

Um Modelo para Análise de Dados Operacionais da Força Aérea Brasileira Utilizando Tecnologias de *Big Data*

(A Model for the Brazilian Air Force Operational Data Analysis Using Big Data Technologies)

André Benzi Baccarin, Marcelo Alexandre Martins da Conceição, Oswaldo Segundo da Costa Neto e Adilson Marques da Cunha

Instituto Tecnológico de Aeronáutica (ITA), São José dos Campos/SP - Brasil

Resumo—Nos últimos anos, a utilização do termo *Grandes Conjuntos de Dados (Big Data)* vêm se expandindo. Grandes empresas vêm criando ferramentas para tratar e analisar dados, aplicando tecnologias de *Big Data*. No âmbito da Força Aérea Brasileira (FAB), ocorrem situações análogas com dados gerados em fluxo contínuo. Este artigo apresenta um modelo para análise de dados operacionais da FAB, utilizando tecnologias de *Big Data*. Define os principais conceitos de interesse da FAB envolvendo estas tecnologias e apresenta os principais *frameworks* utilizados para análise de *Big Data*: Hadoop e Spark, além de potenciais aplicações e vantagens em comparação com ambientes tradicionais de análise de dados. O modelo proposto é fruto de uma pesquisa metodológica e reporta alguns resultados esperados, partindo da aplicação dos principais conceitos envolvendo *Big Data*. A principal contribuição da pesquisa envolve o modelo proposto nos *frameworks* Hadoop e Spark para análise de dados operacionais da FAB.

Palavras-Chave—*Big Data*, Dados Operacionais, Força Aérea.

I. INTRODUÇÃO

Esta pesquisa tem por objetivo apresentar conceitos e tecnologias de *Grandes Conjuntos de Dados (Big Data)*, sua aplicabilidade e capacidade em agregar valor à Força Aérea Brasileira (FAB).

A aplicação desses conceitos e tecnologias pode ser realizada utilizando-se um modelo para análise de dados operacionais gerados por aeronaves, dispositivos móveis, computadores, sensores, provas, entre outros dados gerados por Esquadrões Aéreos.

A. Contextualização

Na última década, o conceito de *Big Data* vem ganhando espaço na comunidade científica e tecnológica [1]. Para exemplificar o momento atual pelo qual a humanidade vem passando em relação à Era da Informação, 90% de todos os dados existentes no mundo foram gerados apenas em 2019 e 2020 [2]. Ferramentas tradicionais para análise de dados vêm sendo consideradas ineficientes para tratar grandes conjuntos de dados originados de diversas fontes, estruturados e não estruturados [3].

E-mail para correspondência com autor: baccarin@ita.br. Agradecimentos ao Instituto Tecnológico de Aeronáutica (ITA) e à Força Aérea Brasileira (FAB) pelo apoio a esta pesquisa.

Na FAB, o cenário não é diferente. As aeronaves vêm se tornando cada vez mais tecnológicas [4] e a diversidade de missões vem crescendo consideravelmente [5]. Na atual conjuntura brasileira, com recursos humanos e financeiros limitados, equipamentos complexos e sofisticados e aumentando de valor [6], armazenar, processar, recuperar e analisar dados operacionais têm se transformado em grandes desafios.

As ferramentas para análise de *Big Data*, por meio da utilização de software livre, e as grandes quantidades de hardware tradicional disponível no mercado envolvendo computadores pessoais [7] vêm representando uma alternativa possível de ser adotada pela comunidade tecnológica para superar esses grandes desafios.

Essa abordagem pode propiciar maior eficiência e menor custo [8], desde que o ambiente de dados seja devidamente preparado, tornando-o pronto para a aplicação de algoritmos de Inteligência Artificial (IA) e suas derivações, mesmo dentro do âmbito de *Big Data*. O objetivo é extrair, de forma eficiente, informações e conhecimentos, a partir de dados armazenados como: horas de voo, notas de cursos, precisões de lançamentos, provas, sensores, atividades operacionais, dentre outros, de diferentes aeronaves e Esquadrões Aéreos.

B. Organização do Artigo

A Seção II define o termo *Big Data*, de acordo com o contexto apresentado. Na Seção III, são apresentados os arcabouços de software (*frameworks*) Hadoop e Spark. Em seguida, a Seção IV mostra os tipos de dados operacionais que a FAB necessita analisar, seguida pela Seção V, que traça um paralelo entre dados operacionais da FAB e as tecnologias citadas nas Seções anteriores, juntamente com as vantagens de sua implantação. Assim, o Modelo Proposto é apresentado na Seção VI. Por fim, a Seção VII sintetiza as conclusões do que foi apresentado nas Seções anteriores, com recomendações e sugestões para trabalhos futuros.

II. CONCEITUAÇÕES

Esta seção apresenta as principais tecnologias e conceitos relevantes utilizados na concepção do modelo proposto desta pesquisa, que vêm sendo empregados por empresas como a Microsoft, Twitter, Facebook, entre outras [9], [10], e que possuem potencial aplicabilidade no âmbito da FAB.

A. O que é Big Data

A Tecnologia de Grandes Conjuntos de dados (*Big Data*) envolve quantidades e variedades de dados que os sistemas computacionais tradicionais não são capazes de armazenar, processar e recuperar.

Esse termo surgiu em 1997 [11], e começou a ser largamente difundido em meados de 2005, acompanhando a popularização da Internet [12], quando grandes companhias como Yahoo e Google perceberam a incapacidade de sistemas e ferramentas tradicionais de Banco de Dados suportarem as grandes quantidades de dados que recebiam e começaram a buscar alternativas de soluções mais viáveis [13].

Desde então, vêm surgindo ferramentas inovadoras como as propiciadas pelo *framework* Hadoop, baseadas em processamento distribuído e algoritmos preparados para serem executados a partir de componentes de hardware agrupados (*clusters*) [13].

Embora algumas definições tenham sido criadas ao longo dos anos, como as sintetizadas na Tabela I, o conceito de *Big Data* deve ser analisado e atualizado de acordo com o contexto em que ele se encontra inserido. As características de Grandes Conjuntos de Dados (*Big Data*) geradas pela Google são diferentes das características de *Big Data* geradas por organizações militares, devido às características únicas de cada Organização.

TABELA I: As definições de *Big Data*.

Autor	Definição
DEMCHENKO (2013) [14]	É um volume massivo de dados estruturados e não estruturados complexo de se processar utilizando apenas técnicas tradicionais de Banco de Dados.
DUMBILL (2012) [15]	São dados que excedem a capacidade de processamento de Sistemas de Banco de Dados convencionais.
GOULARTE; ZILBER; PEDRON (2015) [16]	Além de ser um conjunto de ferramentas, agrega a geração de novas tecnologias e arquiteturas projetadas para se extrair valor econômico de grandes volumes de dados.
MANYIKA (2011) [17]	Refere-se a Banco de Dados cujo tamanho transcende à capacidade de softwares e ferramentas tradicionais em capturar, armazenar, gerenciar e analisar dados.
SCHÖNBERGER-MAYER E CUKIER (2013) [18]	Refere-se a processamentos em grande escala, impossíveis de serem realizados em proporções inferiores para se extrair ideias inovadoras e criar formas originais de geração de valor capazes de alterar: mercados, organizações, relações entre cidadãos e governos, entre outros fatores.

As suas características inerentes são frequentemente denominada de “Vs” do *Big Data* e costumam auxiliar na definição e no entendimento do fenômeno *Big Data*, mostrando sua complexidade e importância para cenários contemporâneos e futuros. Atualmente, cinco características costumam aparecer frequentemente na literatura definidos por [14], conforme mostradas na Fig. 1 e brevemente descritas a seguir:

- 1) **Volume** - característica marcante, envolvendo tamanho de dados;
- 2) **Velocidade** - dados gerados com velocidade, que necessitam de rápidas análises;
- 3) **Variabilidade** - dados coletados em diferentes formatos, estruturados ou não estruturados;
- 4) **Veracidade** - dados reais e confiáveis; e

- 5) **Valor** - dados contendo valores adicionados para propiciar análises.

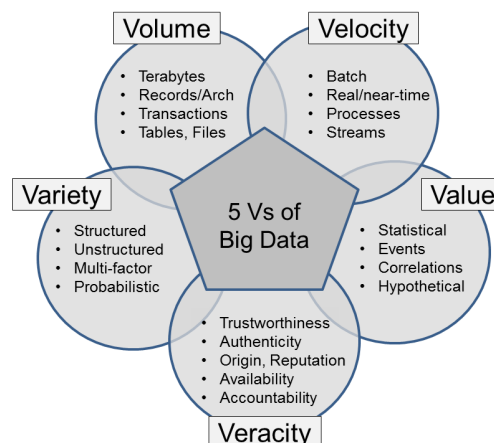


Fig. 1: Os 5 “Vs” do *Big Data*. (Adaptado de [14])

Embora o tamanho de *Big Data* geralmente receba destaque, a falta de estruturação representa o seu principal desafio, dificultando suas análises [19], pois dados costumam se originar de diversas fontes e em diversos formatos.

Alguns autores vêm agregando características às definições de *Big Data*, porém algumas dessas características extras sintetizadas a seguir ainda não possuem consenso na literatura:

- **Viabilidade** - relacionamento entre variáveis e padrões [20];
- **Venalidade** - característica de se projetar os dados para venda [20];
- **Variabilidade** - variações no significado de *Big Data* [21];
- **Visualização** - onde os resultados devem ser de fácil visualização por seus utilizadores [21];
- **Validade** - os dados podem não ser necessários para análises, apesar de sua correteza [22]; e
- **Volatilidade** - os dados perdem seu valor com o passar do tempo [22].

A partir da explosão da Internet e com o crescimento de grandes empresas na área, o fenômeno de *Big Data* se transformou em um conjunto de tecnologias.

Além disso, essas tecnologias vêm aumentando de importância e já se encontram presentes no cotidiano das populações do mundo inteiro a tal ponto que, atualmente, a informação vem se transformando no principal ativo de empresas, tendendo a aumentar cada vez mais sua valia, quando analisada corretamente.

B. Dados Estruturados x Não estruturados

Grandes Conjuntos de Dados, normalmente, possuem dados que podem ser classificados como estruturados e não estruturados.

Os dados estruturados podem ser organizados em tabelas. Um Sistema Gerenciador de Bancos de Dados Relacional (SGBDR) pode hospedar e propiciar o relacionamento de dados estruturados por meio de chaves primárias (que identificam registros) e chaves estrangeiras (que identificam relacionamento entre entidades) [23].

Os dados não estruturados não podem ser organizados em tabelas com relacionamentos. Portanto, eles não podem ser

analisados por ferramentas tradicionais. Eles podem ser originados: de arquivos de texto (como postagens no Facebook); de imagens; de documentos em PDF; ou até mesmo de outras mídias, não assumindo formatos determinados. Para seus armazenamentos, faz-se necessária a utilização de Bancos de Dados Não relacionais [24].

Além das duas classificações anteriores, podem existir também dados semiestruturados. Estes dados são semelhantes aos dados estruturados, porém com formatos diferentes, tornando difícil e onerosa sua classificação em tabelas relacionais. Apesar destes dados semiestruturados poderem ser tratados com o uso de SGBDRs, as ferramentas não relacionais costumam lhes propiciar melhores desempenhos para seus tratamentos [25].

Embora um ambiente de *Big Data* possa ser constituído de qualquer formato de dado relacional ou não relacional, grande parte deste ambiente, normalmente, costuma ser formada de dados semi ou não estruturados, tornando suas análises mais difíceis e trabalhosas, até mesmo impossíveis, se utilizadas apenas ferramentas tradicionais [19].

III. OS *frameworks* DE ANÁLISE DE DADOS

Esta Seção apresenta os *frameworks* e ferramentas de armazenamento, recuperação e análise de *Big Data* utilizados no modelo proposto, com suas principais características, vantagens e limitações.

A. Hadoop

O *framework* Apache Hadoop é composto por uma estrutura que permite processamentos distribuídos de Grandes Conjuntos de Dados (*Big Data*) em *clusters* de computadores, usando modelos relativamente simples de programação. Ele foi projetado e dimensionado para funcionar desde em um único computador até em milhares de máquinas, cada uma oferecendo recursos computação e armazenamento local.

Em vez de depender de grandes capacidades de hardware para fornecer alta disponibilidade, o *framework* foi projetado para detectar e tratar falhas na camada de aplicativo, oferecendo serviços com alta disponibilidade por meio de *clusters* de computadores simples, cada um dos quais sujeito a falhas [26].

O *framework* Hadoop foi criado por Doug Cutting e Mike Cafarella, baseado nas publicações “*The Google File System*” (2003) [27] e “*MapReduce: Simplified Data Processing on Large Clusters*” (2004) [28], que passaram a utilizar o paradigma de programação MapReduce para tratamento de dados em ambientes distribuídos.

Assim, em 2006, foi concebido o Hadoop, considerado atualmente como um *framework* com ferramentas mais eficientes para análises de dados em grande escala, contendo uma estrutura gratuita e de código aberto (*open-source*).

O *framework* Hadoop possui como ferramentas principais: o *Hadoop Distributed File System* (HDFS) e o MapReduce.

O HDFS representa a principal ferramenta e sistema de arquivos do Hadoop. Ele se utiliza do poder de agrupamento de computadores (em *clusters*), é capaz de armazenar dados de forma segura e rápida, além de oferecer tolerância a falhas. Diversas ferramentas atuais de *Big Data* se utilizam do HDFS como base para análises de dados.

O MapReduce é uma ferramenta do Hadoop que propicia a realização de análises e recuperações de dados armazenados

no sistema de arquivos HDFS. Ele baseia-se no paradigma de programação de mesmo nome e propicia utilização eficiente do poder de processamento distribuído do Hadoop.

Desde sua criação, o *framework* Hadoop vem recebendo atualizações e possuindo equipes crescentes de colaboradores, que desenvolvem dezenas de programas utilizando o Hadoop, facilitando coletas de dados, por meio de ferramentas como o HBase (ferramenta de banco de dados não relacional) e o Hive (que utiliza uma linguagem semelhante ao SQL para tratamento de dados estruturados) [29].

Dentre as principais ferramentas do Hadoop, destacam-se os já citados Apache Hive e Apache HBase, que vêm sendo aplicados nesta pesquisa. O Hive é uma ferramenta de Armazém de Dados (*Data Warehouse*) utilizado para armazenar dados no HDFS de forma estruturada, utilizando a linguagem HQL (*Hive Structured Language*), muito semelhante ao SQL e beneficiando, com isto, os profissionais com experiência em bancos de dados tradicionais.

O HBase é uma ferramenta de banco de dados não relacional, orientada a colunas, possuindo tabelas, semelhantes às utilizadas nos bancos de dados relacionais, porém com estruturação distinta e ingestão de diversos tipos de dados. Ambas ferramentas foram desenvolvidas especialmente para o Hadoop e funcionam adequadamente junto a este *framework*.

O Hadoop proporciona melhor desempenho do que sistemas tradicionais de bancos de dados relacionais e não relacionais em contextos em que a quantidade de dados se torna grande e sua estruturação complexa.

B. Spark

O Apache Spark é um *framework* analítico unificado, considerado atualmente como ultrarrápido, utilizado para processamentos de Grandes Conjuntos de Dados (*Big Data*) e de Aprendizado de Máquina (*Machine Learning*) [30]. Ele é compatível com dados do HDFS e pode ser executado em *clusters* Hadoop, utilizando o gerenciador de recursos YARN (*Yet Another Resource Negotiator*) conjuntamente com o Hadoop.

Além disso, o Spark é capaz de processar dados no HBase, Hive e em qualquer outro formato de entrada (*InputFormat*) do Hadoop. Ele foi projetado para executar processamentos em lotes (*batches*), de forma semelhante ao MapReduce, em novas cargas de trabalho, em transmissões (*streamings*), em consultas interativas e em Aprendizado de Máquina (*Machine Learning*) [31].

O *framework* Spark foi criado em 2009 por Matei Zaharia e vem recebendo desde então colaborações de milhares de voluntários, possuindo mais de 300 empresas envolvidas [32]. Assim como o *framework* Hadoop, ele possui código aberto (*open-source*) e seus colaboradores contribuem para economia de custos de implementação e manutenção. Apesar destes benefícios, o Spark, assim como o Hadoop, exige programadores experientes, o que geralmente representa uma desvantagem dos softwares de código aberto (*open-source*) [33].

O Spark surgiu como alternativa ao MapReduce do Hadoop. Diferentemente deste, que realiza o processamento utilizando o disco rígido, o Spark realiza o processamento, sempre que possível, utilizando a memória primária (memória RAM), mais veloz que as memórias secundárias em discos. Em um contexto prático, o Spark propicia a entrega de resultados

mais rapidamente, conforme ilustrado na Fig. 2, que mostra a execução de um algoritmo de contagem de palavras no Hadoop e no Spark.

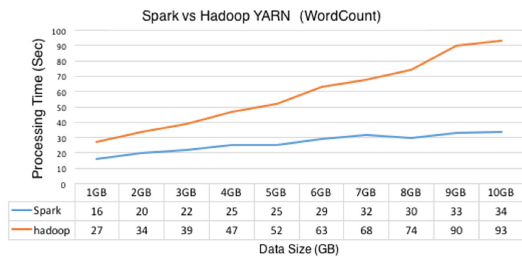


Fig. 2: Algoritmo de contagem de palavras executado no Hadoop e no Spark. ([34])

Dentre as ferramentas do Spark, destaca-se o Spark SQL, utilizado com a linguagem Python (dentre outras), possuindo comandos semelhantes à linguagem SQL. Ele pode ser utilizado com o Hive e com o HBase, expandindo suas capacidades [35]. As bibliotecas MLlib e ML, atuando com as ferramentas do Spark, permitem aplicar mecanismos de *Machine Learning* próprios para processamentos distribuídos.

C. Python

Python representa atualmente uma das linguagens de programação mais utilizadas no mundo [36], devido à sua pequena curva de aprendizado e grande quantidade de bibliotecas, com quase 300.000 projetos de diferentes propósitos [37]. Os *frameworks* Hadoop e Spark podem ser utilizados com Python, embora tenham sido estruturados em Java e Scala, respectivamente. Apesar do Hadoop não possuir suporte direto para Python, a biblioteca MRJob [38] transcreve códigos Python para Java. O Spark possui suporte nativo para Python, embora apresente melhor desempenho utilizando a linguagem Scala [39].

IV. DADOS OPERACIONAIS

Anualmente, a FAB prepara um programa de instruções e requisitos denominado Programa de Instrução e Manutenção Operacional (PIMO), para que Esquadrões Aéreos possam manter a operacionalidade de seus tripulantes. Este programa envolve um conjunto de tarefas como: missões aéreas, provas e cursos, que pilotos e demais tripulantes de aeronaves precisam realizar para serem habilitados a exercer suas funções a bordo de forma correta e segura.

Além dessas tarefas, são realizados em intervalos regulares, Exercícios Operacionais (EXOPs), reunindo Esquadrões Aéreos de diferentes localidades do Brasil. Os EXOPs propiciam o cumprimento de diversas funções em missões das Aviações de Caça, Transporte, Asas Rotativas, Reconhecimento e Patrulha, bem como a realização de simulações em ar e estudos em solo, a fim de aperfeiçoar e criar novas doutrinas e procedimentos operacionais.

Existem, ainda, as missões regulares executadas por Esquadrões, de acordo com horas e tarefas previamente designadas pelo Comando de Preparo (COMPREP) e pelos demais Comandos. Todas essas missões, tarefas e atividades costumam gerar grandes quantidades de dados operacionais.

Eventos de crise envolvendo guerras e calamidades públicas como na crise da falta de oxigênio da Pandemia do Vírus

COVID-19 na cidade de Manaus no estado do Amazonas no início de 2021 [40] ou até mesmo incêndios como o ocorrido na boate Kiss no estado do Rio Grande do Sul em 2013 [41] também costumam gerar grandes quantidades de dados relevantes a serem utilizadas e analisadas em planejamentos de missões atuais e futuras.

Esses Grandes Conjuntos de Dados (*Big Data*) gerados normalmente envolvem horas de voo, missões de lançamento, fórmulas para cálculos de eficiência de lançamentos [42], notas de provas, tipos de missões realizadas [43], testes de condicionamentos físicos [44], entre outras missões, tarefas e atividades realizadas por tripulantes durante o ano.

Todos esses dados de grande valor agregado, necessitam ter suas veracidades comprovadas e possuem as principais características de *Big Data* como: variedades de dados, diversidades de fontes, formatos distintos e necessidades de comunicações, armazenamentos e seguranças. As suas análises necessitam ser realizadas rapidamente e sempre com alto nível de confiabilidade.

V. *Big Data* E A FORÇA AÉREA

A maior parte dos dados operacionais da FAB são estruturados e semi-estruturados. Entretanto, com os avanços tecnológicos dos últimos anos, vêm se constatando cada vez mais a presença de dados não estruturados como: mensagens de status e posição enviadas em tempo real a partir de computadores de aeronaves em missões para os centros de comando, mapas ou filmagens de voos das aeronaves de caça, dados de sensores das novas aeronaves (Gripen e KC-390), entre outros, evidenciando a presença das características de *Big Data* anteriormente mencionadas.

Dentro desse contexto, a utilização de *frameworks* como o Hadoop permite o armazenamento de grandes conjuntos de dados no HDFS, gerando: economia significativa de recursos de toda ordem, uma característica importante para instituições públicas; tolerância a falhas; e armazenamentos e recuperações de dados mais rápidos e seguros, características desejadas por organizações que lidam com dados operacionais, confidenciais e que exigem altos graus de segurança como dados das Forças Armadas.

Também dentro desse contexto, a utilização do *framework* Spark juntamente com o Hadoop, operando sobre o HDFS, consegue aumentar a performance da análise de dados de missões, tarefas e atividades exercidas, tornando os processos mais rápidos, aproveitando melhor ainda cada um dos *frameworks* [45].

A utilização da linguagem Python nesses *frameworks* facilita o aprendizado e o uso do Spark e do Hadoop, por ser uma linguagem já de conhecimento de muitos programadores, gerando também economia de tempo e recurso na formação de usuários e operadores de sistemas militares [46].

Com isso, o modelo proposto nesta pesquisa e descrito a seguir acaba tornando inserções de dados mais eficientes, utilizando menos estruturas e simplificando programações de usuários de computadores em operações nas Forças Armadas, propiciando que grandes conjuntos de dados não estruturados possam ser também utilizados, processados, analisados e conservados em sistemas ainda mais rápidos, seguros e confiáveis.

VI. O MODELO PROPOSTO

Para se utilizar na FAB as funcionalidades existentes em ambientes capazes de propiciar armazenamentos, processamentos, recuperações e análises de Grandes Conjuntos de Dados (*Big Data*), faz-se necessária a especificação e a concepção inicial de um modelo híbrido que envolva tanto ambientes tradicionais, como bancos de dados PostgreSQL, MySQL ou MongoDB, quanto ambientes não tradicionais, como Hadoop e Spark.

O modelo proposto nesta pesquisa envolve armazenamentos, processamentos, recuperações e análises de dados operacionais estruturados e não estruturados originados de diversas fontes como: computadores de Esquadrões Aéreos, celulares, sensores, entre outras.

Os Grandes Conjuntos de Dados (*Big Data*) existentes na FAB podem ser armazenados, processados e recuperados, tanto em bancos de dados estruturados ou não estruturados, locais ou em *clusters*, e depois repassados para sistemas distribuídos, por meio do Sistema de Arquivos do Hadoop (*Hadoop Distributed File System - HDFS*), utilizando ferramentas dos ambientes Hadoop e Spark. Os *frameworks* estão sendo implementados no momento desta pesquisa, para futuros testes e demonstrações de resultados.

Assim, o modelo proposto será capaz de propiciar a utilização de: *clusters* de computadores de baixo custo, para realizar processamentos mais rápidos, seguros e distribuídos; o HDFS, para armazenar e recuperar arquivos; o Hive, para organizar dados estruturados; e o HBase para organizar dados não estruturados.

A partir dos dados armazenados, o modelo proposto será capaz de envolver ferramentas para recuperações e análises de dados e informações, destacando-se entre elas o Spark SQL e as bibliotecas Spark ML e MLLib. Com o Spark SQL, será possível recuperar e analisar dados do Hive e do HBase, por meio de comandos similares à sintaxe de SQL.

Após a recuperação de dados, será possível utilizar as bibliotecas Spark ML e o Spark MLLib para o processamento de dados estruturados e não estruturados, respectivamente, aplicando-se modelos de *Machine Learning* específicos para armazenamento distribuído e com mais desempenho.

A Fig. 3 apresenta o Modelo Proposto para Análise de Dados Operacionais da FAB utilizando tecnologias de *Big Data*.

A partir da especificação e concepção deste modelo proposto, acredita-se que um método para análise de dados operacionais da FAB possa ser desenvolvido, utilizando *Big Data*, visando aumentar eficiência, confiabilidade e segurança e reduzir desperdícios de recursos.

VII. CONCLUSÃO

Esta pesquisa objetivou apresentar apenas a especificação e concepção inicial de um modelo para análise de dados operacionais da Força Aérea Brasileira (FAB), como a parte operacional do Comando da Aeronáutica (COMAER), utilizando ambientes de *Big Data*, como parte de uma pesquisa em nível de mestrado que se encontra em desenvolvimento no Programa de Pós-graduação em Engenharia Eletrônica e Computação na Área de Informática envolvendo o Programa de Pós-Graduação em Aplicações Operacionais (PPGAO). Desta forma, foram apresentadas alternativas viáveis e inovadoras que poderão elevar o nível de proficiência da FAB na área de análise de dados.

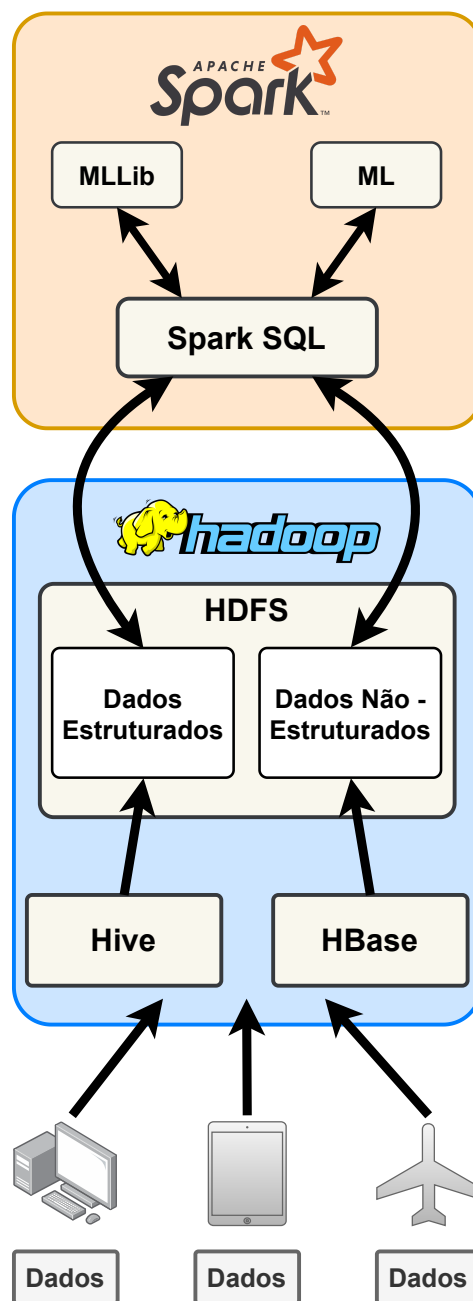


Fig. 3: Modelo Proposto para Análise de Dados Operacionais da FAB utilizando tecnologias de *Big Data*

A. Recomendações

Recomenda-se que as tecnologias utilizadas devam ser adotadas em *clusters* da própria FAB, sem a contratação de servidores em nuvem e, se possível, esses *clusters* possam ser programados e configurados por técnicos devidamente capacitados no âmbito das Forças Armadas, visando a manutenção da segurança de dados e suas instalações. Recomenda-se também que o Comando da Aeronáutica propicie o treinamento ou a contratação de pessoal especializado na área de análise de dados, para melhorar os níveis de capacitação existentes e desenvolver sistemas computadorizados e atualizados envolvendo tecnologias emergentes, de forma a manter a segurança e confidencialidade dos dados.

Para organizações que não necessitam de confidencialidade, existem várias opções de serviços em nuvem, como o Amazon

Web Service (AWS) ou o GoogleCloud. Estas soluções facilitam o uso das ferramentas, além de serem dimensionadas às necessidades dos usuários, gerando economia.

B. Sugestões para Trabalhos Futuros

Sugere-se, para trabalhos futuros, que sejam investigadas implantações de sistemas com interfaces amigáveis, para que os integrantes de Esquadrões Aéreos possam realizar inserções e análises de dados de forma autônoma, isto é, sem a necessidade de uma equipe de TI especializada atuando diretamente, além da extensão do modelo para outros tipos de dados na FAB, como dados financeiros ou de pessoal.

AGRADECIMENTOS

Os autores deste artigo agradecem ao Comando da Aeronáutica (COMAER), ao Instituto Tecnológico de Aeronáutica (ITA) e ao Programa de Pós-Graduação em Aplicações Operacionais (PPGAO), pelo suporte e infraestrutura oferecidos durante o desenvolvimento desta pesquisa.

REFERÊNCIAS

- [1] S. H. Chung, H. L. Ma, M. Hansen, and T. M. Choi, "Data science and analytics in aviation," *Transportation Research Part E: Logistics and Transportation Review*, vol. 134, pp. 1–7, 2020.
- [2] J. Bulao, "How Much Data is Created Every Minute?" 2021. [Online]. Available: <https://techjury.net/blog/how-much-data-is-created-every-day/>
- [3] P. Mannava, "A Study on the Challenges and Types of Big Data," vol. 2, no. August 2013, 2013.
- [4] J. O. Bittencourt, "A Segurança Operacional e os Avanços Tecnológicos: Desafios na Interface Homem/Máquina," Ph.D. dissertation, Universidade do Sul de Santa Catarina, 2020. [Online]. Available: <http://www.riuni.unisul.br/handle/12345/10838>
- [5] COMAER, "DCA 11-45 Conceção Estratégica Força Aérea 100," 2018.
- [6] A. Torrente. (2019) Nem militares escapam da crise : orçamento da Defesa regride ao nível da gestão Lula. [Online]. Available: <https://www.gazetadopovo.com.br/republica/orcamento-defesa-2019-menor-em-quinze-anos-lula/>
- [7] T. W. Dinsmore, *Disruptive Analytics: Charting Your Strategy for Next-Generation Business Analytics*, 1st ed. USA: Apress, 2016.
- [8] M. Y. Santos, C. Costa, J. a. Galvão, C. Andrade, B. A. Martinho, F. V. Lima, and E. Costa, "Evaluating SQL-on-Hadoop for Big Data Warehousing on Not-So-Good Hardware," in *Proceedings of the 21st International Database Engineering Applications Symposium*, ser. IDEAS 2017. New York, NY, USA: Association for Computing Machinery, 2017, p. 242–252. [Online]. Available: <https://doi.org/10.1145/3105831.3105842>
- [9] B. Marr, *Big Data In Practice How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*, 2016.
- [10] B. Arora. (2020) Top Companies Using Apache Hadoop. [Online]. Available: <https://www.w3trainingschool.com/top-companies-using-apache-hadoop>
- [11] N. Mehta and A. Pandit, "Concurrence of big data analytics and healthcare: A systematic review," *International journal of medical informatics*, vol. 114, pp. 57–65, 2018.
- [12] R. Dontha. Who came up with the name Big Data? . [Online]. Available: <https://www.datasciencecentral.com/profiles/blogs/who-came-up-with-the-name-big-data>
- [13] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The Hadoop Distributed File System," in *2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, 2010, pp. 1–10.
- [14] Y. Demchenko, P. Grosso, C. D. Laa, and P. Membrey, "Addressing Big Data Issues in Scientific Data Infrastructure," in *2016 International Conference on Collaboration Technologies and Systems, CTS 2016*, 2016, pp. 48–55.
- [15] E. Dumbill, *Planning for Big Data: O'Reilly Media*, 2012.
- [16] A. d. C. Goularte, S. N. Zilber, and C. D. Pedron, "Big Data: a systematic review of management literature," in *CONTECSI USP - International Conference on Information Systems and Technology Management*, 2015.
- [17] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. H. Byers, "Big data: The next frontier for innovation, competition and productivity," *McKinsey Global Institute*, no. June, p. 156, 2011.
- [18] V. Mayer-Schönberger and K. Cukier, *Big Data: A Revolution That Will Transform How We Live, Work, and Think*, 2013.
- [19] T. H. Davenport, *Big Data @ Work*, 2014.
- [20] M. de Souza, F. G. Almeida, and R. R. Souza, "O TERMO BIG DATA: quebra de paradigma dos n-V's," in *II Workshop de Informação, Dados e Tecnologia, UFPB, Brasil*, 2018.
- [21] T. Chartier, "Vertigo Over the Seven V's of Big Data," *Wiley Periodicals, Inc.*, pp. 11–15, 2016.
- [22] M. A. U. D. Khan, M. F. Uddin, and N. Gupta, "Seven V's of Big Data understanding Big Data to extract value," in *2014 Zone 1 Conference of the American Society for Engineering Education - "Engineering Education: Industry Involvement and Interdisciplinary Trends", ASEE Zone 1 2014*, 2014.
- [23] R. Elmasri and N. S. B., *Sistemas de Banco de Dados*, 4th ed. Pearson, 2005.
- [24] R. P. Padhy, M. R. Patra, and S. C. Satapathy, "RDBMS to NoSQL: reviewing some next-generation non-relational database's," *International Journal of Advanced Engineering Science and Technologies*, vol. 11, no. 1, pp. 15–30, 2011.
- [25] R. D. S. Mello, C. F. Dorneles, A. Kade, V. D. P. Braganholo, and C. A. Heuser, "Dados Semi-Estruturados," *Simpósio Brasileiro De Banco De Dados*, vol. d, pp. 475–513, 2000. [Online]. Available: [http://www.cin.ufpe.br/\\$sim\\$fab/cursos/metodologia-graduacao/2006-1/tutorial.pdf](http://www.cin.ufpe.br/simfab/cursos/metodologia-graduacao/2006-1/tutorial.pdf)
- [26] Apache Hadoop Webpage. [Online]. Available: <https://hadoop.apache.org/>
- [27] S. Ghemawat, H. Gobioff, and S.-T. Leung, "The Google File System," in *Proceedings of the 19th ACM Symposium on Operating Systems Principles, ACM, Bolton Landing, NY (2003)*, 2003, pp. 20–43.
- [28] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," 2004.
- [29] "Hadoop Ecosystem Table." [Online]. Available: <https://hadooecosystemtable.github.io/>
- [30] Databricks. Apache Spark - What is Spark. [Online]. Available: <https://databricks.com/spark/about>
- [31] A. Spark. FAQ — Apache Spark. [Online]. Available: <https://spark.apache.org/faq.html>
- [32] —. Apache Spark. [Online]. Available: <https://spark.apache.org/>
- [33] N. Viorres, P. Xenofon, M. Stavakis, E. Vlachogiannis, P. Koutsabasis, and J. Darzentas, "Major HCI challenges for open source software adoption and development," in *International Conference on Online Communities and Social Computing*. Springer, 2007, pp. 455–464.
- [34] C.-T. Yang, S.-T. Chen, and Y.-Z. Yan, "The implementation of a cloud city traffic state assessment system using a novel big data architecture," *Cluster Computing*, vol. 20, 06 2017.
- [35] J. S. Damji, B. Wenig, T. Das, and D. Lee, *Learning Spark: Lightning-Fast Data Analytics*, second ed., 2020.
- [36] As linguagens de programação mais usadas: saiba quais são elas. [Online]. Available: <https://idocode.com.br/blog/programacao-as-linguagens-de-programacao-mais-usadas-saiba-quais-sao-elas/>
- [37] PyPI, The Python Package Index. [Online]. Available: <https://pypi.org/>
- [38] mrjob: the Python MapReduce library. [Online]. Available: <https://github.com/Yelp/mrjob>
- [39] M. Armbrust, T. Das, A. Davidson, A. Ghodsi, A. Or, J. Rosen, I. Stoica, P. Wendell, R. Xin, and M. Zaharia, "Scaling Spark in the Real World: Performance and Usability," *Proc. VLDB Endow.*, vol. 8, pp. 1840–1843, 2015.
- [40] F. Rocha and O. Lima. FAB ultrapassa 600 horas de voo em apoio ao Amazonas na Operação COVID-19. [Online]. Available: <https://www.fab.mil.br/noticias/mostra/36896/OPERA%C3%87%C3%83%20COVID-19%20-%20FAB%20ultrapassa%20600%20horas%20de%20voo%20em%20apoio%20ao%20Amazonas%20na%20Opera%C3%A7%C3%A3o%20COVID-19>
- [41] GAV, "Esquadrão de transporte da Força Aérea auxilia no transporte de pacientes." [Online]. Available: <https://www.fab.mil.br/noticias/mostra/13991/>
- [42] COMPREP, "Programa de Elevação Operacional (PEVOP) da Aeronave C-105;" 2019.
- [43] —, "Programa de Elevação Operacional (PEVOP) da Aeronave A-29;" 2019.
- [44] —, "Programa de Elevação Operacional (PEVOP) da Aeronave AH-2;" 2019.
- [45] Amol Bansod, "Efficient Big Data Analysis with Apache Spark in HDFS," *International Journal of Engineering and Advanced Technology*, no. 6, pp. 2249–8958, 2015.
- [46] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg et al., "Scikit-learn: Machine learning in Python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.