

SOPA DE LETRAS	
Sigla	Significado
Apache Hadoop	Framework (conjunto de programas) para armazenamento e processamento de BIG DATA, trabalha com clusters (vários computadores ligados que se comportam com se fosse um único aumentando exponencialmente seu desempenho).
Apache Hive	Faz parte do ecossistema Hadoop, trata-se de um Data warehouse com MapReduce. Para consultas e análise de dados usa uma linguagem própria chamada HQL.
Apache Kafka	Faz parte do ecossistema Hadoop, utilizado para gerenciamento de fluxo de dados em tempo real.
Apache Spark	Framework usado como alternativa do Hadoop MapReduce, porém apresenta melhor desempenho.
BI	Significa Business Intelligence, ou seja, Inteligência de Negócio. Conjunto de técnicas, metodologias, processos, tecnologias capazes de transformar dados em informações valiosas para serem usados estrategicamente em organizações.
BIG DATA	Capacidade de uma sociedade obter informações de maneira novas a fim de gerar ideias úteis e bens e serviços de valor significativo. Em outras palavras, é um conceito que envolve uma quantidade massiva de dados sendo produzidos numa velocidade alta, nos mais variados formatos e a produção de insights valiosos sobre eles.
Cognos	Software da IBM para análise de dados baseado em cloud (computação em nuvem).
Data Mining	Processo de explorar grande quantidade de dados a fim de descobrir padrões.
Data Set/ DataFrame	Estrutura de dados semelhante a uma matriz, mas as colunas podem ter tipos diferentes de dados. Linguagem R, Python usam data frames.
Data Warehouse	Depósito de dados com informações específicas, detalhadas de um negócio usadas posteriormente para auxiliar na tomada de decisões.
DBA	Significa 'administrador de banco de dados'. É profissional responsável por gerenciar o banco de dados, instalando, configurando, definindo acessos e restrições, otimizando, criando mecanismos de backups entre outras

	atribuições.
ETL	Significa: E (extração), T (transformação), L (carga). Remete-se ao processo de obter dados de fontes, tratá-los (segundo as regras de negócio, adaptando formatos), carregá-lo em bases como, por exemplo, Data Warehouse, para posterior análise.
Git	Sistema de controle de versões distribuído, Atualmente é utilizado como gerenciador de código-fonte por milhares de desenvolvedores mundialmente.
Hbase	Banco de dados distribuído open-source orientado a coluna, escrito em Java. Tem fácil integração com o Apache Hadoop, sendo uma boa opção para soluções BIG DATA.
HFDS	Módulo do Hadoop para lidar com armazenamento de arquivos. Como temos vários computadores em cluster, cada computador tem seu dispositivo de armazenamento com sua capacidade (ex: HD de 1 TB) e sistema de arquivo (ex: NTFS), o HFDS abstrai essas especificações como se um único 'HD' e um sistema de arquivos próprio.
Keras	Biblioteca (conjunto de funções pré-definidas para serem usadas) para redes neurais, de código aberto, escrita em Phyton.
Kerberos	Protocolo de segurança usado pelo Hadoop para autenticação.
Machine Learning	Significa aprendizado de máquina, fazendo parte da Inteligência Artificial, que consiste em treinar máquinas (para elas fazerem atividades antes executadas por pessoas) baseadas em algoritmos.
MapReduce	Módulo do Hadoop para processamento do BIG DATA.
Matlab	Software voltado para cálculo numérico. Trabalha com análise numérica, matrizes e construção de gráficos.
MongoDB	Um dos principais SGBD noSQL disponíveis no mercado. Orientado a documentos, de código aberto e multiplataforma.
NoSQL	Tecnologia de banco de dados que não somente usa a linguagem SQL. Representa os bancos de dados não relacionais (capazes de armazenar e processar dados não estruturados ou semi-estruturados como, por exemplo, post's no <i>Twitter</i>). Ex de SGBD

	NoSQL é o MongoDB.
Notebook Jupyter	Ambiente computacional web, interativo para criação de programas e serviços. Suporta dezenas de linguagem incluindo R e Python.
NPL	Significa Processamento de Linguagem Natural, que é análise de texto feito por software para descobrir tendências, comportamentos ou padrões. Ex: comentários de um produto para verificar o que clientes pensam.
Pandas	Famosa biblioteca open-source da linguagem Python que fornece alto desempenho, estruturas de dados e ferramentas para análise de dados.
Python	Linguagem de programação muito famosa para análise de dados.
PowerBI	Ferramenta de análise de negócio da Microsoft possibilitando construir insights sobre os dados, multiplataforma (desktop e dispositivos mobile).
Qlik	Ferramenta para visualização de dados do tipo self-service (usuários sem especificação técnica consegue elaborar seus próprios <i>dashboards</i> de maneira rápida).
R	Uma das principais linguagens de programação para análise de dados. Permite: coletar, transformar, limpar, visualizar e aplicar modelos de aprendizado de máquina.
SAS	Conjunto de softwares para análise de dados contemplando: recuperação de arquivos, análise estatística, acesso a banco de dados, geração de gráficos e relatórios.
SGBD	Significa sistema de gerenciar de banco de dados. É um conjunto de programas para gerenciar um banco de dados relacional, responsável pela manipulação, acesso aos dados encontrados na base de dados. Ex de SGBD famosos no mercado: Oracle, MySQL, PostgreSQL, Microsoft SQL Server.
Spotfire	Solução de Análise e Visualização de dados criada pela empresa TIBCO.
SQL	Linguagem usada para gerenciar banco de dados relacionais.
Tableau	Um dos principais softwares de mercado para visualização de dados.
Tensorflow	Biblioteca de código-aberto para aprendizado de máquina compatível com a linguagem Python.

