AWS re:Invent

NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

AIM324-R2

Explore, analyze, and process data using Jupyter notebooks

Sumedha Swamy
Principal Product Manager
AWS

Bobby Lindsey
AI/ML Specialist SA
AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Level 300/400

1. Overview of Amazon SageMaker Studio Notebooks

2. **NEW**: scalable preparation with built-in integration with Amazon EMR
   - Demo: easily create, connect to, and manage Amazon EMR clusters from SageMaker Studio
   - Demo: monitor and debug Spark jobs with one-click from Studio Notebook

3. Demo: end-to-end machine learning workflow on SageMaker Studio Notebook
   - Access data
   - Analyze data
   - Prepare data
   - Train ML model
   - Deploy

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Overview

Amazon SageMaker

PREPARE

SageMaker Ground Truth
Label training data for machine learning

SageMaker Data Wrangler NEW
Aggregate and prepare data for machine learning

SageMaker Processing
Built-in Python, BYO R/Spark

SageMaker Feature Store NEW
Store, update, retrieve, and share features

SageMaker Clarify NEW
Detect bias and understand model predictions

BUILD

SageMaker Studio Notebooks
Jupyter notebooks with elastic compute and sharing

Built-in and bring-your-own algorithms
Dozens of optimized algorithms or bring your own

Local mode
Test and prototype on your local machine

SageMaker Autopilot
Automatically create machine learning models with full visibility

SageMaker JumpStart NEW
Pre-built solutions for common use cases

TRAIN & TUNE

One-click training
Distributed infrastructure management

SageMaker Experiments
Capture, organize, and compare every step

Automatic model tuning
Hyperparameter optimization

Distributed training NEW
Training for large datasets and models

SageMaker Debugger NEW
Debug and profile training runs

Managed Spot training
Reduce training cost by 90%

DEPLOY & MANAGE

One-click deployment
Fully managed, ultra low latency, high throughput

Kubernetes and Kubeflow integration
Simplify Kubernetes-based machine learning

Multi-model endpoints
Reduce cost by hosting multiple models per instance

SageMaker Model Monitor
Maintain accuracy of deployed models

SageMaker Edge Manager NEW
Manage and monitor models on edge devices

SageMaker Pipelines NEW
Workflow orchestration and automation

SageMaker Studio
Integrated development environment (IDE) for ML

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Studio Notebooks

Amazon SageMaker Studio Notebooks

Fast-start sharable notebooks



Easy access with single sign-on (SSO) or AWS Identity and Access Management (IAM)
Access your notebooks in seconds via pre-signed URLs



Fully managed and secure
Administrators manage access and permissions



Quick start
Start your notebooks without spinning up compute resources



Collaborative
Share notebooks with a single click



Elastic
Dial up or down compute resources on the flight

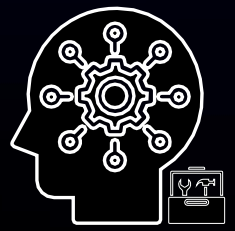


Customizable
Bring your own image with custom packages and libraries; use lifecycle hooks to customize development environments

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data worker personas



Data scientist





Data analyst



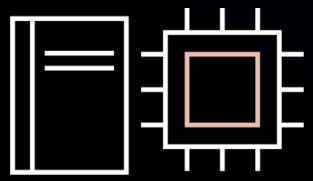
Data engineer

- Data workers are collaborating more closely than ever before
- Using different tools to perform different parts of ML workflow reduces collaboration and workflow efficiency
- Businesses are looking for solutions to unify data analytics and ML workflows to simplify maintenance and increase collaboration

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon SageMaker Studio universal notebook

PERFORM DATA ENGINEERING, ANALYTICS, AND MACHINE LEARNING IN ONE NOTEBOOK



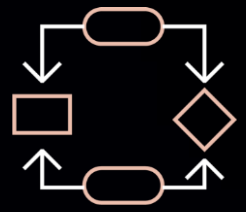
Built-in data and analytics integration

Connect to Spark, Hive, and Presto on Amazon EMR and data lakes on Amazon S3



Interactive data preparation

Interactively access, analyze, and transform a wide range of data



Debugging and monitoring

Visually debug and monitor Spark jobs from the notebook



Build ML workflows

Build, train, and deploy models using preferred framework



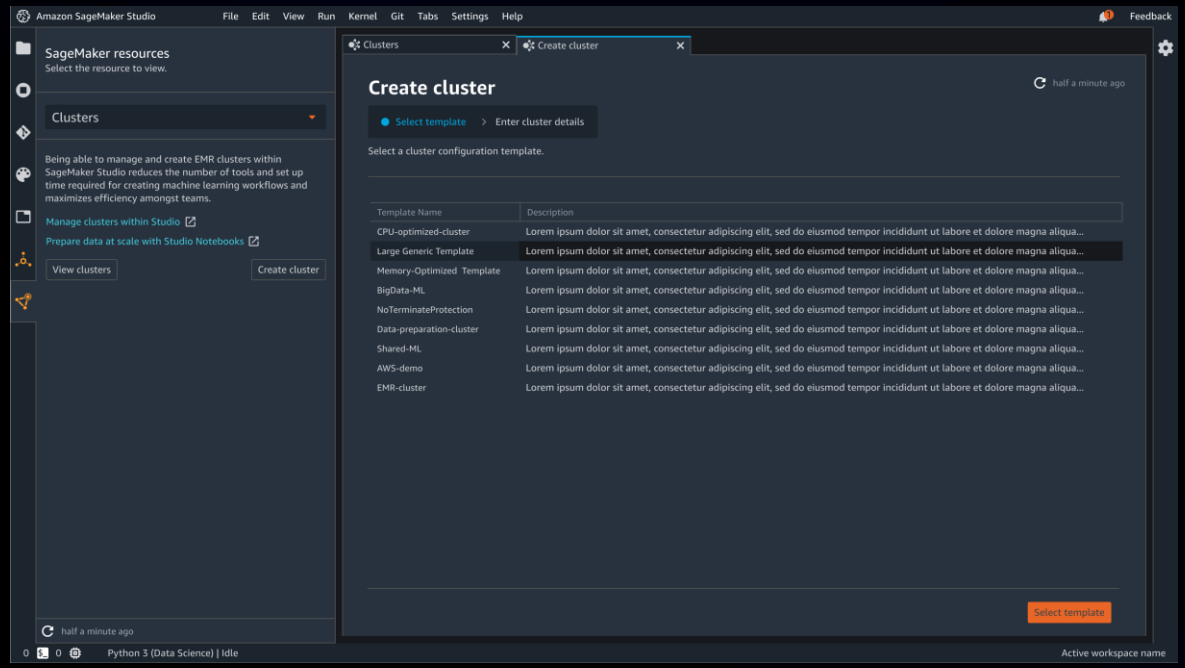
Improve productivity

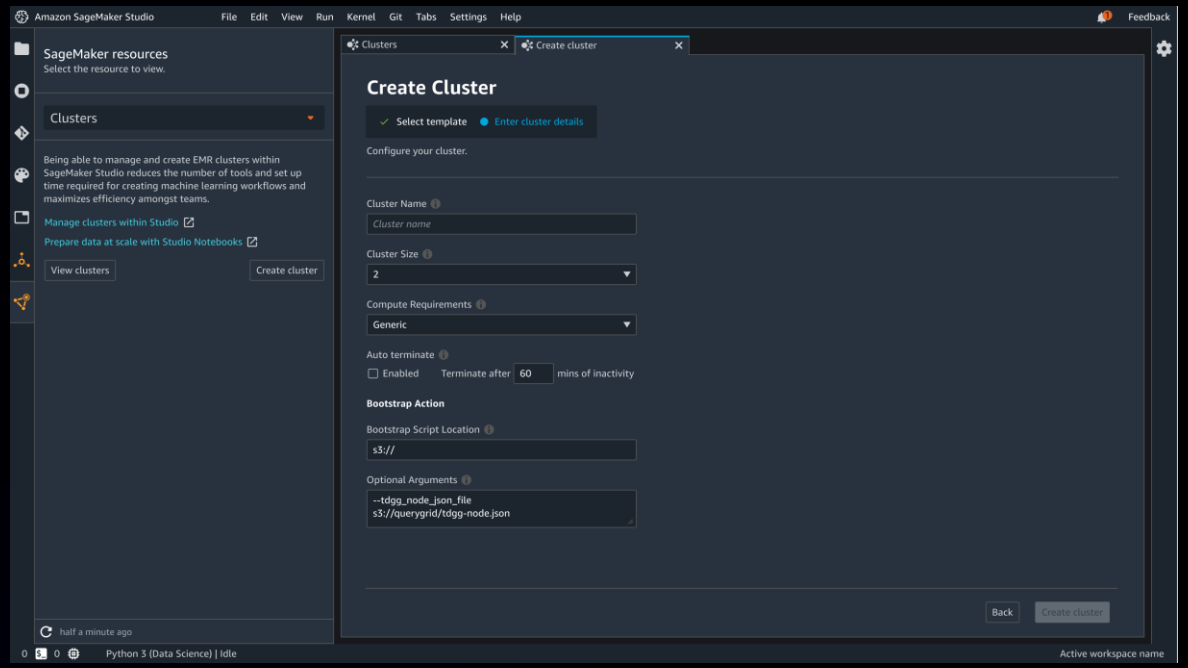
Build end-to-end ML workflows without leaving the notebook

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Studio universal notebook

CREATE EMR CLUSTERS FROM SAGEMAKER STUDIO





Create cluster from template

Use admin-defined templates to enforce configuration guardrails and to enable data workers to easily create clusters

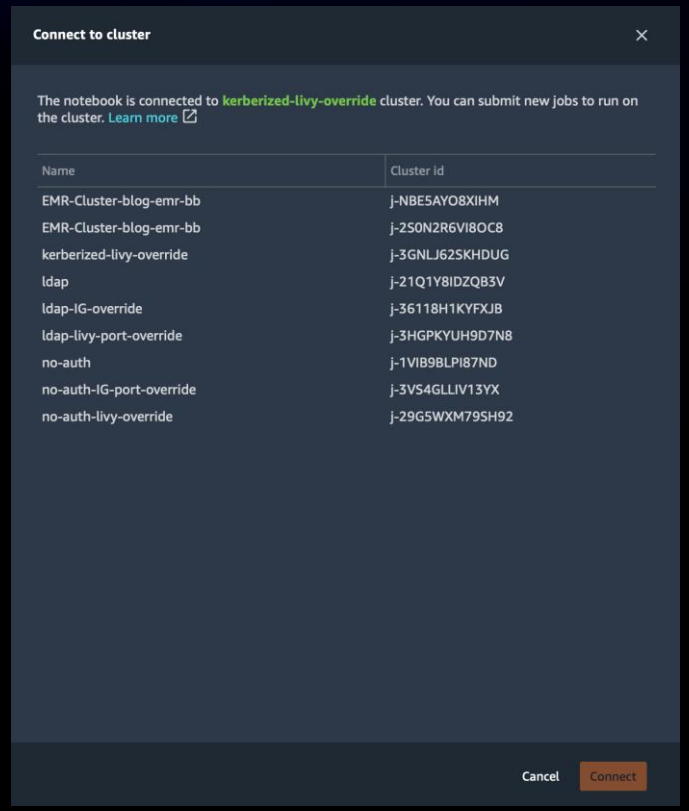
Provide flexibility to users

Provide data workers the flexibility to customize allowed parameters when creating clusters

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

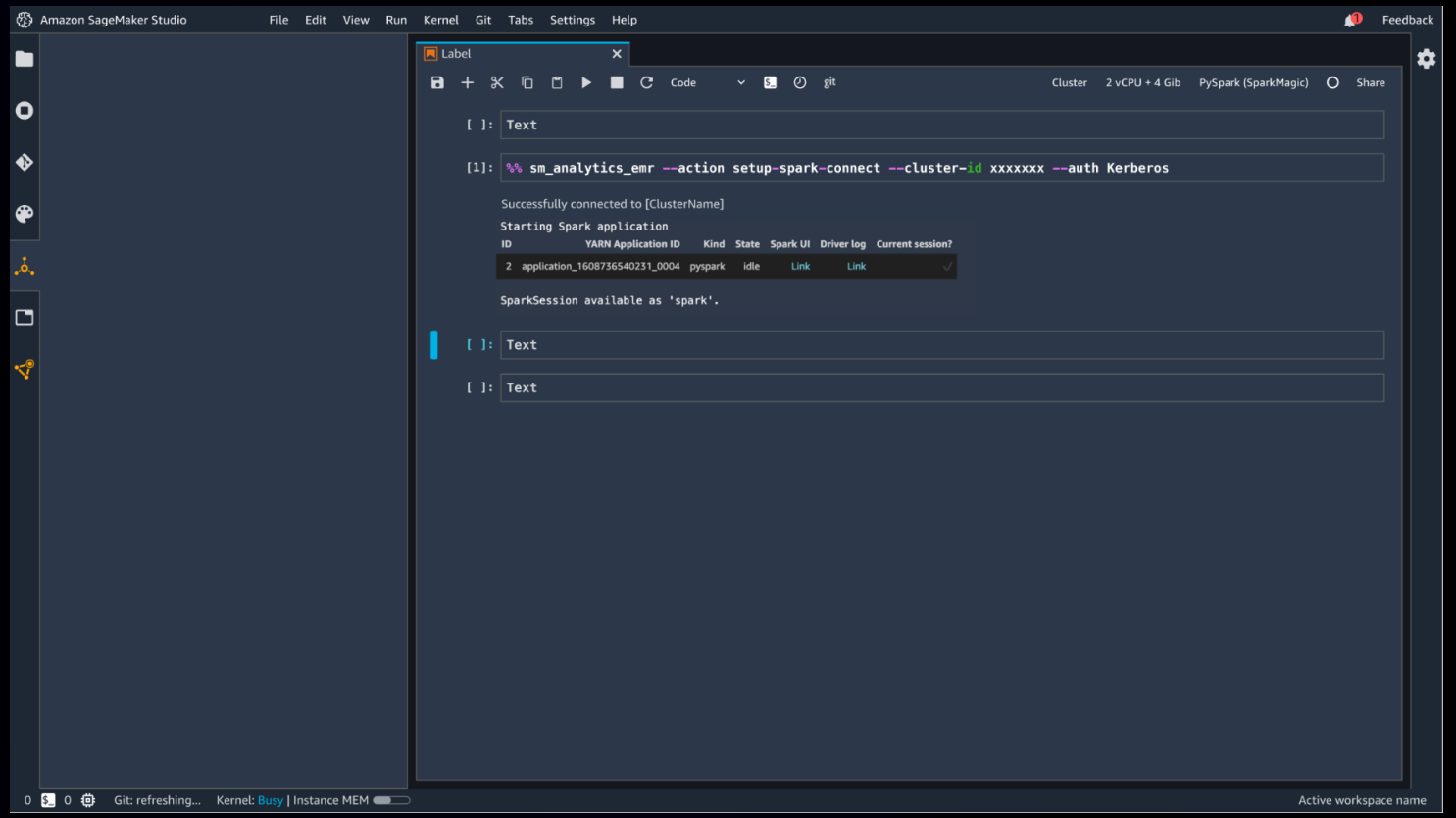
SageMaker Studio universal notebook

CONNECT TO EMR CLUSTERS



Browse clusters

Visual interface to discover EMR clusters



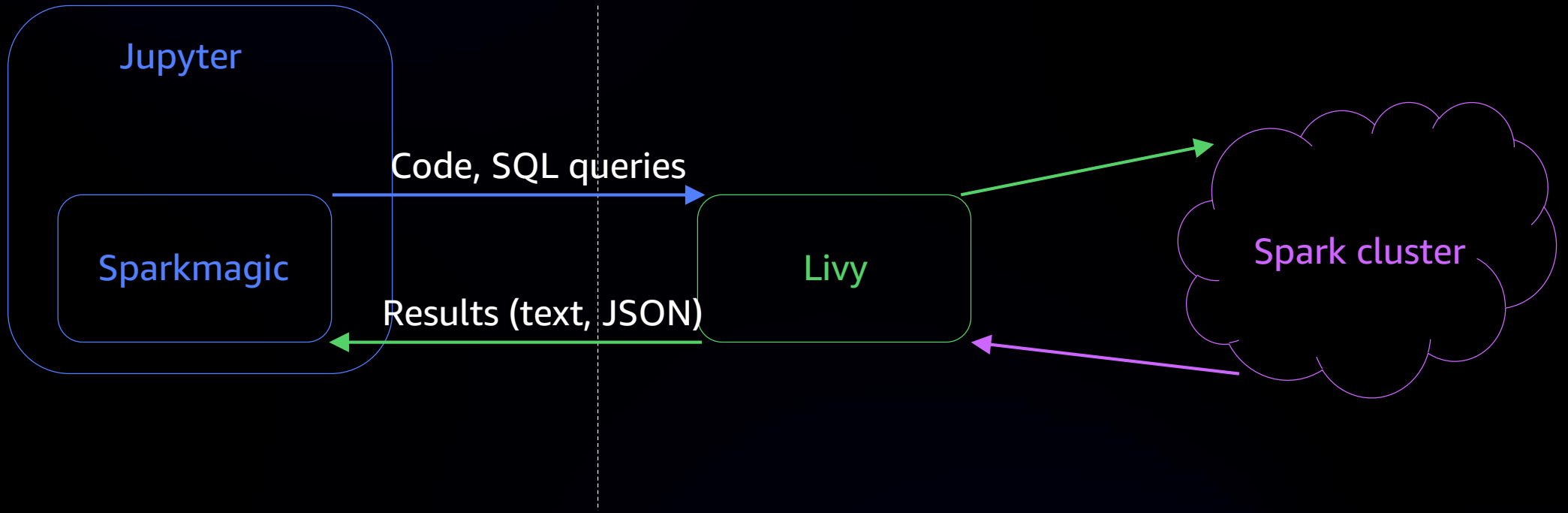
Connect, monitor, and debug

Connect to EMR clusters directly from SageMaker Studio

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SageMaker Studio universal notebook

CONNECT TO EMR CLUSTERS: ARCHITECTURE



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.