Data to Insights with Google Cloud course series

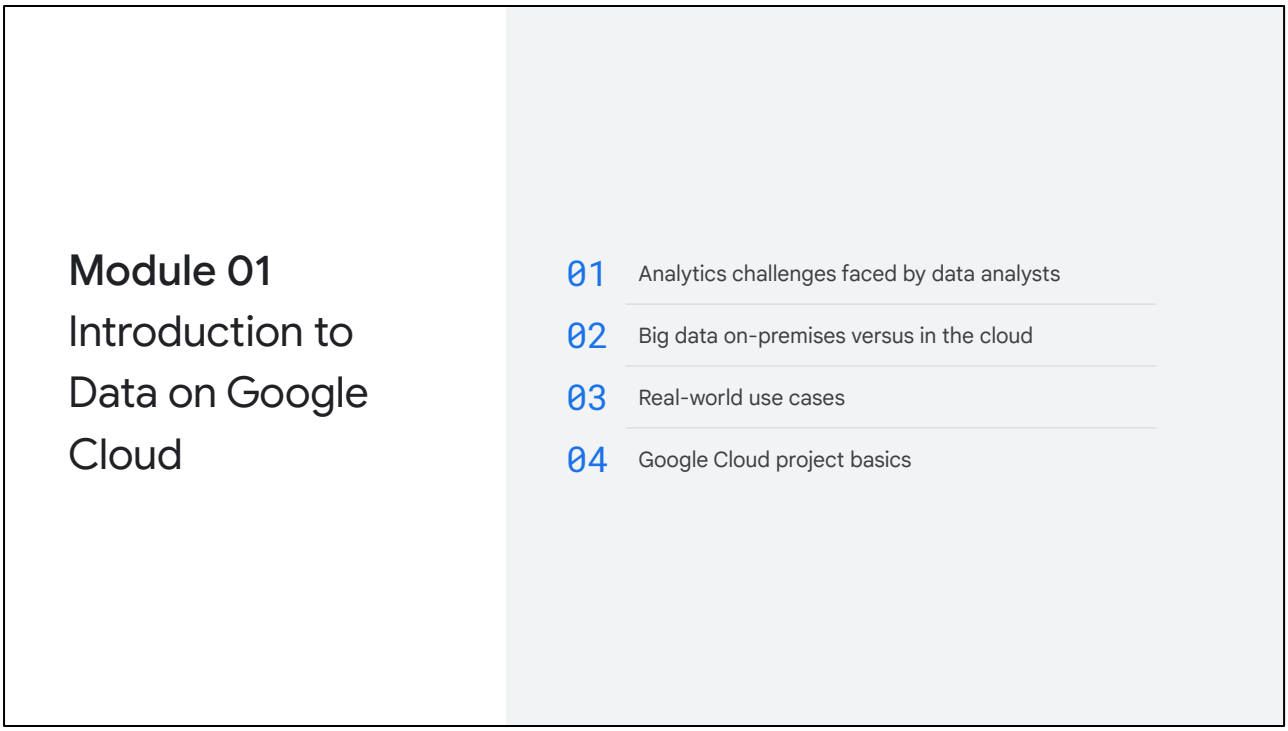
01 Exploring and Preparing your Data with BigQuery

02 Creating New BigQuery Datasets and Visualizing Insights

03 Achieving Advanced Insights with BigQuery

04 Applying Machine Learning to your Data with Google Cloud

Congratulations. You've completed **Exploring and Preparing your Data with BigQuery**, the first course in the Data to Insights with Google Cloud course series. Let's recap all that we've covered so far in this course series.



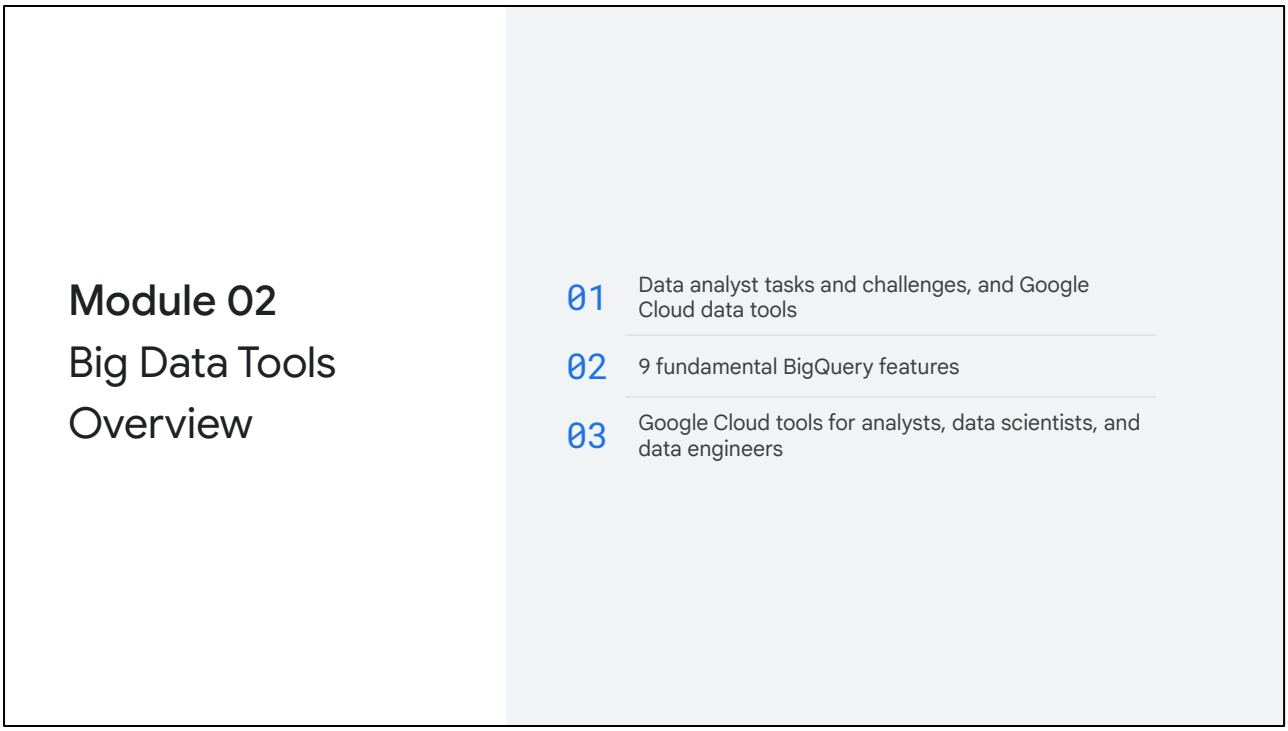
In module 1, data on Google Cloud was introduced.

You looked at the query, infrastructure, and storage challenges that are faced by data analysts. This included queries taking too long to run, there being no easy way to combine and query data collected, on-premises clusters not scaling with analyses, the cost of storing data, and the lack of a central data analytics warehouse or set of tools.

You compared traditional big data on-premises platforms that require a significant investment in infrastructure, with big data on Google Cloud. With Google Cloud, there is no need to focus on infrastructure, allowing you to focus on insights instead. Google Cloud enables efficient resource allocation through the separation of storage and computing power. BigQuery specifically scales automatically and you only pay for what you use.

You were introduced to two real-world use cases, Ocado and Spotify, both of whom turned to BigQuery as a solution for managing and leveraging their big data.

Lastly in module 1, you were introduced briefly to three key components of the Google Cloud Console dashboard: projects, resources, and billing,



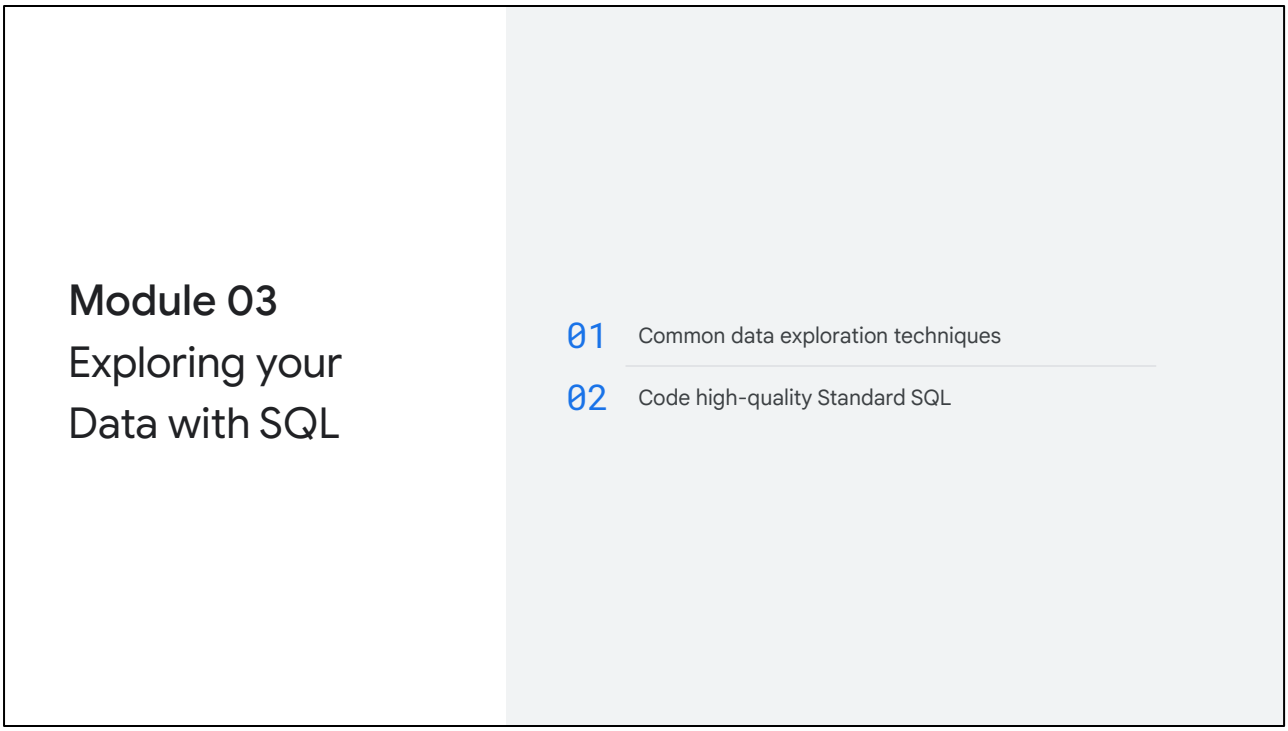
Module 2 provided you with an overview of big data tools in Google Cloud.

You identified 5 key task categories that apply to data analysts, namely ingest, transform, store, analyze, and visualize, as well as the challenges associated with each. You then identified the different scalable big data tools offered by Google Cloud to overcome these data challenges.

You learned about BigQuery, Google Cloud's petabyte-scale data analytics warehouse, and 9 fundamental features that allow you to focus on finding insights instead of managing infrastructure.

You then compared the roles of data analyst, data scientist, and data engineer in terms of what they do, their background, and their need for different Google Cloud tools.

Lastly in module 2, you completed a lab where you explored a BigQuery public dataset.

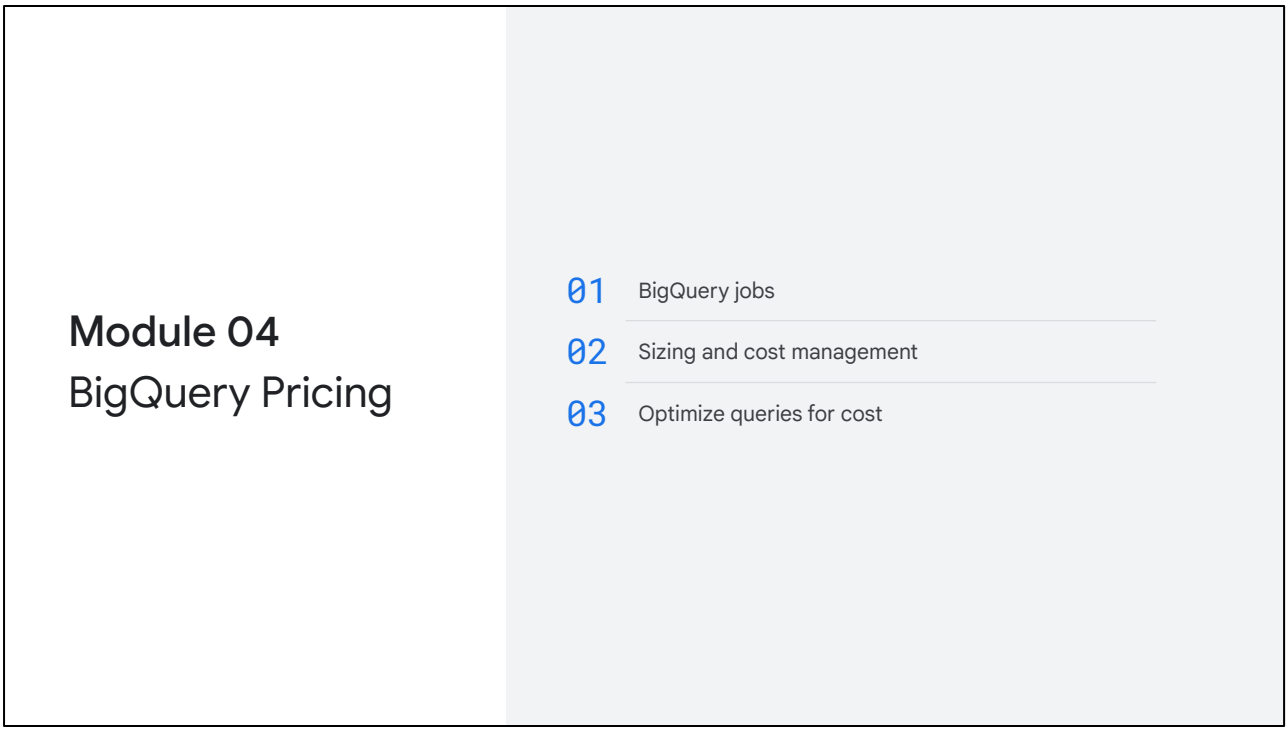


In Module 3, exploring data with SQL was introduced.

You started with 3 fundamental steps to exploring data through SQL, First ask good questions, second, know your data, and third, write good SQL!

You then learned a number of best practices for coding high-quality Standard SQL, including the use of clauses and functions.

You ended the module with a lab, where you troubleshot common SQL errors with BigQuery.



Module 4 discussed BigQuery pricing.

Your learned that the unit of work in BigQuery is called a Job and that there are four different types jobs, namely Query, Load data, Extract, and Copy. You also learned that only Query jobs incur a processing cost!

You then learned how to determine and control the cost of both storage and analytics.

Lastly, you were introduced to a few cost-optimization principles to apply when writing your queries.

Module 05
Cleaning and Transforming Your Data

01 5 principles of dataset integrity

02 Dataset shape and skew

03 Clean and transform data using SQL

04 Clean and transform data using a new UI: Introducing Dataprep

05 Create data pipelines with Data Fusion

In module 5, data cleaning and transformation was discussed.

You were reminded of the saying ‘Garbage in, garbage out’ and learned about the 5 strict integrity rules that high-quality datasets conform to. These are validity, accuracy, completeness, consistency, and uniformity.

You then learned about dataset shape, where ideally you have the right amount of columns and records to make judgements and inferences from your data and insights. You also learned about dataset skew, which is the distribution of values.

You revisited the 5 principles of dataset integrity and how these relate to cleaning and transforming data with SQL.

You were then introduced to Dataprep, a tool that allows you to apply the best practices learned for cleaning data through a drag-and-drop interface. You also had the opportunity to explore the Dataprep UI by building an ecommerce transformation pipeline that will run at a scheduled interval and output results back into BigQuery.

Lastly in module 5, you were introduced to Data Fusion. After exploring the components of the Data Fusion UI, learning how to build a pipeline, and exploring data using Wrangler, you completed a lab where you had the opportunity to build transformations and prepare data with Wrangler in Data Fusion.

Data to Insights with Google Cloud course series

01 Exploring and Preparing your Data with BigQuery

02 Creating New BigQuery Datasets and Visualizing Insights

03 Achieving Advanced Insights with BigQuery

04 Applying Machine Learning to your Data with Google Cloud

We look forward to welcoming you to the next course in the *Data to Insights with Google Cloud* course series, **Creating New BigQuery Datasets and Visualizing Insights**.

In the next course, you will learn how to bring your own datasets into BigQuery and how you can set up your own data analytics warehouse by joining together multiple datasets and visualizing those results inside dashboards.

See you there!