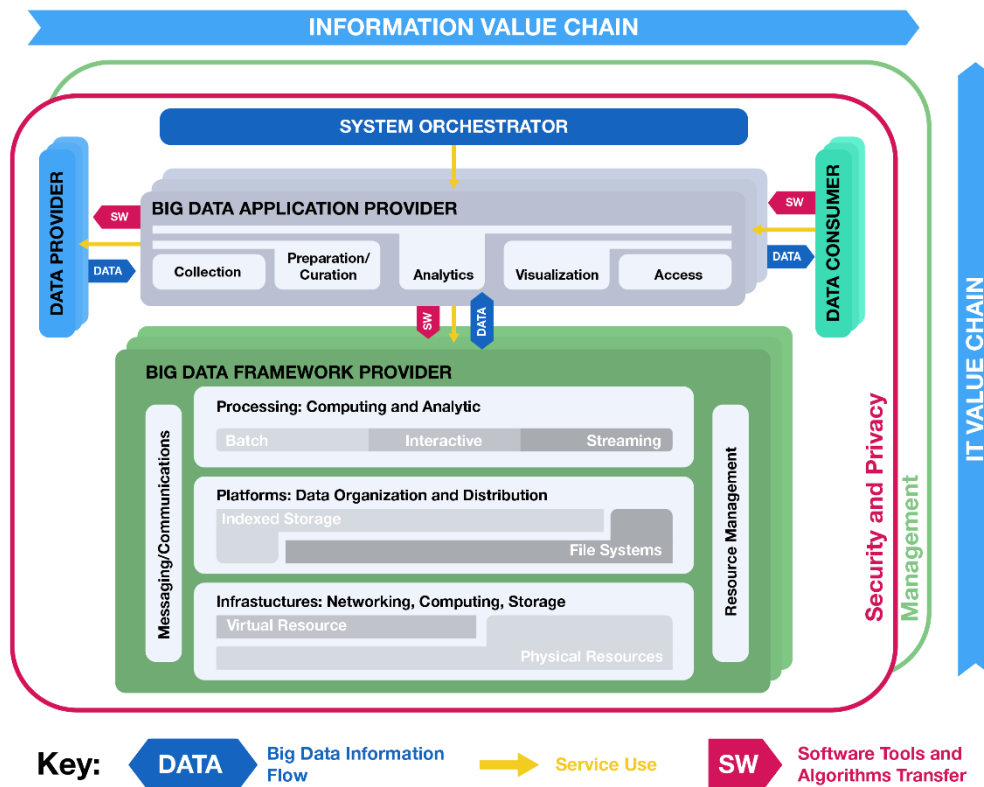


## (English) - Research on the NIST architecture.

The Big Data Reference Architecture, presented by NIST, describes a Big Data system structured into five main components, connected by services for interoperability: System Orchestrator, Data Provider, Application Provider, Big Data Framework Provider, and Data Consumer. These components operate under two essential dimensions: management and security/privacy, which encompass all functions.

The architecture is organized around two value axes: the Information Value axis, which involves creating value through data collection, integration, analysis, and application; and the Information Technology (IT) axis, which provides the infrastructure, platforms, and tools necessary for Big Data operation.



### 1. System Orchestrator

System Orchestrator manages and automatically coordinates the different elements of a Big Data environment, such as applications, data, and infrastructure components. Much like a conductor in an orchestra, the System Orchestrator ensures that all system elements work in harmony. This component uses automated workflows, management tools, and change processes to ensure that different elements within a Big Data environment interact seamlessly. It is essential for keeping the Big Data environment

organized and efficient, coordinating the flow and processing of data within the system.

## **2. Data Provider**

The Data Provider introduces data into the Big Data system from various sources, such as social networks, sensors, banking transactions, and other third-party systems. This function is responsible for ensuring data input from diverse sources and formats (text, images, audio, etc.), maintaining the security and privacy of information. Data is transmitted into the system in three phases: initiation (where authentication and preparation occur), data transfer (where data is effectively introduced into the system), and completion (where the success of the transfer is validated). The Data Provider is responsible for the data variety in the system, ensuring a continuous inflow of information.

## **3. Big Data Application Provider**

This component includes the business logic and functionalities needed to transform data into useful information. Activities within the Application Provider include data collection, preparation, analysis, visualization, and access to processed data. It offers a range of applications (or software) that adapt to the specific needs of each sector. For example, a financial company may have fraud detection or credit scoring software, while a manufacturing company may use applications for inventory management and supply chain optimization. This provider is essential for extracting value from data, turning it into actionable and practical outcomes for the organization.

## **4. Big Data Framework Provider**

The Big Data Framework Provider supplies the necessary infrastructure for storing and processing data. It is composed of three main layers:

- **Infrastructure:** Involves networking, computing, and storage, ensuring that large volumes of data can be stored and transferred efficiently, securely, and scalably.
- **Platform:** Provides tools to organize and distribute data, facilitating high-performance processing. Often, a Hadoop infrastructure is used, enabling distributed storage and processing.
- **Processing:** Performs data analysis using commands and algorithms to process large volumes of information. This layer is where "number crunching" occurs, transforming raw data into insights.

This component is vital for supporting large-scale Big Data operations, offering a robust and scalable structure.

## 5. Data Consumer

The Data Consumer can be either an end user or another system that uses the processed data. Its functions include data search and retrieval, download, local analysis, report generation, and visualization of results for application in internal processes. To obtain data, the Data Consumer uses the interfaces or services provided by the Application Provider, allowing access to reports, data retrieval, and graphical representation. The Data Consumer, therefore, completes the Big Data cycle, utilizing refined information to generate value and insights within its processes.

### Complementary Dimensions

In addition to the five main components, the Big Data Reference Architecture is organized around two essential dimensions:

- **Management:** Ensures coordinated operations and that changes are implemented efficiently.
- **Security and Privacy:** Ensure the protection and confidentiality of data, complying with standards and regulations.

These dimensions are fundamental to any Big Data implementation, ensuring that all components operate securely and efficiently.

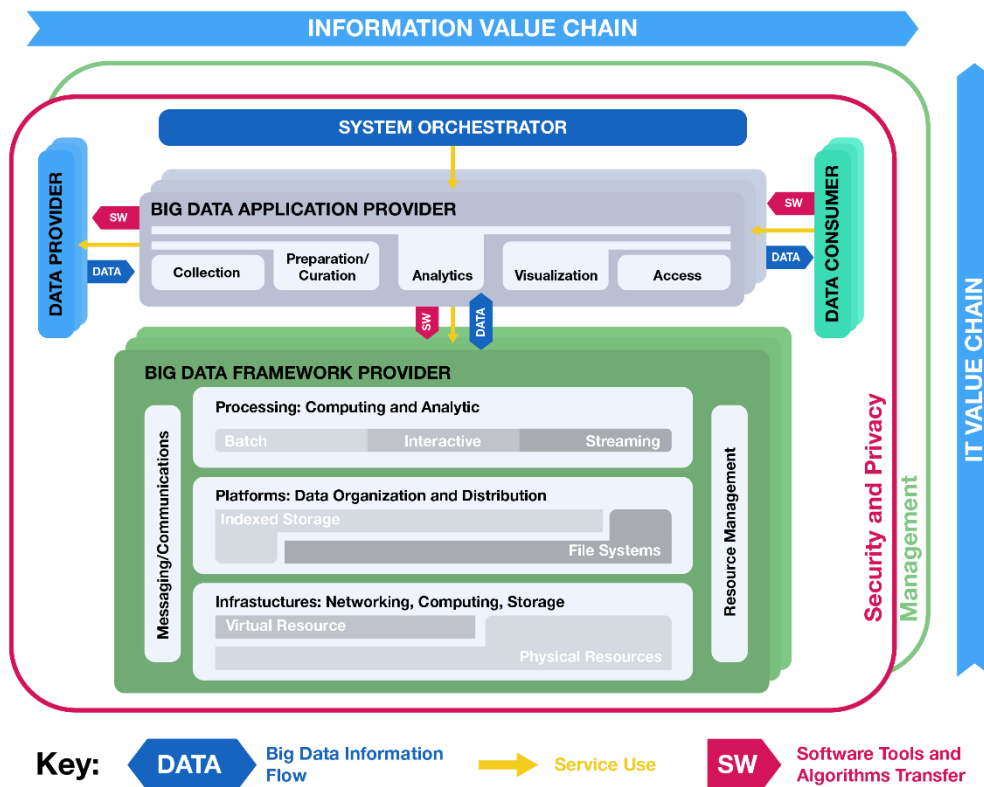
This architecture enables companies to manage large data volumes in an organized, scalable, and secure manner, adding value and providing valuable insights for decision-making.

<https://www.bigdataframework.org/big-data-architecture/>

## (Português) – Pesquisa sobre a arquitetura NIST

A Arquitetura de Referência de Big Data, apresentada pelo NIST, descreve um sistema de Big Data estruturado em cinco componentes principais, conectados por serviços para interoperabilidade: Orquestrador de Sistemas, Provedor de Dados, Provedor de Aplicações, Provedor de Estruturas de Big Data e Consumidor de Dados. Estes componentes operam sob duas dimensões essenciais: gestão e segurança/privacidade, que abrangem todas as funções.

A arquitetura está organizada em dois eixos de valor: o eixo do Valor da Informação, que envolve a criação de valor pela coleta, integração, análise e aplicação de dados; e o eixo de Tecnologia da Informação (TI), que fornece a infraestrutura, plataformas e ferramentas necessárias para a operação de Big Data.



### 1. Orquestrador de Sistemas

O Orquestrador de Sistemas gere e coordena automaticamente os diferentes elementos de um ambiente de Big Data, como aplicações, dados e componentes de infraestrutura. Assim como um maestro numa orquestra, o Orquestrador de Sistemas assegura que todos os elementos do sistema funcionem em sincronia. Este componente utiliza fluxos de trabalho automatizados, ferramentas de gestão e processos de mudança para garantir que os diferentes elementos de um ambiente de Big Data interajam harmoniosamente. É essencial para manter o

ambiente de Big Data organizado e eficiente, coordenando a forma como os dados fluem e são processados no sistema.

## **2. Provedor de Dados**

O Provedor de Dados introduz dados no sistema de Big Data a partir de várias fontes, como redes sociais, sensores, transações bancárias e outros sistemas de terceiros. Esta função é responsável por garantir a entrada de dados de diversas origens e formatos (texto, imagens, áudio, etc.), mantendo a segurança e privacidade das informações. Os dados são transmitidos ao sistema em três fases: iniciação (onde é feita a autenticação e a preparação), transferência de dados (onde os dados são efetivamente introduzidos no sistema) e término (onde é validado o sucesso da transferência). O Provedor de Dados é o responsável pela variedade de dados no sistema, garantindo a entrada contínua de informação.

## **3. Provedor de Aplicações de Big Data**

Este componente inclui a lógica de negócios e as funcionalidades necessárias para transformar dados em informações úteis. As atividades no Provedor de Aplicações incluem a recolha, preparação, análise, visualização e acesso aos dados processados. Ele oferece uma gama variada de aplicações (ou software) que se adaptam às necessidades específicas de cada setor. Por exemplo, uma empresa financeira pode ter software de deteção de fraudes ou de pontuação de crédito, enquanto uma empresa de produção pode utilizar aplicações de gestão de inventário e otimização de cadeias de abastecimento. Este provedor é essencial para extrair o valor dos dados, transformando-os em resultados práticos e acionáveis para a organização.

## **4. Provedor de Estruturas de Big Data**

O Provedor de Estruturas de Big Data fornece a infraestrutura necessária para o armazenamento e processamento dos dados. Ele é composto por três camadas principais:

- **Infraestrutura:** Envolve a rede, computação e armazenamento, garantindo que grandes volumes de dados possam ser armazenados e transferidos de maneira eficiente, segura e escalável.
- **Plataforma:** Fornece as ferramentas para organizar e distribuir dados, facilitando o processamento de alto desempenho. Em muitos casos, utiliza-se a infraestrutura Hadoop, que permite armazenamento e processamento distribuído.
- **Processamento:** Executa a análise de dados através de comandos e algoritmos que processam grandes volumes de informação. Esta camada é

onde a "análise dos números" ocorre, transformando dados brutos em insights.

Este componente é vital para suportar operações de Big Data em grande escala, oferecendo uma estrutura robusta e escalável.

## 5. Consumidor de Dados

O Consumidor de Dados pode ser tanto um utilizador final quanto outro sistema que utiliza os dados processados. Ele desempenha funções como pesquisa e recuperação de dados, download, análise local, criação de relatórios e visualização dos resultados para aplicação em processos internos. Para obter os dados, o Consumidor de Dados usa as interfaces ou serviços fornecidos pelo Provedor de Aplicações, permitindo o acesso a relatórios, recuperação de dados e representação gráfica. O Consumidor de Dados, portanto, completa o ciclo do Big Data, aproveitando a informação refinada para gerar valor e insights dentro de seus próprios processos.

## Dimensões Complementares

Além dos cinco componentes principais, a Arquitetura de Referência de Big Data é organizada em torno de duas dimensões essenciais:

- **Gestão:** Assegura que as operações estejam coordenadas e que as mudanças sejam implementadas com eficiência.
- **Segurança e Privacidade:** Garantem a proteção e confidencialidade dos dados, cumprindo com normas e regulamentos.

Essas dimensões são fundamentais para qualquer implementação de Big Data, garantindo que todos os componentes operem com segurança e eficiência.

Essa arquitetura permite às empresas gerir grandes volumes de dados de forma organizada, escalável e segura, agregando valor e oferecendo insights valiosos para a tomada de decisões.

<https://www.bigdataframework.org/big-data-architecture/>