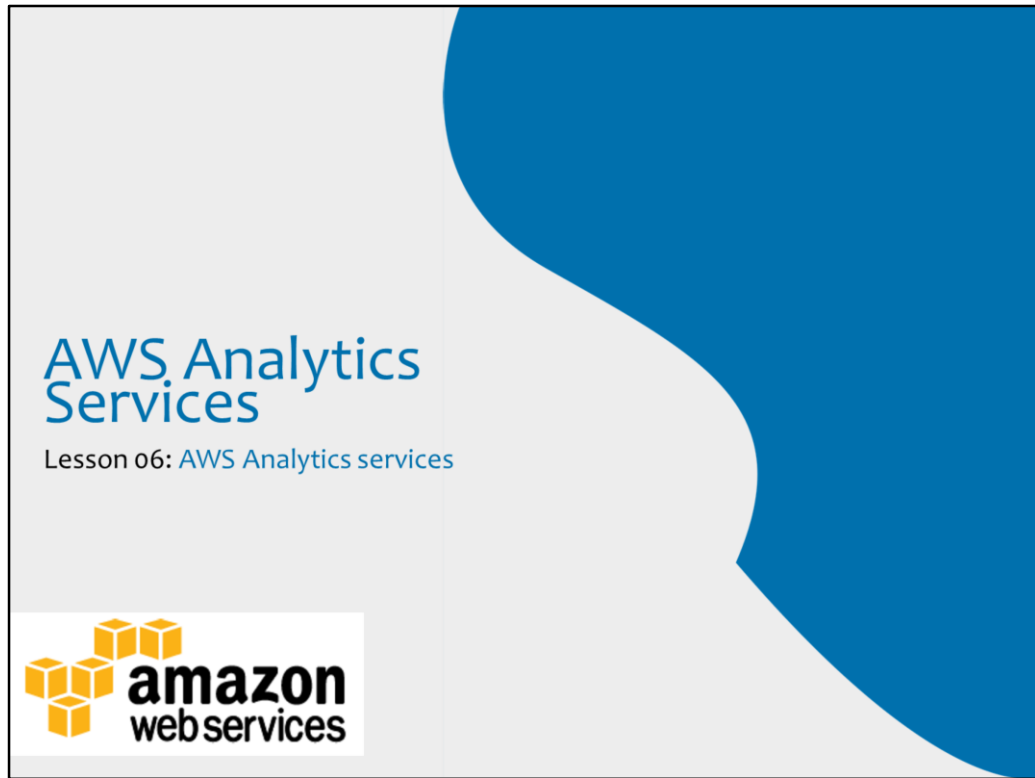


Instructor Notes:

Add instructor notes here.



Instructor Notes:

This lesson is to give an Introduction on Java Server Pages

Lesson Objectives



In this lesson, you will learn:

- Amazon Elastic Map Reduce
- Amazon Red shift
- AWS Data Pipeline
- Amazon CloudSearch

Presentation Title | Author | Date

| © 2017 Capgemini. All rights reserved.

2

Amazon Elastic Map Reduce
Amazon Red shift
AWS Data Pipeline
Amazon CloudSearch

Instructor Notes:

6.1: Amazon EMR
AWS Analytics

Analytics

Amazon Athena

Amazon CloudSearch

Amazon EMR

Amazon ES

Amazon Kinesis

Amazon QuickSight

Amazon Redshift

Search documents

Cluster

EMR engine

Amazon Kinesis Analytics

Amazon Kinesis-enabled app

Dense compute node

EMR engine MapR H3

EMR engine MapR H5

EMR engine MapR H7


HDFS cluster

Amazon Kinesis Firehose

Dense storage node

Amazon Kinesis Streams

Amazon Data Pipeline



Presentation Title | Author | Date

© 2017 Capgemini. All rights reserved.

3

Instructor Notes:

8.1: Amazon EMR Amazon Analytics Services



AWS offers a comprehensive set of services to handle every step of the analytics process chain, that include

- data warehousing
- business intelligence,
- batch processing,
- stream processing,
- machine learning and data workflow orchestration

These services are powerful, flexible, and yet simple to use, enabling organizations to put their raw data to work quickly and easily.



Presentation Title | Author | Date

© 2017 Capgemini. All rights reserved.

4


AWS offers a comprehensive set of services to handle every step of the analytics process chain including data warehousing, business intelligence, batch processing, stream processing, machine learning, and data workflow orchestration. These services are powerful, flexible, and yet simple to use, enabling organizations to put their raw data to work quickly and easily.

Instructor Notes:

6.1: Amazon EMR
AWS Analytics Services


Amazon EMR
(Elastic Map Reduce)

Hosted Hadoop framework




AWS Data Pipeline

Move data among AWS services and on-premises data sources




Amazon Redshift


Petabyte-scale data warehouse service



Amazon Kinesis

Real time processing of streaming data, at any scale





Presentation Title | Author | Date | © 2017 Capgemini. All rights reserved. 5

Instructor Notes:

6.1: Amazon EMR **Amazon EMR**(Elastic Map Reduce)



Product Type --- Hadoop

Amazon EMR simplifies big data processing, providing a managed Hadoop framework that makes it easy, fast, and cost-effective for you to distribute and process vast amounts of your data across dynamically scalable Amazon EC2 instances.

One can also run other popular distributed frameworks such as Apache Spark, Hbase in Amazon EMR.

interact with data in other AWS data stores such as Amazon S3 and Amazon DynamoDB.

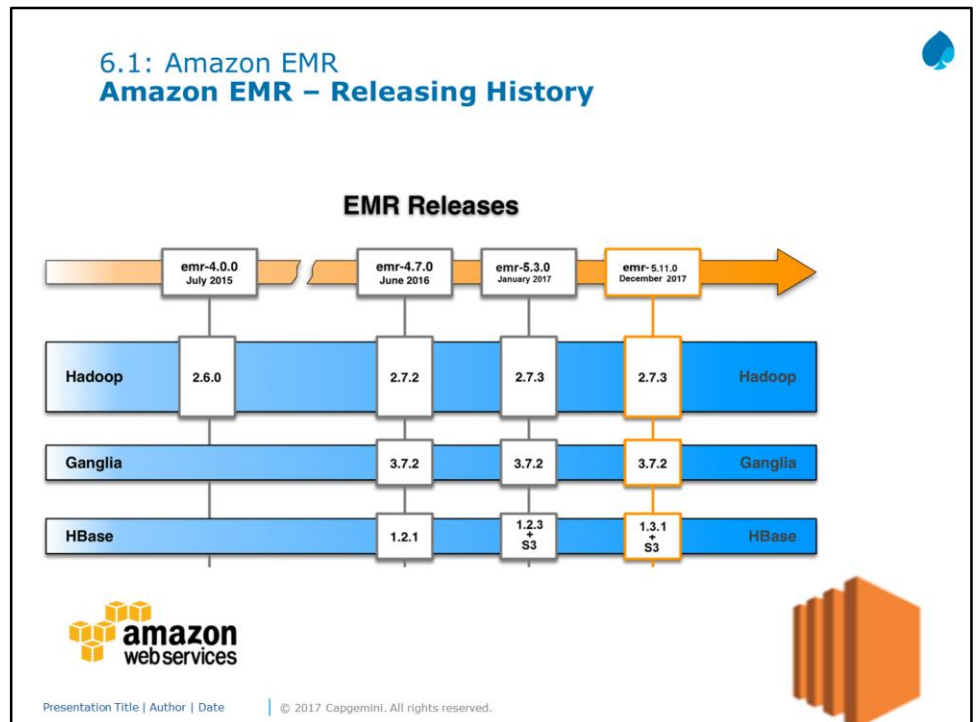


Presentation Title | Author | Date

© 2017 Capgemini. All rights reserved.

6

Provides a managed [Hadoop](#) framework to process vast amounts of data quickly and cost-effectively. Run open source frameworks such as Apache Spark, HBase, Presto, and Flink.

Instructor Notes:

Provides a managed [Hadoop](#) framework to process vast amounts of data quickly and cost-effectively. Run open source frameworks such as Apache Spark, HBase, Presto, and Flink.



Instructor Notes:

6.1: Amazon EMR

Amazon EMR - Benefits

Few benefits are listed below

- Easy to use
- Low cost
- Elastic
- Reliable
- Secure
- Flexible



Presentation Title | Author | Date | © 2017 Capgemini. All rights reserved. 8

Easy to use :-- Amazon EMR cluster in minutes. It takes care of cluster setup, Hadoop configuration, or cluster tuning

Low cost :-- Pay a per-second rate for every second used, with a one-minute minimum charge.

Elastic :- With Amazon EMR, you can provision one, hundreds, or thousands of compute instances to process data at any scale.

Reliable:--- One can spend less time tuning and monitoring your cluster.

Secure :--- Amazon EMR automatically configures Amazon EC2 firewall settings that control network access to instances, and you can launch clusters in an Amazon Virtual Private Cloud .

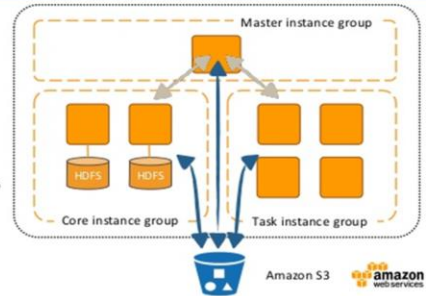
Flexible

Instructor Notes:

6.1: Amazon VPC Amazon EMR

Amazon EMR - Architecture

- Master instance
 - Controls the cluster
- Core instance
 - Life of cluster
 - DataNode and TaskTracker daemons
- Task instances
 - Added or subtracted to perform work (SPOT)
- S3 as underlying 'file system'



Presentation Title | Author | Date

© 2017 Capgemini. All rights reserved.

9

Instructor Notes:

6.2: Amazon Redshift Amazon Redshift



Product Type --- Data Warehouse

- It is a fast, fully managed data warehouse which makes it simple and cost-effective to analyze all your data using standard SQL and your existing Business Intelligence (BI) tools.
- It allows to run complex analytic queries against petabytes of structured data, using sophisticated query optimization, columnar storage on high-performance local disks, and massively parallel query execution.
- Amazon Redshift also includes Redshift Spectrum, allowing you to directly run SQL queries against exabytes of unstructured data in Amazon S3

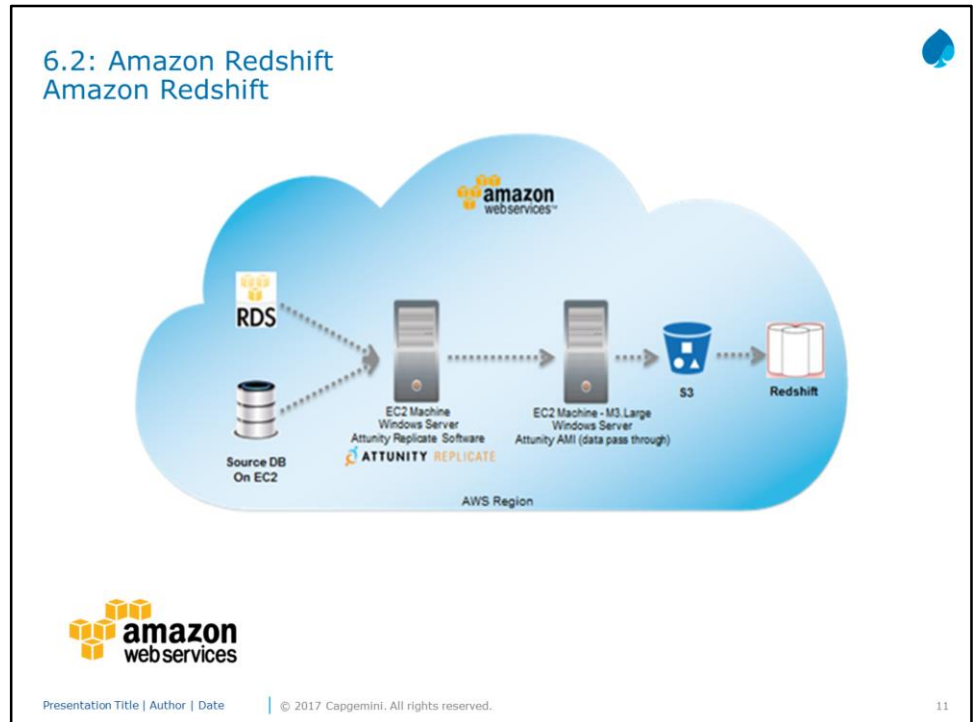


Presentation Title | Author | Date

© 2017 Capgemini. All rights reserved.

10

1. Fast, fully managed, petabyte-scale [data warehouse](#) that makes it simple and cost-effective to analyze all of your data using your existing business intelligence tools.

Instructor Notes:

1. Fast, fully managed, petabyte-scale [data warehouse](#) that makes it simple and cost-effective to analyze all of your data using your existing business intelligence tools.

Instructor Notes:

8..2: Amazon Redshift AWS Data Pipeline



Product Type --- Data Workflow Orchestration.

It is a type of web service which helps reliably process and move data between different AWS compute and storage services, as well as on-premise data sources, at specified intervals.

With AWS Data Pipeline, one can regularly access the data where it's stored, transform and process it at scale, and efficiently transfer the results to AWS services

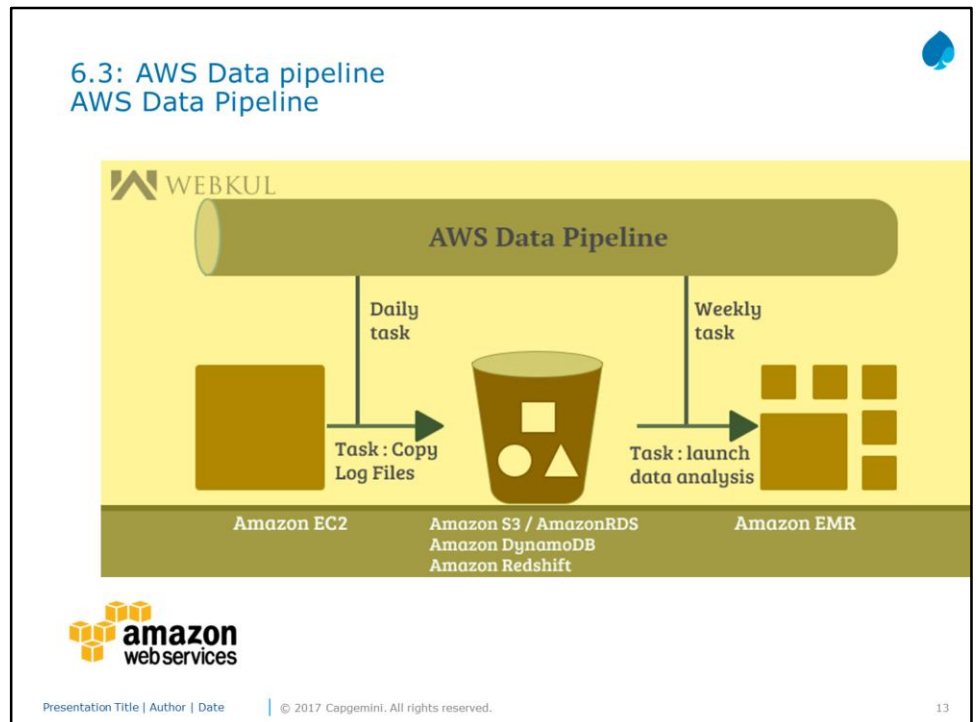
such as Amazon S3, Amazon RDS,
Amazon DynamoDB, and Amazon EMR.



Presentation Title | Author | Date

© 2017 Capgemini. All rights reserved.

12

Instructor Notes:

Instructor Notes:

6.3: AWS Data pipeline

AWS Data Pipeline - Benefits

**Reliable : ---**

- If failures occur in any one activity logic or data sources, AWS Data Pipeline automatically retries the activity. If the failure persists, AWS Data Pipeline sends you failure notifications via Amazon Simple Notification Service (Amazon SNS).

Easy to use :---

- One can check for the existence of an Amazon S3 file by simply providing the name of the Amazon S3 bucket and the path of the file that you want to check for, and AWS Data Pipeline does the rest.

Flexible:--

- One can configure an AWS Data Pipeline to take actions like run Amazon EMR jobs, execute SQL queries directly against databases, or execute custom applications running on Amazon EC2 or in your own datacenter.



Instructor Notes:

6.3: AWS Data pipeline

AWS Data Pipeline - Benefits

**Scalable :-**

- It makes it equally easy to dispatch work to one machine or many, in serial or parallel

Low Cost :-

- AWS Data Pipeline is inexpensive to use

Transparent :--

- Full execution logs are automatically delivered to Amazon S3, giving you a persistent, detailed record of what has happened in your pipeline.



Presentation Title | Author | Date

© 2017 Capgemini. All rights reserved.

15

Instructor Notes:

6.4: Amazon CloudSearch Amazon CloudSearch



- It is a fully managed search service in the cloud which makes it easy to setup , operate and scale a search solution for website and application
- It is fully-managed , full-featured search service
- Automatically scales for data and traffic

It handled both structured and unstructured data

Up and running in less than 1 hour



Instructor Notes:

6.4: Amazon CloudSearch

Amazon CloudSearch - Benefits

The benefits of CloudSearch

- Simple
- Scalable
- Reliable:--
- Secure:--



Presentation Title | Author | Date© 2017 Capgemini. All rights reserved.17

Simple : ---

One can configure and manage an Amazon CloudSearch domain through the AWS Management Console, AWS CLI, and AWS SDKs.

Scalable:---

As your data or query volume changes, Amazon CloudSearch can scale your search domain's resources up or down as needed.

Reliable:--

As your data or query volume changes, Amazon CloudSearch can scale your search domain's resources up or down as needed.

Secure:--

Amazon CloudSearch uses strong cryptographic methods to authenticate users and prevent unauthorized access to your domains.

Instructor Notes:

Summary



In this lesson, you have learnt:

- Amazon Elastic Map Reduce
- Amazon Red shift
- AWS Data Pipeline
- Amazon CloudSearch

Instructor Notes:**Answers for the
Review Questions:**

Answer 1: Amazon
Redshift

Answer 2: Amazon
EMR

Review – Questions

Question 1: _____ is a fast, fully managed data warehouse which makes it simple and cost-effective to analyze all your data using standard SQL and your existing Business Intelligence (BI) tools.

Question 2 : _____ simplifies big data processing, providing a managed Hadoop framework.

Instructor Notes:**Answers for the
Review Questions:****Answer 3:** Option 2**Review – Questions**

- Question 3: Which of the following is not an Amazon Analytic service?

Option 1 : Amazon EMR

Option 2 : Amazon Route 53

Option 3 : AWS Data Pipeline

Option 4 : Amazon Cloudsearch