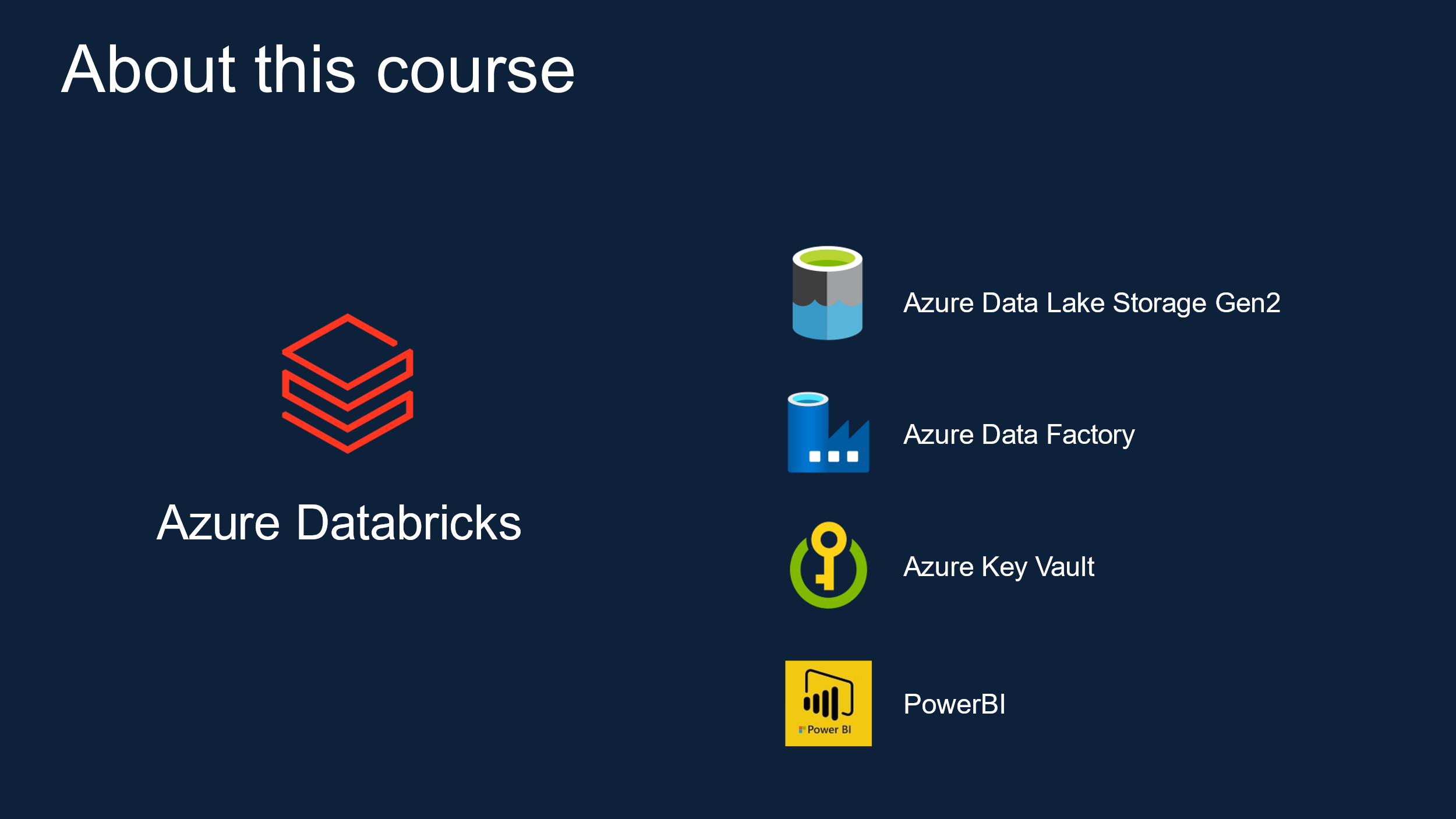
ASSESSMENT-21 Vinutha s

28/12/2023



The document Azure Databricks provides a comprehensive overview of a course focused on Azure Databricks, Azure Data Lake Storage Gen2, Azure Data Factory, Azure Key Vault, Power BI, and the Formula1 Cloud Data Platform. The course is designed for university students, IT developers from other disciplines, AWS/GCP/On-prem data engineers, and data architects, with a hands-on learning approach. It is not intended for those solely interested in Azure Data Engineering Certification, learning Spark ML or Streaming, or learning Scala or Java.

The course structure covers various components of Azure Databricks, including Azure Portal, Databricks, clusters, notebooks, data lake access, securing access, Databricks mounts, jobs, Spark overview, Delta Lake, and orchestration. It emphasizes Azure Databricks' integration with other tools and services such as Azure Data Factory, Azure Active Directory, Azure Dev Ops, and Azure ML.

The introduction to Azure Databricks highlights its integration with Microsoft Azure, Apache Spark, and its architecture, which includes Spark SQL, Spark Streaming, Spark ML, and Spark Graph. It also covers Databricks clusters and their types, configurations, pricing, cost control, and policies.

Furthermore, the document details the components and features of Azure Databricks, including workspaces, notebooks, clusters, jobs, data, models, and cluster types such as all-purpose and job clusters. It also delves into Databricks runtime, auto-termination, auto-scaling, cluster VM types/sizes, and policies for cost control and user interface simplification.

The course emphasizes the pricing factors and calculation for Azure Databricks clusters, including the normalized unit of processing power known as Databricks Unit (DBU). It provides a detailed breakdown of the pricing calculation, considering the workload, tier, VM type, and purchase plan.

The document also offers practical insights into creating and pricing Azure Databricks clusters, including the associated costs for different configurations. It provides links to the Azure Databricks and virtual machine pricing details for further reference.

In conclusion, the course aims to equip learners with a comprehensive understanding of Azure Databricks, its integration with other Azure services, cluster management, and cost-effective utilization. It targets a diverse audience, emphasizing practical, hands-on learning with step-by-step instructions, making it suitable for individuals seeking to enhance their skills in data engineering and analytics using Azure services.