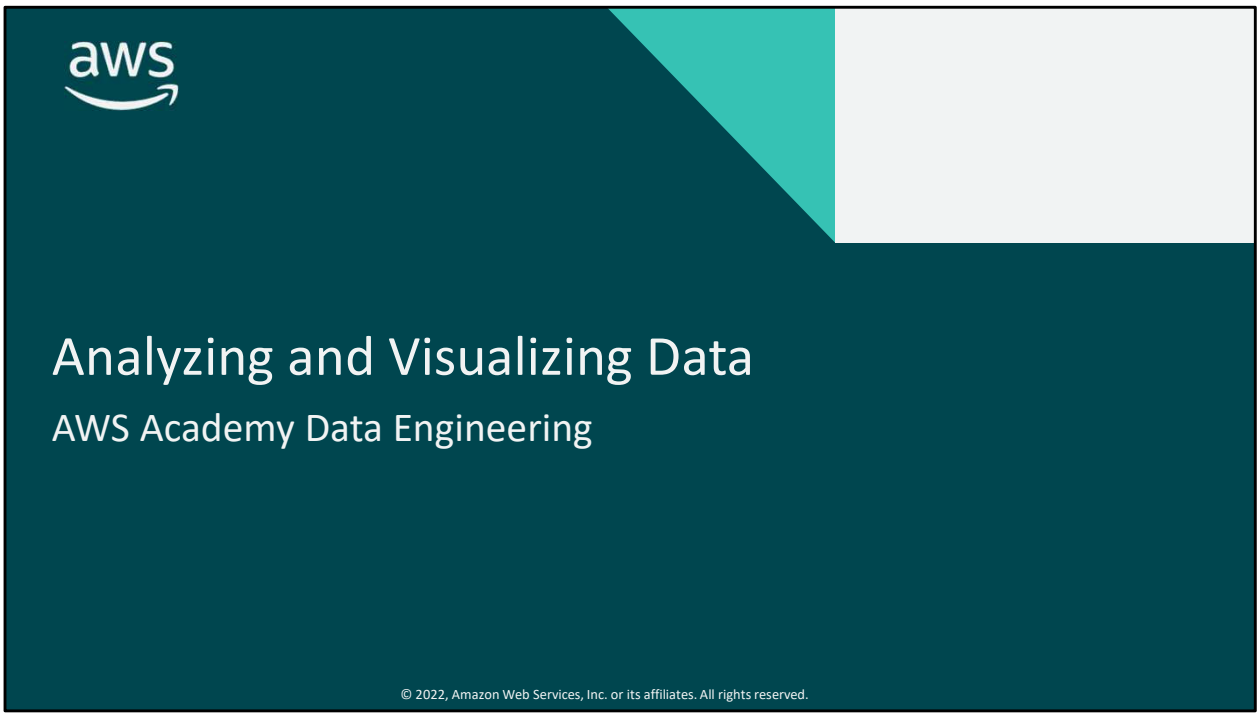
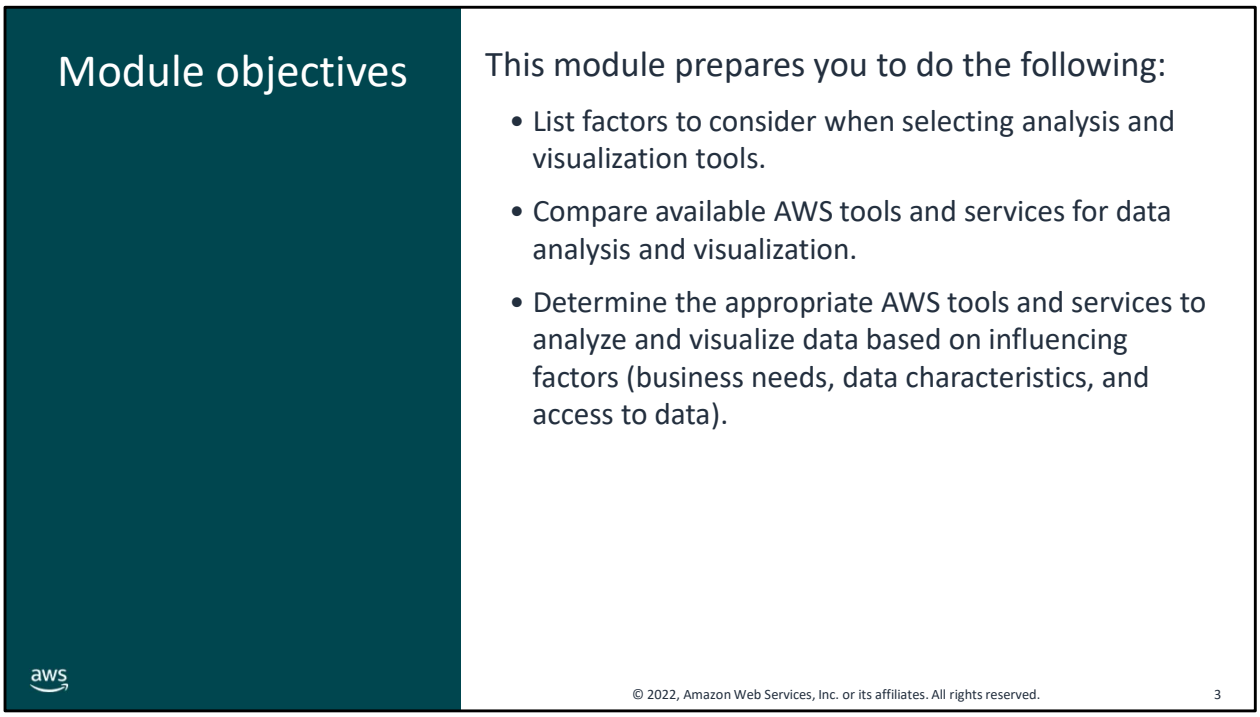


|Slide number 1
|Instructor notes
|This module focuses on the factors involved in determining which AWS solutions to select for analyzing and visualizing data. Examples and use cases of Amazon Athena, Amazon QuickSight, and Amazon OpenSearch Service are provided in this module.
|
|Student notes
Welcome to the Analyzing and Visualizing Data module. This module focuses on AWS solutions to analyze and visualize data and factors to consider when you select a solution.

|Slide number 2
|Instructor notes
|
|Student notes
This introduction section describes the content of this module.



|Slide number 3
|Instructor notes
|The first two learning objectives enable students to accomplish the third and final LO.
|
|Student notes
In this module, the focus is on understanding what to consider when you select tools to analyze and visualize your data. The module compares three AWS tools and services for this purpose. You will also review a use case to examine the influencing factors and the selected AWS solution.

Module overview

Presentation sections

- Considering factors that influence tool selection
- Comparing AWS tools and services
- Selecting tools for a gaming analytics use case

Demo

- Analyzing and Visualizing Data with AWS IoT Analytics and QuickSight

Lab

- Analyzing and Visualizing Streaming Data with Kinesis Data Firehose, OpenSearch Service, and OpenSearch Dashboards

Knowledge checks

- Online knowledge check
- Sample exam question

aws

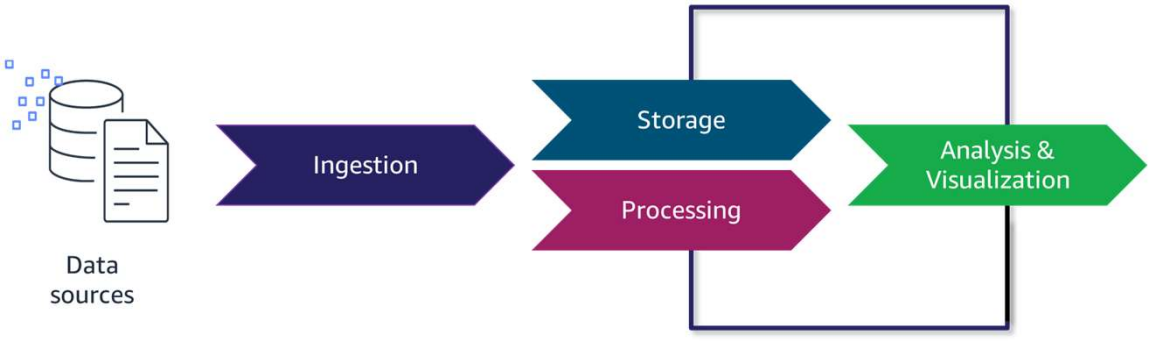
|Slide number 4
|Instructor notes
|Each module has an introduction, content sections, and a wrap-up. The wrap-up for this module contains a sample exam question for you to review with the students.
|
|Student notes
The objectives of this module are presented across multiple sections.

You will view a demo that uses AWS IoT Analytics and Amazon QuickSight to analyze and visualize data. You will also complete a hands-on lab that uses Amazon Kinesis Data Firehose, Amazon OpenSearch Service, and OpenSearch Dashboards to analyze and visualize data.

The module wraps up with a sample exam question and an online knowledge check that covers the presented material.

|Slide number 5
|Instructor notes
|
|Student notes
This section explains the factors to consider when you select analysis and visualization tools.

The simplified data pipeline



|Slide number 6
|Instructor notes
|
|Student notes
For accessibility: Data sources are ingested and move to storage and processing, followed by analysis and visualization. Processing between the storage, processing, and analysis and visualization phases can be iterative. **End of accessibility description.**

Throughout this course, you learned about the layers of a data pipeline. Raw data makes its way through the stages of the pipeline: ingestion, storage, processing, and analysis and visualization to produce insights about the data.

This simplified diagram illustrates where analysis and visualization is in the data pipeline. Although each layer is distinct, how data is analyzed and visualized is tied to how it is stored and processed.

Factors to consider when selecting tools

Business needs

Fully understand the business needs to be able to determine which data analyses and visualizations are needed to help develop insights.

Data characteristics

Understand the type and quality of the data, and how often it is updated and processed.

Access to data

Consider the data pipeline and who needs access to analyze and visualize data.

|Slide number 7
|Instructor notes
|This is the framework used to explain the factors that influence selecting analysis and visualization tools in this section.
|
|Student notes
Data engineers focus on the infrastructure that the data passes through. This three-part framework will guide you through the questions that you need to ask and the factors that you should consider when selecting data analysis and visualization tools and services. To build the correct pipeline, it's necessary to fully understand the desired outcomes based on business needs, the characteristics of the data, and who needs access to analyze and visualize data.

Business needs

- Which analyses are needed to help develop insights?
- What insights can be pulled from the data?
- What visualization would illustrate the insights?
- Does the consumer need to generate a report or interact with a dashboard?

|Slide number 8
|Instructor notes
|This is the first subsection of factors that influence selecting analysis and visualization tools.
|
|Student notes
Keep the end goal in mind as you design the analysis and visualization layer of the data pipeline.

Business needs: Granularity of insight

Industry	Detailed Level	Aggregate Level
Finance	A finance manager wants information (such as revenue, costs, and profit margins) about their line of business.	A CFO wants similar metrics at an aggregate level across all lines of businesses. They want to be able to drill down to any line of business.
Marketing	A marketing manager wants to know about the number of leads, opportunities, and closed deals within an area (such as a postal code or city).	A CMO is interested in related metrics but at a broader level (such as a state or region).
Sales	A sales manager is focused on their sales pipeline and wants to know how long it takes to close an opportunity. They want to assess how many opportunities are needed to achieve quota targets.	A VP of sales wants similar information at an aggregate level. They want to be able to drill down to a sales representative or sales territory.

aws

|Slide number 9

|Instructor notes

|In the list of resources for this course, you can access more business analytics examples if needed.

|

|Student notes

The granularity of insight relates to the level of data that one has access to. A finance manager has access to data for their line of business, whereas the chief financial officer (CFO) has access to all the data.

Consider the grain size of the data that needs to be analyzed and visualized. Is it at a detailed level or an aggregate level?

For example, a marketing manager might want to know how various demand channels are performing. A chief marketing officer (CMO) might be interested in similar metrics at an aggregate level across all lines of businesses, with the ability to drill down into any line of business. The table shows examples of the granularity of insight for finance, marketing, and sales.

Business needs: Visualizing insights





KPIs	Relationships	Comparisons	Distributions	Compositions
Show performance in a particular area or function.	Establish or prove whether a relationship exists between two or more variables.	Show or examine how different variables change over time, or provide a static snapshot of how different variables compare.	Show how your data is distributed over certain intervals (based on clustering or grouping of data).	Highlight the various elements that make up your data—its composition.

|Slide number 10
|Instructor notes
|For more explanation, visit the “Analyzing & Visualizing your Data for Business Analytics” link in the resources section. While data visualization is an important and vast area, this slide provides students with an overview of how the business need connects to the content of the visualization.
|
|Student notes
A key aspect of getting insight from your data is finding patterns. Patterns are often much easier to see in a graph or chart rather than staring at data in a table. The right visualization will help you gain a deeper understanding in a much quicker timeframe. A visualization might be produced as part of a report, or it might be used as part of an interactive dashboard where a user can drill down on something interesting.

Visualizations highlight one of the following five types of insights:

•Key Performance Indicators (KPIs) which are usually a single variable that measures how well you are doing. For example how many sales leads become sales.
•Relationships between two variables, for example, whether or not sales revenue is tied to marketing spending.

•Comparisons of how different variables change over time, for example showing month over month sales and web traffic.
•Distributions of data over certain clusters or groups, for example grouping customers by the number of purchases they've made.
•Compositions of elements that make up your data, for example, sales by region.

For more information, see the Data Visualization section of the aws.amazon.com website. A direct link is in the Content Resources section of the course.

Data characteristics

- How much data is there?
- At what speed and volume does it arrive?
- How frequently is it updated?
- How quickly is it processed?
- What type of data is it?

|Slide number 11
|Instructor notes
|After “business needs”, this is the second subsection of factors that influence selecting analysis and visualization tools.
|
|Student notes
Consider the volume, velocity, variety, veracity, and value of the data.

Data characteristics: Examples of data types

Volume and velocity

- **Historical analysis:** Visualize a year's worth of sales data. Users can drill down by region and salesperson.
- **Streaming Internet of Things (IoT) data:** Visualize the real-time error rates of sensors in a factory.

Variety and veracity

- **Structured data:** A relational database is queried to report on customer service tickets that were submitted in a specific period.
- **Unstructured data:** Sentiment analysis is performed on customer service emails.

Value

- A business analyst uses **periodic reports** to showcase and report results to leadership.
- A DevOps engineer uses **self-service dashboards** to monitor and analyze performance in real time.

aws

|Slide number 12
|Instructor notes
|The goal of this slide is to help students recall information they have been exposed to in previous modules of this course in a new format. Using the 5 Vs, this slide helps students frame the factors they need to consider and understand more about data characteristics.
|
|Student notes
Using the five Vs as an overarching framework, consider what types of data you are dealing with: historical analysis or streaming Internet of Things (IoT) data? Structured, semistructured, or unstructured data? How will the data be visualized—in a report or interactive dashboard?

Consider volume and velocity together as you make infrastructure decisions about how to collect, store, and process data. Consider how much data you need to ingest and how quickly you will ingest it.

Variety and veracity both relate to the data itself—what type of data is it and how trustworthy is it? The way that the data has been formatted and stored might affect how you analyze it.

Value is about ensuring that the outputs from analysis and visualization have business value, and knowing who needs to interact with and present the data visualizations.

Data characteristics: Two fraud detection use cases

Data characteristic considerations	Use case 1: Rule-Based (Batch Pipeline)	Use case 2: ML in Real Time (Streaming Pipeline)
How much data is there?	Millions of transactions (kilobytes to terabytes)	Millions of transactions (bytes to megabytes)
At what speed and volume is it arriving?	In predefined intervals (minutes to multiple days)	In real time (milliseconds to seconds)
How quickly is it processed?	Minutes to hours	Milliseconds to seconds
What type of data is it?	Structured and semistructured	Unstructured and semistructured data
What value do insights from the data provide?	Historical reporting of fraud cases Reactive approach	Ability to detect fraud in real time Proactive approach

|Slide number 13
|Instructor notes
|Students have been exposed to most of this content in previous modules of this course. The goal is encourage learners to make connections between data characteristics and business needs and capabilities.
|
|Student notes
For a fraud detection use case, consider the business needs first. For this purpose, two use cases are shown as examples here: rule-based fraud and machine learning (ML) in real time.

Next, consider the characteristics of the data by asking questions about the data. How much data is there? What value do insights from the data provide?

Rule-based fraud detection takes a reactive approach and provides a historical reporting of fraud cases. Structured data is ingested, analyzed, and reported in predefined intervals.

In contrast, the ML fraud detection example takes a proactive approach. It provides the ability to detect fraud in real time as millions of transactions go through the streaming pipeline.

Access to data

- Where does the data come from?
- Will the data need to be combined from multiple sources?
- Who needs access to the data and at what level? Who can access the tools?

|Slide number 14
|Instructor notes
|After “business needs” and “data characteristics,” this is the third subsection of factors that influence selecting analysis and visualization tools.
|
|Student notes
Keep these questions in mind as you select analysis and visualization tools.

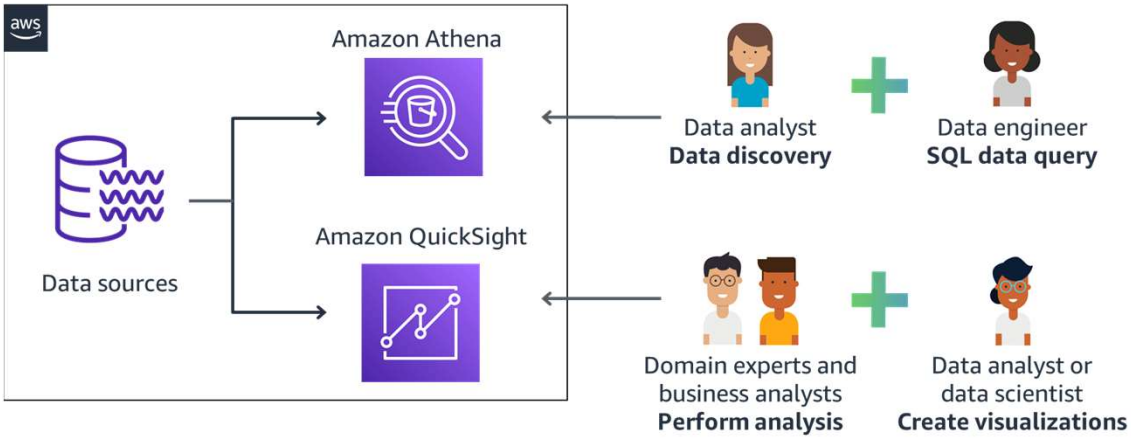
Access to data: **Consider authorization level based on role**

- A user's authorization to access data depends on their role in the organization.
- A business analyst or manager might be authorized to read the output that data engineers or data analysts create but not delete or update it.
- Follow the principle of least privilege. Give users the least amount of access and responsibility that are needed to complete their duties.

|Slide number 15
|Instructor notes
|
|Student notes
Based on the principle of least privilege, a user's authorization to access data depends on their role in the organization. Consider this as you think about selecting tools and services to analyze and visualize data.

The Securing and Scaling the Data Pipeline module and the Data Analytics Lens of the AWS Well-Architected Framework have more details about securing your data analytics workloads.

Access to data: Consider the functions of handling data



|Slide number 16
|Instructor notes
|This slides is meant to help students think about where the data is coming from and who accesses the data through which tools. Using this slide you could engage students to share their experiences and situations they have learned about related to different roles and functions of handling data.
|
|Student notes
A team might not include roles with the specific titles shown on this slide, but someone on the team is likely performing the same functions. A data analyst could work with a data engineer to query data and be able to perform interactive analysis by using Amazon Athena. Domain experts and business analysts could ask questions and explore the data with a data analyst or data scientist to create a visualization by using QuickSight.

Key takeaways: Considering factors that influence tool selection

- When you select analysis and visualization tools, consider the business needs, data characteristics, and access to data.
- Consider the granularity and format of the insights based on business needs.
- Consider the volume, velocity, variety, veracity, and value of your data.
- Consider the functions of individuals who will access, analyze, and visualize the data.

|Slide number 17
|Instructor notes
|
|Student notes
Here are a few key points to summarize this section.

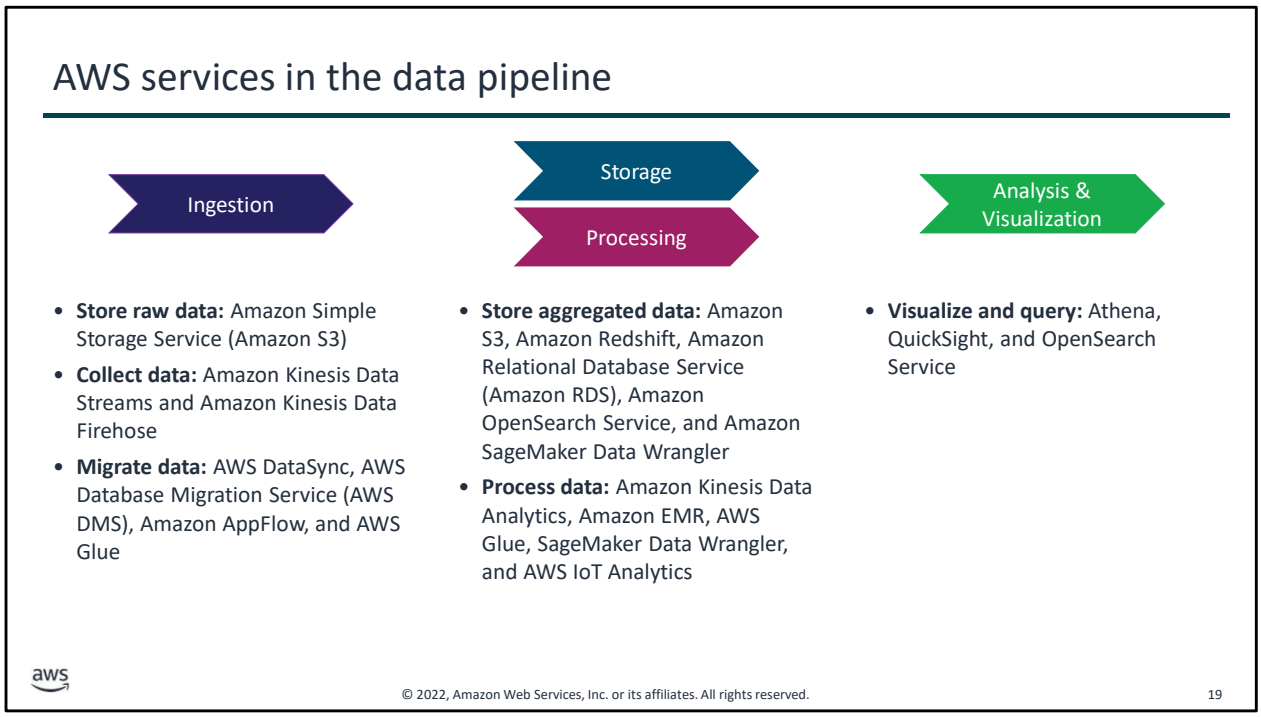
You learned about three factors to consider when you select tools to analyze and visualize data.

For business needs, consider the granularity and format of the insight.

For data characteristics, consider the volume, velocity, variety, veracity, and value of your data.

For access to data, consider the roles and functions of individuals who will access, analyze, and visualize the data.

|Slide number 18
|Instructor notes
|
|Student notes
This section compares AWS tools and services that are available to analyze and visualize data.



|Slide number 19
|Instructor notes
|Students see the most complete version of this slide of the AWS tools and services at this point in Module 11.
|
|Student notes
In this course, you have already encountered the tools and services that are related to the ingestion, storage, and processing layers of the data pipeline.

AWS tools and services that are commonly used to visualize and query data include Athena, QuickSight, and OpenSearch Service. In this section, you will learn about each service's features and compare them.

Amazon Athena

Amazon Athena is an interactive query service that provides the ability to use SQL to analyze data in Amazon S3. Athena includes the following features:

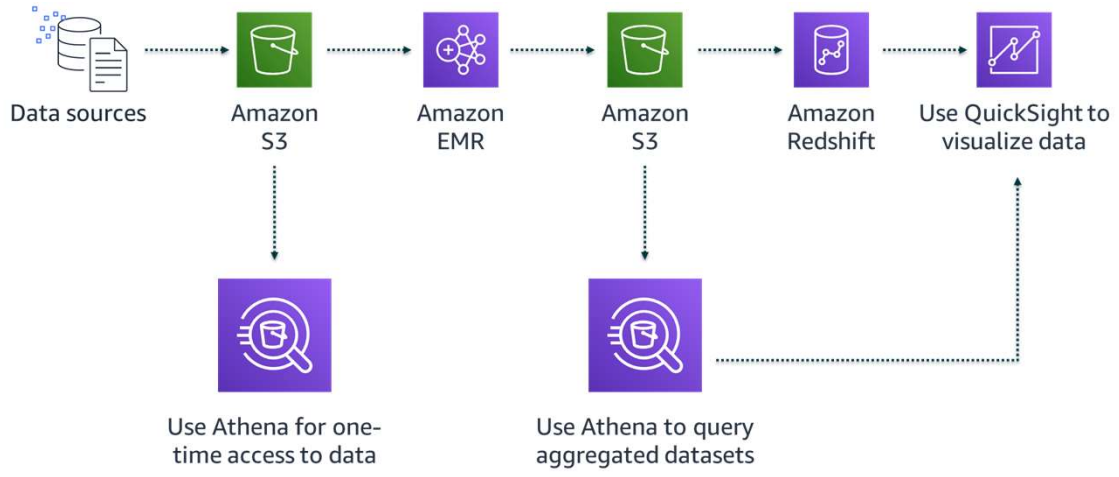
- Is serverless
- Provides the ability to combine data from multiple data sources
- Can be used for one-time queries
- Can be used from your favorite business intelligence (BI) tools (such as QuickSight)
- Can update data stored in Amazon S3 with Apache Iceberg integration

aws

|Slide number 20
|Instructor notes
|If needed, two links in the resources section are shared. The Amazon Athena Workshop Introduction helps provide an overview. The Amazon Athena Overview page has customer case studies listed.
|
|Student notes
Amazon Athena is an interactive query service in which you can use standard SQL to query and analyze data.

Athena is serverless. You don’t need to set up anything, and you don’t need to manage servers or data warehouses. You can use Athena to query data in place and combine data sources. This functionality is helpful for data analysts to perform one-time queries.

Athena: One-time querying for data in Amazon S3



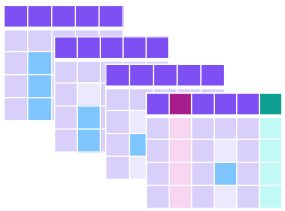
|Slide number 21
|Instructor notes
|If needed, this is the link to more information about this service:
|
|Student notes
For accessibility: Data from multiple sources is put in Amazon S3, where Athena can be used for one-time queries. Amazon EMR aggregates the data and stores the aggregates in S3. Athena can be used to query the aggregated datasets. From S3, the data can be used in Amazon Redshift, where QuickSight can access the data to create visualizations. **End of accessibility description.**

The diagram on the slide illustrates an example of using Athena to perform one-time queries of data that is stored in Amazon Simple Storage Service (Amazon S3).

You can upload data from multiple sources into Amazon S3 and query it using Athena for one-time access to data. You can also use Athena to query aggregated datasets.

Athena: Capability to update data in Amazon S3

- Users can use Athena to insert, update, and delete data that is stored in Amazon S3 (with the Apache Iceberg integration).
- Users can also track data versions automatically.
- The Apache Iceberg integration provides a way for continuous ingestion and updates.



|Slide number 22
|Instructor notes
|
|Student notes
Using Athena, you can query data that is stored in Amazon S3, but you can't update the data.

Consider an example in which survey responses are stored in Amazon S3. Some responses aren't clear and need to be updated during the verification process. With the Apache Iceberg integration, the values are easily updated in Athena.

Amazon QuickSight

- Is a cloud-scale BI service to deliver easy-to-understand insights
- Connects to data in the cloud and combines data from many different sources
- Gives decision-makers the opportunity to explore and interpret information in an interactive visual environment
- Provides forecasting visualization capabilities
- Provides ability to ask questions using natural language with QuickSight Q feature

|Slide number 23
|Instructor notes
|To show an example of forecasting in QuickSight, an example of a line chart that shows multiple variables changing over time and spanning future months is in the resource related to “Forecasting with Amazon Quick Sight”.
|
|Student notes
QuickSight helps decision-makers interact with data in a visual environment. The service is designed to deliver easy-to-understand insights quickly.

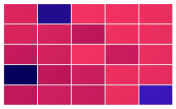
For an introductory video about QuickSight and its interface, see the Amazon QuickSight Overview link in the Content Resources for the course.

QuickSight example use case

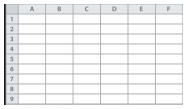
To visualize the sentiments, phrases, and tweets for a specific topic in QuickSight, you will view the following:

A donut chart with the overall sentiment for that specific topic sentiment

A word cloud of the phrases in the most dominant topic



A heat map of tweets that are associated with the dominant topics



A tabular view of related tweets

|Slide number 24

|Instructor notes

|In addition to the implementation guide of the use case shown in this slide, if needed, there is a workshop to use QuickSight with Athena to visualize data in the resources section.

|

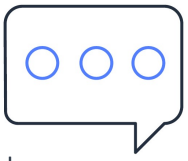
|Student notes

In this case of using QuickSight, the situation is to analyze and visualize tweets that are related to a specific topic. The solution uses QuickSight to create a donut chart, word cloud, heat map, and table of the overall sentiment, most dominant phrases, and associated and related tweets.

For more information about this use case, see “Example Use Cases for Amazon QuickSight” in the Content Resources of the course.

QuickSight Q: Ask questions using natural language

When you need a visualization that isn't already in the dashboard, instead of waiting for the BI team to get back to you, use QuickSight Q to get immediate responses.



Example: You could say "Show me last year's weekly sales in California," and Q would provide an answer in a few seconds.

- Is a BI capability that's powered by ML
- Uses natural language processing
- Doesn't require you to build pre-defined data models or dashboards

|Slide number 25
|Instructor notes
|
|Student notes
Amazon QuickSight Q is powered by machine learning (ML) and provides a method for everyone in your organization to better understand your data. Users ask questions in natural language and receive accurate answers with relevant visualizations to help them gain insights from the data.

Amazon OpenSearch Service

- This managed service helps you deploy, operate, and scale OpenSearch clusters in the AWS Cloud.
- Use this open-source search and analytics engine for use cases such as the following:
  - Log analytics
  - Real-time application monitoring
  - Clickstream analytics

Amazon OpenSearch Service is integrated with visualization tools including OpenSearch Dashboards and Kibana.

|Slide number 26
|Instructor notes
|In the resource section, use the “What is OpenSearch?” link to learn more about the features that OpenSearch provides and how it relates to Amazon OpenSearch.
|
|Student notes
OpenSearch is an open-source search and analytics engine. Amazon OpenSearch Service is a managed service that provisions all the resources for your cluster and deploys it. The cluster is easy to set up, operate, and scale.

Example use case for OpenSearch Dashboards

Situation: Analyze and visualize support calls

Solution:

- Use Amazon S3, Amazon Transcribe, and Amazon Comprehend to get full transcripts of calls, keywords from the transcripts, and an overall sentiment of each call.
- Use OpenSearch Service and OpenSearch Dashboards to search and visualize the data:
  - A pie chart of the overall sentiment (positive, negative, neutral)
  - A bar chart of the keywords that are mentioned in the support calls
  - A histogram of when and how often the calls were made

|Slide number 27
|Instructor notes
|See student notes.
|
|Student notes
The example on this slide uses OpenSearch Dashboards to visualize data from support calls.

The business need is to be able to analyze and visualize support calls, such as what is the subject of each call? How many were positive? How many were negative? How can managers search or review the transcripts of these calls?

The solution uses other AWS services and tools to transcribe the support calls, and then uses OpenSearch Service and OpenSearch Dashboards to search and visualize the data.

For more information about this use case, see “Visualize with QuickSight Using Athena” in the Content Resources page of your course.

Comparison of AWS services for data analysis and visualization

Athena

- Interactive analysis using SQL
- Analyze data directly
- Start querying data instantly
- Serverless

QuickSight

- Dashboards and visualizations
- Build visualizations and dashboards for business analytics
- Serverless

OpenSearch Service

- Operational analytics
- Search, explore, filter, aggregate, and visualize data in near real time
- Fully managed service

|Slide number 28
|Instructor notes
|
|Student notes

Now that we have looked at the features of these AWS services, let's compare them. This comparison can help you know which service to choose for your particular use case.

Athena provides interactive analysis using SQL, while QuickSight is for dashboards and visualizations. OpenSearch Service is for operational analytics.

With Athena, you can start querying data instantly and directly in Amazon S3. With QuickSight, you can visualize data into insights quickly, and with OpenSearch Service, you can visualize data in near real time.

Athena and QuickSight are serverless, and OpenSearch Service is a fully managed service.

Key takeaways: Comparing AWS tools and services

- AWS tools and services that are commonly used to query and visualize data include Athena, QuickSight, and OpenSearch Service.
- Athena is used for interactive analysis with SQL.
- Decision-makers can use QuickSight to interact with data visually and get insight quickly.
- OpenSearch Service is used for operational analytics to visualize data in near real time.

|Slide number 29
|Instructor notes
|
|Student notes
Here are a few key points to summarize this section. You compared three AWS tools and services that are commonly used to query and visualize data.

Athena is used for interactive analysis with SQL. QuickSight is used to interact with data visually and get insight quickly. OpenSearch Service is used for operational analytics to visualize data in near real time.

Demo: Analyzing and Visualizing Data with AWS IoT Analytics and QuickSight

- This recorded demo shows an analysis and visualization solution to monitor remote devices by using AWS IoT Analytics and QuickSight.
- The demo will show how to capture data from the remote devices and store it in a dataset. Then, QuickSight will be used to create an interactive chart to gain insights from the dataset.

|Slide number 30
|Instructor notes
|
|Student notes
You will now view a recorded demonstration of analyzing and visualizing data with AWS IoT Analytics and QuickSight.

|Slide number 31
|Instructor notes
|
|Student notes
In this section, you will apply what you learned about factors to consider when selecting tools, and the AWS tools and services that are available, to a gaming analytics use case.

Applying what you have learned in a use case for gaming analytics

Factors that influence selection of analysis and visualization tools

- Business needs
- Data characteristics
- Access to data

AWS tools and services for analysis and visualization

- Athena
- QuickSight
- OpenSearch Service

Selecting tools for a gaming analytics use case

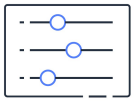
- Solutions based on a particular use case
- Solutions based on personas in the use case

|Slide number 32
|Instructor notes
|This slide explains how the previous sections lead up to this section in which students apply what they have learned in a use case.
|
|Student notes
Understand the use case and business needs to determine the potential solutions using AWS tools and services. Also, consider the personas and their functions within that context as you select AWS tools and services. See the next slides for an example.

Three personas in this use case for gaming analytics

How do different personas in a gaming company use AWS tools and services within different stages of the gaming analytics pipeline?

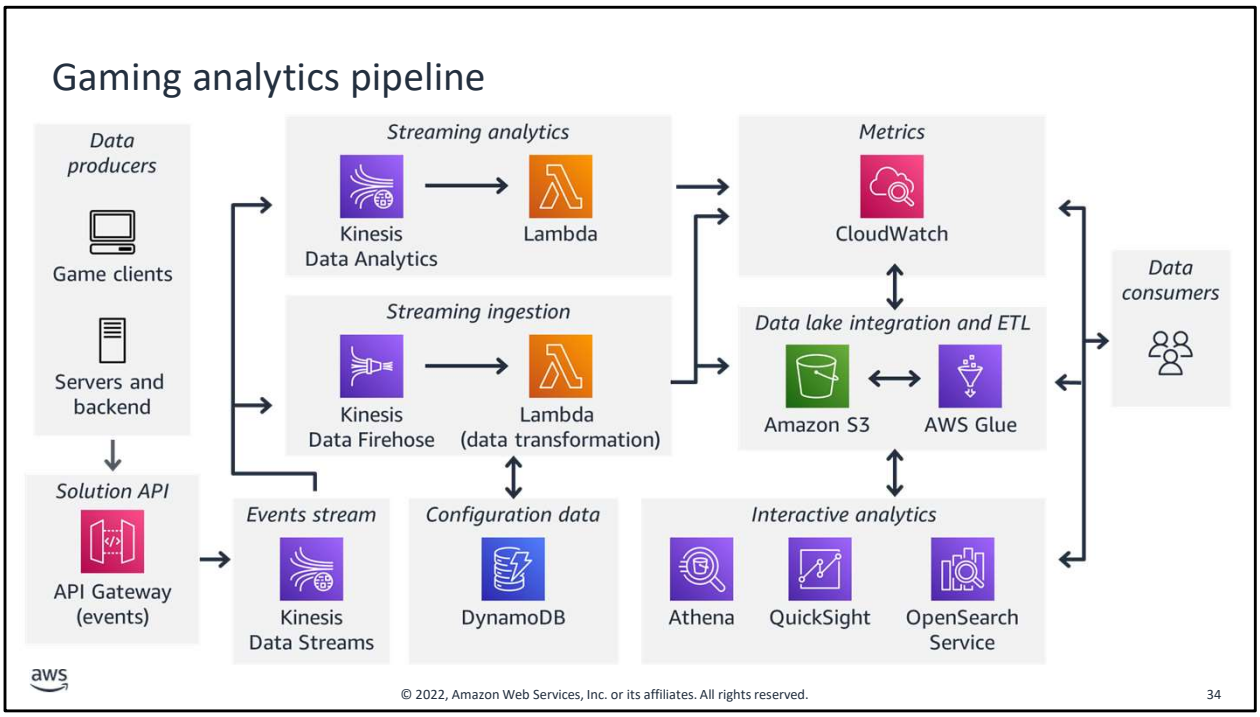
- **Analyst:** Explore and analyze player data
- **Business user:** Showcase and report results to leadership
- **DevOps engineer:** Monitor and analyze performance in real time



|Slide number 33
|Instructor notes
|
|Student notes

A gaming company collects and analyzes data on an ongoing basis. The DevOps engineer monitors the health and performance of gaming servers, and the data analyst continuously analyzes player statistics and explores usage patterns. Business users also analyze data to showcase and visualize important information to leadership periodically.

Let’s look at how three different roles—analyst, business user, DevOps engineer—might use a gaming analytics pipeline within the company.



|Slide number 34
|Instructor notes
|
|Student notes
The servers and backend infrastructure that support the games produce data that Amazon Kinesis Data Streams ingests through Amazon API Gateway. Amazon DynamoDB acts as the store for the API keys. Amazon Kinesis Data Analytics and Kinesis Data Firehose process the data from Kinesis Data Streams in real time by invoking AWS Lambda functions. These functions process and transform the data. Amazon S3 then ingests the data from Kinesis Data Firehose for further batch analysis.

Amazon CloudWatch provides real-time monitoring of the gaming servers, and the logs and metrics are stored in Amazon S3 for further analysis and visualization. Athena queries the log files that are stored in S3 or connects directly to CloudWatch to query the log data with SQL for analysis. QuickSight is used to create in-depth visualizations of server and backend performance metrics. These capabilities enable predictive analysis of the infrastructure burden and help identify areas of potential concern. Finally, engineers use OpenSearch Service to perform real-time additional monitoring as well as debugging and other troubleshooting.

Using Athena to explore and analyze player data

- **Business need:** Generate insights by querying daily aggregates of player usage data
- **Data characteristics:** Batch data for financial and geographical usage patterns
- **Data accessed:** Player purchase history, play history, and geographic information

Data analysts

|Slide number 35
|Instructor notes
|
|Student notes
The analyst in this scenario is interested in helping developers improve gameplay experience by analyzing player statistics, such as identifying which types of weapons and ammunition the players use. Similar methods can be used to visualize data about the geographic locations of the game's player base. Doing so would help developers determine regions that might need additional infrastructure to support a growing gaming population. This information could also help to identify possible cost savings in regions where the base isn’t as large as initially predicted.

Using QuickSight to showcase and report results

- **Business need:** Visualize KPIs, such as average revenue per user, average revenue per paying user, retention rate, and conversion rate and forecasting
- **Data characteristics:** Combined data from multiple sources and aggregated to a high granularity level
- **Data accessed:** Player purchase history, play history, and geographic information

Business users

|Slide number 36
|Instructor notes
|
|Student notes
The business user's focus in this scenario is on summarizing, visualizing, and presenting metrics that are related to management concerns. The business user needs the capability to tailor reports to management's desired scope, which makes QuickSight an excellent tool for data visualization.

KPIs for gaming companies include metrics such as conversion rate. The conversion rate measures the number of users who have made a purchase during a specific time period. The ability to turn free users into paying users is a high priority for investors. QuickSight can help to showcase such metrics, including forecasting, effectively.

Using OpenSearch Service to monitor and analyze performance in real time

- **Business need:** Monitor health and performance, and analyze performance for predictive load balancing
- **Data characteristics:** Large volumes of streaming telemetry data and server logs, including structured and unstructured data
- **Data accessed:** Access logs and performance data for game servers

DevOps engineers

|Slide number 37
|Instructor notes
|
|Student notes
The DevOps engineers perform real-time monitoring of the gaming servers and backend infrastructure to ensure that gameplay is smooth and uninterrupted. They use OpenSearch Service to visualize the health of gaming servers and backend performance metrics in real time. These capabilities enable predictive analysis of the infrastructure burden and help identify areas of potential concern.

Key takeaways: Selecting tools for a gaming analytics use case

- This use case showcased the granularity of visualized insights:
  - Daily batch aggregates of client usage patterns (Athena)
  - Consolidated aggregate KPIs for leadership (QuickSight)
  - Continuous health and performance monitoring (OpenSearch Service)
- Keep the influencing factors in mind when you select AWS tools and services. Multiple solutions exist to meet the business needs of data analysis and visualization.

|Slide number 38
|Instructor notes
|
|Student notes
Here are a few key points to summarize this section.

The use case showcased the granularity of visualized insights. Daily batch aggregates of client usage patterns were visualized by using Athena. Consolidated aggregate KPIs were visualized for leadership through QuickSight. Continuous health and performance monitoring was visualized with OpenSearch Service.

Keep the influencing factors in mind when you select AWS tools and services. Multiple solutions exist to meet the business needs of data analysis and visualization.

|Slide number 39
|Instructor notes
|
|Student notes
You will now complete a lab. The next slide summarizes what you will do in the lab, and you will find the detailed instructions in the lab environment.

Lab introduction:
Analyzing and Visualizing Streaming Data with Kinesis Data Firehose, OpenSearch Service, and OpenSearch Dashboards

- In this lab, you will analyze user activity and access patterns for a university bookstore website by using streaming log data.
- You will build an index for CloudWatch log events in OpenSearch Service.
- Then, you will use OpenSearch Dashboards to visualize the data:
  - Create a pie chart to show the operating systems and browsers that visitors use to consume the website.
  - Create a heat map to show how users are referred to product pages.
- Open your lab environment to start the lab and find additional details about the tasks that you will perform during this lab.

|Slide number 40
|Instructor notes
|
|Student notes
Access the lab environment through your online course to get additional details and complete the lab.

Debrief: Analyzing and Visualizing Streaming Data with Kinesis Data Firehose, OpenSearch Service, and OpenSearch Dashboards

- In this lab, you had to log in to OpenSearch Service with credentials other than your AWS credentials.
  - Which service managed user authentication to OpenSearch Service?
  - Which service managed user authorization?
  - Why were AWS services needed as part of the architecture in this lab?
- What is the purpose of index patterns in OpenSearch?
- What was the purpose of the Lambda function named **os-demo-lambda-function**?
- What was the purpose of the **agent.json** file on the Apache web server?
- Why was CloudWatch needed as part of the overall solution?

aws

|Slide number 41
|Instructor notes
|Q1 - **Answer:** Amazon Cognito; IAM; **example strong response:** OpenSearch Service depends on Amazon Cognito for authentication and IAM for authorization. For example, Amazon Cognito hosts the login screen for OpenSearch.
|Q2 - **Example strong response:** OpenSearch uses index patterns to identify the OpenSearch indices that you want to use to create visualizations. In the lab, we used the datetime field to filter for the data that we were interested in.
|Q3 - **Example strong response:** This Lambda function enriched the web server logs with more information. The function determined the site visitor's geographical location by converting the visitor's IP address to a location.
|Q4 - **Example strong response:** The file was used to connect the web server to the Kinesis Data Firehose delivery stream.
|Q5 - **Example strong response:** CloudWatch can be used to capture custom logs from non AWS services, such as web access logs from an Apache web server. CloudWatch is integrated with Kinesis Data Firehose, which is integrated with Lambda. Lambda was used to enrich the access log data. After enrichment, the data was returned to Kinesis Data Firehose and then sent to CloudWatch for later analysis. The flexibility of CloudWatch to capture custom logs from non AWS services made the solution possible.
|

|Student notes

Your instructor might review these questions with you, or you might review them on your own. Use this opportunity to extend your thinking about the tasks that you performed during the lab.

|Slide number 42
|Instructor notes
|
|Student notes
This section summarizes what you have learned and brings the Analyzing and Visualizing Data module to a close.

Module summary

This module prepared you to do the following:

- List factors to consider when selecting analysis and visualization tools.
- Compare available AWS tools and services for data analysis and visualization.
- Determine the appropriate AWS tools and services to analyze and visualize data based on influencing factors (business needs, data characteristics, and access to data).

|Slide number 43
|Instructor notes This is a good opportunity to ask students to reflect on what they have learned using an online group or discussion board. You might ask the students to recall a point from the module that aligns to one of the objectives listed. This provides a good segue to the knowledge check and sample question.
|
|Student notes
This module described factors that influence selecting analysis and visualization tools. You compared AWS tools and services for data analysis and visualization. Ultimately, the module has prepared you to determine appropriate AWS tools and services to analyze and visualize data based on influencing factors (business needs, data characteristics, and access to data).

Module knowledge check

- The knowledge check is delivered online within your course.
- The knowledge check includes 10 questions based on material presented on the slides and in the slide notes.
- You can retake the knowledge check as many times as you like.

|Slide number 44
|Instructor notes
|
|Student notes
Use your online course to access the knowledge check for this module.

Sample exam question

A company's web application produces 250 GB of clickstream data per day, which is stored in Amazon S3. The company wants to use the data to find out how quickly the application's webpages loaded during the last month. They want to be able to compare month-to-month data to gain insights. Which data analysis and visualization solution would provide these capabilities while also minimizing the cost and complexity?

Identify the key words and phrases before continuing.

The following are the key words and phrases:

- 250 GB of **clickstream data** per day
- **How quickly** the application's webpages loaded during the **last month**
- **Compare month-to-month data** to gain insights
- **Minimizing the cost and complexity**

aws

|Slide number 45
|Instructor notes
|The key words section is animated to be revealed on click.
|
|Student notes
Select the most cost effective data analysis and visualization solution that you would consider as a data engineer.

Sample exam question: Response choices

A company's web application produces 250 GB of **clickstream data** per day, which is stored in Amazon S3. The company wants to use the data to find out **how quickly** the application's webpages loaded during the **last month**. They want to be able to **compare month-to-month** data to gain insights. Which data analysis and visualization solution would provide these capabilities while also **minimizing the cost and complexity**?

Choice	Response
A	Use OpenSearch Service to analyze and OpenSearch Dashboards to visualize the data.
B	Use Apache Spark to analyze the data and Amazon EMR Hive to visualize the data.
C	Use Kinesis Data Analytics to analyze the data and Microsoft Excel to visualize the data.
D	Use Athena to analyze the data and QuickSight to visualize the data.

aws

|Slide number 46
|Instructor notes
|
|Student notes
Use the key words that you identified on the previous slide, and review each of the possible responses to determine which one best addresses the question.

Sample exam question: Answer

The correct answer is D.

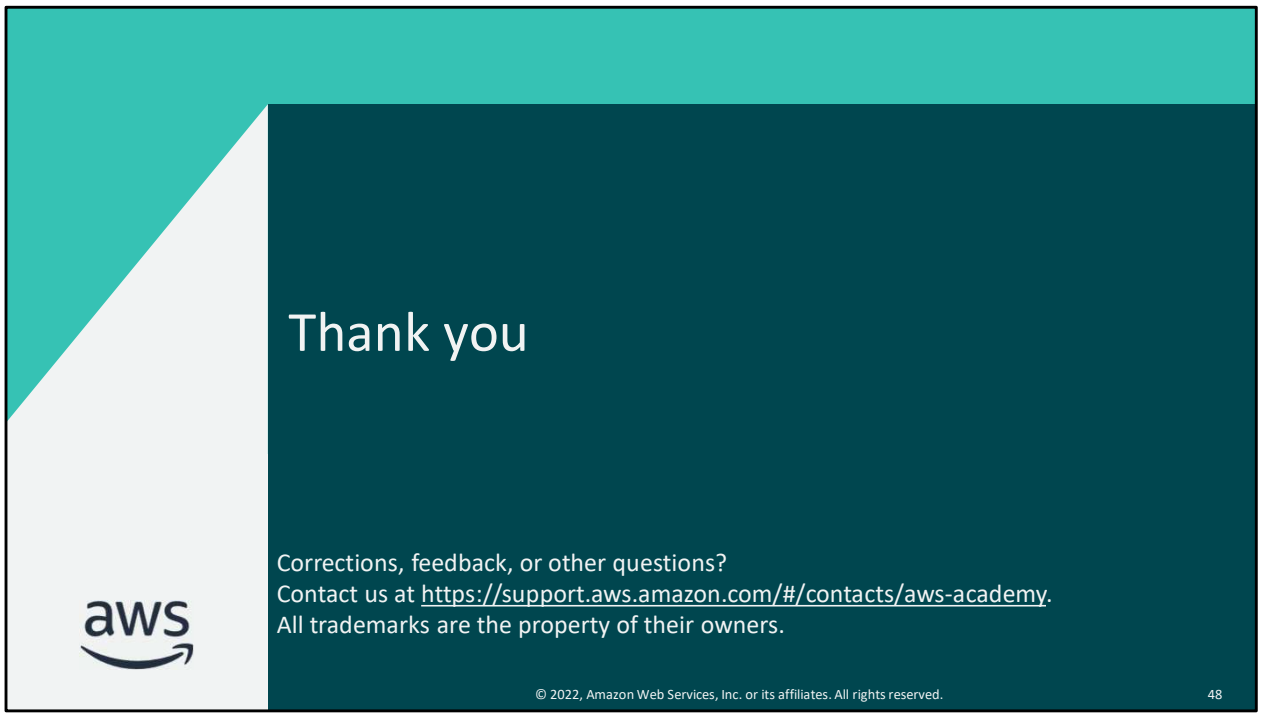
Choice	Response
D	Use Athena to analyze the data and QuickSight to visualize the data.

|Slide number 47
|Instructor notes
|
|Student notes
Choice A (Use OpenSearch Service and OpenSearch Dashboards) is not the most cost-effective solution. It isn't cost-effective to collect and keep large amounts of historical data to use with OpenSearch Service and OpenSearch Dashboards.

Choice B (Use Apache Spark and Amazon EMR Hive) is not an appropriate solution for data analysis and visualization. While Spark can be used to process the data, Amazon EMR Hive does not offer capabilities for data visualization.

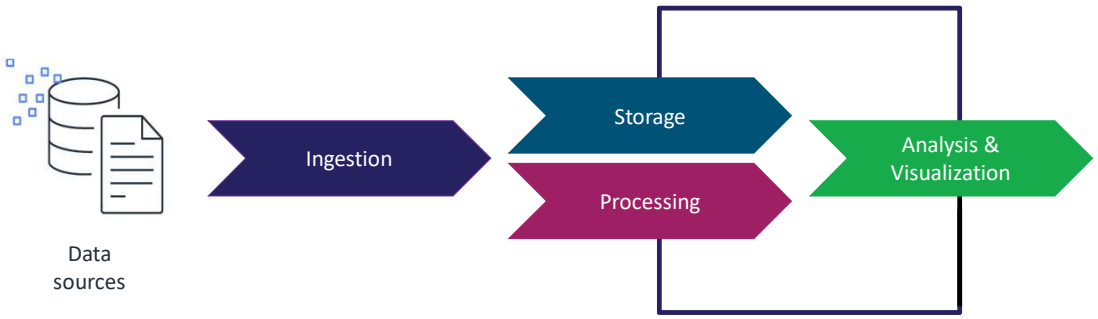
Choice C (Use Kinesis Data Analytics and Microsoft Excel) does not meet the requirement to minimize complexity. Needing to download the data increases complexity and prevents the solution from being scalable.

Choice D (Use Athena and QuickSight) is the most cost-effective and appropriate solution for the analysis and visualization of historical, clickstream data over time.



|Slide number 48
|Instructor notes
|
|Student notes
That concludes this module. The Content Resources page of your course includes links to additional resources that are related to this module. Thank you for completing this module.

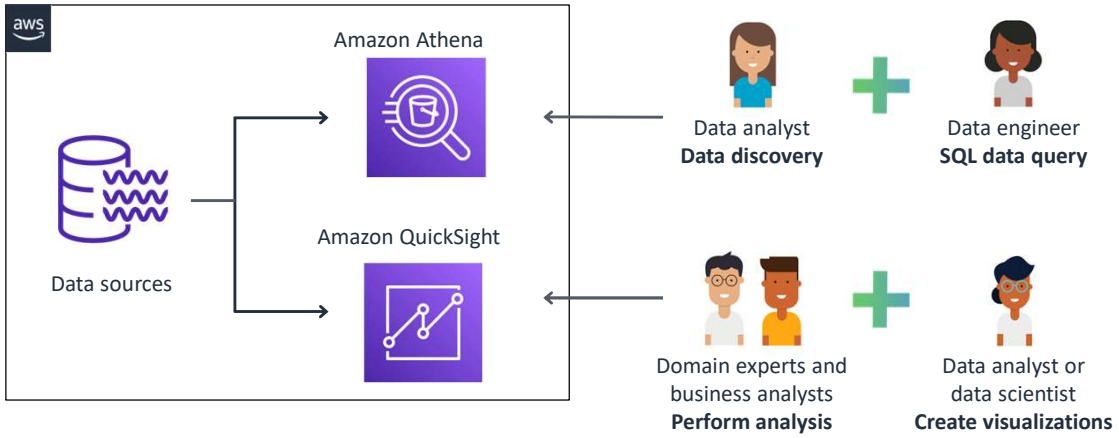
The simplified data pipeline



|Slide number 49
|Instructor notes
|
|Student notes
Throughout this course, you learned about the layers of a data pipeline. Raw data makes its way through the stages of the pipeline: ingestion, storage, processing, and analysis and visualization to produce insights about the data.

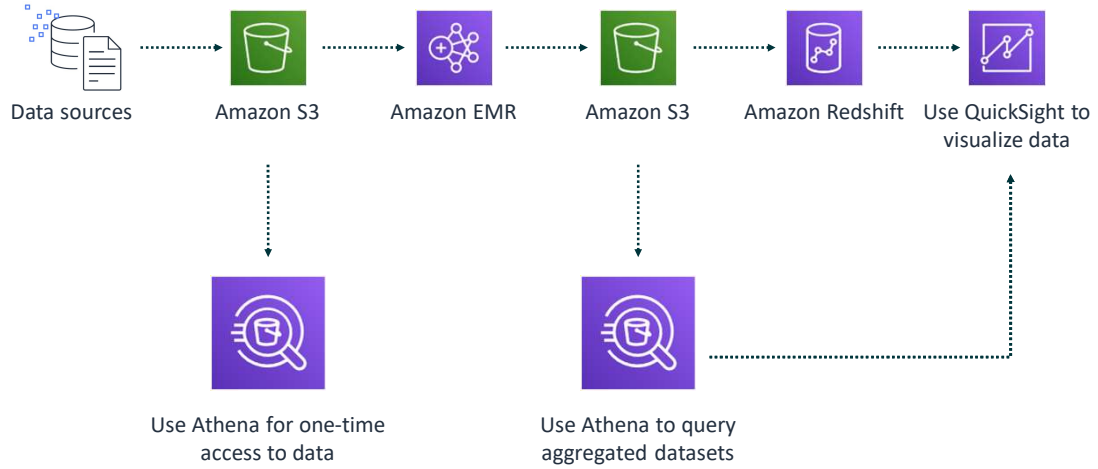
This simplified diagram illustrates where analysis and visualization is in the data pipeline. Although each layer is distinct, how data is analyzed and visualized is tied to how it is stored and processed.

Access to data: Consider the functions of handling data



|Slide number 50
|Instructor notes
|This slides is meant to help students think about where the data is coming from and who accesses the data through which tools. Using this slide you could engage students to share their experiences and situations they have learned about related to different roles and functions of handling data.
|
|Student notes
A team might not include roles with the specific titles shown on this slide, but someone on the team is likely performing the same functions. An analyst could work with a data engineer to query data and be able to perform interactive analysis by using Amazon Athena. Domain experts and business analysts could ask questions and explore the data with a data analyst or data scientist to create a visualization by using QuickSight.

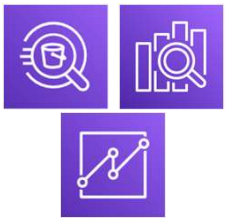
Athena: One-time querying for data in Amazon S3



|Slide number 51
|Instructor notes
|If needed, this is the link to more information about this service:
|
|Student notes
The diagram on the slide illustrates an example of using Athena to perform one-time queries of data that is stored in Amazon Simple Storage Service (Amazon S3).

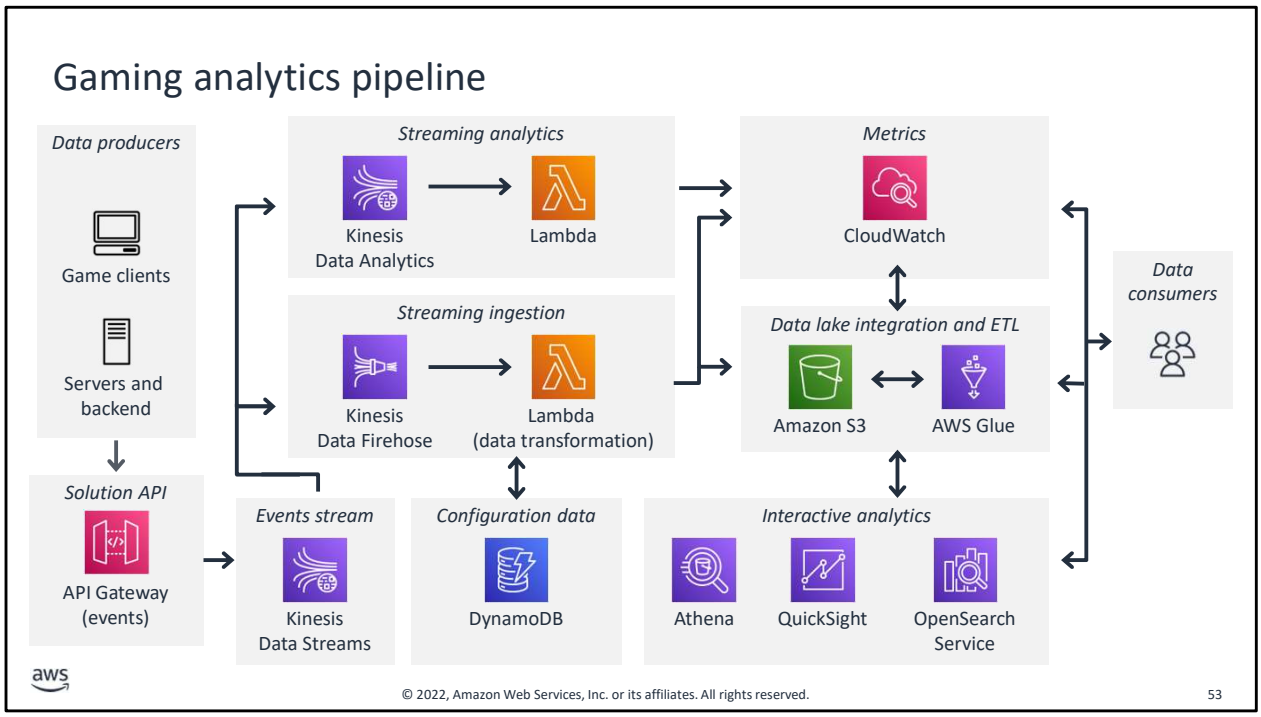
You can upload data from multiple sources into Amazon S3 and query it using Athena for one-time access to data. You can also use Athena to query aggregated datasets.

Applying what you have learned in a use case for gaming analytics



Factors that influence selection of analysis and visualization tools	AWS tools and services for analysis and visualization	Selecting tools for a gaming analytics use case
• Business needs • Data characteristics • Access to data	• Athena • QuickSight • OpenSearch Service	• Solutions based on a particular use case • Solutions based on personas in the use case

|Slide number 52
|Instructor notes
|This slide explains how the previous sections lead up to this section in which students apply what they have learned in a use case.
|
|Student notes
Understand the use case and business needs to determine the potential solutions using AWS tools and services. Also, consider the personas and their functions within that context as you select AWS tools and services. See the next slides for an example.



|Slide number 53
|Instructor notes
|
|Student notes
The servers and backend infrastructure that support the games produce data that Amazon Kinesis Data Streams ingests through Amazon API Gateway. Amazon DynamoDB acts as the store for the API keys. Amazon Kinesis Data Analytics and Kinesis Data Firehose process the data from Kinesis Data Streams in real time by invoking AWS Lambda functions. These functions process and transform the data. Amazon S3 then ingests the data from Kinesis Data Firehose for further batch analysis.

Amazon CloudWatch provides real-time monitoring of the gaming servers, and the logs and metrics are stored in Amazon S3 for further analysis and visualization. Athena queries the log files that are stored in S3 or connects directly to CloudWatch to query the log data with SQL for analysis. QuickSight is used to create in-depth visualizations of server and backend performance metrics. These capabilities enable predictive analysis of the infrastructure burden and help identify areas of potential concern. Finally, engineers use OpenSearch Service to perform real-time additional monitoring as well as debugging and other troubleshooting.