



Data Science & Machine Learning

Aula 02

**BOAS VINDAS**

Consultor: Daniel Omar



# Daniel Omar

*Consultor de IA*  
**IBM**



Projetos com Volkswagen,  
Sul-américa, Bradesco,  
Sicredi e BB

---

DNC Group – Sponsor DEX,  
Consultor MBP e IP

# O que veremos neste módulo:

**01**

Definindo  
Plataformas de  
Data Science

**02**

Criando conta  
na AWS

**03**

Serviços  
para Data  
Engineering

**04**

Serviços para  
EDA

**05**

Modeling e  
Deploy com Sage  
Maker

# O que veremos neste módulo:

**06**

Criando conta  
na IBM Cloud

**07**

Serviços para  
Data  
Engineering

**08**

Serviços para  
EDA

**09**

Modeling e  
Deploy com  
Watson Studio

**10**

Resumo de  
plataformas



Data Science & Machine Learning

Aula 02

# **DEFININDO PLATAFORMAS DE DATA SCIENCE**

Consultor: Daniel Omar

# Definindo Plataformas

Softwares que ajudam cientistas de dados a desenvolver e deployar suas próprias soluções

---

Notebooks, storages, modelos prontos, autoML, ferramentas de monitoramento, batch job, APIs

---

Benefícios corporativos de plataformas também consistem em padronização corporativa, colaboração em projeto e centralização do desenvolvimento.

# On-Premise x Cloud

On-premise é quando o software é instalado e executado em computadores nas instalações da pessoa ou empresa.

---

A maioria das empresas utilizam plataformas de DSML disponibilizadas em Cloud.

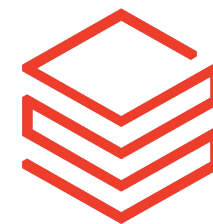
---

Redução de custo e manutenção de infra, disponibilização de recursos, evolução constante da plataforma.

# PLATAFORMAS DSML



Google Cloud



**databricks**







As of January 2021

© Gartner, Inc  
**Gartner.**

# AWS DSML



S3  
Bucket

Data  
Pipeline

Kinesis  
Data  
Stream

Glue

Athena

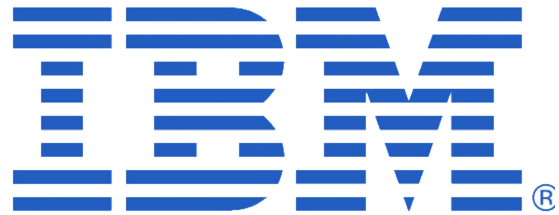
**Sage  
Maker**

EMR

Data  
Migration  
Service

Quick  
Insight

# IBM Cloud DSML



Data Stage

Cloud  
Object  
Storage

Cloud Pak  
for Data

SQL  
Query

**Watson  
Studio**

Analytics  
Engine

Streaming  
Analytics

Watson  
Open Scale

Cognos  
Dashboard

# RESUMO



**Plataformas de Data Science ajudam o cientista disponibilizando ferramentas**

---



**Agilidade, padronização e automação do ciclo de vida de ciência de dados**

---



**On-premise x Cloud**



## **3- Evaluate Results**

Consultor: Daniel Omar

# **Passo a passo para criar conta na AWS**

<https://aws.amazon.com/pt/premiumsupport/knowledge-center/create-and-activate-aws-account/>



Data Science & Machine Learning

## **4- Serviços para Data Engineering**

Consultor: Daniel Omar

# **S3 Bucket**



**Simple Storage Service, armazenamento de objetos de qualquer formato**

---



**Muito utilizado como um Data Lake**

---



**Espinha dorsal de storage da plataforma de ciência de dados da AWS, o SageMaker**



# **Criando um S3 na AWS**

Como criar uma instância de S3 na AWS

---

Subir um documento no S3

---

Visualizar um documento no S3

# **AWS Data Pipeline**



**Serviço para migrar dados de um lugar para outro, ETL**

---



**Destinos podem ser bancos MySQL, PostgreSQL, MongoDB e S3**

---



**Os data sources podem estar On-premise**

# **AWS Step Functions**



**Serviço visual para criar uma cadeia de passos**

---



**Para executar qualquer função**

---



**Construir uma Pipeline completa desde o banco de dados até o Sage Maker**

# **AWS Batch**



**Executa Batch Jobs como imagens Docker**

---



**Para executar qualquer função**

---



**Os data sources podem estar On-premise**



Data Science & Machine Learning

dnc.group

## **5- Serviços para EDA**

Consultor: Daniel Omar

# Athena



**Serviço de consulta aos dados do S3**

---



**Permite executar queries SQL**

---



**Alto desempenho para consultas de grandes bases de dados**

# AWS QuickSight



**Business Analytics e visualização na Cloud**

---



**Friendly User para não-técnicos**

---



**Vários formatos como CSV, Excel ou conectado a bancos SQL e principalmente ao S3 via Athena**

# **AWS Sage Maker**



**Fornece Jupyter Notebooks dentro da Cloud AWS**

---



**Permite toda a construção de um EDA com a liberdade das bibliotecas de python**





Data Science & Machine Learning

dnc.group

# **6- Preparation, Modeling e Deploy com Sage Maker**

Consultor: Daniel Omar

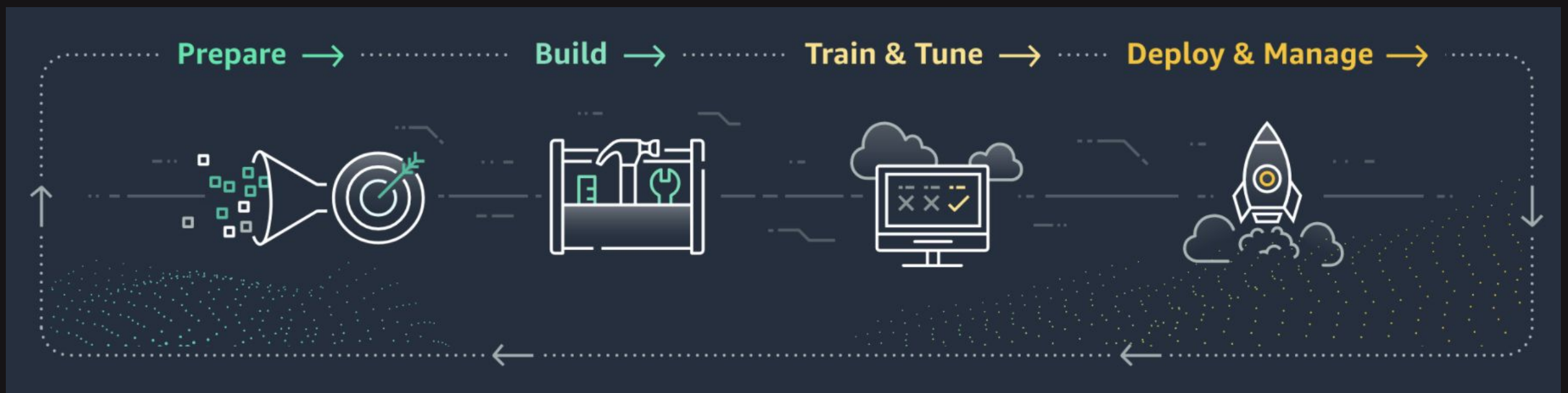
# AWS SageMaker



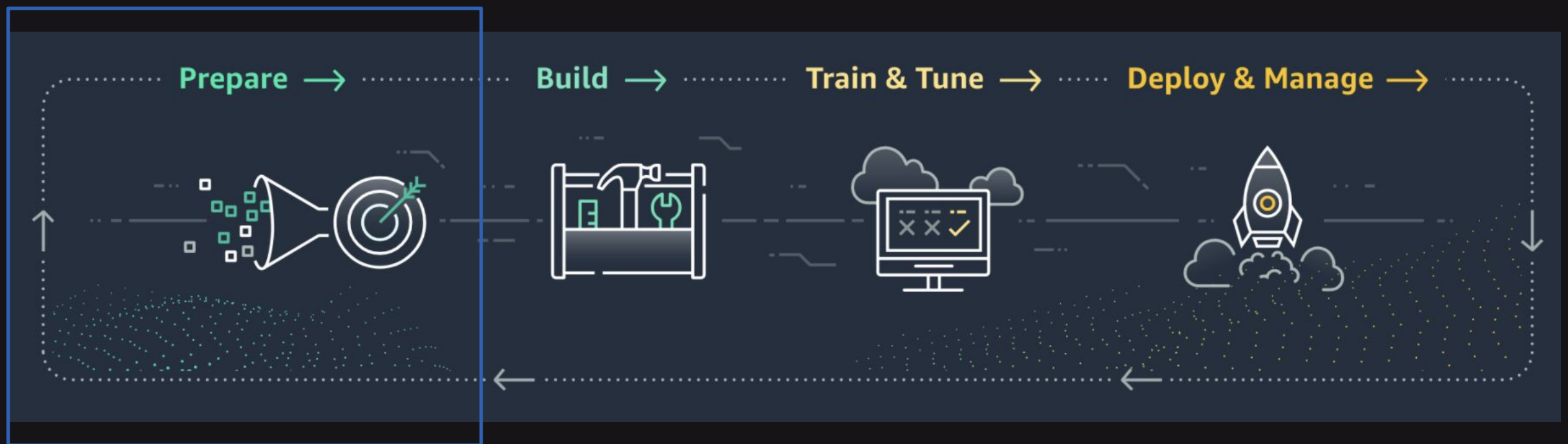
**Coração do DSML da AWS**



**Permite toda a construção do ciclo de vida de um projeto de Ciência de Dados**



# SageMaker Prepare



O Data Preparation pode ser construído e executado via Jupyter Notebook

# Data Wrangler



**Ferramenta Low-code**

---



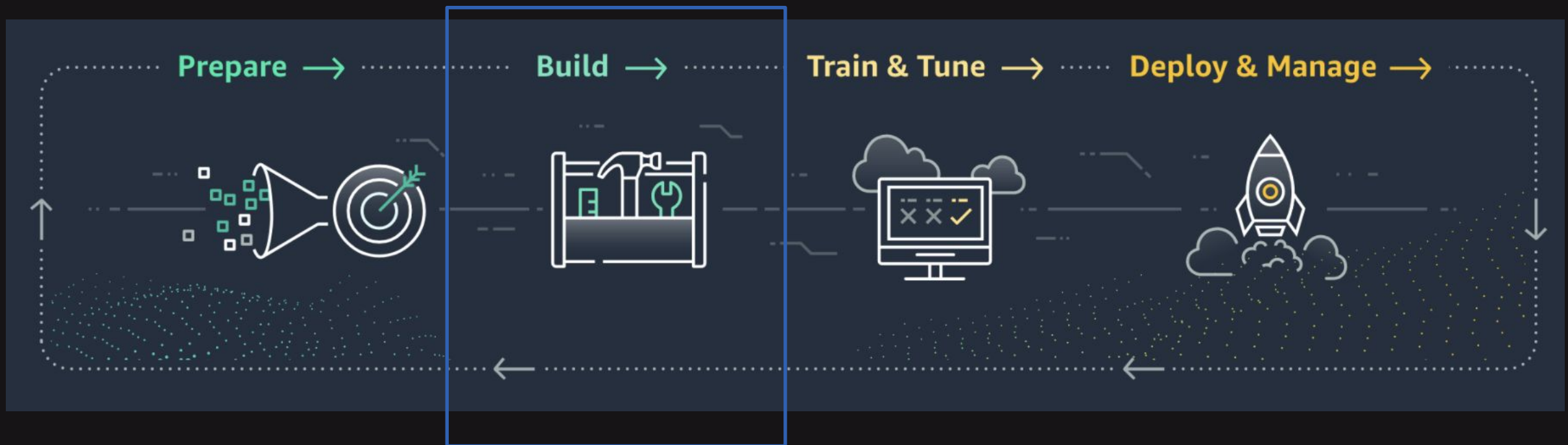
**Agregação e preparação dos dados**

---



**Mais de 300 transformações built-in**

# SageMaker Build



Todo o desenvolvimento do modelo pode ser realizado via Jupyter Notebook utilizando também de modelos built-in do SageMaker

---

Sage Maker também fornece serviços de AutoML que realiza o build e teste do modelo automaticamente

# **SageMaker Studio**



**Notebook Jupyter**

---



**Construção do seu modelo próprio**

# **SageMaker AutoPilot**



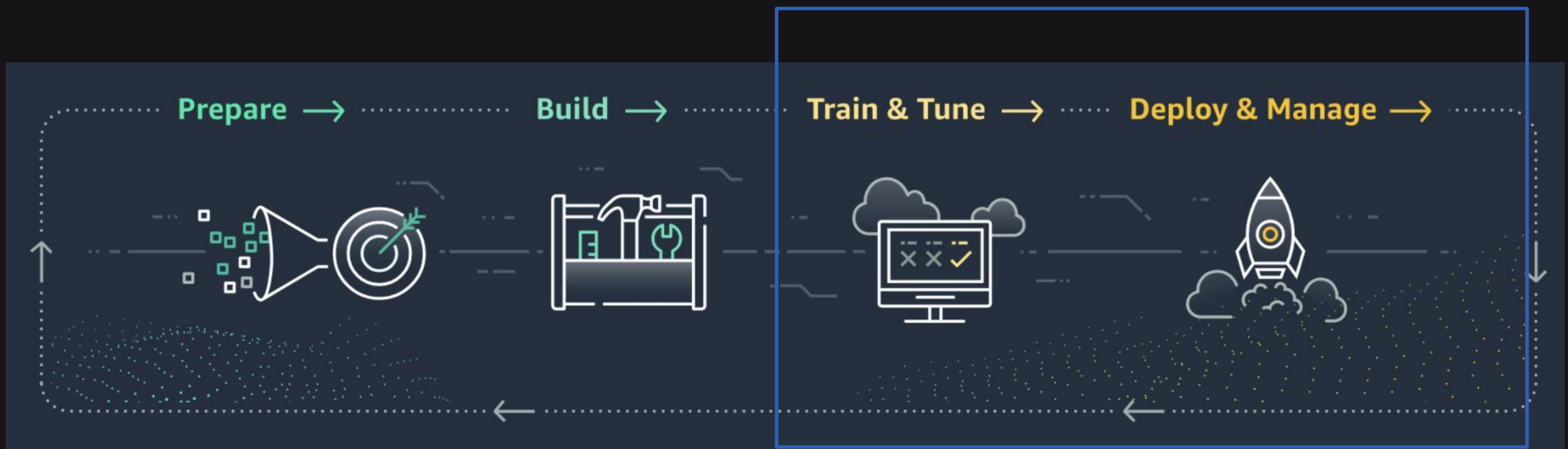
**AutoML da AWS**

---



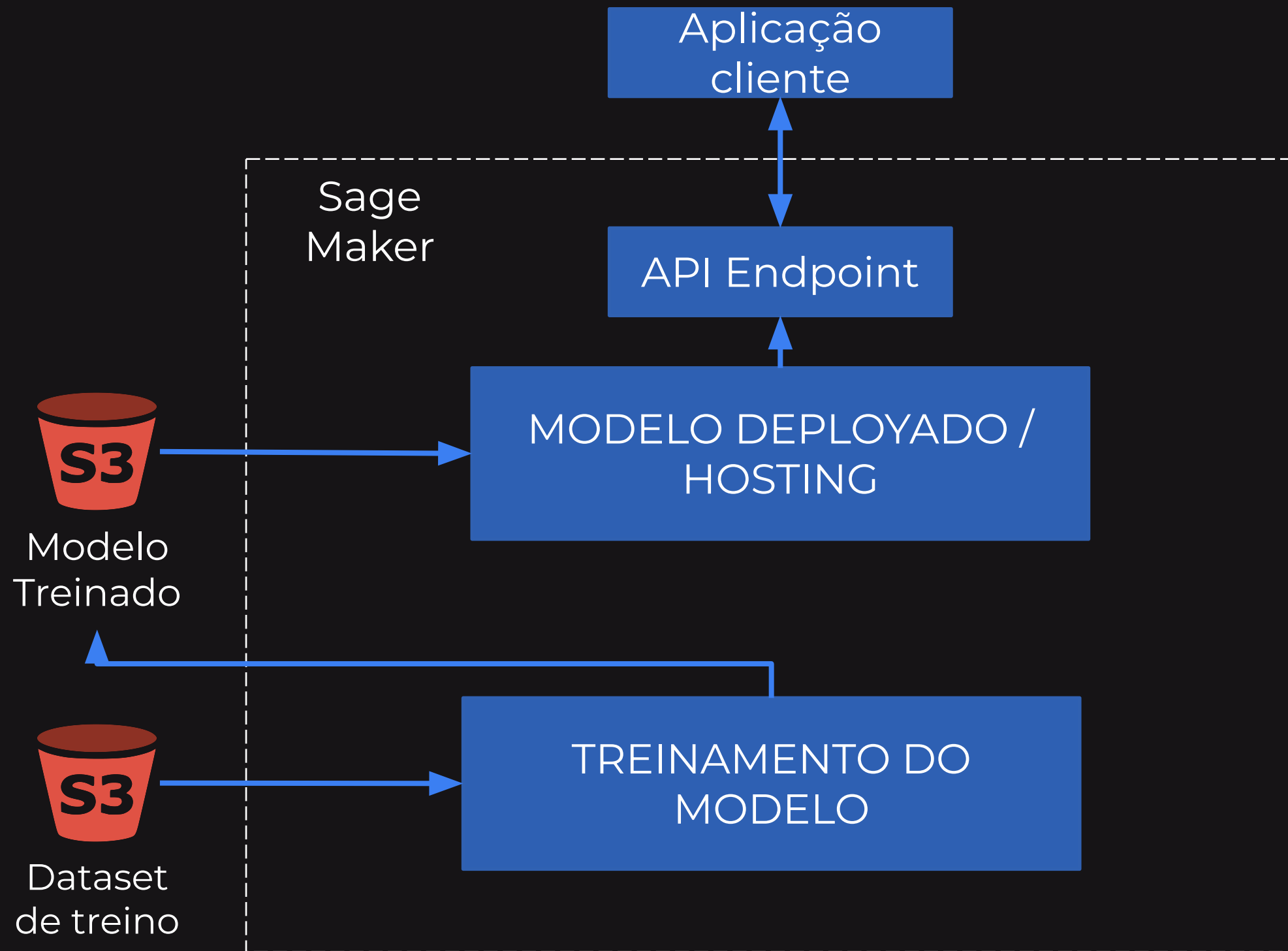
**Seleção e construção de um modelo**

# Train & Deploy





# Sage Maker Train & Deploy





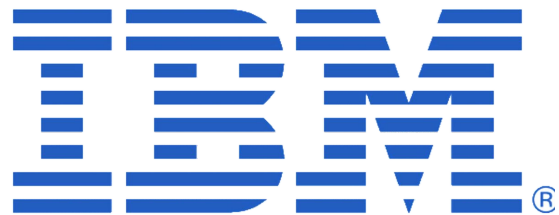
Data Science & Machine Learning

dnc.group

## **7- Criando conta na IBM Cloud**

Consultor: Daniel Omar

# IBM Cloud DSML



Data Stage

Cloud  
Object  
Storage

Cloud Pak  
for Data

SQL  
Query

**Watson  
Studio**

Analytics  
Engine

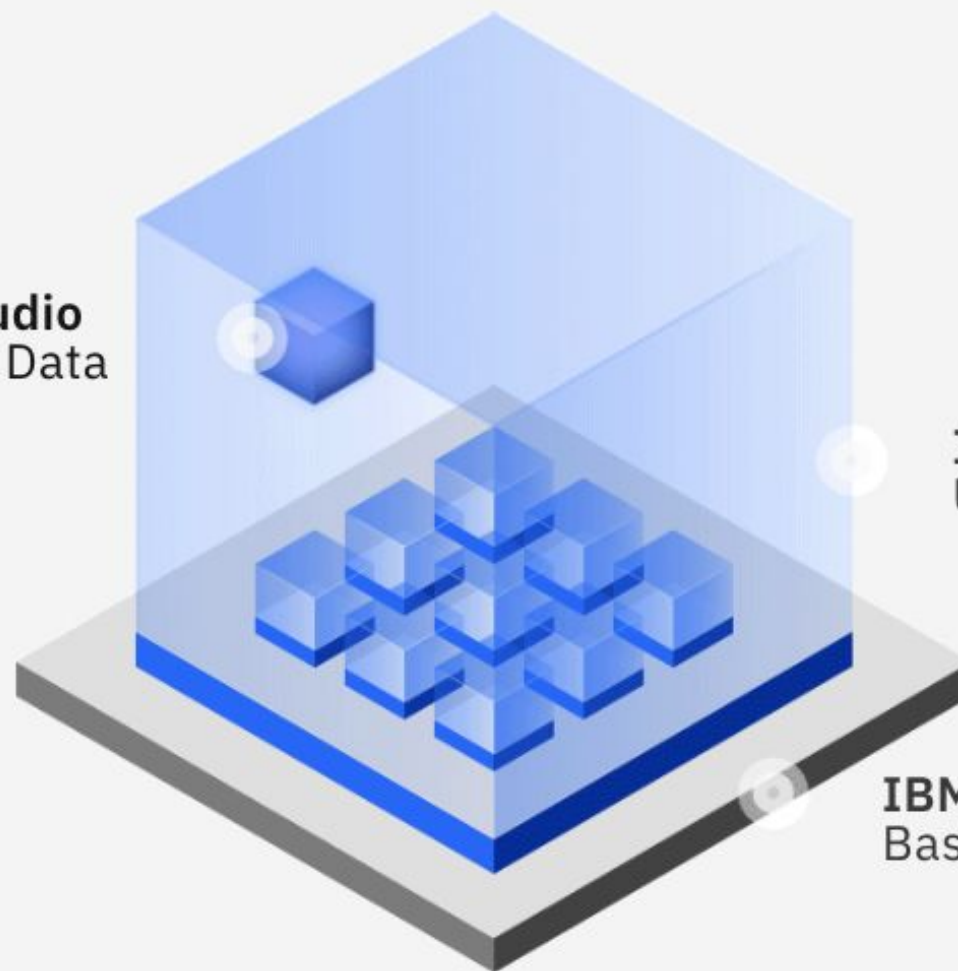
Streaming  
Analytics

Watson  
Open Scale

Cognos  
Dashboard

# IBM Cloud Pak for Data

**IBM Watson Studio**  
in Cloud Pak for Data



**IBM Cloud Pak for Data**  
Unifying platform

**IBM Cloud**  
Base cloud infrastructure

IBM Watson Studio is part of IBM Cloud Pak for Data and serves as the data science capability of the data fabric architecture.



Data Science & Machine Learning

dnc.group

## **8- Serviços de Data Engineering**

Consultor: Daniel Omar

# CoS



**Cloud Object Storage, armazenamento de objetos de qualquer formato**

---



**Muito utilizado como um Data Lake**

---



**Espinha dorsal de storage da plataforma de ciência de dados da IBM, o Watson Studio**

# Data Stage



**Serviço para migrar dados de um lugar para outro, ETL**

---



**Destinos podem ser bancos MySQL, PostgreSQL, MongoDB e COS**

---



**Permite a realização de transformações durante a migração**

# Data Virtualization



**Serviço para visualizar, acessar, manipular e analisar os dados.**

---



**Sem migrar os dados**

---



**Permite acesso a bancos On-premisse**





Data Science & Machine Learning

dnc.group

## **9- Serviços de EDA**

Consultor: Daniel Omar

# SQL Query



**Serviço de consulta aos dados do COS**

---



**Permite executar queries SQL**

---



**Alto desempenho para consultas de grandes bases de dados**

# Cognos Dashboard



**Business Analytics e visualização na Cloud**

---



**Friendly User para não-técnicos**

---



**Vários formatos como CSV, Excel ou conectado a bancos SQL e COS**

# Watson Studio



**Fornece Jupyter Notebooks dentro do IBM Cloud Pak for Data**

---



**Permite toda a construção de um EDA com a liberdade das bibliotecas de python**



dnc.group

# **10 – Modeling e Deploy com Watson Studio**

Consultor: Daniel Omar

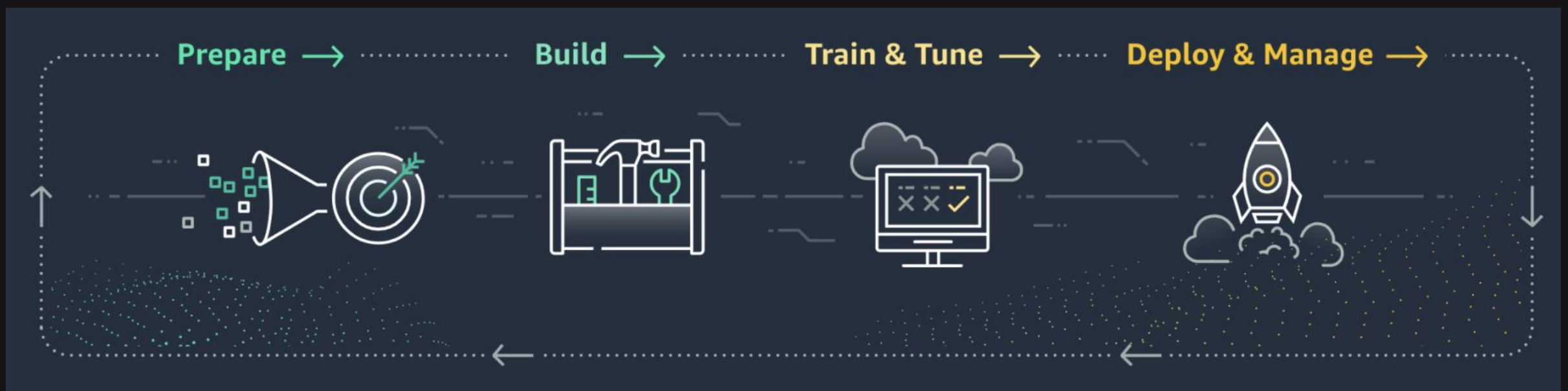
# Watson Studio



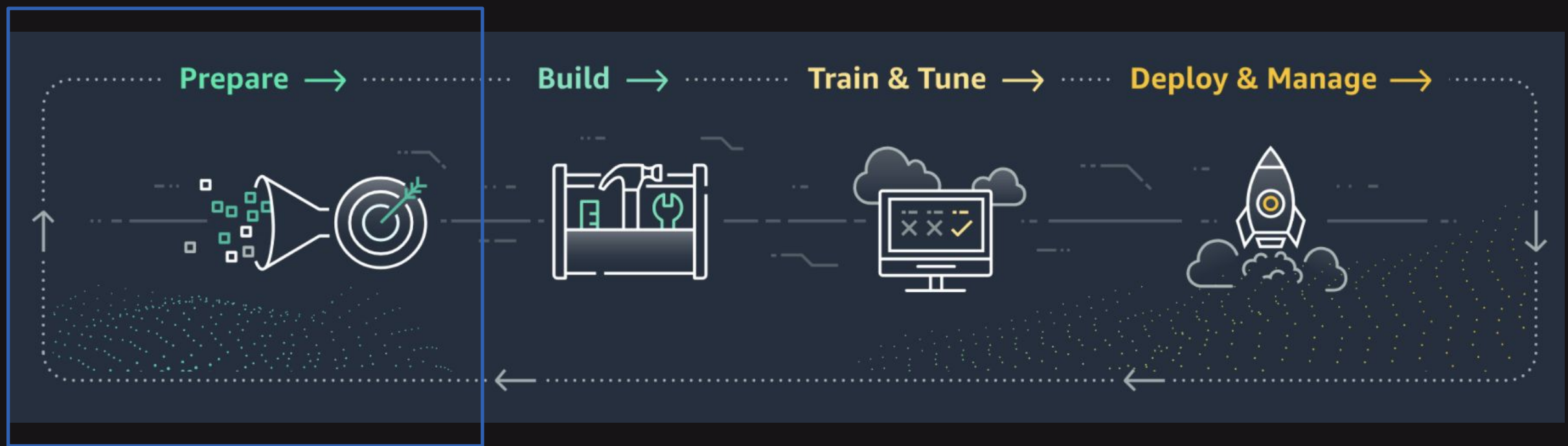
**Coração do Cloud Pak for Data**



**Permite toda a construção do ciclo de vida de um projeto de Ciência de Dados**



# Cloud Pak for Data Prepare



○ Data Preparation pode ser construído e executado via Jupyter Notebook

---

# Data Stage Flow



**Ferramenta Low-code**

---



**Diversas transformações built-in**

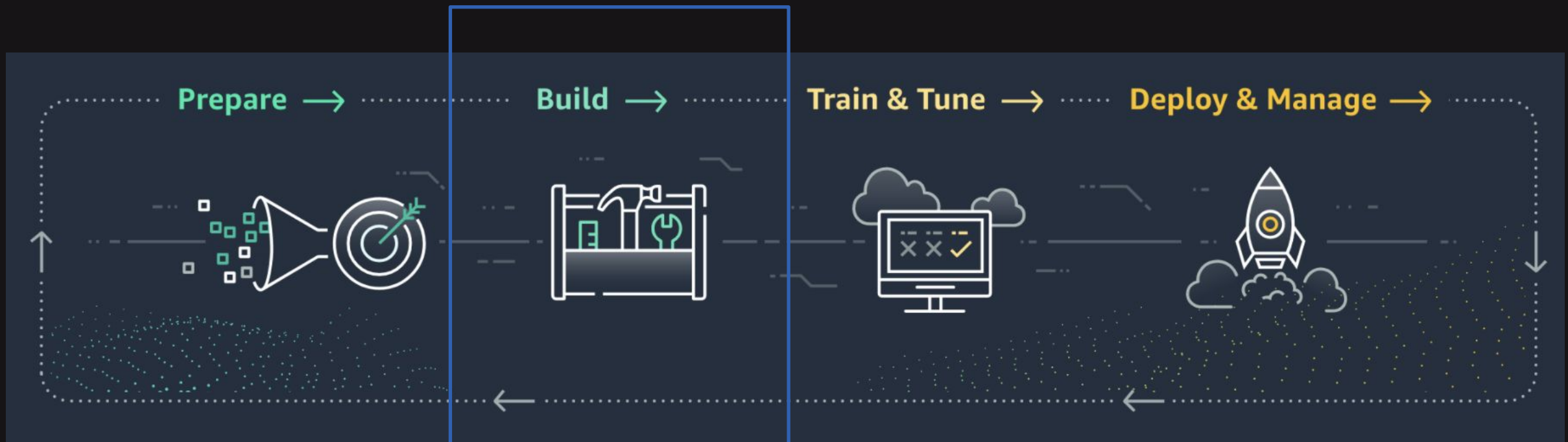
---



**Agregação e preparação dos dados de diversas bases diferentes**



# Cloud Pak for Build



Todo o desenvolvimento do modelo é realizado via Jupyter Notebook ou utilizando modelos built-in com o SPSS Modeler

---

Cloud Pak for Data também fornece serviços de AutoML que realiza o build e teste do modelo automaticamente

# Watson Studio



**Notebook Jupyter**

---



**Construção do próprio Data Preparation**

# **AutoML Experiment**



**AutoML do Cloud Pak for Data**

---



**Seleção e construção de um modelo**

# SPSS Modeler



**Ferramenta de análise estatística avançada**

---



**Incorporou etapas de preparação em Low Code**

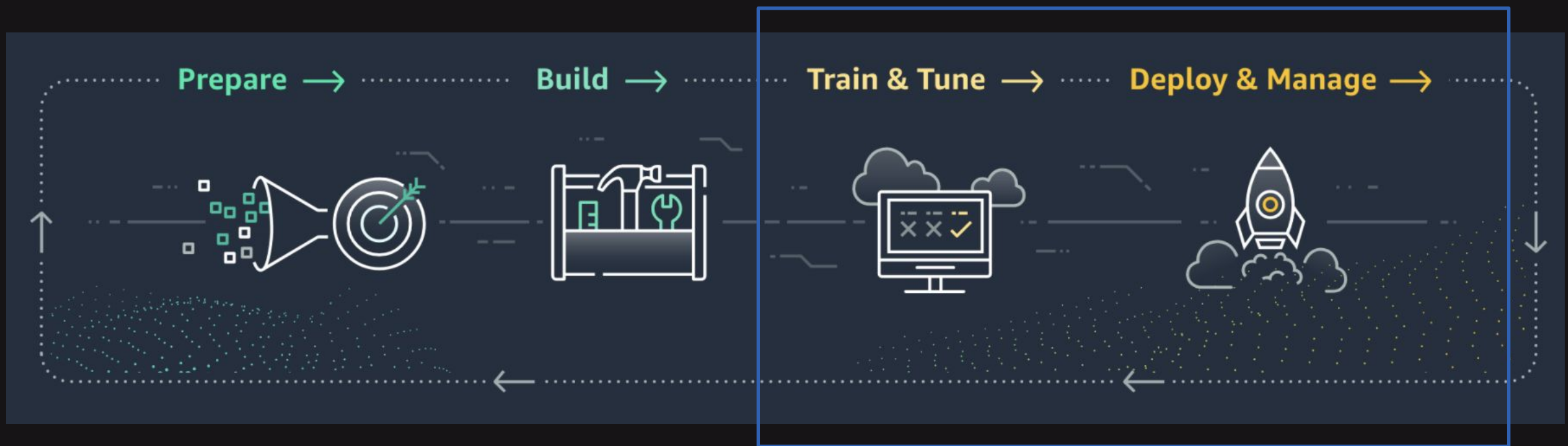
---



**Construção do ciclo completo de ML**

# Cloud Pak for Data

## Train & Deploy



# Watson Machine Learning



**Serviço que transforma o modelo em API**

---



**Utilizado também para o autoML**

# Pipelines de Deploy



**Construção de uma Pipeline para o Deploy**

---



**Deploy em API ou em Batch**

# Watson Open Scale



**Monitoramento de modelos**

---



**Análise de vieses e fairness**

---

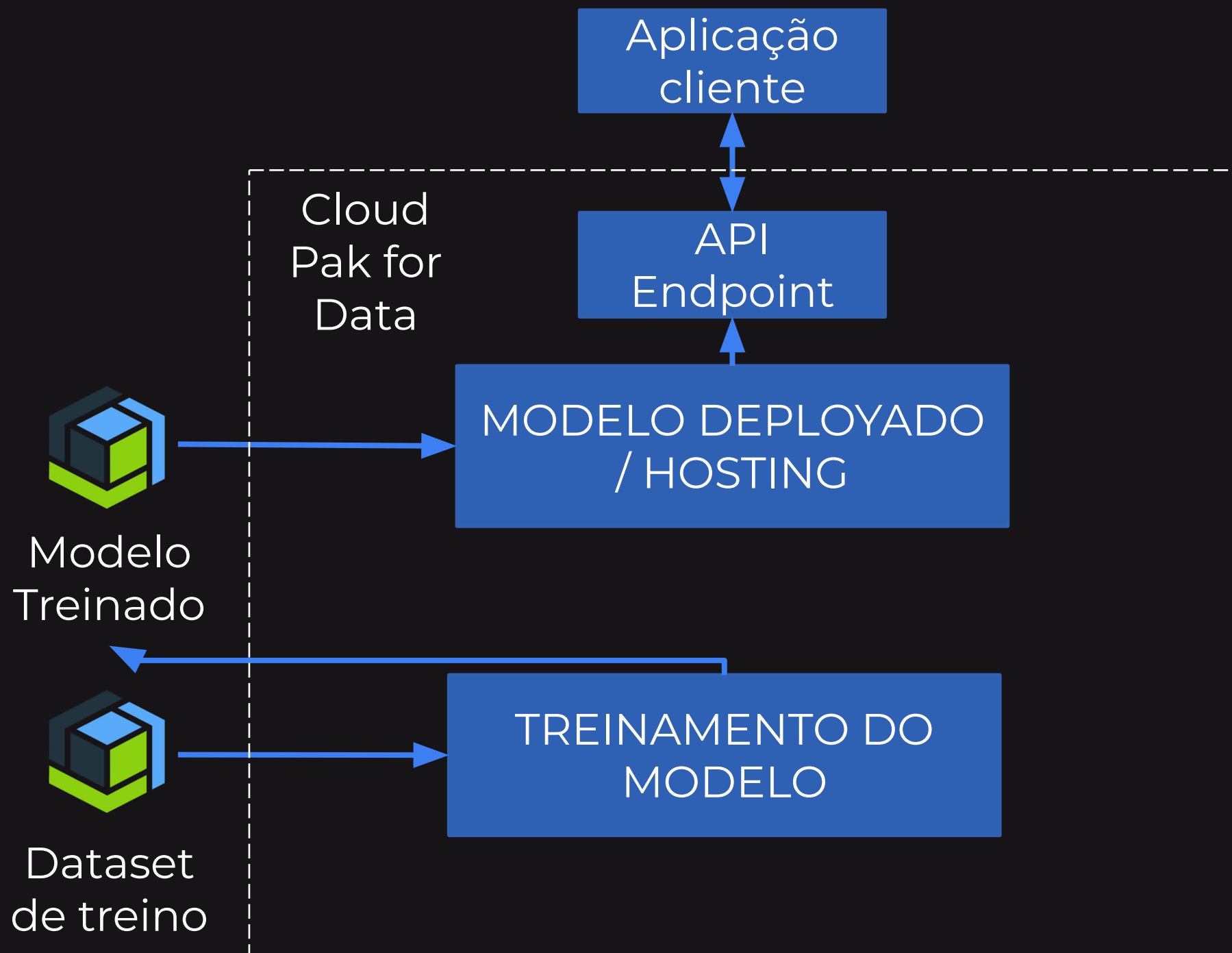


**Manter o modelo e gerar confiança na utilização**



# Cloud Pak for Data

## Train & Deploy





dnc.group

# **11 – Resumo plataformas**

Consultor: Daniel Omar

# RESUMO



**Plataforma de DSML e vantagens**

---



**AWS Sage Maker**

---



**Cloud Pak for Data e Watson Studio**

**Modelos de slides pra ajudar!**

**FAÇA UMA CÓPIA E NÃO  
MODIFIQUE DAQUI PRA BAIXO!!!**

# Título vai aqui assim

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce finibus nunc pulvinar placerat varius. Vivamus mi nisi, placerat aliquam erat sed, auctor ultricies.



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.

# Título vai aqui assim



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.



Lorem ipsum  
dolor sit amet,  
consectetur  
adipiscing elit.

# **Título vai aqui assim**



**Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce finibus nunc pulvinar placerat varius.**

---



**Vivamus mi nisi, placerat aliquam erat sed, auctor ultricies.**

---



**Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce finibus nunc pulvinar placerat varius.**

**Título vai aqui**

**se forem duas linhas**

**Lorem ipsum dolor sit amet, consectetur  
adipiscing elit. Fusce finibus nunc pulvinar  
placerat varius. Vivamus mi nisi, placerat  
aliquam erat sed, auctor ultricies.**

---

**Lorem ipsum dolor sit amet, consectetur  
adipiscing elit. Fusce finibus nunc pulvinar  
placerat varius. Vivamus mi nisi, placerat  
aliquam erat sed, auctor ultricies.**



**Título vai aqui**

**se forem duas linhas**

**Lorem ipsum dolor sit amet, consectetur adipiscing elit. Fusce finibus nunc pulvinar placerat varius. Vivamus mi nisi, placerat aliquam erat sed, auctor ultricies.**

- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Fusce finibus nunc pulvinar placerat varius.
- Vivamus mi nisi, placerat aliquam erat.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Fusce finibus nunc pulvinar lorem ipsum

# **Título vai aqui**

## **se forem duas linhas**

- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Fusce finibus nunc pulvinar placerat varius.
- Vivamus mi nisi, placerat aliquam erat.
- Lorem ipsum dolor sit amet, consectetur adipiscing elit.
- Fusce finibus nunc pulvinar lorem ipsum



# **Título vai aqui**

## **se forem duas linhas**

Lorem ipsum dolor sit amet,  
consectetur adipiscing elit.

---

Fusce finibus nunc pulvinar  
placerat varius.

---

Vivamus mi nisi, placerat  
aliquam erat.

---

Lorem ipsum dolor sit amet,  
consectetur adipiscing elit.

---

Fusce finibus nunc pulvinar  
lorem ipsum



**ÍCONES**

