AWS re:Invent

ANT306-R

Simplifying the data science experience with EMR notebooks

Samit Chaudhuri

Senior Manager, Amazon EMR

Amazon Web Services

AWS re:Invent

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws

Agenda

Why should I use Amazon EMR notebooks?

What are Amazon EMR notebooks?

What about the latest Amazon EMR notebooks?

Demo

Q&A

Why should I use Amazon EMR notebooks?

Zero setup to explore data at scale

Uncover patterns and insights in data with preinstalled tools

Create line-of-business (LOB) reports from large datasets

Run your big-data analytics and render reports with graphical plots from a single source

Build scalable ETL processes and ML pipelines

Python for data scientists, Spark SQL for analysts, and Scala for data engineers

Leverage the speed and flexibility of Apache Spark

On-demand access to compute clusters

What are Amazon EMR notebooks?

Serverless notebooks for big-data computing



Amazon EMR notebooks



Amazon EMR



Databases



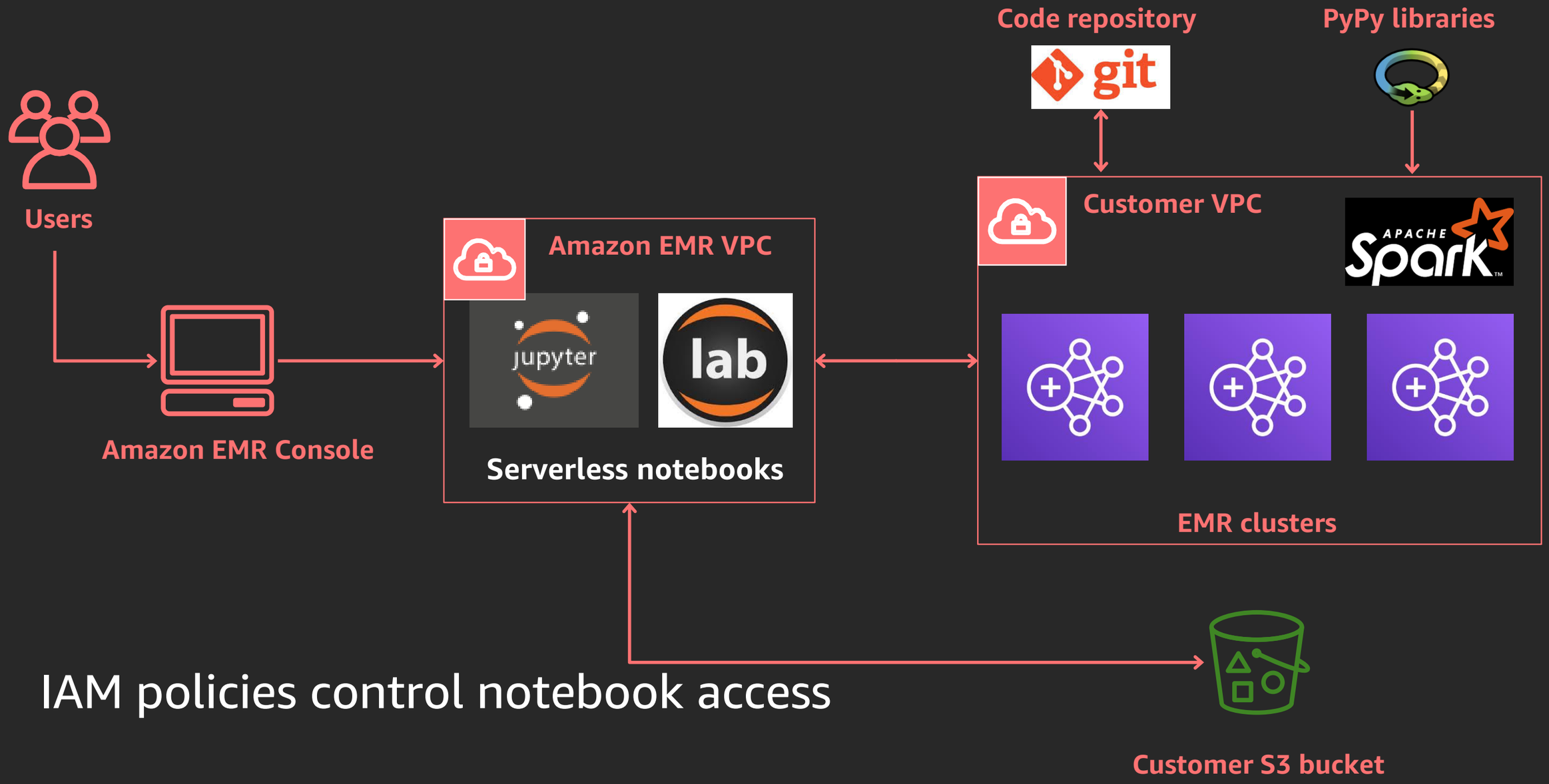
Amazon S3

Develop code on serverless notebooks

Run Apache Spark

Connect to a variety of data sources

What are Amazon EMR notebooks?



IAM policies control notebook access

What about the latest Amazon EMR notebooks?

Managed workspace

- Save and share multiple notebooks in a single workspace

Bring your own libraries

- Install and import custom libraries on a running EMR cluster

Use Git version control

- Record code changes
- Commit and retrieve code to and from remote repositories

Use JupyterLab

- IDE experience for data science

Demo

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Related breakouts

ANT308-R – Deep dive into running Apache Spark on Amazon EMR

ANT327-R – Migrating Apache Spark and Hive from on-premises to Amazon EMR

ANT202-R – Turbocharge your Spark performance with Amazon EMR

ANT301-R – Build & optimize Spark data pipelines for incremental data processing

ANT227-S – Preparing data for reporting, ML, and AI on AWS

ANT223 – Spark access control on Amazon EMR with AWS Lake Formation

ANT226 – Lower costs on Amazon EMR: AWS Auto Scaling and Spot pricing

Resources for migrating to Amazon EMR

- Detailed migration guide
- Free on-site two-day workshop
  - Deconstruction of workloads
  - Future-state architecture
  - Migration plan and recommended next steps
  - Professional Services available for implementation

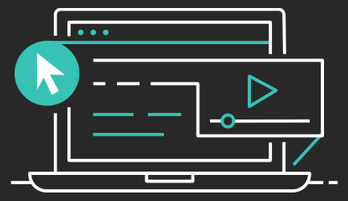


- Visit aws.amazon.com/emr/emr-migration/
- Email emr-migration-help@amazon.com



Learn big data with AWS Training and Certification

Resources created by the experts at AWS to help you build and validate data analytics skills



New free digital course, Data Analytics Fundamentals, introduces Amazon S3, Amazon Kinesis, Amazon EMR, AWS Glue, and Amazon Redshift



Classroom offerings, including Big Data on AWS, feature AWS expert instructors and hands-on labs



Validate expertise with the **AWS Certified Big Data - Specialty** exam or the new **AWS Certified Data Analytics - Specialty** beta exam

Visit aws.amazon.com/training/paths-specialty/

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws training and certification

Thank you!

Samit Chaudhuri

csamit@amazon.com

AWS re:Invent

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws



Please complete the session survey in the mobile app.

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.