



AZ-201 Designing Data

Microsoft AI+Tour LATAM



AZ-201 Designing Data

Consejos y Trucos para
Ingenieros de Datos



Microsoft
AI+Tour
LATAM



- **Resumen de actualización del examen DP-201**

- Este examen se modificó para diferenciar al ingeniero de datos de un administrador de base de datos, con énfasis en la idoneidad del ingeniero de datos en *Diseño* de soluciones para almacenamiento de datos en formas, estructuras y fines variables.
 - Enfoque eliminado en HDInsight
 - Enfoque reducido en casos de datos híbridos
- + información en
 - [Documentos de cambios](#)



Learning path for Azure Data Engineer

Azure Data Engineers design and implement the management, monitoring, security, and privacy of data using the full stack of Azure data services to satisfy business needs.



aka.ms/AzureDataEngLearning



Classroom

Implementing an Azure Data Solution DP-200T01

Digital skilling: free interactive training
content on Microsoft Learn



Azure fundamentals

8H 59M – 11 Modules

1. Principles of cloud computing
2. Azure architecture and service guarantees
3. Create an Azure account
4. Manage services with the Azure portal
5. Azure compute options
6. Azure data storage options
7. Azure networking options
8. Security, responsibility and trust in Azure
9. Apply and monitor infrastructure standards with Azure Policy
10. Control and organize Azure resources with Azure Resource Manager
11. Predict costs and optimize spending for Azure



Azure for the Data Engineer

1H 51M – 3 Modules

1. Understand the evolving world of data
2. Survey the services on the Azure Data Platform
3. Identify the tasks of a data engineer in a cloud-hosted architecture



Store data in Azure

3H 50M – 5 Modules

1. Choose a data storage approach in Azure
2. Create an Azure Storage account
3. Connect an app to Azure Storage
4. Secure your Azure Storage account
5. Store application data with Azure Blob storage



Work with relational data in Azure

3H 20M – 4 Modules

1. Provision an Azure SQL database to store application data
2. Create an Azure Database for PostgreSQL server
3. Scale multiple Azure SQL Databases with SQL elastic pools
4. Secure your Azure SQL Database



Work with NoSQL data in Azure Cosmos DB

5H 55M – 8 Modules

1. Create an Azure Cosmos DB database built to scale
2. Choose the appropriate API for Azure Cosmos DB storage
3. Insert and query data in your Azure Cosmos DB database
4. Store and access graph data in Azure Cosmos DB with the Graph API
5. Store and access NoSQL data with Azure Cosmos DB and the Table API
6. Build a .NET Core app for Azure Cosmos DB in Visual Studio Code
7. Optimize the performance of Azure Cosmos DB by using partitioning and indexing strategies
8. Distribute your data globally with Azure Cosmos DB



Large Scale Data Processing with Azure Data Lake Storage Gen2

2H 5M – 3 Modules

1. Introduction to Azure Data Lake Storage
2. Upload data to Azure Data Lake Storage
3. Secure your Azure Storage Account



Implement a Data Streaming Solution with Azure Streaming Analytics

1H 14M – 2 Modules

1. Work with data streams by using Azure Stream Analytics
2. Transform data by using Azure Stream Analytics



Implement a Data Warehouse with Azure SQL Data Warehouse

3H 21M – 4 Modules

1. Design a data warehouse with Azure SQL Data Warehouse
2. Query data in Azure SQL Data Warehouse
3. Import data into Azure SQL Data Warehouse using PolyBase
4. Data warehouse security



Perform data engineering with Azure Databricks

7H 49M – 9 Modules

1. Introduction to Azure Databricks
2. Access SQL Data warehouse instances with Azure Databricks
3. Data ingestion with Azure data factory
4. Read and write data by using Azure Databricks
5. Perform basic data transformations in Azure Databricks
6. Perform advanced data transformations in Azure Databricks
7. Create data pipelines by using Databricks Delta
8. Work with streaming data in Azure Databricks
9. Create data visualizations using Azure Databricks and Power BI



Extract knowledge and insights from your data with Azure Databricks

4H 21M – 6 Modules

1. Introduction to Azure Databricks
2. Read and write data by using Azure Databricks
3. Perform exploratory data analysis with Azure Databricks
4. Train, evaluate, and select machine-learning models with Azure Databricks
5. Deep learning with Azure Databricks
6. Perform text analytics with Azure Databricks



Exam DP-200:
Implementing an Azure Data Solution

Become a Microsoft Certified Professional

Certification helps establish your credentials with current and future employers.

Pass certification exams [DP-200](#) and [DP-201](#) to become a Microsoft Certified: Azure Data Engineer Associate.



Exam DP-201:
Designing an Azure Data Solution

Information subject to change. For the latest information, visit Microsoft.com/certification.

Free digital skilling:
Microsoft.com/Learn

Assess your skills at Pluralsight:
Azure.com/Pluralsight

Find a Learning Partner:
aka.ms/LearningPartner

Microsoft Certification:
Microsoft.com/Certification

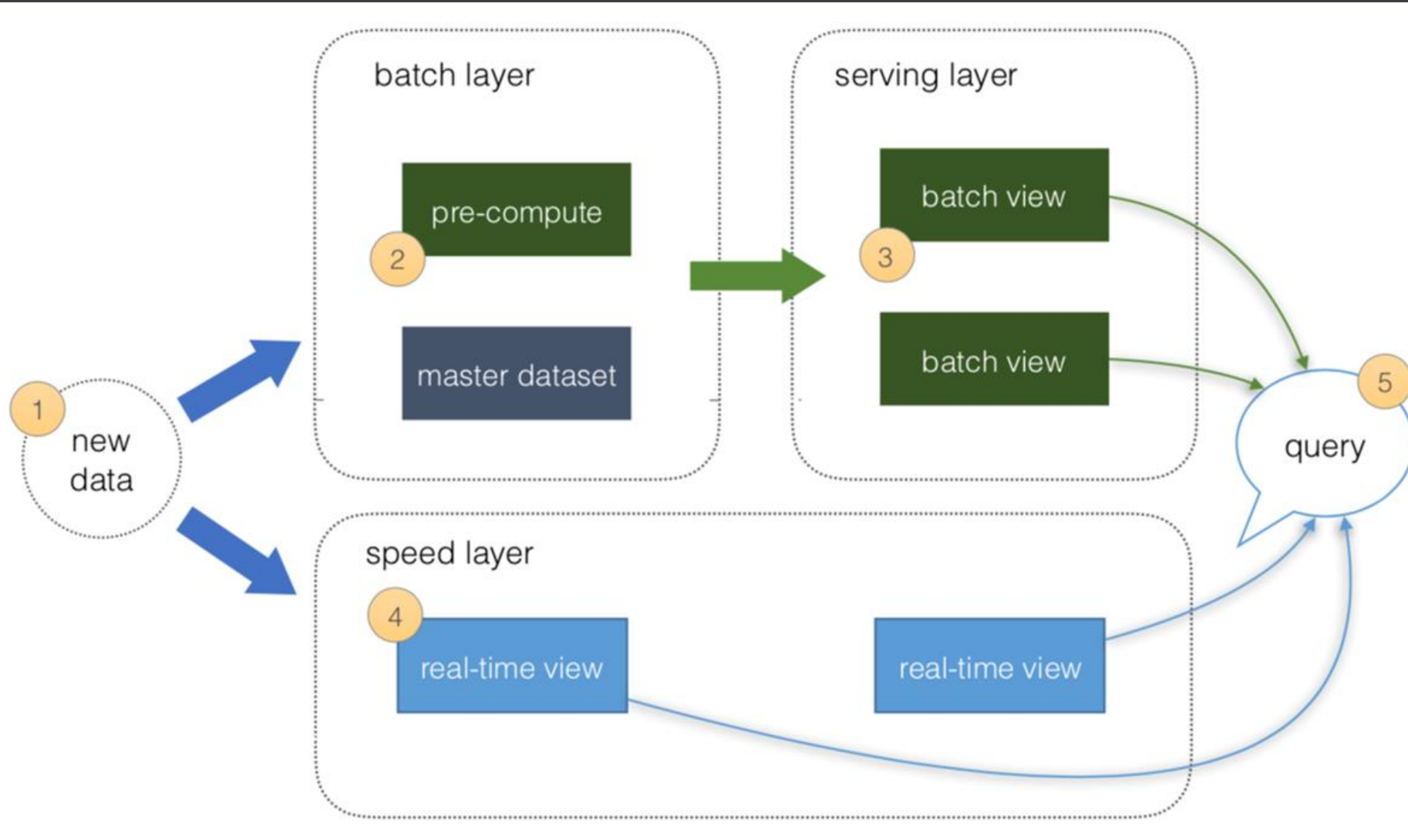


¿Que habilidades mide?



- **Diseño de soluciones de almacenamiento de datos de Azure**
 - 40-45%
- **Soluciones de procesamiento de datos de diseño**
 - 25-30%
- **Diseño para Seguridad de Datos y Cumplimiento**
 - 25-30 %

Arquitectura LAMBDA



Arquitectura LAMBDA



- Las arquitecturas Lambda nos ayuda a solucionar las necesidades de procesar datos para poder realizar análisis en información masiva, de una manera más rápida y cada vez más variable y de diferentes plataformas
- Se apoya del procesamiento por lotes, procesamiento de flujos y de una capa de servicio para minimizar la latencia al momento de consultar macrodatos

Arquitectura LAMBDA



- La arquitectura plantea 3 bloques/capas:
 - Capa batch (batch layer):
 - Almacena y procesa datos masivos para incluirlos en nuestros análisis históricos, es la ingesta para modelos los predictivos o de analítica avanzada
 - Capa veloz (speed layer)
 - Tomas los datos y los procesa en tiempo real/casi para análisis de tipo operacional
 - Capa de servicio (serving layer):
 - Se encarga de unifica y mostrar la información de las 2 capas anteriores
 - Ayuda a los usuarios a interpretar la información combinada
 - Le da una visión mayor sobre lo que está pasando



Arquitectura LAMBDA



Capa de lote

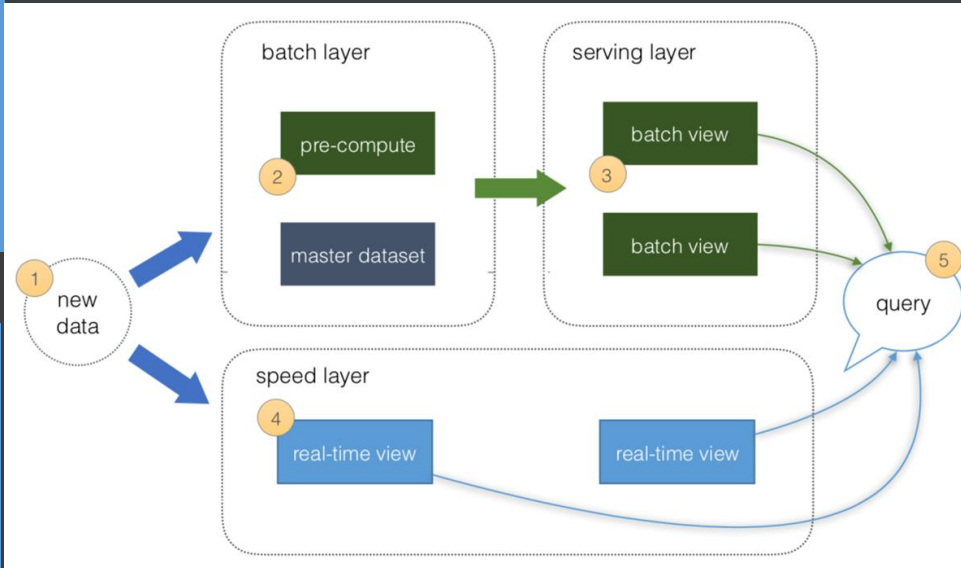
La capa Batch se usa para calcular previamente los resultados en grandes volúmenes de datos utilizando un sistema de datos distribuido.

1. Nuevos datos ingeridos desde las fuentes
2. Los datos se calculan previamente

Capa de servicio

La capa de servicio actúa como la salida de almacenamiento de la capa Batch o Speed que utilizan las aplicaciones cliente para acceder a los resultados de los conjuntos de datos.

3. Las vistas por lotes se pueden usar como vistas precalculadas



Arquitectura LAMBDA



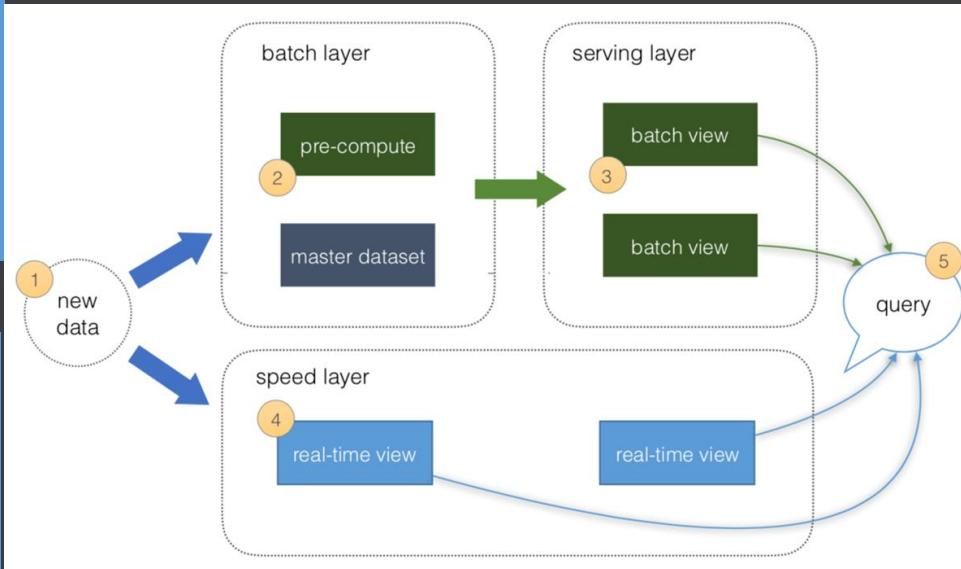
Capa de velocidad

La capa de velocidad procesa los flujos de datos en tiempo real o casi real. Esto funciona bien cuando el objetivo es minimizar la latencia de la ingestión de datos para el análisis..

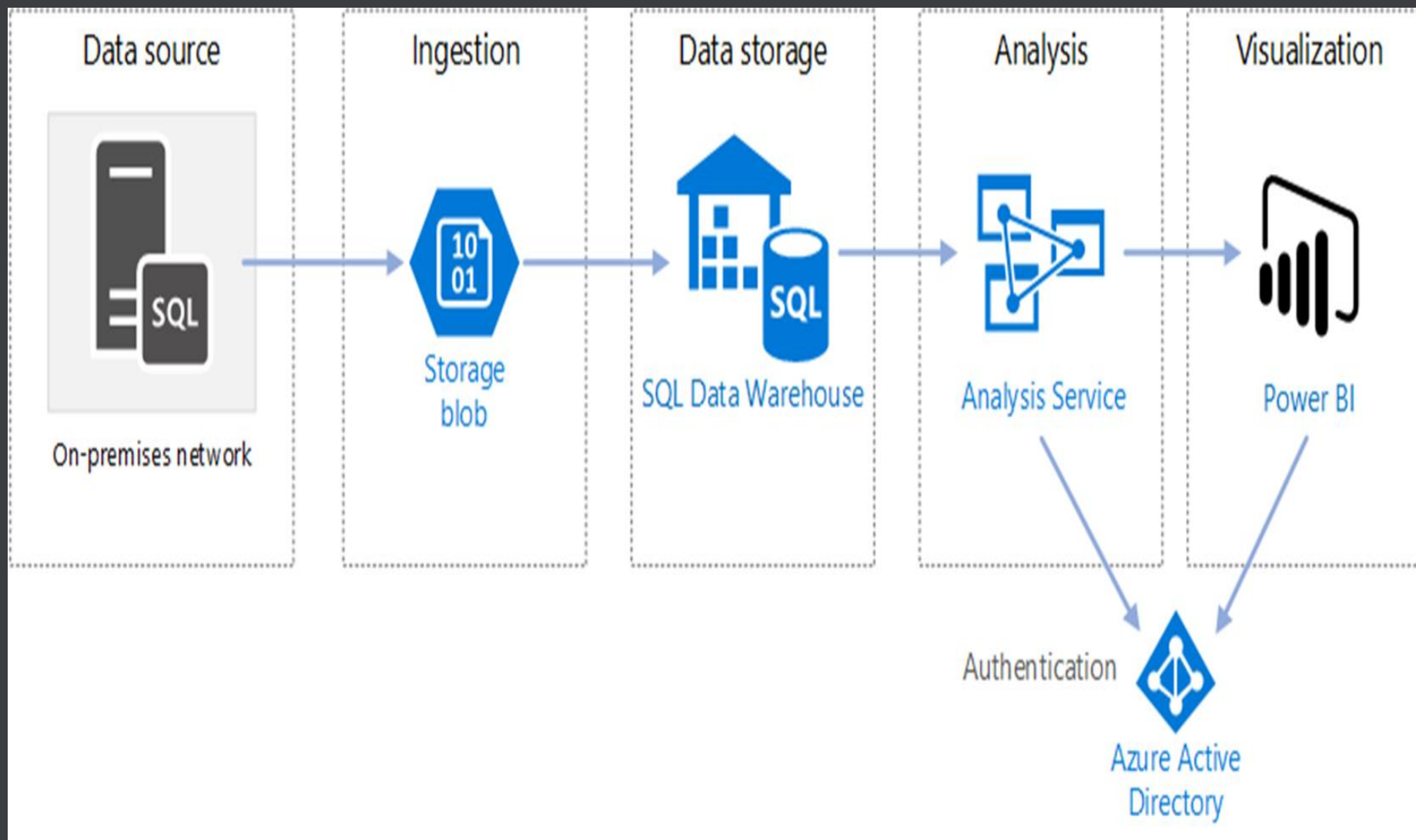
1. Nuevos datos ingeridos de fuentes
4. Vistas en tiempo real de los datos creados

Capa de servicio

La capa de servicio es opcional en la arquitectura en tiempo real y actúa como salida de almacenamiento de la capa Batch o Speed que utilizan las aplicaciones cliente para acceder a los resultados de los conjuntos de datos.



Diseñar una arquitectura de inteligencia empresarial



Diseñar soluciones de almacenamiento de datos Azure



- Recomendar una solución de datos de Azure basada en requerimientos
 - Elegir la solución de almacenamiento de datos correcta para cumplir con los requisitos técnicos y comerciales.
 - Elegir el tipo de distribución de particiones



Diseñar soluciones de almacenamiento de datos Azure



- Diseñar almacenes de datos de nube no relacionales
 - Diseñar distribución de datos y particiones
 - Diseñar para escalar, incluyéndose múltiples regiones, latencia y rendimiento
 - Diseñar una solución que usa:
 - Cosmos DB
 - Data Lake Storage Gen2 o
 - almacenamiento blob
 - Seleccionar la API de Cosmos DB adecuada
 - Diseño de una estrategia de recuperación ante desastres
 - Diseño para alta disponibilidad



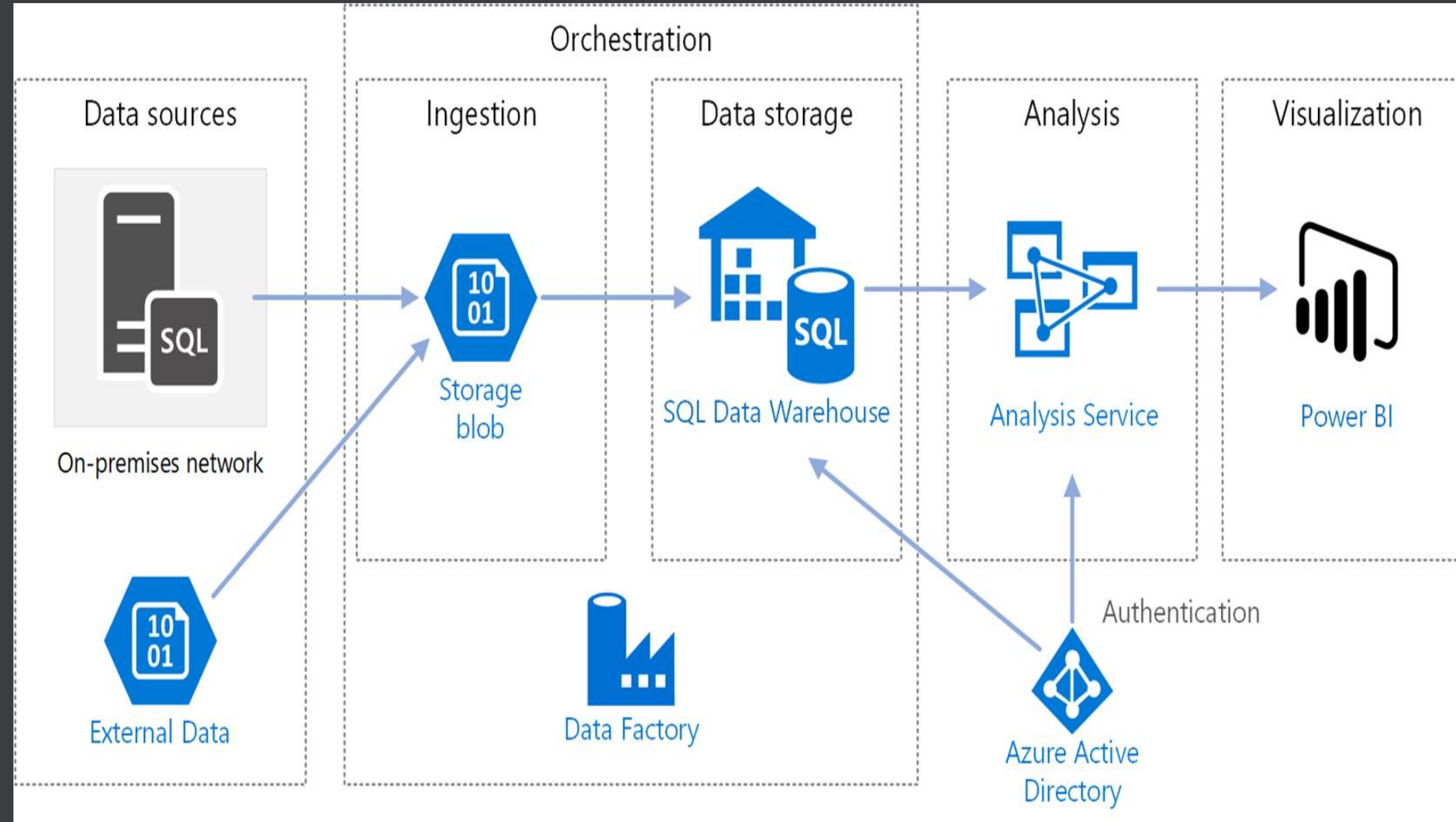
Diseñar soluciones de almacenamiento de datos Azure



- Diseñar almacenes de datos de nube relacionales
 - Diseñar distribución de datos y particiones
 - Diseñar para escalar, incluyéndose múltiples regiones, latencia y rendimiento
 - Diseñar una solución que usa
 - Bases de datos de SQL y
 - Almacén de datos de SQL
 - Diseño de una estrategia de recuperación ante desastres
 - Diseño para alta disponibilidad



Automatica una arquitectura de inteligencia empresarial



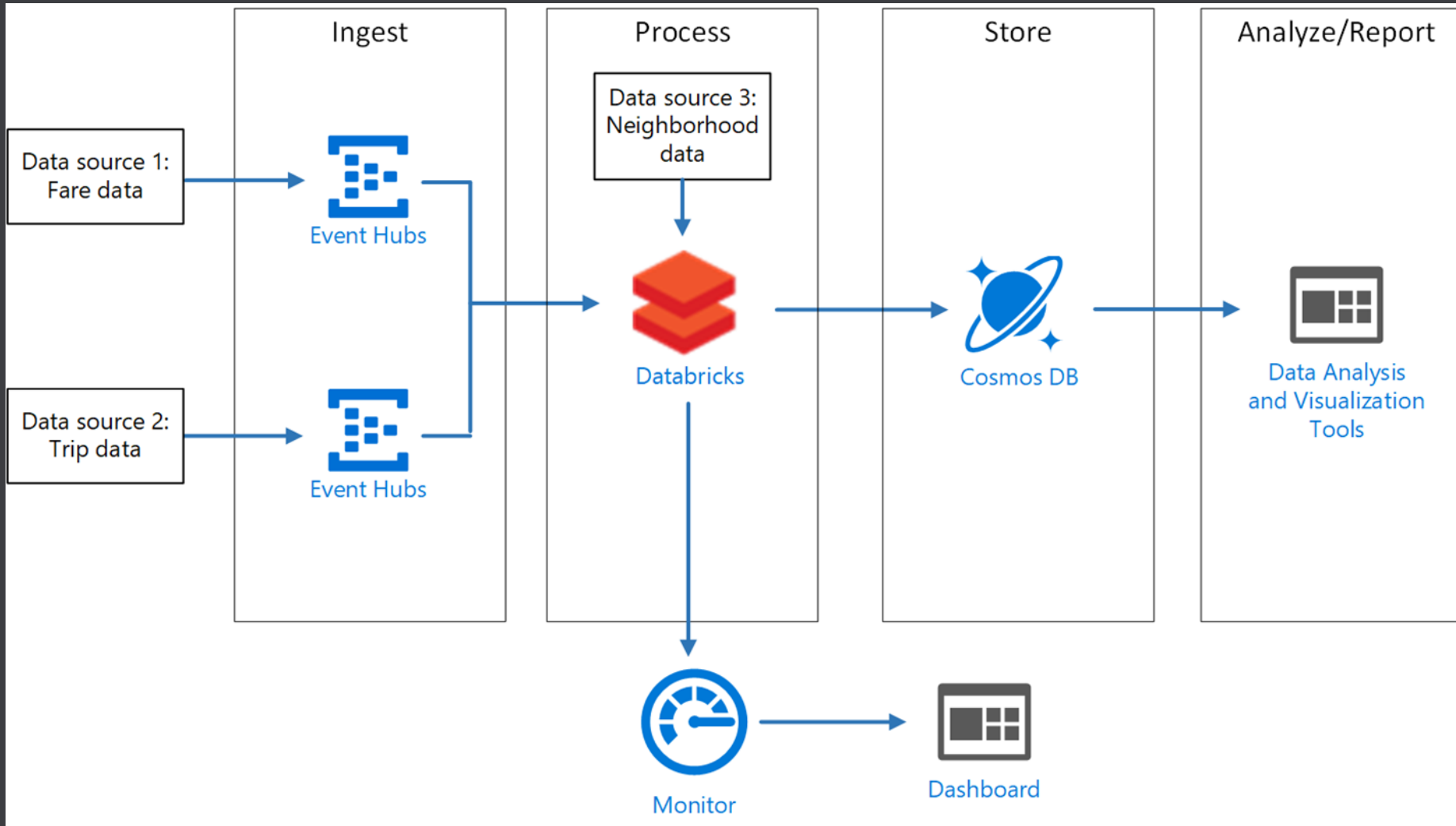
Diseñar soluciones de procesamiento de datos



- Diseñar soluciones de procesamiento por lotes.
 - Diseñar soluciones de procesamiento por lotes con
 - Data Factory y
 - Azure Databricks
 - Identificar el método óptimo de entrada de datos para una solución de procesamiento por lotes (batch layer)
 - Identificar el lugar donde se debe realizar el procesamiento como en la fuente, en el destino o en tránsito



Diseñe un stream processing pipeline con Azure Databricks



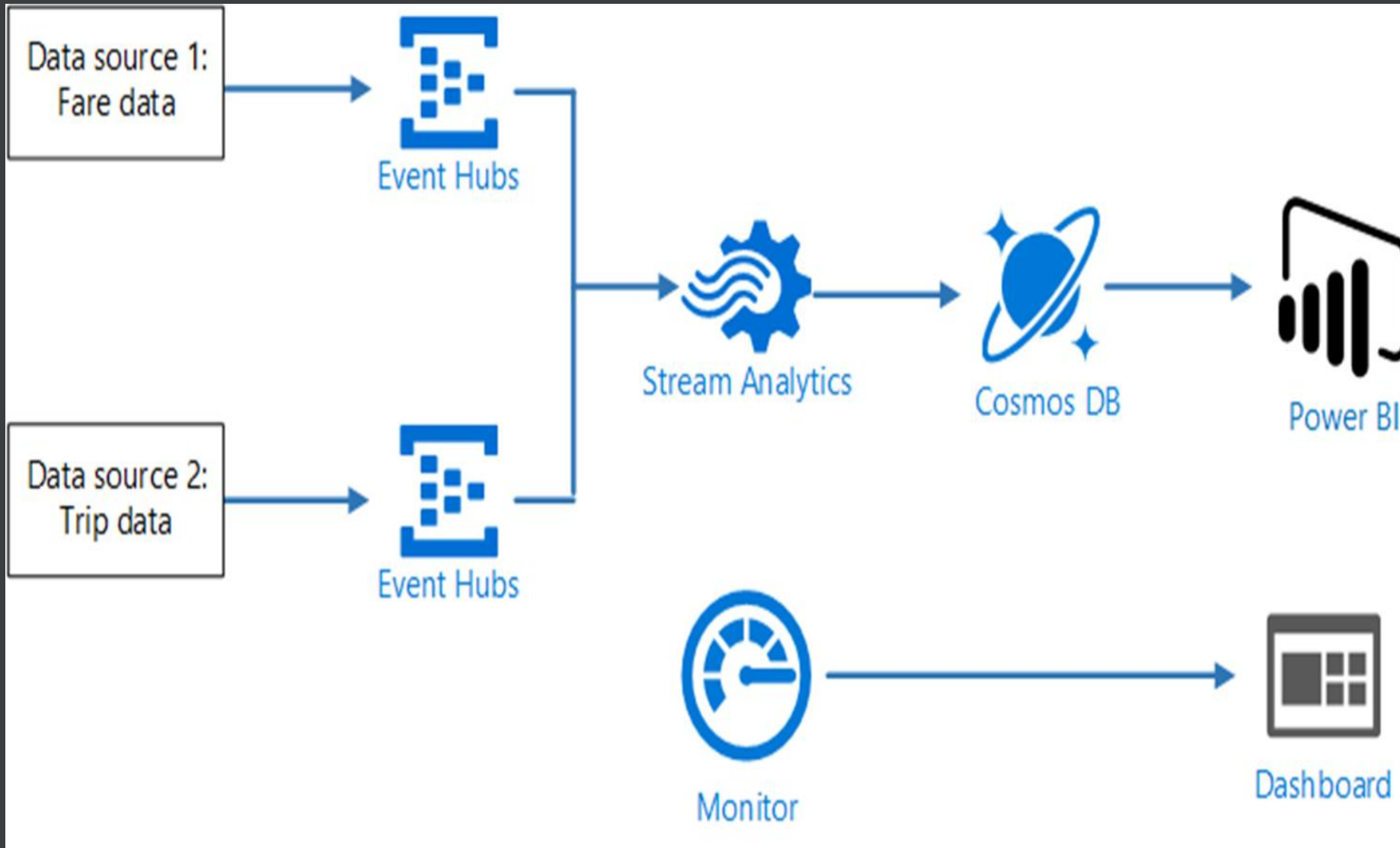
Diseñar soluciones de procesamiento de datos



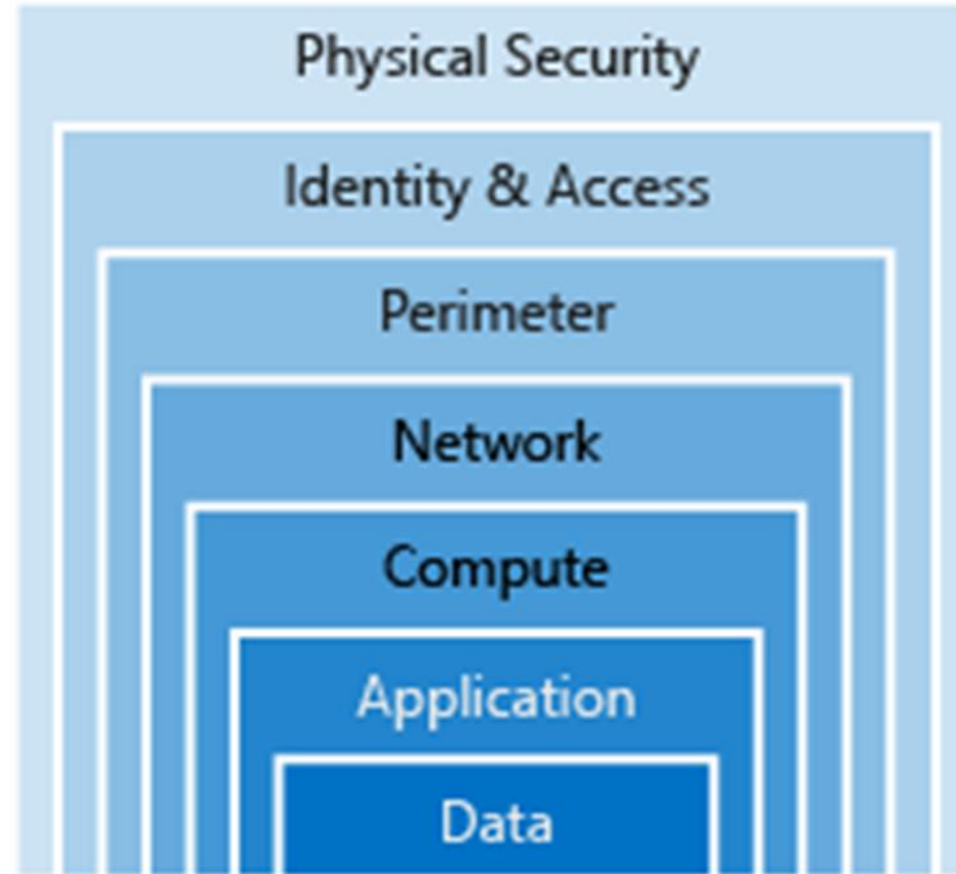
- Diseñar soluciones de procesamiento en tiempo real
 - Diseñar para procesamiento en tiempo real mediante
 - El uso de Stream Analytics y
 - Azure Databricks
 - Diseñar y proveer recursos de cómputo.



Diseñe un stream processing pipeline con Azure Stream Analytics



Enfoque de
defensa en
profundidad.



Diseño para seguridad de datos y cumplimiento



- Diseñar seguridad para acceso de datos de origen
 - Planificar para puntos finales seguros
 - Elegir el mecanismo adecuado de autenticación
 - Claves de acceso
 - Firmas de acceso compartido (SAS) y
 - Azure Active Directory (Azure AD)



Diseño para seguridad de datos y cumplimiento



- Diseñar la seguridad para directivas y estándares de datos
 - Diseñar para cifrado de datos en reposo y en tránsito
 - Diseñar para auditoría y enmascaramiento de datos
 - Diseñar para privacidad y clasificación de datos
 - Diseñar una directiva de retención de datos
 - Planear una estrategia de archivado
 - Planear para purgar datos basados en requerimientos comerciales

Referencias



- [Azure Tips and Tricks](#)
- [Examen DP-201: Diseñar una solución de datos de Azure](#)
- [Learning path for Azure Data Engineer](#)
- [Microsoft Azure certification exams](#)
- [Azure Cosmos DB: Implementación de una arquitectura lambda en la plataforma Azure](#)
-