evitando falhas, e com o tempo reduzido para 5 horas, possibilitando que os resultados do mapeamento de simetria sejam disponibilizados logo após coleta dos dados, com melhores recursos de cruzamento de dados e resultados mais robustos.

Os resultados obtidos a partir do presente estudo também foram utilizados no desenvolvimento de um novo artigo intitulado ‘*On the Asymmetry of Internet eXchange Points - What You See Is Not Always What You Get*’ organizado pelo grupo de pesquisa da universidade de twente, onde nele foi investigado a simetria do tráfego sob a perspectiva dos IXPs, o qual foi submetido ao CNSM2022 (*18th International Conference on Network and Service Management*).

Referências

APNIC. Critical Infrastructure. 2021. Disponível em: <<https://help.apnic.net/s/article/Critical-Infrastructure#are-ixp-critical-infrastructure>> Acesso em: 12 dez. de 2021.

BERTHOLDO, Leandro. et al. **Forecasting the Impact of IXP Outages Using Anycast**. 978-3-903176-27-0 © 2021 IFIP. Acesso em: 23 out. de 2021.

BRITO, S. H. B., Santos, M. A. S., dos Reis Fontes, R., Perez, D. A. L., and Rothenberg, C. E. (2015). Anatomia do ecossistema de pontos de troca de tráfego públicos na internet do Brasil. XXXIII Simpósio Brasileiro de Redes de Computadores (SBRC).

CANDELA, Massimo. et al. **Impact of the covid-19 pandemic on the internet latency: A large-scale study**. Computer Networks. vol. 182, p. 107495, 2020. Acesso em: 10 dez. de 2021.

COSTA, Marcelo; DE SALLES, Fábio. Pentaho Data Integration - ETL em Software Livre. **INFOQ**. 2017. Disponível em: <<https://www.infoq.com/br/articles/pentaho-pdi/>>. Acesso em: 06 dez. 2021.

FERREIRA, Joaquim. O que é um PTT (e qual é exatamente sua importância)?. **Eletronet**. 2020. Disponível em: <<https://www.eletronet.com/o-que-e-um-ptt-e-qual-e-exatamente-sua-importancia/>>. Acesso em: 25 out. 2021.

FERREIRA, João; et al. **O Processo ETL em Sistemas Data Warehouse**. 2010. Universidade do Minho. Departamento de Informática. Disponível em: <<http://inforum.org.pt/INForum2010/papers/sistemas-inteligentes/Paper080.pdf>>. Acesso em: 22 nov 2021.

FERREIRA, hugo. et. al. **Encaminhamento anycast em redes IPv6 : uma proposta**, 2012. Acesso em: 21 dez. de 2021.

HOESCHELE, Thomas. et al. **Importance of internet exchange point (ixp) infrastructure for 5g: Estimating the impact of 5g use cases**. Telecommunications Policy, vol. 45, no. 3, p. 102091, 2021.

KIMBALL, Ralph; ROSS, Margy. The Data Warehouse Toolkit: the definitive guide to dimensional modeling. 3. ed. Indianapolis: Wiley, 2013. Acesso em: 03 dez. 2021.

NOMIKOS, G.; et al, X. Dimitropoulos, and V. Giotsas, “O peer, where art thou? uncovering remote peering interconnections at ixps,” in Proceedings of the Internet Measurement Conference 2018, 2018, pp. 265–278

NICBR. Como funciona a Internet? Parte 1: O protocolo IP. NICBrrvídeos. **Youtube**. 2014. 5min44s. Disponível em: <<https://www.youtube.com/watch?v=HNQD0qJ0TC4>>. Acesso em: 23 out. de 2021.

NICBR. Como funciona a Internet? Parte 2: Sistemas Autônomos, BGP, PTTs. NICbrvideos. Youtube. 2014. 7min36s. Disponível em: <https://www.youtube.com/watch?v=C5qNAT_j63M>. Acesso em: 23 out. de 2021.

PRESSMAN, Roger S. Software engineering: a practitioner's approach. Palgrave macmillan, 2005. Acesso em: 22 jun. 2022.

SHARDA, Ramesh; et. al. Business Intelligence, Analytics, and Data Science: a managerial perspective. 4. ed. Boston: Pearson, 2016. Acesso em: 03 dez. 2021.

SHERMAN, Rick. Business Intelligence Guidebook: from data integration to analytics. San Francisco: Morgan Kaufmann Publishers Inc., 2014. Acesso em: 03 dez. 2021.

SILVA, André. BGP - Border Gateway Protocol. Devmedia. 2013. Disponível em <<https://www.devmedia.com.br/bgp-border-gateway-protocol/29123>>. Acesso em: 25 out. 2021.

W. B. de Vries, R. de O. Schmidt, W. Hardaker, J. Heidemann, P.-T. de Boer, and A. Pras, “Broad and Load-Aware Anycast Mapping with Verfploeter,” in Proceedings of the 2017 Internet Measurement Conference, ser. IMC '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 477–488. [Online]. Disponível em: <<https://doi.org/10.1145/3131365.3131371>> . Acesso em: 30 abr. 2022.

WERMANN, A. G. (2016). Dynamix: um mercado de acordos dinâmicos em pontos de troca de tráfego. Acesso em: 24 nov. de 2021.

WEXLER, Steve; et. al. The Big Book of Dashboards. Hoboken: John Wiley & Sons, Inc, 2017. Acesso em: 03 dez., 2021