

MAYNA HELENA AZEVEDO

**EXPLORANDO UM FILE SERVER COMPARTILHADO COM BI E
SOCIAL NETWORK ANALYSIS**

CAMPINAS

2023

MAYNA HELENA AZEVEDO

**EXPLORANDO UM FILE SERVER COMPARTILHADO COM BI E SOCIAL
NETWORK ANALYSIS**

Trabalho de Conclusão de Curso apresentado como parte dos requisitos para obtenção do diploma do Curso Tecnologia em Análise e Desenvolvimento de Sistemas do Instituto Federal de Educação, Ciência e Tecnologia Campus Campinas.

Orientador: Prof. Dr. Andreiuid Sheffer Correa.

CAMPINAS

2023

Ficha Catalográfica
Instituto Federal de São Paulo – Campus Campinas
Biblioteca – Pedro Augusto Pinheiro Fantinatti
Rosângela Gomes – CRB 8/8461

Azevedo, Mayna Helena
A994e Explorando um file server compartilhado com BI e social network analysis./
Mayna Helena Azevedo. – Campinas, SP: [s.n.], 2023.
46 f. : il.

Orientador: Andreiwid Sheffer Corrêa.
Trabalho de Conclusão de Curso (graduação) – Instituto Federal de Educação, Ciência e Tecnologia de São Paulo - Campus Campinas. Curso de Tecnologia em Análise e Desenvolvimento de Sistemas, 2023.

1. SNA (Arquitetura de rede de computador). 2. Business intelligence. 3. Computação em nuvem. 4. Servidor de arquivos. 5. Python. I. Instituto Federal de Educação, Ciência e Tecnologia de São Paulo Campus Campinas. Curso de Tecnologia em Análise e Desenvolvimento de Sistemas. II. Título.

ATA N.º 11/2023 - TADS-CMP/DAE-CMP/DRG/CMP/IFSP

Ata de Defesa de Trabalho de Conclusão de Curso - Graduação

Na presente data, realizou-se a sessão pública de defesa do Trabalho de Conclusão de Curso intitulado **EXPLORANDO UM FILE SERVER COMPARTILHADO COM BI E SOCIAL NETWORK ANALYSIS**, apresentado pela aluna **MAYNA HELENA AZEVEDO (CP3010015)** do Curso **SUPERIOR DE TECNOLOGIA EM ANÁLISE E DESENVOLVIMENTO DE SISTEMAS** (Campus Campinas). Os trabalhos foram iniciados às **18h00** pelo Professor presidente da banca examinadora, constituída pelos seguintes membros:

Membros	Instituição	Presença (Sim/Não)
ANDREIWID SHEFFER CORRÊA (Presidente/Orientador)	IFSP-CMP	SIM
ZADY CASTANEDA SALAZAR (Examinador 1)	IFSP-CMP	SIM
JOSÉ AMÉRICO DOS SANTOS MENDONÇA (Examinador 2)	IFSP-CMP	SIM

Observações:

A banca examinadora, tendo terminado a apresentação do conteúdo da monografia, passou à arguição da candidata. Em seguida, os examinadores reuniram-se para avaliação e deram o parecer final sobre o trabalho apresentado pela aluna, tendo sido atribuído o seguinte resultado:

☒ [X] Aprovado(a) ☐ [] Reprovado(a)

Proclamados os resultados pelo presidente da banca examinadora, foram encerrados os trabalhos e, para constar, eu lavrei a presente ata que assino em nome dos demais membros da banca examinadora.

Campus Campinas, 10 de novembro de 2023

Documento assinado eletronicamente por:

- Andreiwid Sheffer Correa, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 10/11/2023 18:43:49.
- Jose Americo dos Santos Mendonca, PROFESSOR ENS BASICO TECN TECNOLOGICO, em 10/11/2023 18:44:29.
- Zady Castaneda Salazar, COORDENADOR(A) - FUC1 - TADS-CMP, em 10/11/2023 18:44:37.

Este documento foi emitido pelo SUAP em 09/11/2023. Para comprovar sua autenticidade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifsp.edu.br/autenticar-documento/> e forneça os dados abaixo:

Código Verificador: 646126
Código de Autenticação: b4be5d7ff6



" Todas as famílias felizes se parecem, cada família infeliz é infeliz à sua maneira."

Anna Kariênina

Léon Tolstoi

RESUMO

Estudos sobre a migração de arquivos de servidores “on premise” para a nuvem proporcionam uma maneira segura e eficiente de transferir informações para empresas de diferentes tamanhos, incluindo micro, pequenas, médias e grandes empresas. Ferramentas de BI, R e Python auxiliam no tratamento de grande volume de dados na tomada de decisão e extração de informação, que auxiliam a equipe de Tecnologia da Informação durante a migração. O objetivo deste trabalho é explorar modelos de “Business Intelligence” para análise exploratória dos arquivos que foram adicionados, modificados, movidos, renomeados e excluídos em um servidor “on premise”. Em complemento, identificar o comportamento da troca de informação entre os departamentos da empresa, utilizando “Social Network Analysis”. Foram selecionados dados diários de julho a dezembro de 2022, fornecidos por uma empresa terceirizada durante o monitoramento dos arquivos manipulados no servidor “on premise”. A análise conjunta com “Social Network Analysis” permitiu identificar departamentos, horários e pastas com maior volume de manipulação nos dados. Ao final, foi sugerido um plano de migração dos dados do servidor “on premise” para a nuvem.

Palavras-chave: SNA; BI; On Premise; File Server; Python.

ABSTRACT

Studies on the migration of on-premise server files to the cloud provide a secure and efficient way to transfer information for companies of different sizes, including micro, small, medium, and large enterprises. Business Intelligence (BI), R, and Python tools assist in handling large volumes of data in decision-making and information extraction, aiding the Information Technology team during the migration process. This work aims to explore Business Intelligence models for exploratory analysis of files that have been added, modified, moved, renamed, and deleted on an on-premise server. Additionally, to identify the information exchange behaviour between the company's departments using Social Network Analysis (SNA). Daily data was selected daily data from July to December 2022, provided by a third-party company during the monitoring of files manipulated on the on-premise server. The joint analysis with Social Network Analysis allowed for the identification of departments, timeframes, and folders with the highest data manipulation volume. Finally, a data migration plan from the on-premise server to the cloud was suggested.

Keywords: SNA; BI; On Premise; File Server; Python.

SUMÁRIO

1 INTRODUÇÃO	8
1.1 JUSTIFICATIVA	9
1.2 OBJETIVO	10
1.2.1 Objetivo Geral	10
1.2.2 Objetivos Específicos	10
2 FUNDAMENTAÇÃO TEÓRICA	11
2.1 SOCIAL NETWORK ANALYSIS	11
2.2 ON PREMISE	13
2.3 COMPUTAÇÃO EM NUVEM	14
2.4 BUSINESS INTELLIGENCE	15
3 METODOLOGIA	16
3.1 MATERIAIS	16
3.1.1 Social Network Analysis	16
3.1.2 Business Intelligence	17
3.2 MÉTODOS	18
3.2.1 Análises SNA	19
3.2.2 Análises BI	20
4 RESULTADOS	23
4.1 ANÁLISE DE SOCIAL NETWORK ANALYSIS	23
4.2 ANÁLISE DE BUSINESS INTELLIGENCE	27
5 CONSIDERAÇÕES FINAIS	32
REFERÊNCIAS	34
APÊNDICE A – ENTREVISTA REALIZADA COM OS FUNCIONÁRIOS	37

1 INTRODUÇÃO

Nas empresas, independentemente do porte, é comum compartilhar arquivos internamente e, em alguns casos, com clientes e outras empresas. Esses arquivos podem conter diversos formatos, como planilhas, textos, imagens, vídeos e bancos de dados, com diferentes níveis de segurança. Com a evolução da tecnologia e o uso cada vez mais frequente de equipamentos eletrônicos, a forma como criamos e gerenciamos informações vem mudando significativamente. Embora a tendência seja abandonar o papel físico e adotar o mundo digital, gerenciar documentos eletrônicos pode ser trabalhoso e requer profissionais de TI especializados na gestão de arquivos.

Felizmente, plataformas como Microsoft e Google oferecem maneiras rápidas e fáceis de criar e gerenciar arquivos na nuvem, através de serviços como Azure, OneDrive, Drive, Google Cloud e AWS. O armazenamento em nuvem permite que o cliente acesse e gerencie seus arquivos de qualquer lugar, com acesso à internet e equipamentos pessoais.

A nuvem está cada vez mais popular devido à facilidade de gerenciamento dos arquivos e aos preços acessíveis. Além disso, as empresas não precisam investir em equipamentos físicos de armazenamento, como servidores. É importante destacar que o uso da nuvem requer profissionais especializados para garantir uma gestão eficiente dos documentos e serviços.

No entanto, ainda existem muitas empresas que valorizam o uso "on premise", especialmente aquelas que precisam manter informações altamente sigilosas em instalações com serviços e arquivos altamente confidenciais (Kaseb et al., 2018). Para empresas mais tradicionais, é necessário adquirir servidores, switches, firewalls e outros equipamentos e softwares para criar uma rede para todos os funcionários, muitas vezes utilizando sistemas operacionais como Windows ou Linux (Diamond, 2020). O compartilhamento de arquivos e dados é feito por meio de dispositivos conectados à mesma rede.

O uso "on premise" permite que a equipe de TI centralize as informações e crie regras de acesso restrito para usuários e departamentos. No entanto, o uso "on premise" também tem desvantagens, como o alto custo de equipamentos como servidores e switches, a necessidade de espaço de instalação próprio (com controle de temperatura e umidade) e energia elétrica 24 horas, além da compra de licenças de equipamentos de fornecedores como Dell e Cisco (Magalhães, 2020).

Para empresas que utilizam "on premise", cabe ao departamento de TI gerenciar o acesso às pastas e arquivos internos da empresa, bem como monitorar o comportamento dos

usuários ao manipular esses documentos para identificar qualquer má utilização ou comportamento malicioso.

O objetivo deste estudo é realizar um estudo de caso sobre o gerenciamento de dados de uma empresa que utiliza um servidor de arquivos físico e planeja migrar para a plataforma Microsoft SharePoint. A dificuldade em migrar se deve à grande quantidade de dados "on premise", com mais de 100 mil arquivos em quase 12 terabytes, que são modificados constantemente, 24 horas por dia, 7 dias por semana.

O presente estudo utiliza o modelo de "Business Intelligence" (BI) para análise exploratória dos dados de adição, modificação, movimentação, renomeação e exclusão no servidor de arquivos. Além disso, a análise de "Social Network Analysis" (SNA) foi usada para estudar o comportamento do uso do servidor de arquivos por departamento da empresa.

A SNA estuda relacionamentos entre pessoas, partes relacionadas ou situações, onde a prioridade da análise é o próprio relacionamento. Sua função é conceitualizar estruturas sociais em rede conectadas entre si. Com isso, é possível entender a estrutura e as consequências dos relacionamentos.

As tecnologias usadas para o tratamento dos dados foram Power BI da Microsoft e R Studio. Power BI é uma plataforma criada pela Microsoft focada em Business Intelligence, capaz de conectar-se facilmente aos dados e criar relatórios personalizados.

1.1 JUSTIFICATIVA

Com o avanço tecnológico, e com os efeitos pós pandemia, a informação digital assumiu uma grande importância para o mundo corporativo. Plataformas de fácil acesso, alta velocidade e intuitivas estão popularizando-se, principalmente em grandes organizações, estimulando e promovendo o trabalho dinâmico entre todos os setores. Muitas empresas ainda utilizam "on premise", conhecidos como "File Server", que tendem a ter velocidade inferior a uma plataforma online quando Acesso por VPN (para funcionários que trabalham em Home Office), e com diversos riscos, como interrupção de acesso, arquivos corrompidos pelo servidor ou equipamento de armazenamento queimado. Além disso, manter uma instalação com formato "on premise" é caro e trabalhoso.

Manter servidores "on premise" exige investimento de hardware, software, manutenção e equipe de TI capacitada para gerenciar e manter a estrutura. O local deve ser refrigerado e

sem umidade, para evitar danos aos equipamentos. Já com o cloud, as empresas não se preocupam com a manutenção dos equipamentos.

Além disso, provedores cloud oferecem garantia de disponibilidade, além de acesso a aplicativos e arquivos em qualquer lugar do mundo, desde que exista conexão à internet. Para empresas onde os funcionários trabalham remotamente ou de forma híbrida, torna-se vantajoso.

Diversas plataformas surgiram com o intuito de gerenciar arquivos corporativos online, desde os mais básicos, como arquivos de texto (em Word, Pdf), planilhas eletrônicas (Excel), imagens, até os mais complexos, como bancos de dados. A plataforma SharePoint, da Microsoft, foi a plataforma escolhida pela Empresa analisada para gerenciar os conteúdos online. Por questão de confidencialidade, a companhia será referenciada como “Empresa”.

O SharePoint permite gerenciar os acessos e permissões pela intranet, e seu uso pode ser feito na versão web ou em aplicativo desktop. Além disso, os usuários podem gerar grupos internos de trabalho, personalizando com ferramentas como Power Apps e Power Automate (Microsoft, 2023a).

1.2 OBJETIVO

As subseções a seguir descrevem os objetivos gerais e específicos do presente trabalho.

1.2.1 Objetivo Geral

O objetivo deste estudo de caso é desenvolver um plano de migração dos arquivos do “File Server on premise” para cloud da plataforma SharePoint da Microsoft.

1.2.2 Objetivos Específicos

- Levantar dos dados disponíveis para o “File Server” em estudo;
- Analisar os dados de forma exploratória;
- Desenvolvimento de um modelo de BI visando melhores formas de visualização de dados;
- Analisar os dados em séries temporais;
- Mostrar o relacionamento entre os departamentos da Empresa utilizando SNA.

2 FUNDAMENTAÇÃO TEÓRICA

Nas seções seguintes, serão apresentadas as fundamentações teóricas que sustentam este estudo. Serão abordados conceitos como “Social Network Analysis”, “On Premise”, Computação em Nuvem e “Business Intelligence”.

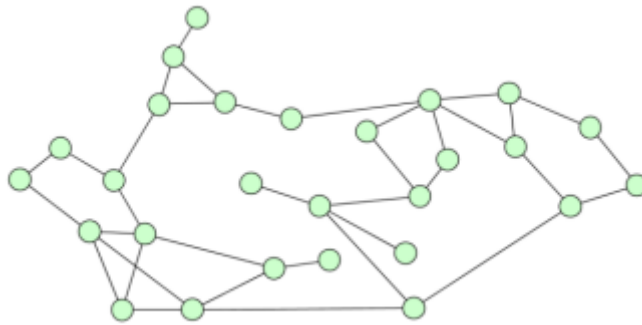
2.1 SOCIAL NETWORK ANALYSIS

“Social Network Analysis” (SNA) é uma metodologia que estuda as relações sociais entre indivíduos, grupos ou organizações (Otte; Rousseau, 2002). Sua função é conceitualizar estruturas sociais em rede conectadas entre si. Com isso, é possível entender a estrutura e as consequências dos relacionamentos, identificando padrões e estruturas. Suas áreas de atuação são inúmeras, sendo aplicadas nas próprias redes sociais, como Facebook e Instagram, sociologia, antropologia, ciências políticas etc. É comumente usada para identificar grupos com as mesmas características ou interesses e identificar a difusão de informações (Otte; Rousseau, 2002).

Sua principal estrutura contém nodos, laços e estruturas. Nodos podem ser qualquer indivíduo na rede social estudada, variando entre pessoa (ou animais), grupo ou organização. Cada nodo é considerado como uma unidade distinta dentro da rede estudada, criam conexões com outro nodo (Marteleto, 2001), e contém suas próprias características. Um exemplo de nodo são pessoas dentro de uma rede social, como o Facebook, ou um cão, dentro de um “pet shop”.

Quando um nodo se relaciona um com outro, eles criam laços entre si. Cada nodo pode conter laços entre um ou mais nodos, que podem variar em diferentes tipos de intensidade (Butts, 2008). Os laços podem conter direção ou não, dependendo dos dados estudados.

Ao se ter nodos e laços, cria-se uma estrutura na rede, que possui padrões entre as conexões (Butts, 2008). Frequentemente, as estruturas são exemplificadas através de grafos, como mostra a Figura 1. Cada círculo simboliza o nodo, e a conexão entre os nodos são os laços.

Figura 1 – Grafo.

Fonte: Borba (2013)

Existem diversos parâmetros de análise dentro do SNA, mas para simplificação desse trabalho, foram abordados apenas grau e centralidade de proximidade (em inglês, “closeness”). Grau representa o número total de laços que um nodo contém na estrutura. Nodos com maior quantidade de laços tendem a ter maior influência na rede (Otte; Rousseau, 2002).

A centralidade de proximidade mede o quanto um específico nodo é conectado com os outros nodos e é um indicador importante para informar os nodos mais relevantes para espalhar informações de forma eficiente, como se comunica (Otte; Rousseau, 2002). Para calcular a centralidade de proximidade de um nodo, realiza-se a inversa da soma das distâncias geodésicas. Seu resultado varia entre 0 e 1, e quanto mais próximo de 1, maior é a centralidade de proximidade do nodo. Portanto, quanto menor a distância, maior a centralidade. Nodos com alta centralidade são capazes de acessar e disseminar informação com maior facilidade (Grunspan et al., 2014). Veja abaixo a equação.

$$CP_i = \frac{1}{\sum_i d_{ij}}$$

onde:

CP_i : centralidade de proximidade do nodo i

d_{ij} : distância geodésica entre os nodos i e j

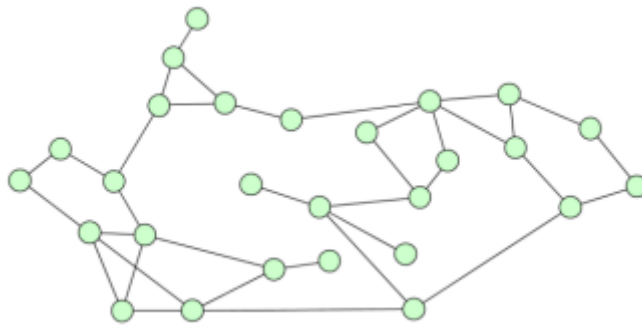
Esse cálculo é computacionalmente custoso, principalmente para grandes redes. Entretanto, como o presente trabalho contém uma rede pequena, essa restrição não é uma limitação.

Outra análise comumente usada em SNA é a de comunidade, definido como um conjunto de nodos próximos entre si, mais densamente conectados entre si do que com o restante da rede (Huberman; Adamic, 2003). Esse aglomerado que os nodos criam são muitas vezes chamados de “cluster”, ou em português, clusterização. Essas comunidades indicam

características semelhantes entre os nodos, como localização geográfica, relações, atividades compartilhadas, etc. (Recuero, 2006). As comunidades podem ser detectadas por diferentes métodos, mas para facilitação, foi selecionado o Método de Louvain, muito comum na literatura.

O Método de Louvain é um algoritmo de detecção de comunidades em redes, que usa como técnica a otimização da modularidade da rede (Neo4j, 2023a). O algoritmo age em duas etapas, de forma hierárquica e iterativa, até que todos os nodos sejam classificados nas comunidades criadas (Aires; Nakamura, 2017). Como resultado, pode-se gerar grafos, onde os nodos são distintos por cor, e cada cor representa uma comunidade.

Figura 2 – Grafo aplicado o Método de Louvain.



Fonte: Aires e Nakamura (2017)

2.2 ON PREMISE

Atualmente, ainda é muito comum empresas manterem suas aplicações, arquivos e bancos de dados em servidores físicos locais, conhecidos como “on premise” (Osti; Pereira, 2021). As empresas que optam por terem servidores “on premise” responsabilizam-se por comprar equipamentos como servidores, switches, firewall, entre outros. Esses equipamentos são caros, requerem profissionais especializados, e além de tudo, necessitam de local com controle de temperatura e umidade, para evitar danos, quedas de energia, incêndio e falhas aos equipamentos.

Todo serviço, para funcionar, necessita de uma infraestrutura “on premise”, por mais que não seja própria do usuário (Abreu, 2022). Por exemplo, ao acessar plataformas de “streaming”, necessita-se apenas de internet, navegador ou aplicativos. Entretanto, por mais que o usuário não tenha uma instalação própria com servidores, cabos, etc., a empresa fornecedora do serviço contém uma infraestrutura local completa e complexa, com

computadores criados especialmente para esse serviço (Abreu, 2022). Qualquer empresa que disponibiliza um produto ou plataforma online, utiliza uma infraestrutura computacional, seja direta ou indiretamente.

Empresas que necessitam de controle total sobre seus dados e sistemas optam pelo modelo “on premise”, e.g.: bancos, instituições financeiras e governamentais e empresas com fábricas. Dependendo do serviço prestado, algumas empresas sofrem com regulamentos rígidos de segurança e privacidade, o que justifica criar uma estrutura computacional “on premise”, ao invés da nuvem (Abreu, 2022).

Resumindo, existem diversas vantagens e desvantagens nessa infraestrutura (Abreu, 2022). As vantagens são: privacidade das informações e sistemas; customização dos serviços e segurança. Entre suas desvantagens, estão: alto custo para montar e manter uma instalação; orçamento disponível para manutenção e aquisições inesperadas (e.g., novos equipamentos) e equipe técnica preparada para qualquer imprevisto.

2.3 COMPUTAÇÃO EM NUVEM

Nos últimos anos, com o aumento da computação em nuvem, muitas empresas estão migrando para modelos de hospedagem de software em nuvem, que oferecem maior escalabilidade, flexibilidade e eficiência de custos (Osti; Pereira, 2021). Diferentemente da hospedagem “on premise”, o armazenamento e processamento em nuvem é feito externamente. Dessa forma, não é necessário investir em equipamentos e espaço físico.

Na nuvem, os dados e serviços podem ser acessados e gerenciados pela internet. E por estar utilizando um serviço de terceiro, toda a infraestrutura e segurança fica a cargo da empresa fornecedora (Abreu, 2022; Osti; Pereira, 2021). Com isso, as empresas pagam conforme a necessidade, podendo aumentar e diminuir o uso de acordo com a demanda (Cleo, 2023). Além disso, os serviços oferecidos por empresas fornecedoras (e.g, AWS, Azure, IBM Cloud, etc), são quase instantâneos, permitindo que qualquer software integrado ao seu ambiente esteja pronto para uso imediato, após a assinatura e pagamento (Cleo, 2023).

Resumindo, existem diversas vantagens e desvantagens nessa infraestrutura (Abreu, 2022). As vantagens são: custo fixo; baixa manutenção de equipamentos físicos; suporte técnico especializado pela empresa contratada; escalabilidade sob demanda e facilidade de uso. Entre suas desvantagens, estão o armazenamento de dados sigilosos em servidores de terceiros e falta de conhecimento técnico sobre a infraestrutura. Além disso, diante de tantas facilidades, existe

a possibilidade que profissionais do TI negligenciem a otimização dos custos, especialmente se eles não estiverem alinhados com as práticas recomendadas ou não tiverem uma compreensão aprofundada dos serviços em nuvem (Global Data Solution, 2021).

2.4 BUSINESS INTELLIGENCE

BI é um conjunto de tecnologias, metodologias e processos que transformam dados brutos em informações úteis para auxiliar na tomada de decisões de negócios (Neves, 2023). Essas informações normalmente incluem análises e dashboards que fornecem compreensão e visibilidade sobre o desempenho de negócios e a eficácia de estratégias, podendo ser utilizadas em todos os departamentos de uma empresa.

O BI é essencial para empresas de todos os tamanhos e setores, pois ajuda a identificar oportunidades de crescimento, otimizar custos e melhorar a eficiência operacional (Neves, 2023). As informações fornecidas pelo BI também podem ajudar as empresas a identificar e mitigar riscos, melhorar a satisfação do cliente e otimizar a cadeia de suprimentos. Além disso, o BI permite que as empresas coletem, analisem e apresentem dados de diversas fontes (Microsoft, 2022b).

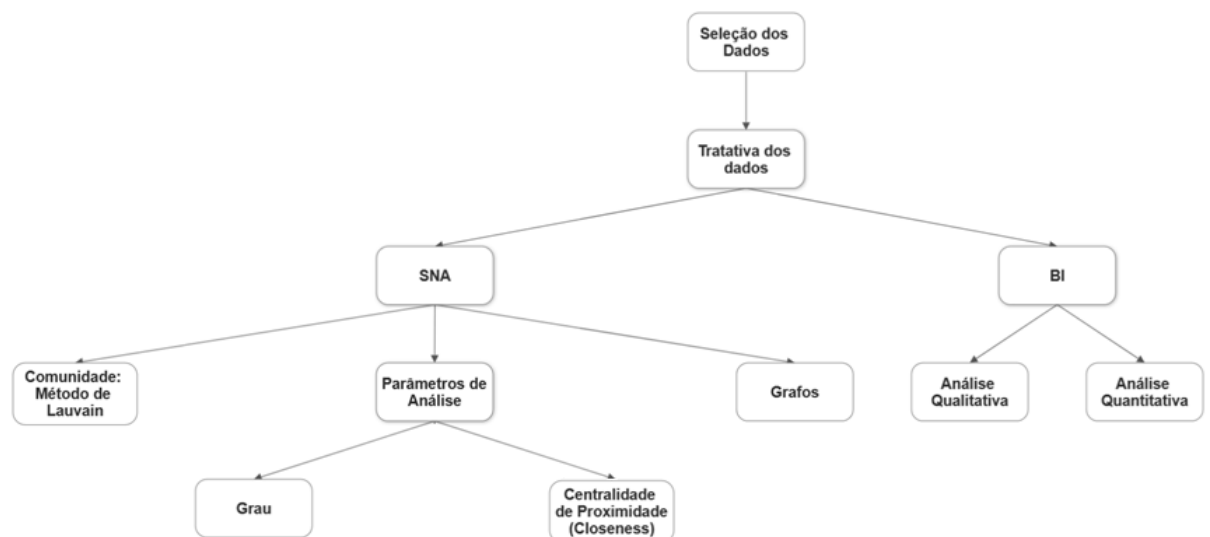
O uso do BI nas empresas pode gerar vantagens como otimização de preços e custos, informações acessíveis e de fácil compreensão e descobertas de novas oportunidades de negócio (Neves, 2023).

3 METODOLOGIA

O trabalho encontra-se dividido nas seguintes etapas principais (Figura 3):

- Seleção e tratamento dos dados
- Aplicação da técnica SNA aos dados
- Análise qualitativa e quantitativa dos dados com BI

Figura 3 – Estrutura Analítica de Projetos resumindo a metodologia. No centro, encontra-se a seleção e tratamento dos dados, seguidos de análises SNA e BI.



Fonte: Elaborado pela autora (2023)

3.1 MATERIAIS

Os tópicos a seguir detalham os materiais utilizados nas pesquisas de SNA e BI.

3.1.1 Social Network Analysis

Os dados para esta etapa foram coletados através de entrevistas internas com membros de cada departamento, a fim de identificar quais pastas, além do seu departamento, são utilizadas. Essas entrevistas foram conduzidas para identificar e mapear as pastas que são

compartilhadas e utilizadas entre diferentes áreas da empresa. O resultado da entrevista pode ser visto no Apêndice A.

Foram selecionados de dois a três funcionários por departamento, com funções distintas entre si, como gerentes, líderes de equipe ou colaboradores-chave, que possuem conhecimento sobre as pastas e arquivos utilizados em suas respectivas áreas de atuação.

Durante as entrevistas, as pessoas foram questionadas sobre as pastas específicas que são acessadas por suas equipes. Os entrevistados foram solicitados a descrever os tipos de informações contidas nessas pastas, a frequência de acesso e os usuários autorizados.

3.1.2 Business Intelligence

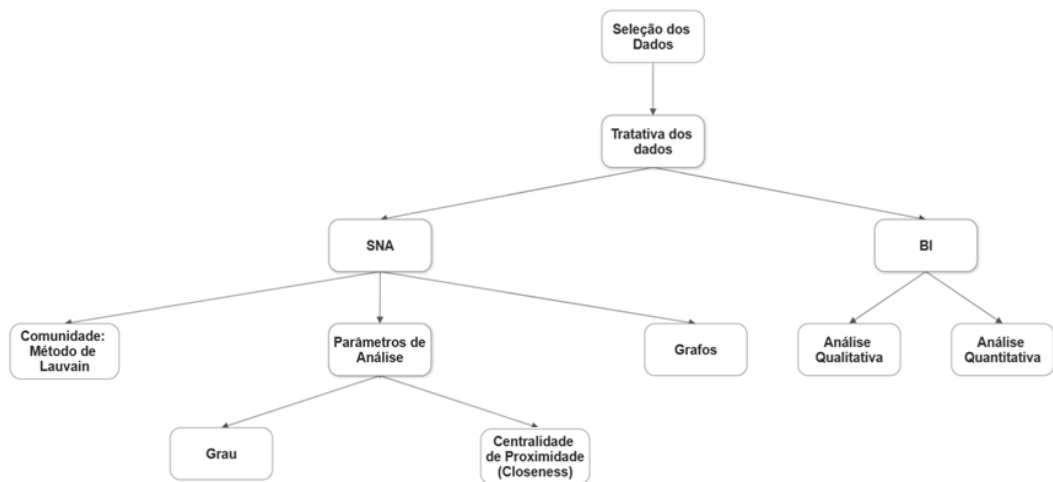
Para essa etapa, foram selecionados dados do período de seis meses, de julho a dezembro de 2022. Os dados são disponibilizados diariamente por uma empresa terceira, em arquivos CSV, que é um formato comum para armazenamento e troca de dados tabulares.

Dentre os registros contidos nos arquivos CSV, destacam-se as seguintes categorias: Ação, Tipo de Objeto (indicando pasta ou arquivo), Caminho (representando o local do arquivo ou pasta manipulada), Servidor Empregado, Editor (identificando o responsável pela edição), Equipamento (indicando o dispositivo utilizado na edição) e Data. Neste estudo, foram foco de análise as colunas Ação, Caminho e Data, fundamentais para o desenvolvimento da pesquisa.

Os dados selecionados abrangem informações sobre quaisquer ações realizadas nos arquivos presentes no File Server "on premise" (servidor local) da empresa. Essas ações incluem renomear (mudança de nome), modificar (alterações no conteúdo dos arquivos), adicionar (inclusão de novos arquivos), excluir (remoção de arquivos) e movimentar (transferência de arquivos para pastas diferentes) arquivos "on premise". Esses eventos são registrados nas colunas Ação, Caminho e Data, fornecendo insights cruciais para a análise exploratória que visa compreender os padrões de uso e manipulação dos dados ao longo do tempo.

Os arquivos contidos "on premise" são compartilhados entre os funcionários por meio de uma pasta mapeada, a W:/, que contém subpastas para cada departamento, como Aplicação, Diretoria, Equus, Financeiro, Industrial, IT, Jurídico, LATAM, Marketing, Qualidade, R&D, Regulatórios, RH, Síntese, Supply Chain, Transferência e Vendas. A Figura 4 apresenta o conteúdo desta pasta.

Figura 4 – Pasta W:/, com suas devidas pastas internas.



Fonte: Elaborado pela autora (2023)

A pasta W:/Transferência é a única acessível a todos os funcionários, enquanto as demais exigem permissão de acesso concedida pelo TI com autorização do gestor correspondente. Dependendo do cargo, o usuário pode necessitar de acesso a várias pastas, concedido mediante autorização.

Na empresa, vários funcionários optam por utilizar o método "on premise" para salvar arquivos pessoais, mantendo-os em uma pasta privada mapeada como O:/. Embora a maioria dos colaboradores já tenha migrado para o serviço do Onedrive para armazenar seus arquivos particulares - indicado a todos os funcionários da companhia -, ainda existem alguns que continuam utilizando a pasta O:/, que foi criada e utilizada antes mesmo da implementação do Onedrive na empresa.

3.2 MÉTODOS

Este trabalho foi dividido em duas etapas principais: a primeira é a aplicação do método SNA, utilizado para captar o relacionamento dos departamentos da Empresa, visando compreender como é a troca de informações entre os setores; a segunda etapa explora a análise quantitativa, extraindo métricas para retirar informações relevantes, visando identificar um grupo de colaboradores que mais influência no fluxo de informações da Empresa.

3.2.1 Análises SNA

Nas análises SNA, serão utilizados os dados obtidos nas entrevistas internas, realizadas para cada setor com pelo menos três funcionários. Entre os funcionários, foram selecionados funcionários entre os cargos de analista, coordenador e supervisor. Com as respostas obtidas, e utilizando o software Studio R, o grafo das respostas será criado utilizando a biblioteca *igraph*.

Essa biblioteca conta com a função *graph_from_data_frame()*, que auxilia na criação de gráficos utilizando um ou dois conjuntos de dados, gerando grafos direcionados ou não (Csardi, 2023a). Como as entrevistas fornecem referência de todas as pastas acessadas por cada setor, será utilizado o grafo direcionado. As setas indicam qual setor acessa quais pastas, apontando se ocorre reciprocidade de acesso. Portanto, o grafo será gerado pelo código 1, e sua plotagem pelo código 2.

```
(1) grafo <- graph_from_data_frame(d = dados, directed = F)
```

```
(2) plot(grafo)
```

Em SNA, suas conexões são representadas pela frequência em que os relacionamentos acontecem entre os nodos, e é conhecido na literatura como grau. No R, é possível extrair suas medidas através da função *degree()*, também da biblioteca *igraph* (Csardi, 2023b). Os graus consideram a direção do laço (relacionamento) em direção ao nodo, podendo distinguir entre laço entrando ou saindo do nodo (Newman, 2010). Para essa pesquisa, usaremos grau não direcionado, portanto, será somados todos os laços do nodo, somando suas entradas (“in”) e saídas (“out”). Seu código pode ser encontrado no código 3.

```
(3) V(grafo)$grau <- degree(grafo)
```

Na análise de centralidade de proximidade, a função *closeness()* mede quantos passos são necessários para acessar todos os outros nodos (vértices) de um determinado nodo (Csardi, 2023c). Seu resultado varia entre 0 e 1, e quanto maior o valor, mais importante é esse nodo na estrutura, pois contém maior influência (Neo4j, 2023b). O código pode ser visto abaixo

```
(4) V(grafo)$centProx <- closeness(grafo, mode = "all", weights = NA, normalized = T)
```

onde:

mode = "all" define que será usada a soma das entradas e saída dos laços no nodo, sem importar a direção;

weights = NA determina que os laços não têm peso no cálculo;

normalized = T calcula a proximidade normalizada (distância média inversa para todos os nodos).

Todas as saídas das funções acima resultam em uma matriz, que será exibida em forma de tabela.

Ao introduzir o estudo de comunidade aos dados, utilizando o método Louvain, possibilita-se verificar comunidades fortes entre si. Comunidade refere-se a grupos de nodos conectados entre si, formando sub-redes, com alta densidade de conexões (Newman, 2010). Existem diversos métodos de detecção de comunidade; entretanto, para esse trabalho, foi selecionado o algoritmo de Louvain, por ser um método de detecção hierárquica, dividindo comunidades em subcomunidades sucessivas (Newman, 2010). Sua construção acontece na execução de 3 códigos diferentes.

Seu primeiro código (5) divide as comunidades em conjuntos ("cluster"), aplicando a função *cluster_louvain()*. A função implementa um algoritmo de otimização multinível de modularidade para identificar a estrutura da comunidade (Gregorovic; Nepusz, 2023).

(5) *comunidade* < - *cluster_louvain(grafo)*

O segundo código (6) retorna como resultado um objeto da classe de comunidade (Csardi, 2023d). E para finalizar, o código 7 plota o gráfico de comunidades.

(6) *V(grafo)\$id_comunidade* < - *membership(comunidade)*

(7) *plot(comunidade, grafo)*

3.2.2 Análises BI

Nessa etapa da pesquisa, foram utilizados os dados fornecidos pela empresa terceirizada. Esses dados, selecionados do período de julho a dezembro de 2022, fornece informações de qualquer manipulação de arquivos que tenha ocorrido no servidor físico da Empresa, relatando dia, horário, local, e tipo de modificação que esse arquivo sofreu. Esse relatório é fornecido

diariamente, constando qualquer renomeio, modificação, adição, exclusão ou movimentação dos arquivos no File Server “on premise” da Empresa.

A preparação dos dados foi predominantemente conduzida utilizando a linguagem de programação Python. Uma vez que a empresa terceirizada fornece um arquivo CSV diariamente, a consolidação desses dados foi executada por meio do ‘Command Prompt’ no ambiente Windows, com o comando *copy*.

Para a etapa de tratamento e limpeza dos dados, foi empregada a biblioteca Pandas do Python. O Pandas é uma poderosa ferramenta de código aberto voltada para análise e manipulação de dados, amplamente reconhecida e utilizada na área de Ciência de Dados (Almeida, 2023).

No início do processo, foram removidas todas as colunas que não possuíam relevância para a análise, utilizando o método `drop()` (8). Isso resultou em dados limpos e devidamente organizados.

(8) *`dados.drop(["coluna1", "coluna2", ..., "colunan"])`*

Além das pastas de cada departamento disponíveis no diretório W:/, também existem pastas de usuários no diretório O:/, onde cada pasta contém o nome e sobrenome do funcionário proprietário. Para realizar a análise do volume de dados manipulados por todos os funcionários, as pastas foram renomeadas para um único rótulo, como 'Usuário'. Essa etapa foi realizada utilizando o método `replace()` (9). Caso fosse identificado um nome de pasta diferente das pastas dos departamentos, essa célula teria seu nome substituído por 'Usuário'. Dessa maneira, obtemos uma única variável para investigar os dados manipulados nas pastas pessoais dos colaboradores da empresa.

(9) *`dados['colunapasta'].replace([value for value in dados['colunapasta']
if value not in ["Aplicação", "Diretoria", ..., "Vendas"]], "Usuário",
inplace = True)`*

Assim, os dados tratados foram salvos em um novo arquivo CSV (10).

(10) *`dados.to_csv(nome_novo_arquivo, index = False)`*

Dentre as análises escolhidas para este estudo, destaca-se a contabilização das ações realizadas em cada pasta no diretório W:/, permitindo a identificação do setor com maior volume de manipulações nos arquivos. A fórmula utilizada consiste na soma de todas as ações ocorridas durante o período de estudo, como exemplificado a seguir. É relevante ressaltar que essa soma é efetuada individualmente para cada pasta, e não para o conjunto total.

$$\sum_{i=1}^n x_i = x_1 + x_2 + \dots + x_n$$

Ao obter a contagem das ações efetuadas em cada pasta, torna-se viável a identificação dos departamentos que requerem maior focalização durante a transição dos arquivos para a nuvem. Isso possibilita direcionar uma atenção mais precisa e direcionada durante todo o processo.

A análise da contagem de ações ao longo das horas do dia tem como objetivo investigar os períodos de maior atividade na manipulação de arquivos. A identificação dos horários de pico de manipulação proporciona uma compreensão mais profunda dos padrões de comportamento dos usuários, oferecendo insights valiosos para o planejamento de recursos e a otimização do File Server. Além disso, para complementar essa avaliação, o estudo da manipulação mensal dos arquivos visou identificar se existe uma tendência de uso semelhante de um mês para outro, ou se há variações significativas nos padrões de utilização das informações.

Adicionalmente, é importante discernir o tipo de ação que se desencadeia nos arquivos. Nesta avaliação, foram quantificadas as renomeações, modificações, adições, exclusões e movimentações de arquivos ocorridos ao longo do período de análise.

Essas análises proporcionam um entendimento mais profundo do padrão de manipulação de arquivos no ambiente "on premise" da Empresa. Adicionalmente, elas oferecerão insights cruciais para a identificação dos principais setores envolvidos nas atividades de arquivo, bem como para a detecção de tendências temporais e horários de maior atividade. A partir desses resultados, será possível embasar decisões mais precisas, com o objetivo de aprimorar a gestão de documentos e otimizar a utilização dos recursos no contexto "on premise".

4 RESULTADOS

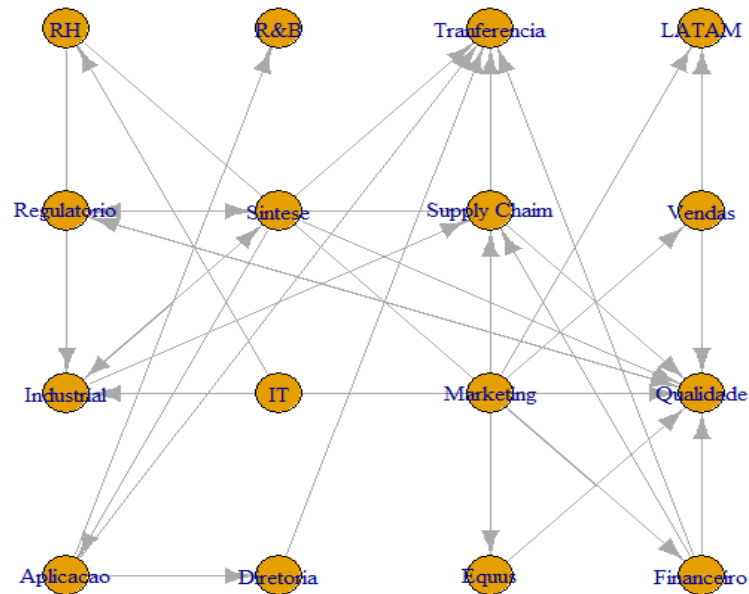
O seguinte tópico explora os resultados obtidos durante o desenvolvimento deste trabalho, dividido em dois subtópicos, uma para Análise de SNA e outro para BI.

4.1 ANÁLISE DE SOCIAL NETWORK ANALYSIS

Nesta primeira seção dos resultados, será demonstrada a forma como as informações estão interligadas dentro dos departamentos por meio de análises de SNA. Esses resultados provêm das entrevistas realizadas com os funcionários, abrangendo de dois a três colaboradores de cada departamento. A partir das respostas obtidas, foram criados grafos que visualizam de maneira clara os vínculos entre os diferentes departamentos (Figura 5).

Na Figura 5 verifica-se que o departamento Jurídico não apresenta conexões com nenhum outro departamento. Essa discrepância é atribuída à natureza sigilosa do conteúdo em sua pasta, acessada apenas pelos próprios membros da equipe. Além disso, nota-se que o departamento de Qualidade registra o maior número de conexões com outros departamentos. Essa discrepância é devida à sua responsabilidade pelos procedimentos de segurança, que são utilizados por diversos setores, especialmente aqueles envolvidos em atividades de maior risco.

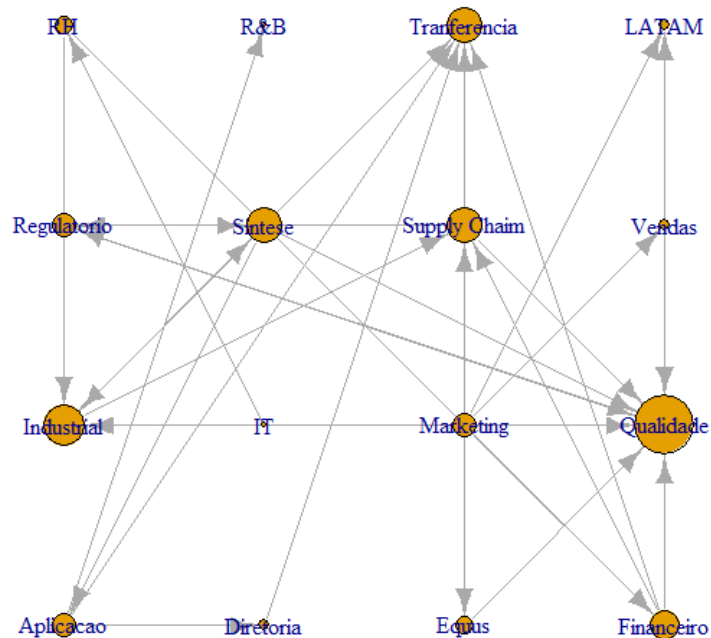
Figura 5 –Grafo, nodo representa os departamentos e laços os relacionamentos.



Fonte: Elaborado pela autora (2023)

Na Figura 6, é ressaltado o volume de interações entre os departamentos. A representação dos nodos no grafo é proporcional ao número de conexões, o que nos permite reconhecer os departamentos de Qualidade, Industrial, Síntese e Supply Chain como aqueles com um grau mais significativo de relacionamentos. A pasta denominada Transferência se destaca por ser de acesso aberto a todos os funcionários, sendo frequentemente utilizada para a troca de arquivos não confidenciais. Isso facilita a comunicação e compartilhamento de informações dentro da empresa.

Figura 6 – Grafo com nodos destacados pelo grau de relacionamento.



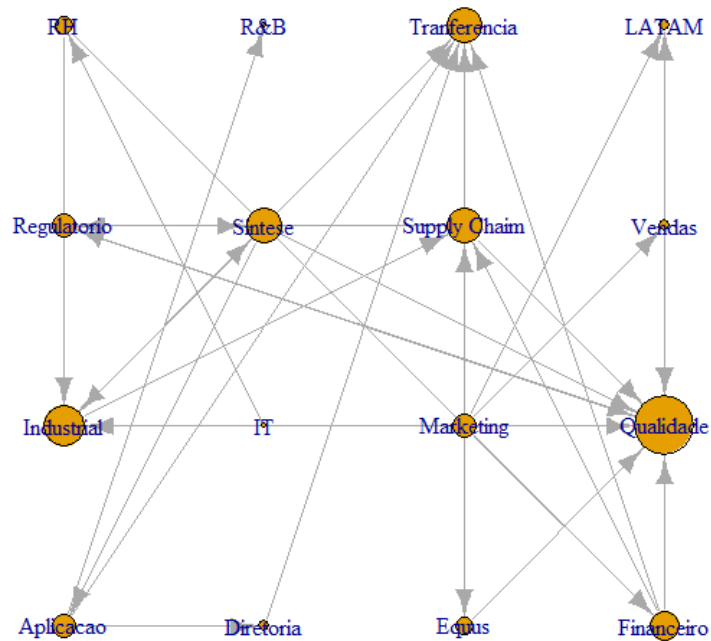
Fonte: Elaborado pela autora (2023)

Ao aplicarmos o estudo de comunidade aos dados, e utilizando o método de Louvain, pudemos identificar conjuntos de nodos com maior relacionamento entre si do que com outros nodos da rede. A detecção de comunidade tem como finalidade avaliar como os nós se agrupam em conjuntos (Newman, 2010). O método de Louvain demonstra de maneira eficaz as divisões de tarefas dentro da empresa, ao criar grupos com características particulares em cada conjunto, mas com similaridades notáveis dentro de cada grupo. Por exemplo, os departamentos Latam, Marketing e Vendas são responsáveis diretos pela divulgação e vendas dos produtos aos clientes, enquanto os departamentos Supply Chain, Equus, Síntese e Qualidade são encarregadas pela qualidade e satisfação dos clientes pós-venda.

A Figura 7 exibe as comunidades formadas em quatro grupos distintos. As áreas mais voltadas para a indústria são representadas em verde claro, compreendendo os departamentos diretamente envolvidos na produção, controle de qualidade e atividades laboratoriais. Os grupos de suporte, incluindo Financeiro, Recursos Humanos e Tecnologia da

Informação, são representados em azul. Em vermelho, destacam-se os setores administrativos, que fornecem suporte aos diretores. Por fim, em roxo, encontram-se os departamentos voltados para atendimento ao cliente e vendas.

Figura 7 – Comunidades das pastas pelo método Louvain.



Fonte: Elaborado pela autora (2023)

A Tabela 1 atesta que a pasta Qualidade registra o maior número de acessos (10), seguidos de Industrial (7), Síntese (6) e Supply Chain (6), todos pertencentes ao mesmo grupo na análise de comunidade. Os resultados da Tabela 1 expõem as pastas ordenadas de grau mais elevado para o menor.

Além disso, a coluna de centralidade de proximidade mede o quanto um nodo é conectado a outro nodo na rede (Telatnik, 2020). Quanto maior o valor obtido, maior a influência do nodo dentro da estrutura. Portanto, as pastas Qualidade (0,60), Industrial (0,57), Síntese (0,57), Transferência (0,57) e Financeiro (0,57) são as pastas consideradas mais importantes para o conjunto.

Tabela 1 –Pastas, com seus valores de grau e centralidade de proximidade.

Pastas	Grau	Centralidade de proximidade
Qualidade	10	0.6000000
Industrial	7	0.5769231
Síntese	6	0.5769231
Supply Chain	6	0.5000000
Transferência	6	0.5769231
Financeiro	5	0.5769231
Aplicação	4	0.4545455
Marketing	4	0.5000000
Regulatório	4	0.4545455
Equus	3	0.3947368
RH	3	0.4285714
Diretoria	2	0.3947368
Vendas	2	0.4285714
LATAM	2	0.4285714
IT	1	0.3061224
R&B	1	0.3191489

Fonte: Elaborado pela autora (2023)

4.2 ANÁLISE DE BUSINESS INTELLIGENCE

Foram selecionados dados diários de julho a dezembro de 2022 para as análises de BI, totalizando 1.631.444 ações sofridas nos arquivos internos. Na Figura 8 observamos as ações realizadas em cada pasta do W:/, sendo que a pasta Usuários é a soma de todas as pastas particulares dos usuários da companhia, localizada na pasta O:/.

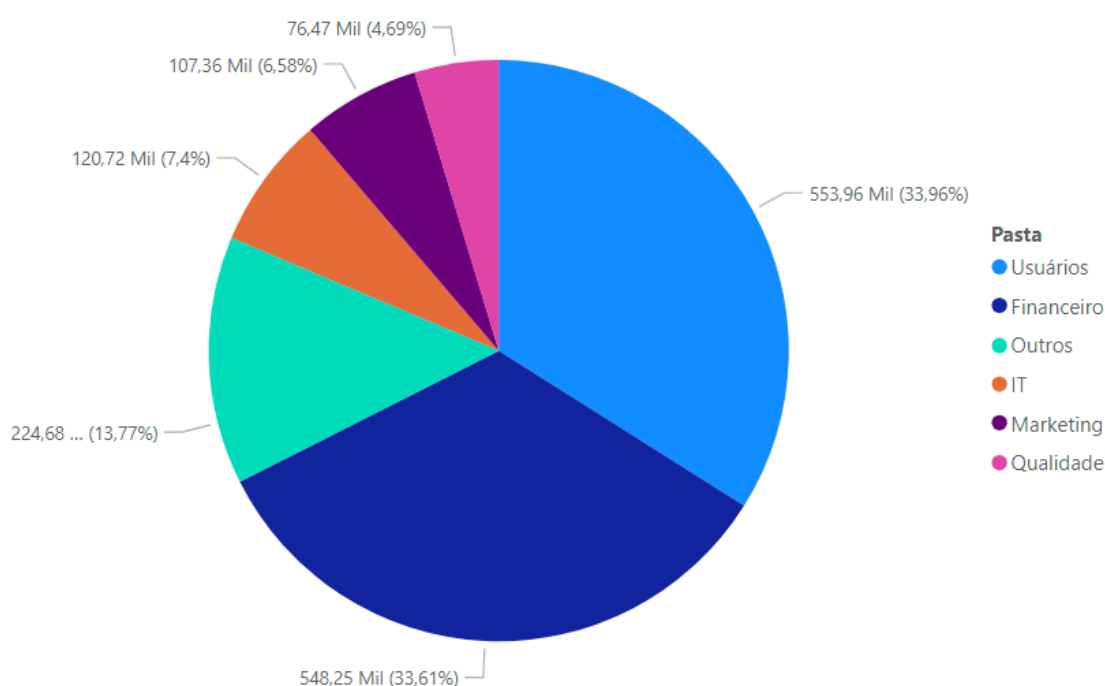
A pasta Usuários e a pasta Financeiro juntas representam mais da metade das operações e, quando adicionadas às pastas de TI, Marketing e Qualidade, essas áreas são responsáveis pelo controle da maioria dos arquivos na empresa, totalizando aproximadamente 86% (1.406.761 ações).

A pasta de TI não sofre ação humana na manipulação dos seus arquivos, pois ela contém apenas os dados de backup dos softwares da fábrica (como *MyLims* e *Intouch Aveva*),

que são atualizados diariamente. Dessa forma, somente as pastas Usuários, Financeiro, Marketing e Qualidade, que são as mais frequentemente acessadas, necessitam de atenção especial durante o processo de migração dos arquivos para a nuvem, pois o backup pode ser direcionado para outra localidade. Como as demais pastas não acumulam um alto valor de ações, seus valores somados são representados pelo item Outros na Figura 8.

Figura 8 – Pastas e seus devidos valores e porcentagens de ações.

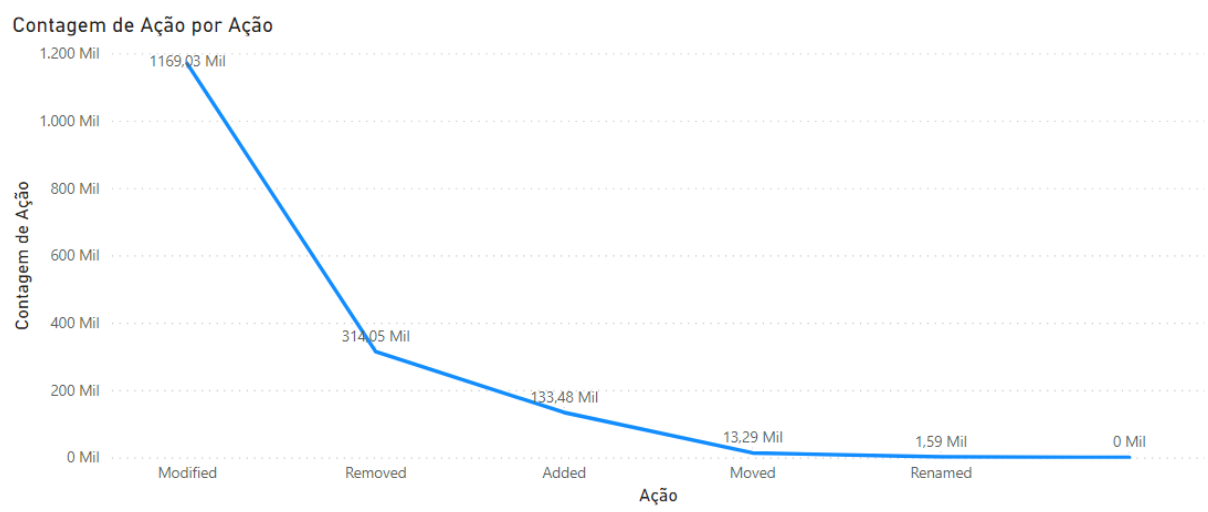
Contagem de Ações por Pasta



Fonte: Elaborado pela autora (2023)

A Figura 9 apresenta a quantidade de ações por sua classificação: modificado, removido, adicionado, movido e renomeado. Observa-se que os arquivos Modificados dominam o tipo de manipulação sofrida pelos arquivos, com aproximadamente 70% (1169030 ações), seguido de Removido (19%, 314053 ações), Adicionado (8%, 133481 ações), Movido (1%, 13294 ações) e Renomeado (1%, 1585 ações).

Figura 9 – Ações realizadas nas pastas.



Fonte: Elaborado pela autora (2023)

Os dados detalhados podem ser observados na tabela 2.

Tabela 2 –Pastas e ações realizadas em cada pasta.

Pastas	Ações	Modified	Removed	Added	Moved	Renamed
Aplicação	22.213	5.983	2.232	12.058	1.782	158
Diretoria	3.368	2.213	67	1.068	6	14
Equus	563	8	129	342	72	12
Financeiro	548.251	521.547	4.745	19.040	2.545	374
Industrial	46.635	22.218	8.255	14.537	1.201	424
IT	120.716	104.699	16	16.001	-	-
Jurídico	364	60	34	190	76	4
Latam	285	170	9	98	8	-
Marketing	107.356	87.388	18.084	1.258	570	56
Qualidade	76.474	4.991	64.752	5.752	904	75

R&D	9	6	-	2	1	-
Regulatórios	178	13	83	67	2	13
RH	16.141	5.286	2.680	6.224	1.814	137
Síntese	6.696	3.399	1.254	1.798	201	44
Supply Chain	23.440	16.392	1.417	5.252	277	102
Transferência	40.892	12.224	5.184	22.335	1.038	111
Usuários	553.965	370.035	183.606	309	-	14
Vendas	63.898	12.398	21.506	27.150	2.797	47
Total	1.631.444	1.169.030	314.053	133.481	13.294	1.585

Fonte: Elaborado pela autora (2023)

Analisando a manipulação dos arquivos por mês, de julho a dezembro de 2022, não identificou-se padrão na manipulação dos arquivos (Figura 10). Observamos picos altos de ações no mês de agosto e novembro, e baixo uso em setembro e dezembro. O único mês que ocorre atividade atípica é dezembro, pois contém férias coletivas dos funcionários na segunda metade do mês, o que justificaria o pequeno volume de dados usados neste mês.

Figura 10 – Ações realizadas nas pastas por mês.

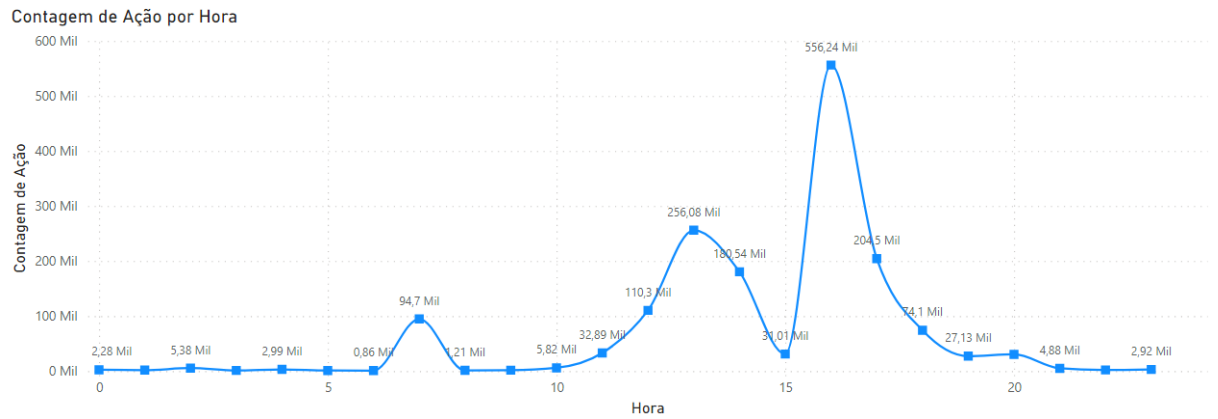


Fonte: Elaborado pela autora (2023)

Entretanto, analisando as ações por hora do dia, percebe-se que o horário administrativo obtém a maior quantidade de manipulação, entre as 6 e 20 horas. Por se tratar de

uma empresa que contém três turnos, os horários das 7 e 16 horas contém maior manuseio dos dados, por ser a hora logo após a troca de turno, que ocorre às 6 e 15 horas.

Figura 11 – Ações realizadas nas pastas por hora.



Fonte: Elaborado pela autora (2023)

5 CONSIDERAÇÕES FINAIS

Técnicas e metodologias comuns em análises de ciência de dados e seus derivados são fundamentais para a tomada de decisões nos departamentos de tecnologia, e estão sendo amplamente implementados à cultura empresarial. Os insights extraídos são valiosos e permitem aos gestores aprimorar estratégias e otimizar o desempenho do departamento.

O objetivo deste trabalho foi apresentar um estudo de caso capaz de auxiliar na decisão de como migrar dados de um servidor “on premise” para a nuvem. As limitações em tais ações encontram-se no alto volume de dados e no uso constante do mesmo, o que dificulta a suspensão do acesso dos funcionários aos arquivos para realizar a migração. Os dados utilizados foram fornecidos em relatórios diários por uma empresa terceirizada, e foi selecionado o período de julho a dezembro de 2022.

Primeiramente, efetuou-se o levantamento dos dados disponíveis para o "File Server" em estudo, o que proporcionou uma visão abrangente dos arquivos e pastas existentes. Em seguida, realizou-se uma análise exploratória dos dados, identificando tendências e padrões de uso que se mostraram cruciais em nosso processo decisório.

O uso de um modelo de BI e análises de SNA, em conjunto com ferramentas como Power BI, Python e R foram fundamentais para a extração de informações valiosas do grande volume de dados disponíveis. Essas ferramentas permitiram apresentar as informações de maneira didática e dinâmica, proporcionando uma visão clara aos gestores e facilitando a tomada de decisões.

Dos resultados, observou-se que as pastas Usuários e Financeiro equivalem a mais da metade das ações realizadas, e se somadas as pastas de TI, Marketing e Qualidade, esse valor chega a quase 90%. Da mesma forma, as pastas Qualidade e Financeiro contém alto nível de acesso por parte de outros departamentos.

Este trabalho propôs criar um plano de migração dos dados por conjuntos de pastas, utilizando o resultado obtido pela análise de comunidade. A migração será realizada por etapas baseando-se nos conjuntos que apresentam similaridades, sendo eles: Financeiro, RH e IT; Marketing, Vendas e LATAM; Transferência, Aplicação, Diretoria, R&B; e Regulatório, Supply Chain, Qualidade, Industrial, Síntese e Equus. Esses grupos de pastas compartilham características semelhantes entre si, refletindo satisfatoriamente as divisões internas da empresa, assim como seus relacionamentos.

As análises temporais de horários revelaram que o uso máximo dos arquivos ocorre durante o horário administrativo, das 8h às 18h, com picos menores antes ou após a mudança

de turno da fábrica, que ocorre às 6h, 14h e 22h. Em geral, o período compreendido das 21h às 6h apresenta o menor uso dos arquivos, tornando o melhor horário para interromper o acesso e realizar a migração com o mínimo impacto nas operações.

No entanto, não foi possível identificar padrão no estudo temporal mensal. Uma alternativa viável seria aumentar a janela de tempo da análise, para dois ou três anos. Dessa forma, com um maior volume de informação, seria possível identificar se existe padrão no uso dos dados ao longo dos meses do ano. Essa abordagem de análise de longo prazo pode revelar tendências sazonais ou variações que não são aparentes em períodos mais curtos.

De forma geral, esta pesquisa identificou pastas que apresentam uma maior manipulação de dados, bem como horários e padrões de uso, por meio do estudo de SNA. Essas informações desempenharam um papel fundamental durante o processo de migração dos arquivos do servidor “on premise” para a nuvem. Acredita-se que os insights obtidos não apenas auxiliarão na tomada de decisões estratégicas, mas também garantirão um processo de migração mais eficiente, minimizando interrupções nas operações.

REFERÊNCIAS

ABREU, L. **O que é on premise, quais são as suas características e diferenças do sistema cloud?** Rock Content. 2022. Disponível em: <https://rockcontent.com/br/blog/o-que-e-on-premise>. Acesso em: 07 de maio de 2023.

AIRES, V., NAKAMURA, F. **Aplicação de Medidas de Centralidade ao Método Louvain para Detecção de Comunidades em Redes Sociais.** XLIX Simpósio Brasileiro de Pesquisa Operacional. Blumenau, 2017. Disponível em: <http://www.din.uem.br/sbpo/sbpo2017/pdf/169162.pdf>. Acesso em: 11 de março de 2023.

ALMEIDA, M. **Pandas Python: o que é, para que serve e como instalar.** Alura. 2023. Disponível em: <https://www.alura.com.br/artigos/pandas-o-que-e-para-que-serve-como-instalar>. Acesso em: 12 de agosto de 2023.

BORBA, E. M. **Medidas de Centralidade em Grafos e Aplicações em redes de dados.** 2013. Dissertação (Mestrado em Matemática Aplicada) - Universidade Federal do Rio Grande do Sul. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/86094/000909891.pdf>. Acesso em: 01 de abril de 2023.

BUTTS, C. T. **Social network analysis: A methodological introduction.** Asian Journal of Social Psychology. 2008. p. 13–41. DOI: 10.1111/j.1467-839X.2007.00241.x. Disponível em: <https://courses.washington.edu/ir2010/readings/butts.pdf>. Acesso em: 01 de abril de 2023.

CLEO. **On Premise vs. Cloud: Key Differences, Benefits and Risks.** c2023. Disponível em: <https://www.cleo.com/blog/knowledge-base-on-premise-vs-cloud>. Acesso em: 07 de maio de 2023.

CSARDI, G. **Creating igraph graphs from data frames or vice-versa.** R Documentation. c2023a. Disponível em: https://search.r-project.org/CRAN/refmans/igraph/html/graph_from_data_frame.html. Acesso em: 26 de maio de 2023.

CSARDI, G. **Degree and degree distribution of the vertices.** R Documentation. c2023b. Disponível em: <https://search.r-project.org/CRAN/refmans/igraph/html/degree.html>. Acesso em: 26 de maio de 2023.

CSARDI, G. **Closeness centrality of vertices.** R Documentation. c2023c. Disponível em: <https://search.r-project.org/CRAN/refmans/igraph/html/closeness.html>. Acesso em: 26 de maio de 2023.

CSARDI, G. **Functions to deal with the result of network community detection.** R Documentation. c2023d. Disponível em: <https://search.r-project.org/CRAN/refmans/igraph/html/communities.html>. Acesso em: 26 de maio de 2023.

DIAMOND, P. **Compartilhamento de arquivos de forma segura para pequenas empresas.** 2020. Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/business->

insights-ideas/resources/secure-file-sharing-for-small-businesses. Acesso em: 02 de outubro de 2023.

GLOBAL DATA SOLUTION. **Migração para cloud:** conheça os (X) desafios e como enfrentá-los. 2021. Disponível em: <https://globaldata.com.br/migracao-para-cloud/>. Acesso em: 03 de junho de 2023.

GREGOROVIC, T., NEPUSZ, T. **Finding community structure by multi-level optimization of modularity.** R Documentation. c2023. Disponível em: https://search.r-project.org/CRAN/refmans/igraph/html/cluster_louvain.html. Acesso em: 26 de maio de 2023.

GRUNSPAN, D.Z., WIGGINS, B.L., GOODREAU, S.M. **Understanding Classrooms through Social Network Analysis:** A Primer for Social Network Analysis in Education Research. CBE Life Sci Educ. 2014 Summer;13(2):167-79. DOI: 10.1187/cbe.13-08-0162. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4041496/>. Acesso em: 23 de março de 2023.

HUBERMAN, B.; ADAMIC, L. **Information Dynamics in the Networked World.** Complex Networks. Lecture Notes in Physics, Springer, 2003. Disponível em: <http://www.hpl.hp.com/research/idl/papers/infodynamics/infodynamics.pdf>. Acesso em: 04 abril de 2023.

KASEB, M. R., KHAFAGY, M. H., ALI, I. A., SAAD, E. M. **Redundant Independent Files (RIF):** A Technique for Reducing Storage and Resources in Big Data Replication. Trends and Advances in Information Systems and Technologies. 2018. Disponível em: https://doi.org/10.1007/978-3-319-77703-0_18. Acesso em: 23 de março de 2023.

MAGALHÃES, L. **File Server / Servidor de Arquivos:** físico ou em nuvem, qual escolher? Brasil Cloud Nuvem Corporativa. 2020. Disponível em: <https://blog.brasilcloud.com.br/file-server-servidor-de-arquivos-fisico-ou-em-nuvem-qual-escolher/>. Acesso em: 23 de março de 2023.

MARTELETO, R. M. **Análise de redes sociais - aplicação nos estudos de transferência da informação.** Ci. Inf., Brasília, v. 30, n. 1, p. 71-81, jan./abr. 2001. Disponível em: <https://www.scielo.br/j/ci/a/6Y7Dyj4cVd5jdRkXJVxhxqN/?format=pdf&lang=pt>. Acesso em: 23 de março de 2023.

MICROSOFT. **SharePoint.** Microsoft. c2023a. Disponível em: <https://www.microsoft.com/pt-br/microsoft-365/sharepoint/collaboration>. Acesso em: 23 de março de 2023.

MICROSOFT. **O que é Business Intelligence?** Microsoft. c2023. Disponível em: <https://powerbi.microsoft.com/pt-br/what-is-business-intelligence/>. Acesso em: 26 de março de 2023.

NEO4J. **Louvain.** Neo4j, Inc. c2023a. Disponível em: <https://neo4j.com/docs/graph-data-science/current/algorithms/loouvain/>. Acesso em: 07 de maio de 2023.

NEO4J. **Closeness Centrality**. Neo4j, Inc. c2023b. Disponível em: <https://neo4j.com/docs/graph-data-science/current/algorithms/closeness-centrality/>. Acesso em: 07 de maio de 2023.

NEVES, D. **Power BI**: Diferença entre dashboards e relatórios. 2023. Disponível em: <https://www.alura.com.br/artigos/dashboards-vs-relatorios>. Acesso em: 07 de maio de 2023.

NEWMAN, M. E. J. *Networks: An Introduction*. New York: Oxford. 2010. 772 p.

OSTI, E. W. A., PEREIRA, R. B. O. **Análise de Desempenho do banco de dados SQL Server em Infraestruturas On Premise e Cloud**. Revista de Tecnologia Invest. v.5, n.1. 2021. ISSN: 2595-8178. Disponível em: <http://revista.institutoinvest.edu.br/index.php/revistainvest/article/view/58/32>. Acesso em: 07 de maio de 2023.

OTTE, E., ROUSSEAU, R. **Social network analysis**: a powerful strategy, also for the information sciences. *Journal of Information Science*. 2002. 28(6), 441–453. Disponível em: <https://doi.org/10.1177/016555150202800601>. Acesso em: 26 de março de 2023.

RECUERO, R. C. **Comunidades em Redes Sociais na Internet**: Proposta de Tipologia baseada no Fotolog.com. Tese de Doutorado - Porto Alegre: Universidade Federal do Rio Grande do Sul. 2006. Disponível em: <https://www.lume.ufrgs.br/bitstream/handle/10183/8614/000582681.pdf?sequence=1&isAlloved=y>. Acesso em: 26 de março de 2023.

TELATNIK, M. **How To Get Started with Social Network Analysis**. Towards Data Science. 2020. Disponível em: <https://towardsdatascience.com/how-to-get-started-with-social-network-analysis-6d527685d374>. Acesso em: 27 de julho de 2023.

APÊNDICE A – ENTREVISTA REALIZADA COM OS FUNCIONÁRIOS

Durante a entrevista com os funcionários da empresa A, foi questionado qual(is) pasta(s) é(são) acessada(s) fora do próprio departamento, selecionando de dois a três funcionários de cada departamento. Seguem as respostas e nível do cargo de cada funcionário.

- Aplicação
 - Analista: R&B e Transferência
 - Coordenador: R&B
 - Gerente: Diretoria
- Diretoria
 - Analista: Transferência
 - Coordenador: Transferência
- Equus
 - Analista: Qualidade e Supply Chain
 - Coordenador: Qualidade e Supply Chain
 - Gerente: Qualidade e Supply Chain
- Financeiro
 - Analista: Qualidade e Transferência
 - Coordenador: Qualidade, Transferência e Supply Chain
 - Gerente: Qualidade e Supply Chain
- Industrial
 - Analista: Transferência, Qualidade e Síntese
 - Coordenador: Qualidade, Supply Chain e Síntese
 - Gerente: Qualidade, Supply Chain e Síntese
- IT
 - Analista: RH
 - Gerente: RH
- Marketing
 - Analista: Transferência e LATAM
 - Coordenador: Transferência, LATAM, Financeiro e Vendas
 - Gerente: Financeiro e Vendas
- Qualidade
 - Analista: Regulatório e Industrial

- Coordenador: Regulatório, Industrial e LATAM
- Regulatório
 - Analista: Síntese e Qualidade
 - Coordenador: Síntese e Qualidade
- Síntese
 - Analista: Industrial, Qualidade e Aplicação
 - Coordenador: Industrial, Qualidade e Aplicação
 - Gerente: Industrial, Qualidade e Aplicação
- Supply Chain
 - Analista: Regulatório e Equus
 - Coordenador: Qualidade, Regulatório e Equus
 - Gerente: Regulatório e Equus
- Vendas
 - Coordenador: Qualidade
 - Gerente: Qualidade
- RH
 - Analista: Industrial e Financeiro
 - Coordenador: Industrial e Financeiro