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

ANT331

Understanding TCO for different Amazon EMR deployment models

Francis McGregor-Macdonald

ANZ Analytics SA Leader
AWS

Srikanth Sopirala

Principal Analytics Specialist Solutions Architect
AWS

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

Agenda

Introduction to Amazon EMR

EMR Deployment Models

Calculating TCO

Use cases

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

Amazon EMR

BIG DATA ANALYTICS USING OPEN SOURCE FRAMEWORKS
APACHE SPARK, PRESTO, TRINO, HADOOP, HIVE, HBASE & FLINK



Over 3X faster than open source

Performance-optimized runtime for popular frameworks like Spark, Hive, and Presto with 100% open-source API compatibility



Latest open-source features

New open-source features available within 60 days of release in open source



Cost effective for large scale analytics

Reduce costs using EC2 Spot, EMR managed scaling, performance-optimized runtime, and per-second billing



Run on Kubernetes or serverless

New deployment options to run applications on Amazon EKS or Serverless



Interactive data science

EMR Studio IDE with notebooks and diagnostic apps, integrated experience with SageMaker Studio



S3 Data lake scenarios

Integrates with AWS Lake Formation, Apache Ranger, Apache Iceberg, and Apache HUDI for row-level updates, time travel, and data access control scenarios

aws

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

Amazon EMR deployment options



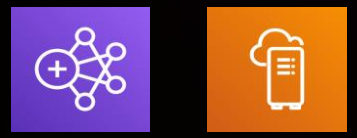
Amazon EMR on Amazon EC2

Choose instances that offer the best price performance for your workload



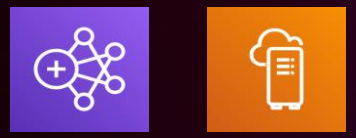
Amazon EMR on Amazon EKS

Automate provisioning, management, and scaling of Apache Spark jobs on Amazon EKS



Amazon EMR on AWS Outposts

Set up, manage, and scale Amazon EMR in your on-premises environments just as you would in the cloud



Amazon EMR Serverless

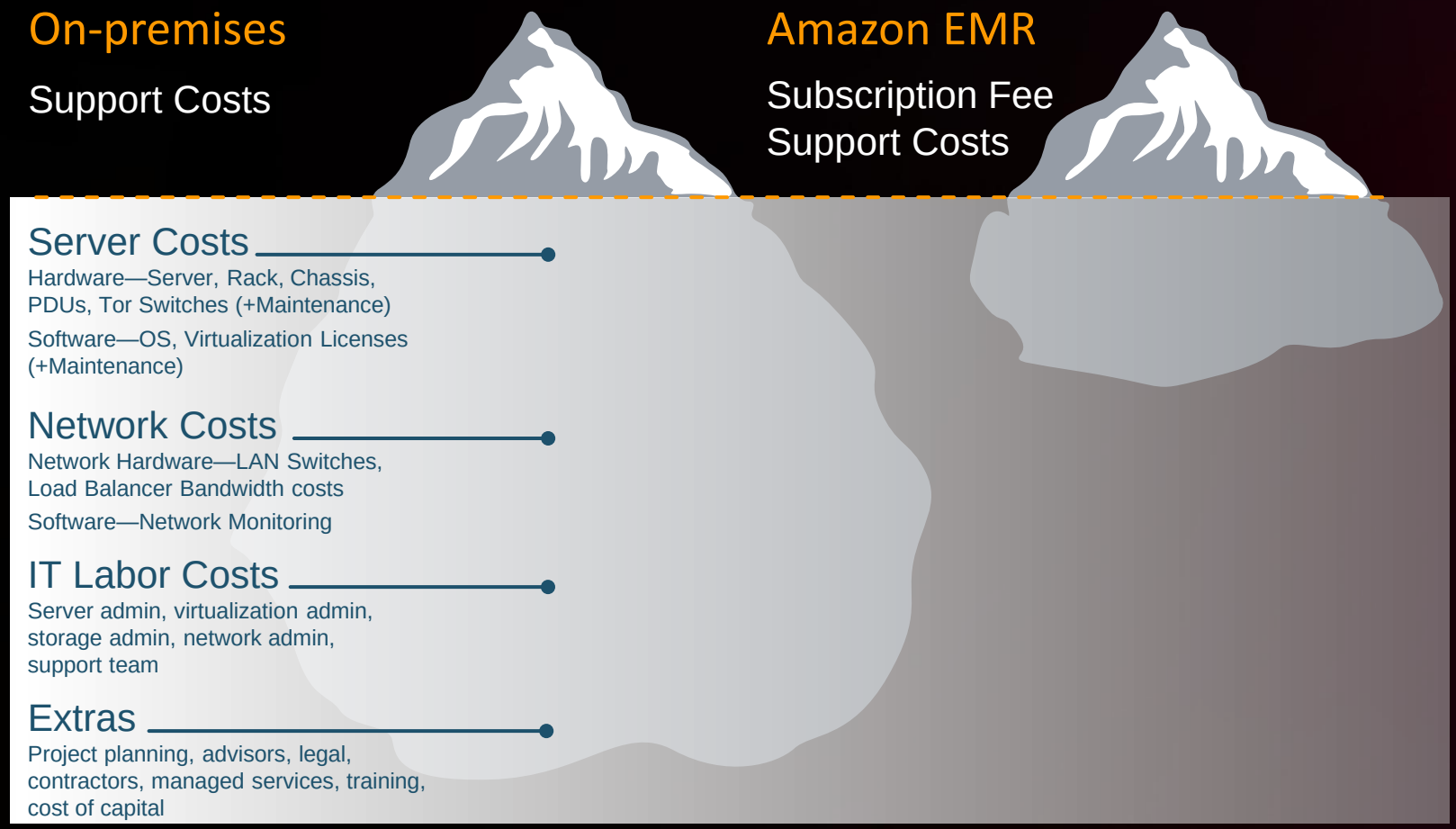
Run petabyte-scale data analytics in the cloud without managing and operating clusters

aws

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

Lowest Cost

Lowest TCO



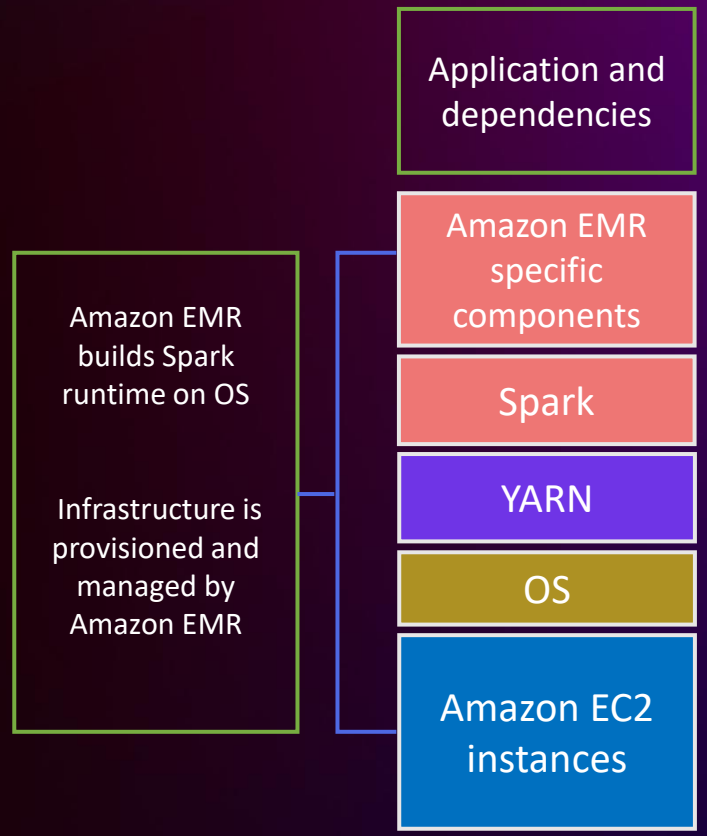
- Less admin time to manage, and support Hadoop clusters
- No up-front costs—hardware acquisition, installation
- Save on operating costs—data center space, power, cooling
- Business Value: Cost of delays, Risk Premium, Competitive abilities, governance, etc.

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

Amazon EMR on Amazon EC2

Control over instances drives a cluster-centric model

- Flexibility in cluster configuration
- Ability to run multiple frameworks
- Optimize for cost by leveraging: RIs, Savings plan
- Cost effective when cluster is highly utilized, optimized with spot instances



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

Amazon EMR on EC2 – TCO Calculation

- Cost of Services consumed
  - Amazon EMR
  - Amazon EC2
  - Amazon EBS volumes
- Other Costs
  - Cluster Administration
  - Over provisioning
  - Cluster startup time

Example TCO Calculation

Scenario: Cluster with 1 master (m5.2xlarge) and 20 worker nodes (m5.12xlarge) running 100 hours on-demand[1]

Amazon EMR Costs : $545

Amazon EC2 Costs : $4,710 (including EBS)

Total Costs : $5,255

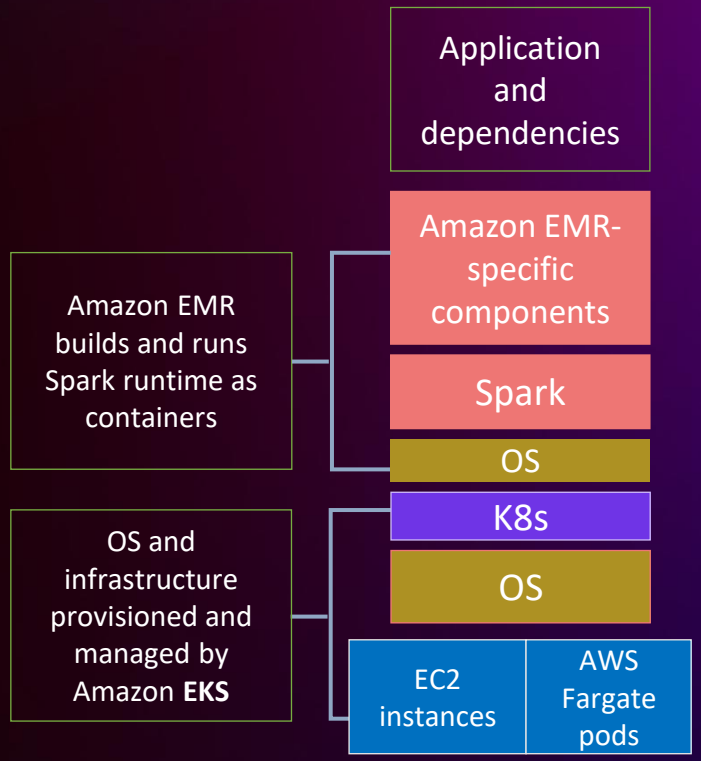
[1]vCPU hours (provisioned) = 96,000
Memory-hours ((provisioned) = 384,000

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

Amazon EMR on EKS

Consolidate analytics workloads with other workloads on EKS

- Simplify infrastructure management
- Consolidate multiple versions of Spark on same EKS cluster
- Simplify Spark application upgrades
- Add Multi-AZ resiliency by EKS with worker nodes across multiple AZs



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

Amazon EMR on EKS – TCO Calculation

- Cost of Services consumed
  - Amazon EMR on EKS
  - Amazon EC2
  - Amazon EKS
  - Amazon EBS volumes
- Other Costs
  - Amazon EKS cluster Configuration
  - Administration

Example TCO Calculation

Scenario: Same workloads as the EMR on EC2 cluster running at 70% Utilization

vCPU hours = 67,200

Memory-hours = 268,800

Amazon EMR Costs : $979

Amazon EC2 Costs: $4,668 (including EKS and EBS)

Total Costs : $5,647

aws

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

Amazon EMR Serverless



No clusters to manage, automatically applies updates



Multi-AZ for high availability



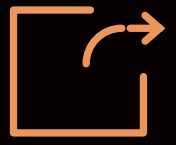
Dynamic worker allocation (you specify min/max workers)



Pay based on memory and CPU consumed



Uses Amazon EMR's performance-optimized runtime for Spark, Hive, and Presto



Latest OSS features within 60 days of release – you choose the version required by your applications



Specify different IAM role for each job, allocate costs by department, and apply cost controls



Amazon EMR Studio IDE for integrated debugging and diagnostics

aws

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

Amazon EMR Serverless – TCO Calculation

- Cost is based on resources consumed
  - vCPU
  - Memory
  - Storage
- Other Costs
  - None

Example TCO Calculation

Scenario: Same workloads as the Amazon EMR on EC2 cluster running at **70%** Utilization

vCPU hours = 67,200

Memory-hours = 268,800

Total Costs : $5,090

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

Comparing TCO – Amazon EMR deployment options

Amazon EMR on EC2

Scenario: Cluster with 1 master (m5.2xlarge) and 20 worker nodes (m5.12xlarge) running 100 hours on demand

Amazon EMR Costs : $545

AmazonEC2 Costs: $4,710

Total Costs : $5,255

Amazon EMR Serverless

Scenario: Same workloads as the Amazon EMR on EC2 cluster running at 70% Utilization

vCPU hours = 67,200
Memory-hours = 268,800

Total Costs : $5,090

Amazon EMR on EKS

Scenario: Same workloads as the Amazon EMR on EC2 cluster running at 70% Utilization

vCPU hours = 67,200
Memory-hours = 268,800

Amazon EMR Costs : $979
Amazon EC2 Costs: $4,668

Total Costs : $5,647

aws

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

Thank you!



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

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