
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

PEX114

Supercharge your data analytics with AWS and ExitCertified

Myles Brown (he/him)

Senior Cloud and DevOps Advisor
ExitCertified

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.



ExitCertified is a leading IT training provider with a 97% student satisfaction rating thanks to its award-winning instructors, unparalleled virtual learning, and comprehensive customer care.

ExitCertified has been an AWS Training Partner since 2014 and was awarded the AWS Training Partner of the Year – Americas when it was last awarded in 2019.

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

Why the cloud is the perfect place for data analytics

Why the AWS analytics stack is the best place for data analytics

How to learn the skills needed for data analytics in AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data analytics timeline

1970s: Relational databases started

1980s: Private and public sector data moved into a variety of RDBMS

1990s: Data warehousing and business intelligence really took off

2000s: Big data and Hadoop came to the forefront

2010s: Analytics really started moving to the cloud, data lakes popular

2020s: Lake house architecture

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data analytics in the cloud

- In the late 2000s, organizations of all sizes were running into Big Data problems as the volume of machine-generated data grew
  - This growth doesn’t look like it will slow – no one has less data than they used to
- Housing this data in on-premises data centers caused scalability issues
- Obviously, moving this data to the cloud, where you have seemingly limitless, seamless scalability, made a lot of sense
  - A lot of the data was already being generated in the cloud anyway
- Amazon S3 is the perfect place to hold analytics data because it is scalable, durable, and cost-effective at holding data in any format

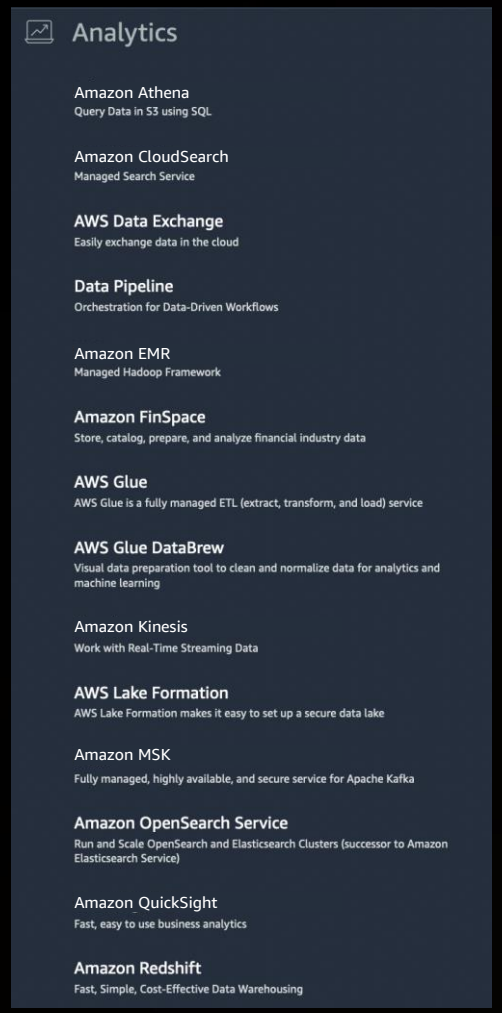
aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Hadoop workloads in the cloud

- Traditional Hadoop workloads really benefit from the parallelism and economics of the cloud
  - If a job can be sufficiently parallelized and it takes 10 machines 10 hours, then it would take 100 machines just 1 hour. In the cloud, it's the same price.
- AWS innovates and makes running Hadoop workloads easier by first creating a managed Hadoop service, Amazon EMR, then a serverless ETL service, AWS Glue
- A complex data analytics pipeline that takes data through ingestion, storage, processing, and visualization can be easily built and automated using the various AWS analytics services

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS analytics services

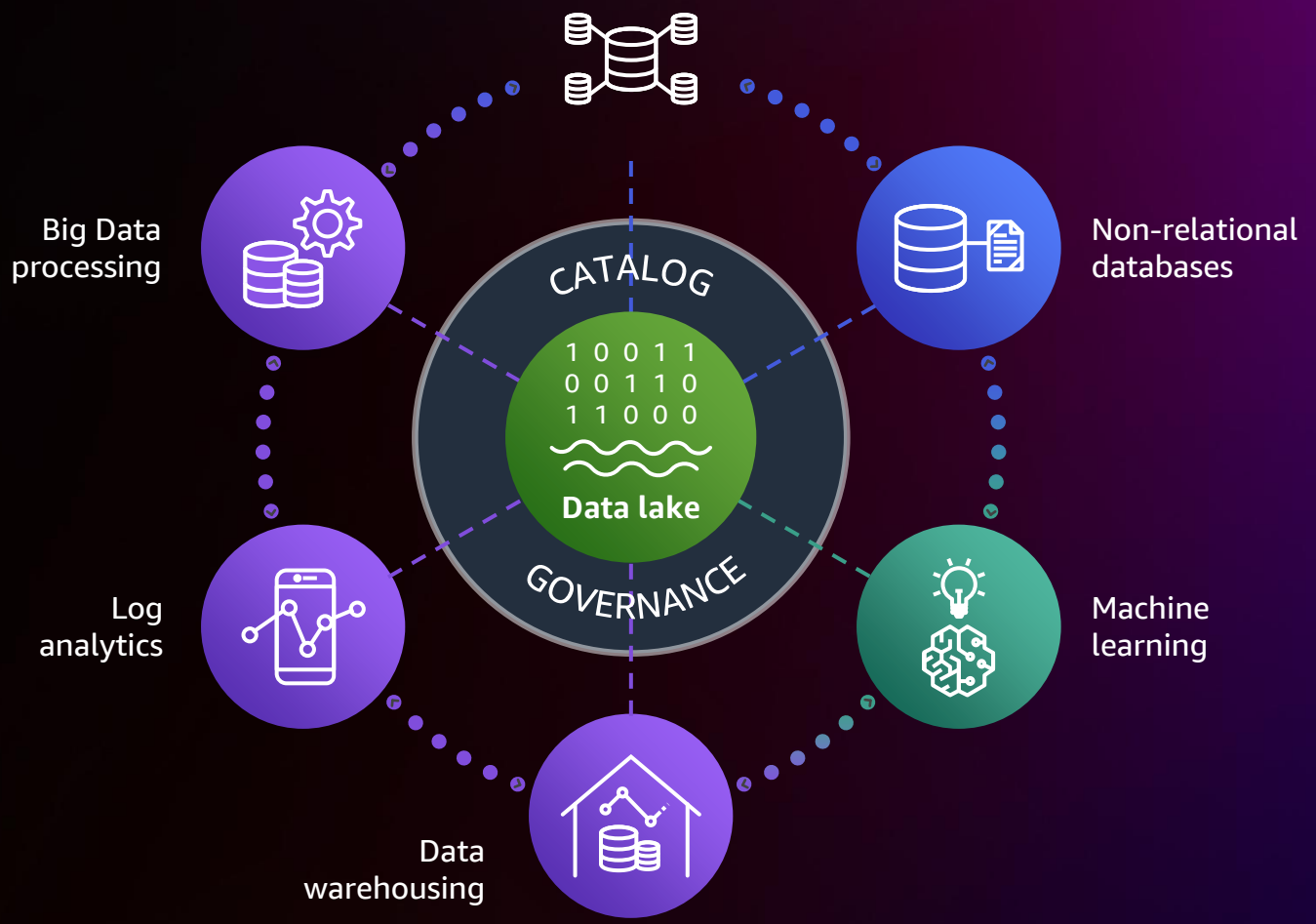


- The number of AWS services in the analytics category is always growing
- Amazon S3 continues to be the basis for most analytics platforms
- Once the data is stored and cataloged in a data lake, it can be accessible to anyone with permissions
- AWS has many machine learning services that can also make use of this data

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS lake house architecture

- A modern data strategy makes use of data wherever it lives, be it in a data lake, a data warehouse, or any other purpose-built data store
- The lake house architecture lets you use the *right tool for the right job,* but still provides a coherent data plan with security and governance



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS lake house architecture

- A modern data strategy makes use of data wherever it lives, be it in a data lake, a data warehouse, or any other purpose-built data store
- The lake house architecture lets you use the *right tool for the right job,* but still provides a coherent data plan with security and governance



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How to get your team prepared for AWS data analytics

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Authorized AWS training courses

- There are four one-day AWS courses specifically on data:
  - **Building Data Lakes on AWS**: Covers the basics of establishing and using an S3 data lake, including information on using AWS Lake Formation and AWS Glue
  - **Building Batch Data Analytics Solutions on AWS**: Covers the use of Amazon EMR to run Apache Spark jobs to perform batch analysis
  - **Building Data Analytics Solutions Using Amazon Redshift**: Covers the use of Amazon Redshift to build a modern data warehouse
  - **Building Streaming Data Analytics Solutions on AWS**: Covers streaming solutions using Amazon Kinesis and Amazon Managed Streaming for Apache Kafka (Amazon MSK)

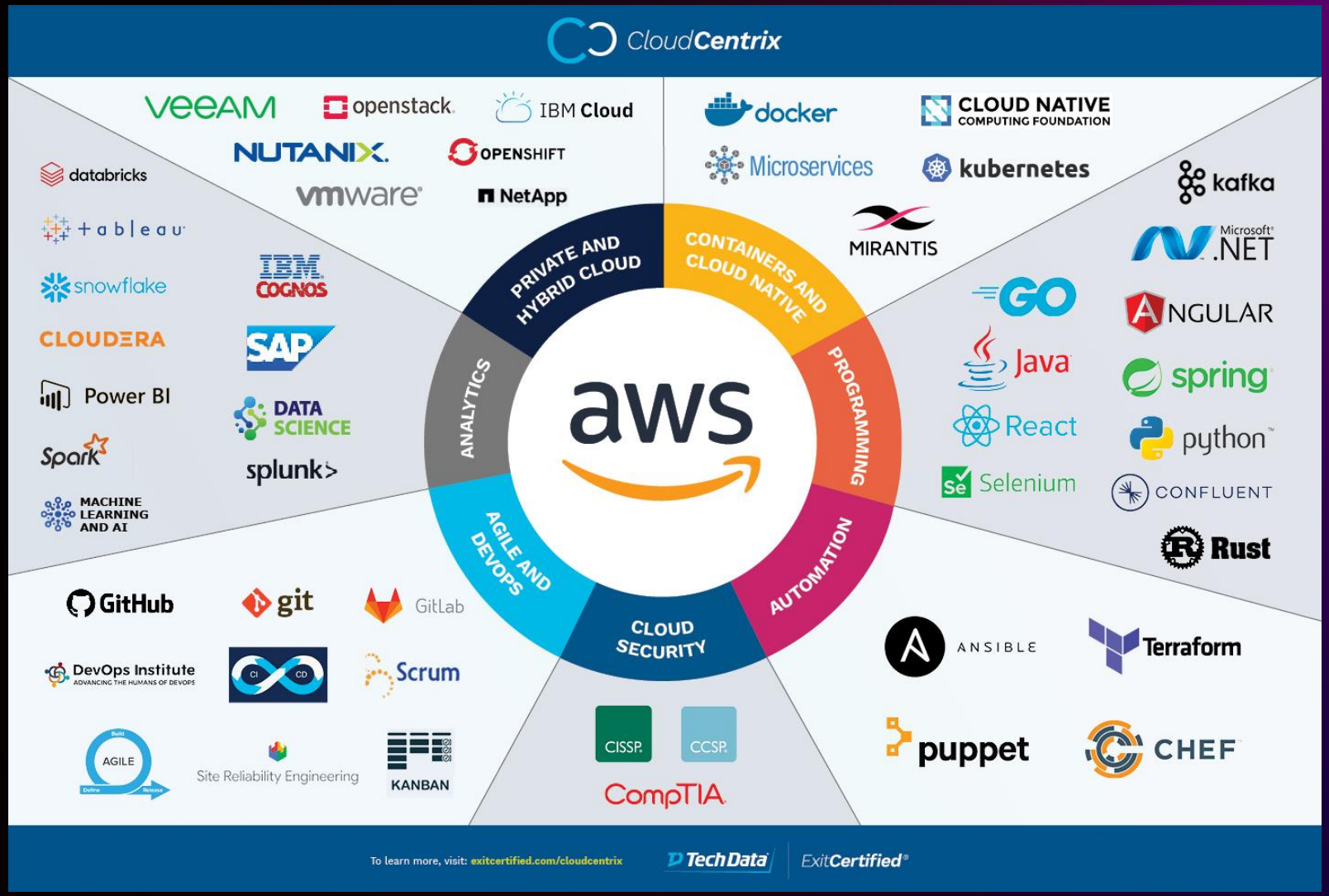
A good prerequisite to these courses, if you're new to AWS, is either the one-day **AWS Technical Essentials** or even the three-day **Architecting on AWS**

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Preparing for a cloud journey

- A cloud journey comprises technologies from a variety of vendors
- The CloudCentrix portfolio addresses training across open-source and commercial products used to create cloud-native solutions



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Connect with ExitCertified

- Visit us at Booth 146 for more information on training and certification
- www.exitcertified.com



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Myles Brown
myles@exitcertified.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.