

## Do Relacional ao Big Data: como integrar MySQL em arquiteturas modernas de dados e inteligência artificial

Explorando a integração de MySQL em arquiteturas de dados e inteligência artificial, com foco em Big Data e análise preditiva.



# Introdução

Nesta palestra, vou compartilhar minha experiência de mais de 29 anos em tecnologia, começando como DBA Oracle e evoluindo para projetos de ciência de dados e Inteligência Artificial. Farei uma análise prática de como integrar MySQL em arquiteturas modernas de dados, destacando a importância de pipelines de Big Data e análises preditivas.





# Quem sou eu

Sou Alexandre Barcelos, com mais de 29 anos de experiência em grandes empresas de outsourcing e mais de 20 anos no ensino superior. Minha especialização abrange Banco de Dados e Big Data, atuando como professor e coordenador do curso de Engenharia de Software na FIAP, além de ser autor de um livro sobre Modelagem de Banco de Dados e reconhecido pela Oracle por minhas contribuições na área.





### Servidores em Cloud

A arquitetura de dados inclui dois servidores em cloud: um com MySQL e o outro com Python, e o Azure com Hadoop e Spark. Essa configuração permite uma integração eficiente entre fontes de dados relacionais e ambientes de Big Data.



### Armazenamento de Dados em Hadoop

No servidor Hadoop, os dados são armazenados em arquivos CSV, como 'clientes', 'produtos.csv' e 'vendas'. Esses arquivos são a base para a extração e análise preditiva.



### Exportação de Dados

Exportação de dados do Hadoop para o MySQL, permitindo a integração de dados de Big Data com análises preditivas realizadas em Python.



### Análise Preditiva com Python

A análise preditiva é realizada no servidor com Anaconda Python, utilizando bibliotecas como Logistic para treinar modelos de Machine Learning que ajudam a prever o churn.

## Arquitetura de Pipeline de Dados



# Servidores em Cloud

A arquitetura de dados inclui dois servidores em cloud: um com MySQL 8 e Anaconda Python, e o outro com Hadoop e NiFi. Essa configuração permite uma integração eficiente entre bancos de dados relacionais e ambientes de Big Data.





# Armazenamento de Dados em Hadoop

No servidor Hadoop, os dados são armazenados em arquivos CSV. Esses arquivos são a base para a extração e análise preditiva.



# Exportação de Dados

Exportação de dados do Hadoop para o MySQL, permitindo a integração de dados de Big Data com análises preditivas realizadas em Python.



# Análise Preditiva com Python

A análise preditiva é realizada no servidor com Anaconda Python, utilizando bibliotecas como Scikit-learn para treinar modelos de Machine Learning que ajudam a prever o churn.



# Servidores em Cloud



Dataserver: MySQL 8



Dataserver: Anaconda  
Python



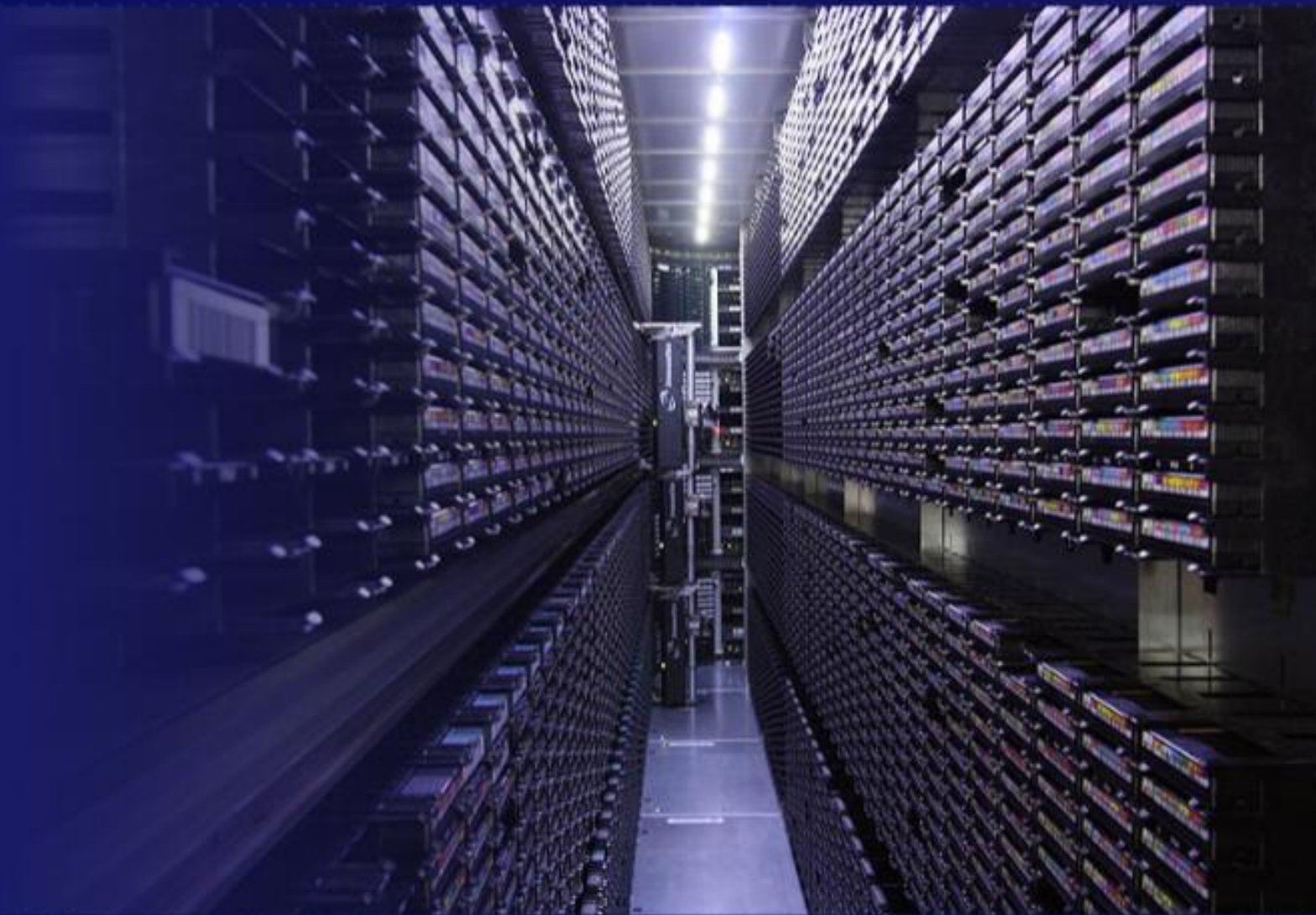
Hadoop: Hadoop



Hadoop: NiFi

# Big Data com Hadoop

Os dados estão armazenados no Hadoop, permitindo acesso eficiente para processamento e análise. Com essa estrutura, é possível integrar dados de várias fontes, facilitando a construção de pipelines de dados robustos e escaláveis.





## Exportação de Dados do Hadoop

A exportação de dados do Hadoop para MySQL é realizada utilizando a ferramenta NIFI, que facilita a transferência de grandes volumes de dados entre sistemas de Big Data e bancos de dados relacionais.

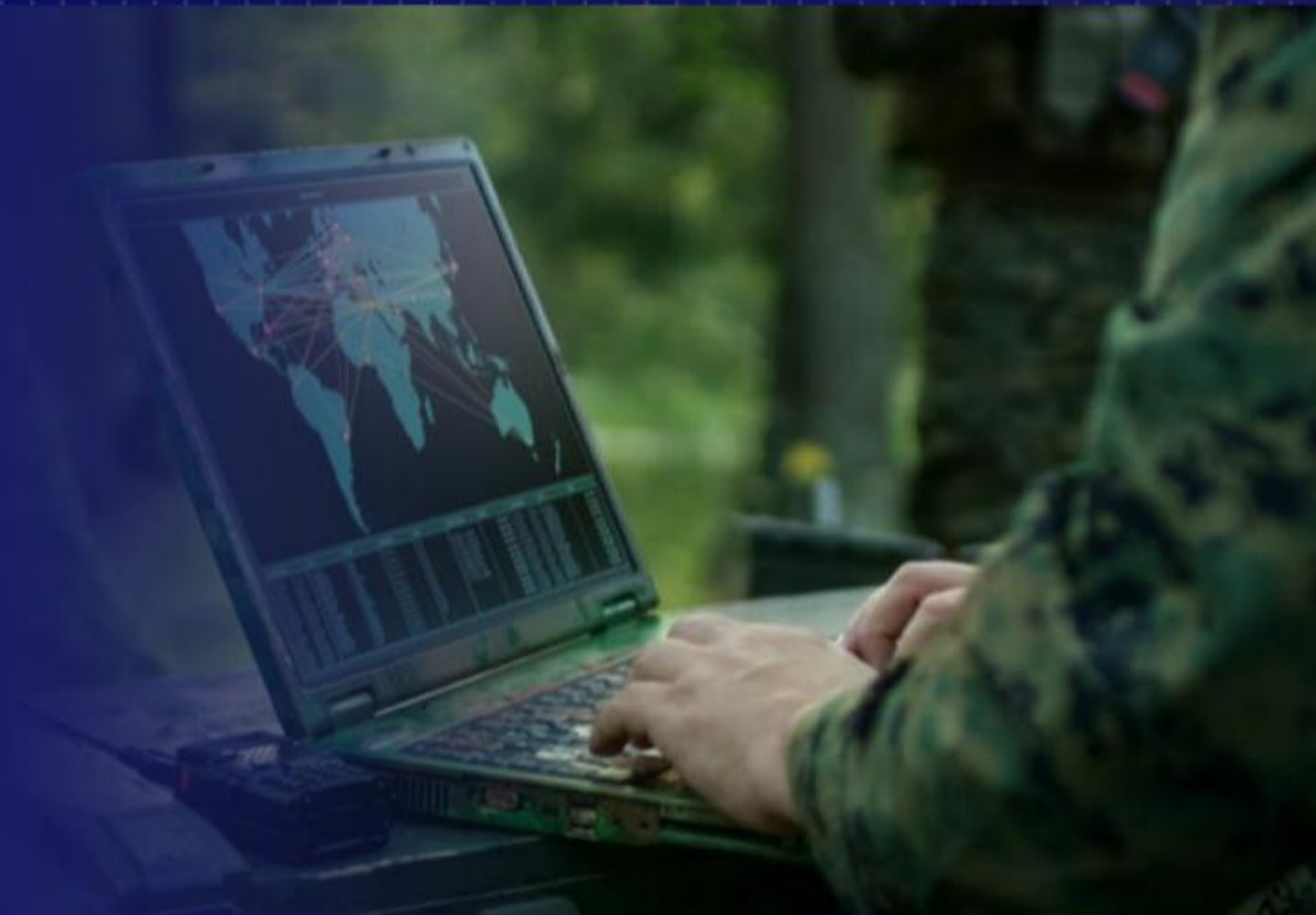


# Exportação de Dados



# Análise Preditiva com Python

Utilizamos o Anaconda Python para desenvolver análises preditivas a partir de dados extraídos do MySQL. O processo abrange desde a conexão ao banco de dados até a implementação de modelos de machine learning, permitindo previsões eficazes e insights valiosos.



# Projeto de Previsão de Churn com Machine Learning

Este projeto visa prever a rotatividade de clientes utilizando um pipeline de dados que integra MySQL e técnicas avançadas de aprendizado de máquina. Através da análise de dados históricos, o modelo identifica padrões que possibilitam prever quais clientes estão em risco de cancelamento, otimizando estratégias de retenção.





## Pipeline de Pré- processamento





# Treinamento e Avaliação de Modelos

Processo de construção e avaliação de modelos preditivos no contexto de análise de churn.

## Fase 1: Treinamento

Os dados de treinamento são preparados e os modelos de machine learning são treinados usando algoritmos como Regressão Logística, Random Forest e Gradient Boosting.

## Fase 2: Avaliação

Os modelos treinados são avaliados em um conjunto de dados de teste para medir sua acurácia, AUC-ROC e gerar relatórios de classificação.

## Fase 3: Seleção do Melhor Modelo

O modelo com melhor desempenho é selecionado com base nas métricas de avaliação, pronto para ser utilizado na produção.

# Fase 1:

# Treinamento

Os dados de treinamento são preparados e os modelos de machine learning são treinados usando algoritmos como Regressão Logística, Random Forest e Gradient Boosting.

## Fase 2: Avaliação

Os modelos treinados são avaliados em um conjunto de dados de teste para medir sua acurácia, AUC-ROC e gerar relatórios de classificação.



## Fase 3: Seleção do Melhor Modelo

O modelo com melhor desempenho é selecionado com base nas métricas de avaliação, pronto para ser utilizado na produção.

# Resultados e Melhor Modelo

Analizamos o desempenho dos diferentes modelos de Machine Learning aplicados aos dados de churn. O modelo de 'Gradient Boosting' apresentou o melhor desempenho, com uma AUC-ROC de XX % e uma acurácia de YY%, elevando a precisão das previsões de churn.





# Produção e Previsões

No servidor com Anaconda Python, conectamos ao MySQL para extrair dados e realizar análises preditivas. Após o pré-processamento, o modelo de machine learning é aplicado para prever churn, permitindo decisões informadas com base nos resultados gerados.





# O Papel do MySQL em Arquiteturas de Big Data

O MySQL se destaca como uma solução robusta para a integração em pipelines de dados modernos, possibilitando a análise e a previsão de churn de forma eficiente. À medida que mais empresas adotam Big Data e inteligência artificial, o MySQL se torna um componente essencial, permitindo a manipulação de grandes volumes de dados com facilidade e agilidade.





# Obrigado



[www.linkedin.com/in/alexandrebarcelos/](https://www.linkedin.com/in/alexandrebarcelos/)  
[profalexandre.barcelos@fiap.com.br](mailto:profalexandre.barcelos@fiap.com.br)

