Dominando o Amazon SageMaker: Coleta de Dados e Preparação com o Amazon SageMaker

Carolina Junqueira Ferreira

Startup Solutions Architect
AWS

Agenda

- Desafios de dados em ML
- Amazon Sagemaker
- Amazon Sagemaker Canvas
- Amazon Sagemaker Studio Notebooks

Tour de recursos do Amazon SageMaker

PREPARE DADOS E CRIE, TREINE E IMPLANTE MODELOS DE ML PARA QUALQUER CASO DE USO

PREPARAR

Geoespacial
Visualize dados geoespaciais

Ground Truth
Crie conjuntos de dados de alta qualidade para ML

Data Wrangler
Agregue e prepare dados para ML

Spark integrado
Integração integrada com o Amazon EMR e o AWS Glue

Processing
Python embutido, BYO R/Spark

Feature Store
Armazene, catalogue, pesquise e reutilize features

Clarify
Detecte viés e entenda as previsões do modelo

CONSTRUIR

Notebooks Studio e instâncias de notebooks
Notebooks Jupyter totalmente gerenciados com computação elástica

Editor de código
Editor de código baseado no VS Code Open Source (Code-OSS)

RStudio
FRStudio totalmente gerenciado

Studio Lab
FAmbiente de desenvolvimento de ML gratuito

Algoritmos integrados
Algoritmos tabulares, de PNL e de visão integrados

Jump Start
Descoberta, treinamento e implantação de modelos, soluções e exemplos baseados em UI

Autopilot
Crie automaticamente modelos de ML com visibilidade total

Traga o seu
Traga seu próprio contêiner e algoritmos

Modo local
Teste e crie protótipos em sua máquina local

TREINAR E SINTONIZAR

Treinamento totalmente gerenciado
Amplas opções de hardware, fáceis de configurar e escalar

Bibliotecas de treinamento distribuídas
Treinamento de alto desempenho para grandes conjuntos de dados
e modelos

Compilador de treinamento
Treinamento mais rápido em modelos de aprendizado profundo

Ajuste automático do modelo
Otimização de hiperparâmetros

Treinamento local gerenciado
Reduza o custo de treinamento em até 90%

Depurador e Profiler
O treinamento de depuração e perfil é executado

Experiments
Rastreie, visualize e compartilhe artefatos do modelo
entre equipes

Suporte de personalização
Integre-se com estruturas populares de código aberto
e bibliotecas

IMPLANTAR E GERENCIAR

Implantação totalmente gerenciada
Latência ultrabaixa e inferência de alto rendimento

Inferência em tempo real
Para padrões de tráfego estáveis

Inferência sem servidor
Para ipadrões de tráfego intermitente

Inferência assíncrona
Para cargas úteis grandes para longos tempos de processamento

Transformação em lote
Fou inferência off-line em lotes de grandes conjuntos de dados

Endpoints de vários modelos
Reduza os custos hospedando vários modelos por instância

Endpoints de vários contêineres
Reduza os custos hospedando vários contêineres
por instância

Teste de sombra
Valide o desempenho do modelo na produção

Recomendador de inferência
UMSelecione automaticamente uma instância de computador e configuração

Model monitor
Mantenha a precisão dos modelos implantados

Operadores e componentes do Kubernetes
Gerencie e monitore modelos em dispositivos periféricos

Gerente de borda
Gerencie e monitore modelos em dispositivos periféricos

MLOPs: Pipelines|Project| Model Registry
Automação do fluxo de trabalho, CI/CD para ML, catálogo central de modelos

Amazon CodeWhisperer
Acelere a construção comSugestões de código baseadas em IA

Canvas
Gere previsões precisas de aprendizado de máquina, sem necessidade de código

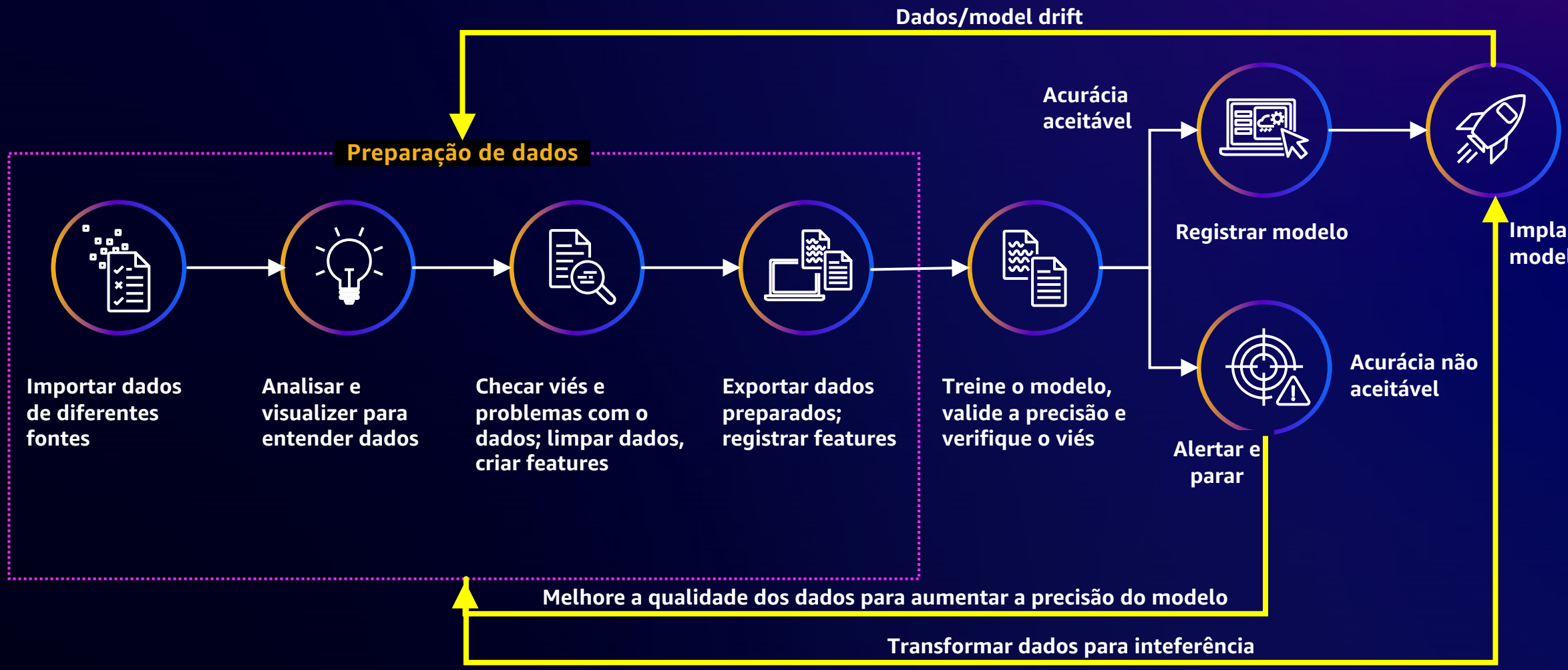
Studio
Plataforma com váriosambientes de desenvolvimento integrados (IDE) para ML e ciência de dados

Governança
Cartões modelo | Painel | Permissões

Amazon SageMaker ajuda as organizações a aproveitar o ML

A preparação de dados ocupa de 60% a 80% do tempo do projeto de ML

A TAREFA MAIS CRÍTICA E DEMORADA



Superando as barreiras do ML

Construtores de ML insuficientes

Ferramentas de ML no-code

Faça previsões de ML independentemente da experiência em ML

Processamento e rotulagem de grandes volumes de dados para ML

Ferramentas de preparação de dados específicas

Rotule e prepare dados para ML

Ferramentas diferentes de ciência de dados

Ferramentas de ML integradas em uma única interface

Crie, treine e implante modelos usando IDEs

Operações manuais de ML

Recursos integrados de MLOps

Implemente práticas de MLOps para agilizar o ciclo de vida do ML

O Amazon SageMaker traz ferramentas para cada etapa do ciclo de vida do ML em uma interface de usuário visual unificada

Amazon Sagemaker Canvas

A maneira mais rápida e fácil de preparar dados para ML no SageMaker Canvas

Preparação de dados ponta a ponta

Importação de dados, análise, limpeza, engenharia de recursos e produção

No code/low code

Análises visuais de dados, mais de 300 transformações integradas, biblioteca de códigos personalizados

Fácil de escalar

Apoiado por trabalhos de processamento Apache Spark sem servidor

Fácil de produzir

Treine diretamente no Canvas, agende trabalhos e automatize fluxos de trabalho de preparação de dados com SageMaker Pipelines

Como funciona a preparação de dados no Canvas

Amazon SageMaker Canvas

A maneira mais fácil e rápida de preparar visualmente os dados para o aprendizado de máquina

Importe dados

de mais de 50 fontes de dados usando interface visual ou SQL

Analise

Explore e analise dados com visualizações e análises integradas

Limpe e enriqueça

Execute a engenharia de recursos com mais de 300 transformações de dados integradas

Melhore a precisão

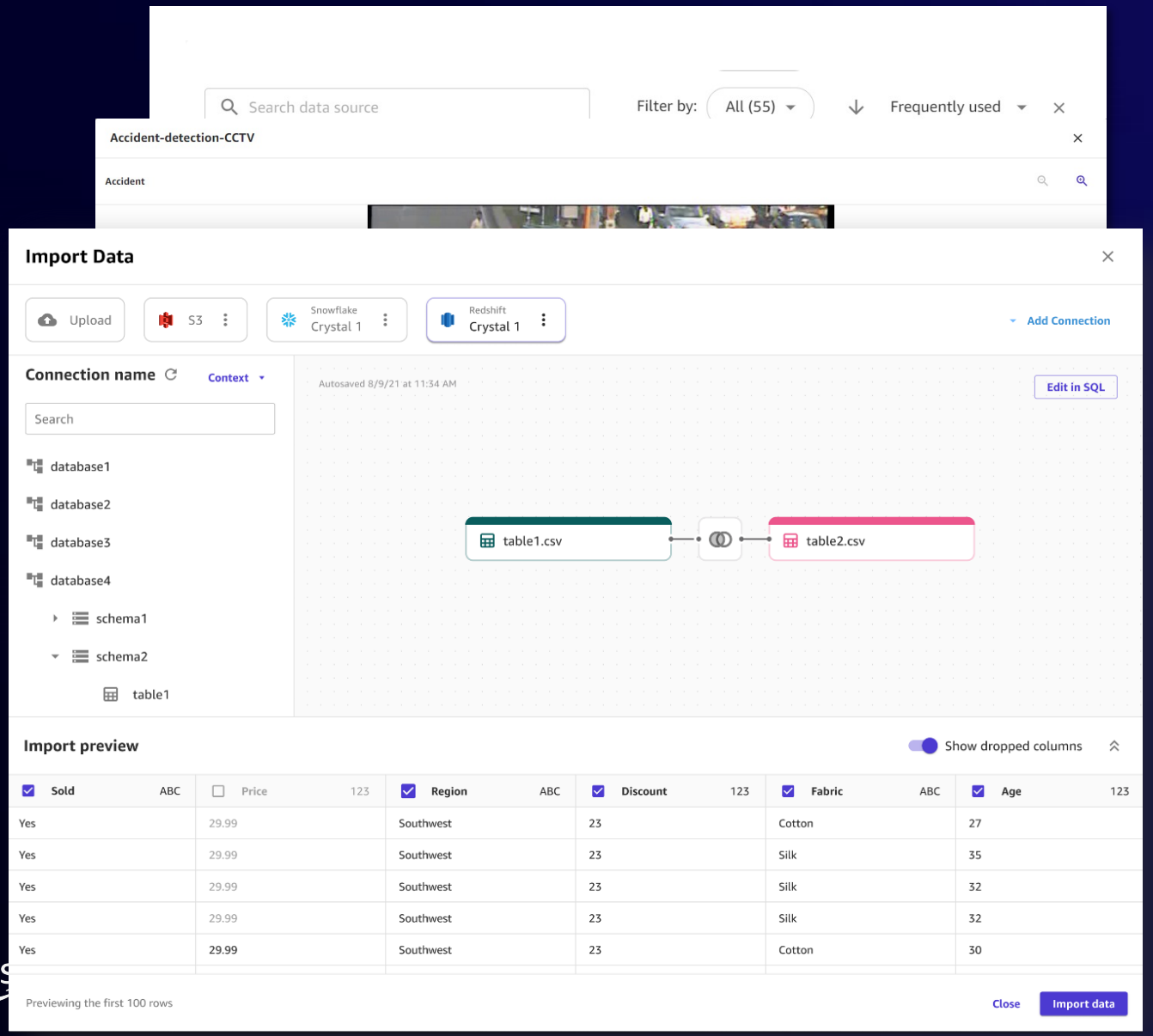
Estime o desempenho e a precisão do modelo e diagnostique possíveis problemas

Treine e automatize

Treine o modelo com o AutoML ou integre-o aos componentes do SageMaker

Produção interativa

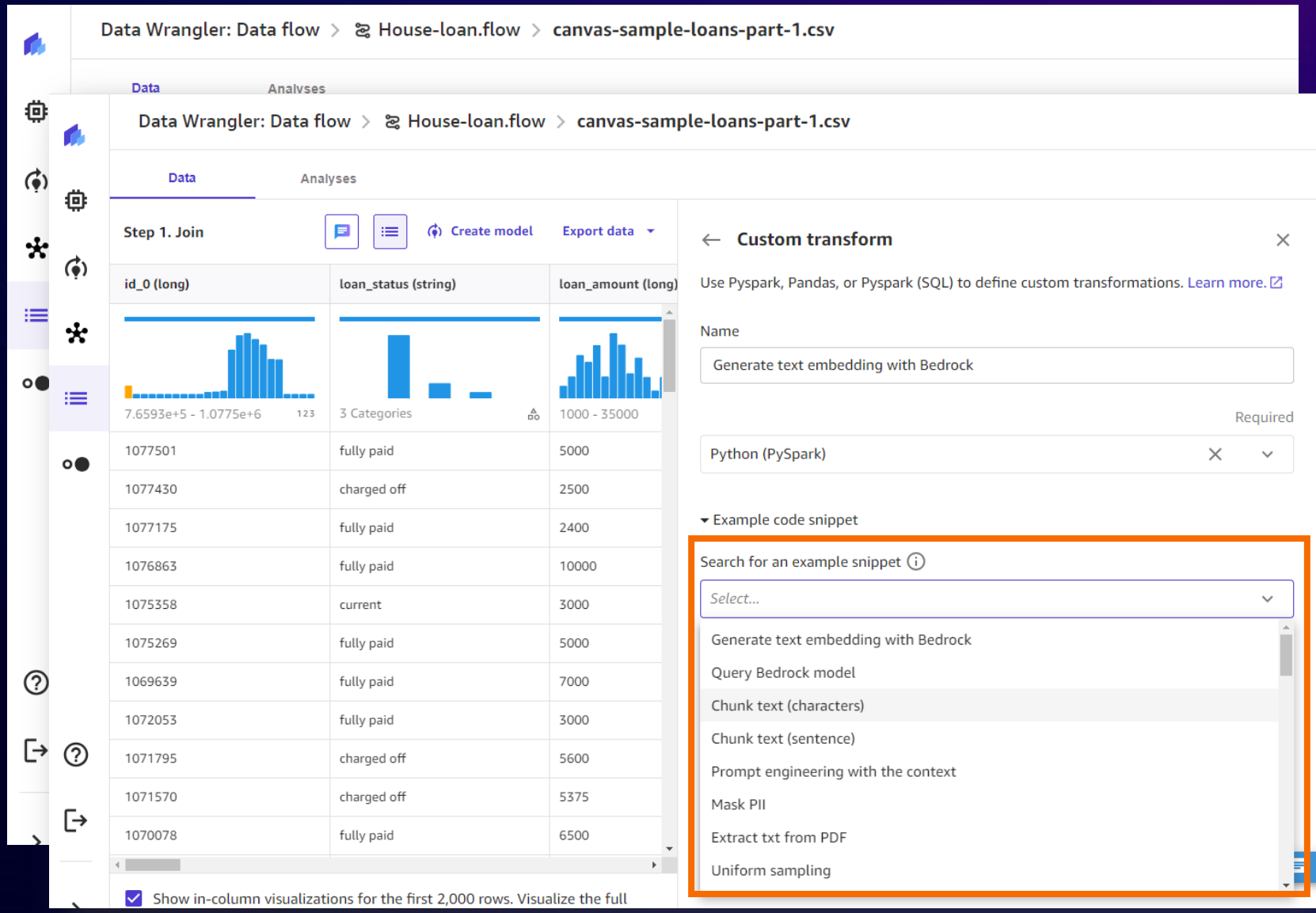
Interface unificada para agregar dados



- **Conecte-se a várias fontes de dados**
  - **Amazon DocumentDB**, **Salesforce Data Cloud**, Amazon S3, Amazon Athena, Amazon Redshift, Amazon RDS, Snowflake, and 50+ data sources
- **Multimodal**
  - CSV, Parquet, tabelas de banco de dados
  - JPG ou PNG
- **Explore e selecione dados com o visual, constructor de consultas ou consultas SQL**

Limpe dados e recursos de engenharia com facilidade

- **Mais de 300 transformações** apoiadas pelo Spark
- **Transformações comuns e avançadas de ML**
  - Séries temporais, divisão de trem/teste, redução de dimensionalidade e muito mais
  - Transformações de texto e imagem
- Flexibilidade e potência de **trechos de código personalizados**



Demo – Amazon Sagemaker Canvas

Amazon Sagemaker Studio Notebooks

Notebooks SageMaker Studio

TOTALMENTE GERENCIADO
Jupyter Lab

Configuração com um clique
Configure identidade, acesso, controle, rede e armazenamento no
um clique

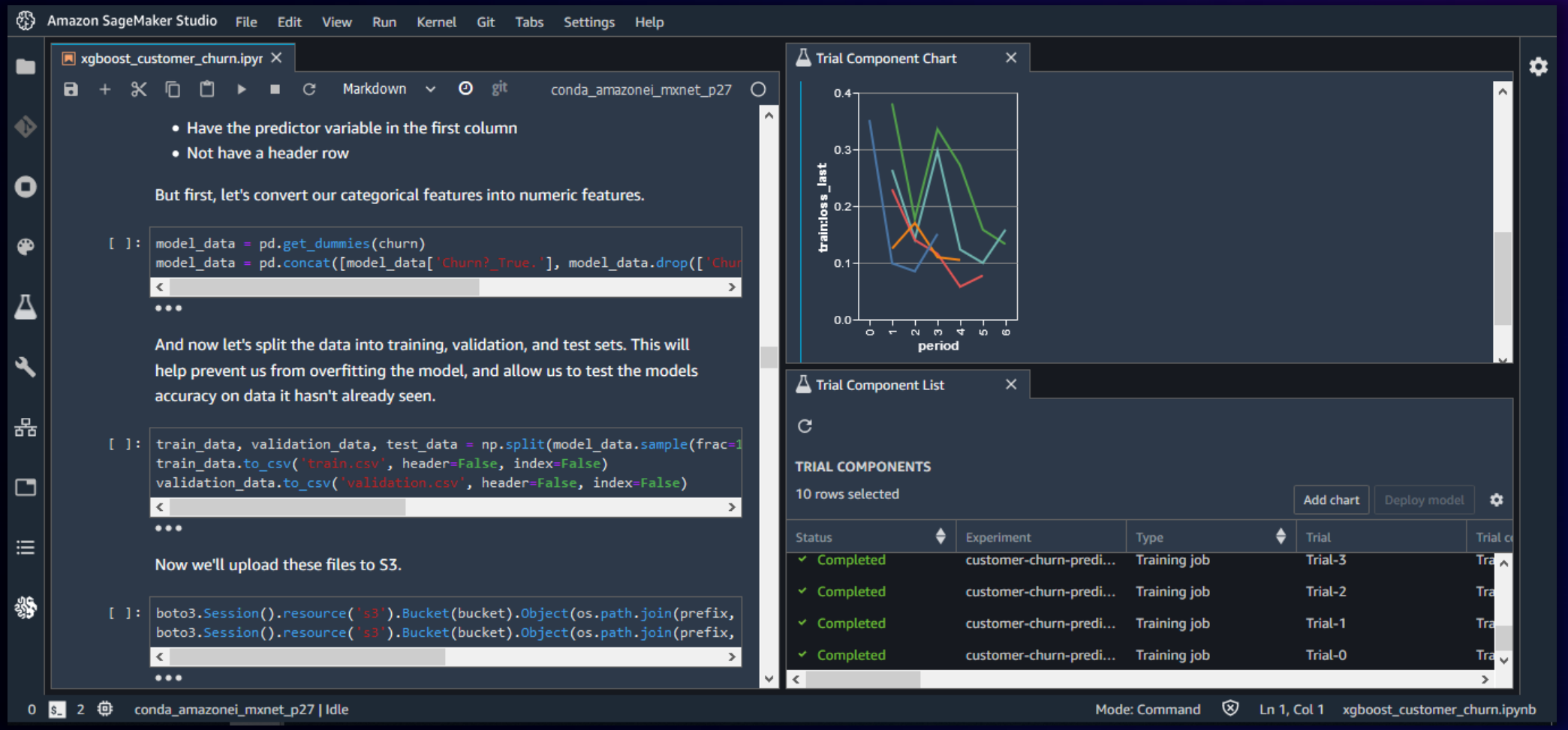
Prepare dados em grande escala
Crie, navegue e conecte-se aos clusters do Amazon EMR; Crie, teste e execute aplicativos interativos de preparação e análise de dados com sessões interativas do AWS Glue

Notebooks de início rápido
Lançamento totalmente gerenciado Jupyter Lab em segundos; redimensione o
computação e armazenamento em qualquer lugar

SageMaker Distribution
Imagens pré-configuradas que incluem estruturas populares para aprendizado de máquina, ciência de dados e visualização

Personalizável
Traga seu próprio ambiente de notebook para o SageMaker Studio usando uma imagem personalizada do Docker

Use o Amazon SageMaker Studio para atualizar modelos e ver imediatamente o impacto na qualidade dos modelos



Preparação de dados em notebooks Studio

Amazon EMR

O Amazon EMR é uma plataforma de cluster gerenciada que simplifica a execução de estruturas de big data na AWS para processar e analisar grandes quantidades de dados.

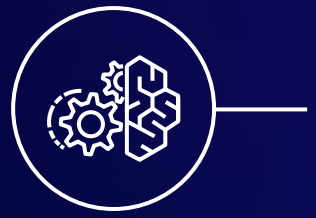
AWS Glue

O AWS Glue é um serviço de integração de dados sem servidor que facilita aos usuários de análises descobrir, preparar, mover e integrar dados de várias fontes

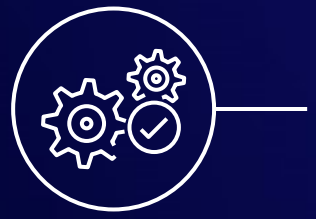
Demo - Amazon Sagemaker Studio

Amazon Sagemaker Feature Store

Desafios de separar features stores



Feature drift



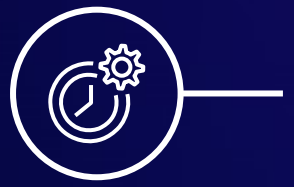
Feature duplicados

Desenvolvimento/implantação lento de modelo

Amazon SageMaker Feature Store

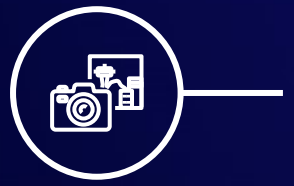
Armazene, descubra e compartilhe com segurança recursos para ML

Online e off-line

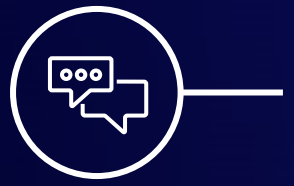


Latência de milisegundo

Features consistentes



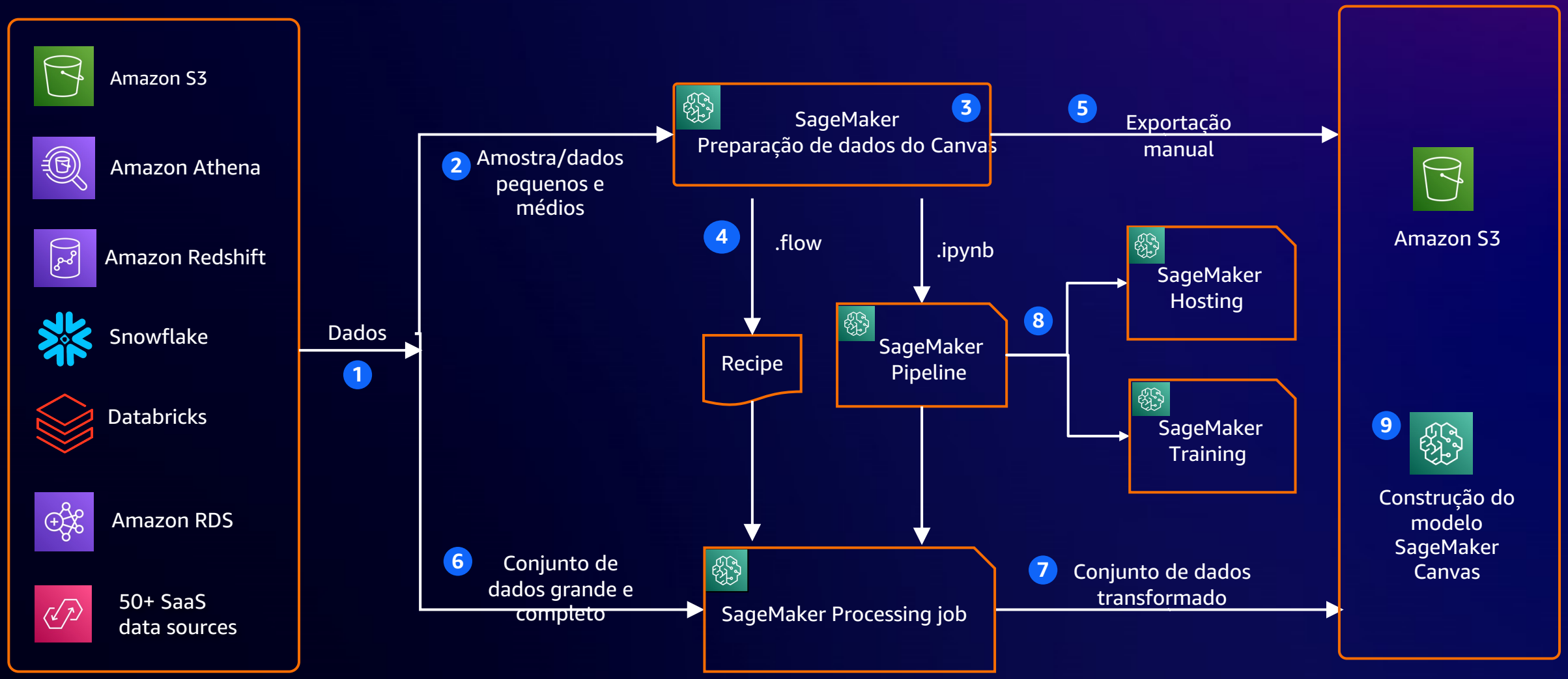
Busca visual



Compartilhar e colaborar

Resumo

Acelere a preparação de dados com o SageMaker Canvas



Perguntas?

Interessado em encurtar seu ciclo de inovação?

Visão geral do SageMaker Canvas

Workshop SageMaker Canvas

Obrigada!

Carolina Junqueira Ferreira
caroljf@amazon.com