

# Um Modelo para Análise de Dados Operacionais da Força Aérea Brasileira

André Benzi Baccarin e Adilson Marques da Cunha

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

**Resumo** — O Termo *Big Data* vem ganhando importância com o passar dos anos. Surge junto com a popularização da internet, representando grandes conjuntos de dados que não podem ser analisados por ferramentas tradicionais devido ao seu tamanho e complexidade. Na Força Aérea Brasileira (FAB), a quantidade e variedade de dados gerados vem tornando a análise de Dados Operacionais um desafio, principalmente com a chegada de novas aeronaves e a evolução dos tipos de missão. Desta forma, um Modelo para Análise de Dados Operacionais baseado na utilização de tecnologias de *Big Data*, sobretudo com a utilização dos *frameworks* Hadoop e Spark, pode facilitar a análise destes dados de forma eficiente. Assim, este trabalho apresenta a proposição de um Modelo de Análise de Dados Operacionais utilizando *Big Data*.

## I. INTRODUÇÃO

Esquadrões Aéreos anualmente geram Grandes Conjuntos de Dados (*Big Data*) e Informações Operacionais tais como: horas de voo, notas de provas, precisões de lançamentos, acertos de bombas, entre outros, além de dados de sensores eletrônicos instalados nas aeronaves. Para isso, necessita-se de armazenamentos, processamentos, recuperações, análises rápidas e seguras de dados e informações, bem como de utilização de soluções tecnológicas adequadas para se analisar e criar, apropriadamente, indicadores operacionais para a Força Aérea Brasileira (FAB) [1].

## II. CONTEXTUALIZAÇÃO

A FAB vem preparando, anualmente, um programa de instruções e requisitos para que Esquadrões Aéreos possam manter a operacionalidade de seus tripulantes, chamado de Programa de Instrução e Manutenção Operacional (PIMO). Trata-se de um conjunto de missões operacionais, provas de conhecimento, cursos e demais requisitos que todo tripulante tem que fazer durante o ano para a manutenção de suas operacionalidades. Eventualmente, vários Esquadrões Aéreos de diferentes Aviações se unem para aplicar e testar os conhecimentos adquiridos, manter operacionalidades e desenvolver doutrinas, em Exercícios Operacionais (EXOPs).

São realizadas, ainda, conforme a necessidade, grandes operações aéreas para apoiar governos no enfrentamento de crises como: o incêndio na boate Kiss, a distribuição de vacinas contra a COVID-19, ajudas em alagamentos regionais, auxílios na crise de oxigênio em Manaus, combate aos tráfico de drogas, eventuais conflitos armados, entre outros eventos de interesse federal.

Todas essas missões, tarefas e atividades geram grandes conjuntos de dados (*Big Data*) que necessitam ser analisados com rapidez, correte e eficiência, e que tendem a aumentar ainda mais de volume com a chegada de novas aeronaves e novas missões [2].

## III. OBJETIVO DA PESQUISA

Desenvolver (conceber, especificar, modelar e projetar) um Modelo de Análise de Dados Operacionais para a FAB, visando aumentar a eficiência, qualidade, confiabilidade e segurança (*safety*) da análise de dados, relacionais ou não-relacionais.

Este Modelo baseia-se em tecnologias de *Big Data* que vêm sendo utilizadas por grandes empresas, como Microsoft, Twitter e Facebook [3], entre outras, em *clusters* de computadores, envolvendo diversos computadores de pequeno porte trabalhando juntos, de forma coordenada e transparente, agregando o poder de processamento de máquinas de baixo custo. Para isso, o Apache Hadoop e o Apache Spark podem ser utilizados com ferramentas específicas como o *Hadoop Distributed File System* (HDFS) e o Hive, do *framework* Hadoop, e o Spark SQL e Spark ML, do *framework* Spark.

Com esse modelo proposto, concebido, projetado e implantado, a análise dos dados operacionais, sejam eles dados relacionais tais como horas de voo, consumo de combustível, precisão de lançamentos de pessoal ou de disparos de mísseis, ou não-relacionais, como sensores de aeronaves que medem velocidade, posição GPS ou necessidade de manutenção, torna-se mais eficiente [4], auxiliando Esquadrões Aéreos da FAB em suas tomadas de decisões e definições de indicadores operacionais relacionados aos seus tripulantes e aeronaves.

## IV. FERRAMENTAS UTILIZADAS

A biblioteca de software Apache Hadoop consiste de uma estrutura (*framework*) que permite processamentos distribuídos de grandes conjuntos de dados em *clusters* de computadores simples utilizando modelos relativamente simples de programação. O Apache Hadoop foi projetado para ser utilizado em servidores únicos ou até em milhares de máquinas, cada uma oferecendo recursos de computação e de armazenamento local. Em vez de depender de grandes capacidades de hardware para fornecer alta disponibilidade, a própria biblioteca foi projetada para detectar e tratar falhas nos computadores automaticamente, disponibilizando serviços mais estáveis [5].

O Apache Spark é um *framework* analítico unificado utilizado para processamentos de grandes conjuntos de dados (*Big Data*) e de aprendizado de máquina (*machine learning*). 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. É capaz de processar dados no HBase, Hive e em qualquer outro formato de entrada (*InputFormat*) do Hadoop. 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*), utilizando memória RAM, sendo mais eficiente que o Hadoop em processamento de dados [6].

## V. MODELO PROPOSTO

Ao se utilizar o Modelo Proposto, os dados operacionais gerados pelos sistemas computadorizados presentes na FAB são, inicialmente, salvos em HDFS, utilizando-se as ferramentas Hive, para dados estruturados, e HBase, para dados não-estruturados. Após armazenados, esses dados são selecionados e pré-processados na ferramenta Spark SQL, nativa do Spark, que propicia a realização de processamentos mais eficientes do que ao se utilizar o Hadoop. Depois de selecionados e limpos, os dados são processados, por duas bibliotecas nativas do Spark para *Machine Learning*: o Spark ML e o Spark MLlib, ambas bibliotecas preparadas para uso em ambiente distribuído e para processamentos de dados estruturados e não-estruturados, respectivamente.

Desta forma, utilizando o modelo estruturado, que aproveita as vantagens de ambos os *frameworks* [7], informações ocultas em dados operacionais, de acordo com a necessidade de cada Esquadrão, podem ser descobertas, utilizando tanto dados históricos de voo quanto dados de sensores de aeronaves.

## VI. CONCLUSÃO

Este Pôster apresentou uma síntese da pesquisa envolvendo a proposta de um Modelo para Análise de Dados Operacionais utilizando tecnologias de *Big Data* para a Força Aérea Brasileira. Com este modelo, espera-se que a análise de dados operacionais, em geral, possa ser realizada com mais eficiência, visando a obtenção de melhorias no processo de avaliação e criação de Indicadores Operacionais.

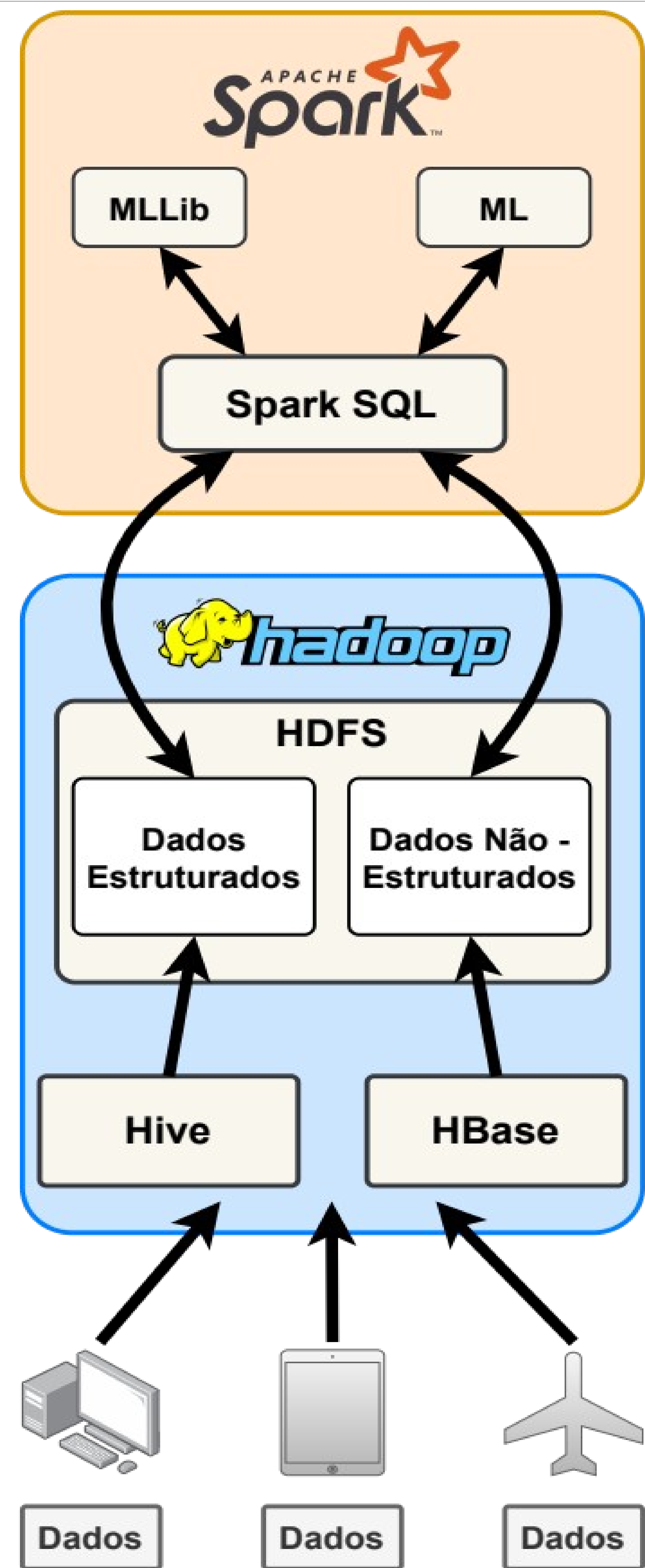


Figura 1 – Modelo Proposto para Análise de Dados Operacionais da FAB utilizando tecnologias de *Big Data*

## REFERÊNCIAS

- [1] P. Mannava, "A Study on the Challenges and Types of Big Data," vol. 2, no. August 2013, 2013.
- [2] COMAER, "DCA 11-45 Concepc,ao Estrat `egica Forc,a A `erea 100," 2018.
- [3] B. Marr, Big Data In Practice How 45 Successfull Companies Used Big Data Analytics to Deliver Extraordinary Results, 2016.
- [4] M. Y. Santos, C. Costa, J. a. Galvao, 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.
- [5] Apache Hadoop Webpage. [Online]. Available: <https://hadoop.apache.org/>
- [6] Databricks. Apache Spark - What is Spark. [Online]. Available: <https://databricks.com/spark/about>
- [7] Amol Bansod, "Efficient Big Data Analysis with Apache Spark in HDFS," International Journal of Engineering and Advanced Technology, no. 6, pp. 2249–8958, 2015.