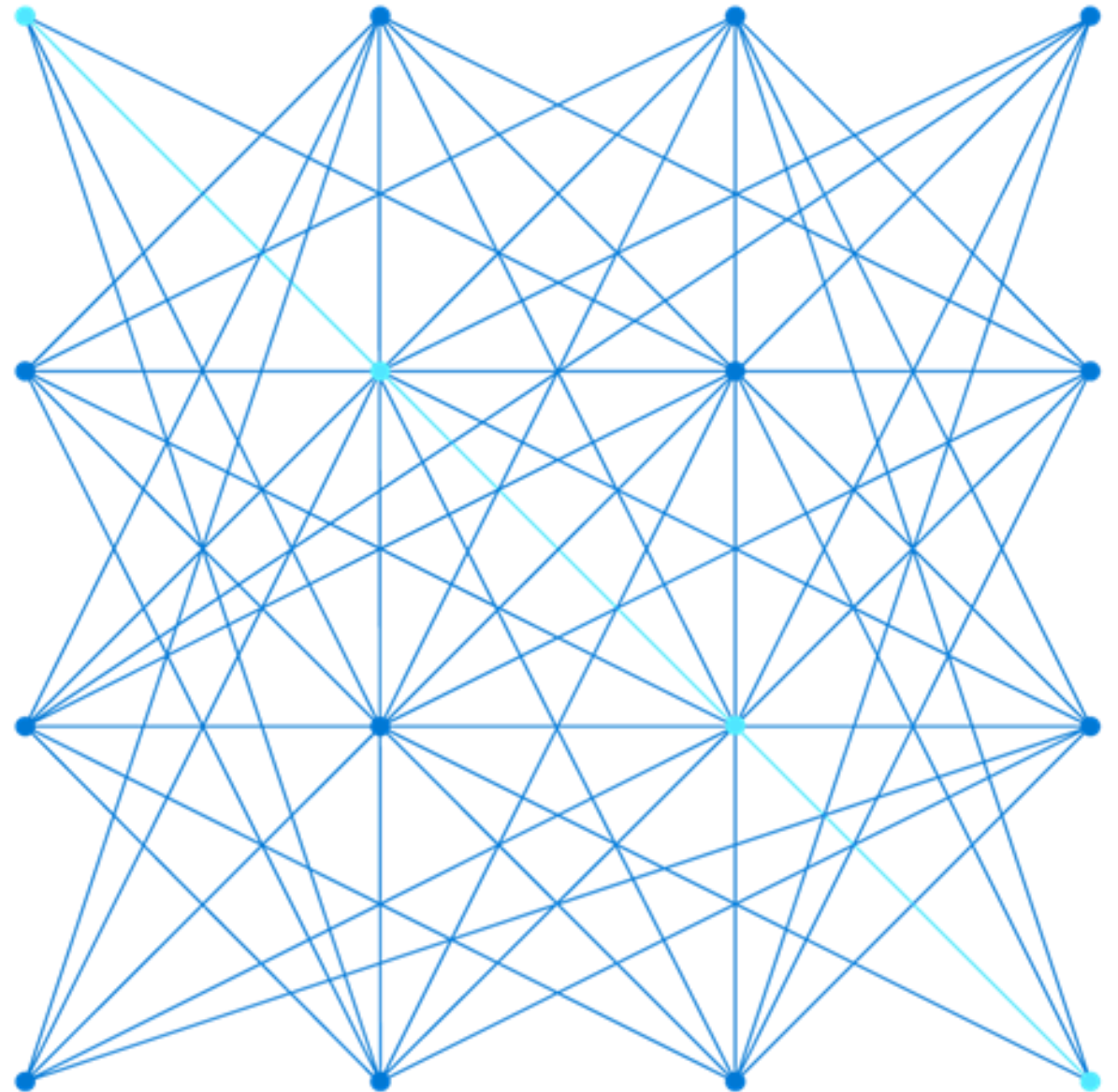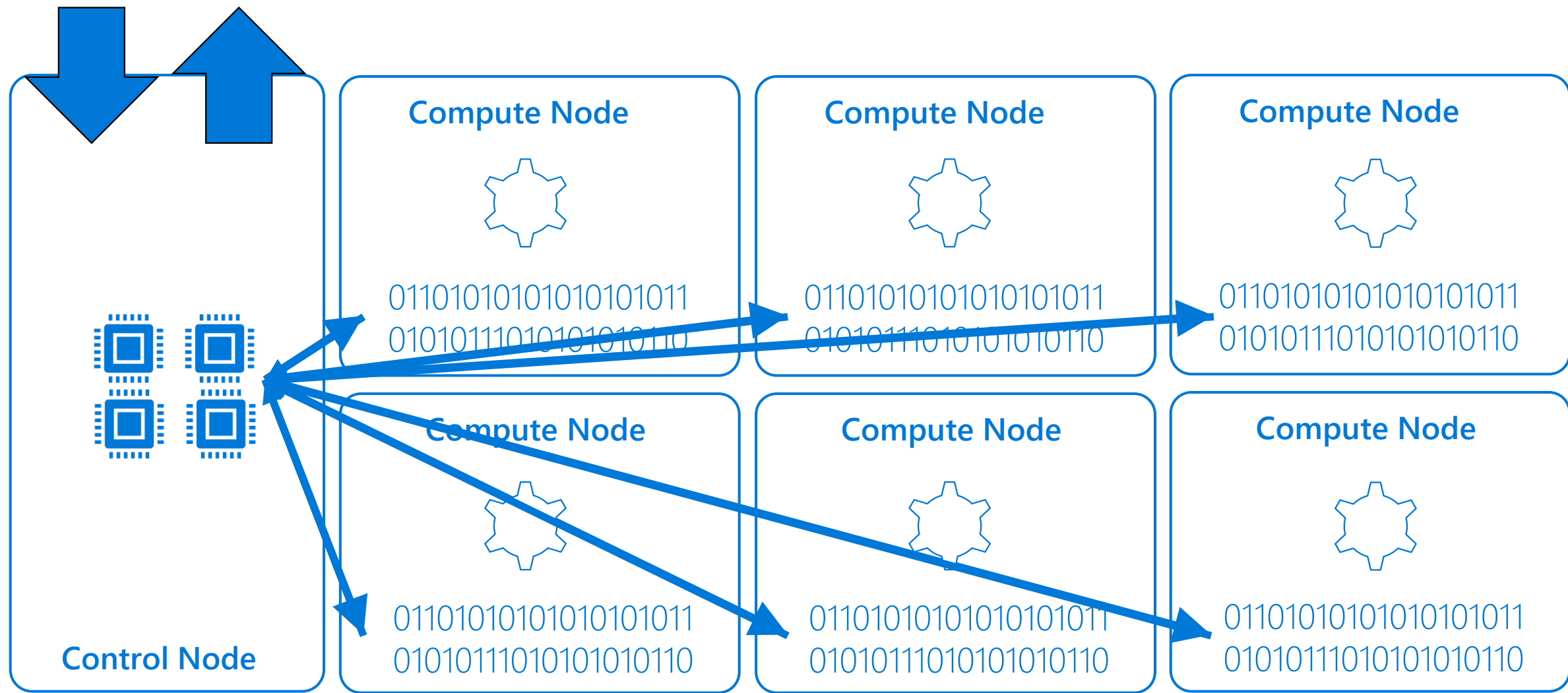Microsoft Azure

DP-203T00: Ingest and load data into the data warehouse

Lesson 01: Use data loading best practices in Azure Synapse Analytics

Dedicated SQL Pool architecture revision

Understand data load design goals

> Where is my data coming from?

> Is the data nett new? or do you receive changes from existing datasets?

> How often is the data being refreshed, added to or replaced?

> What formats are the data coming in?

> Is the data ingestible as-is? or are transformations and cleansing tasks required?

> Which takes priority, loading or querying/analysis?

Manage singleton updates

Set-up dedicated data loading accounts

Manage concurrent access to Azure Synapse Analytics

Control Node

Compute Node
0110101010101010101011
0101011101010101010110

Compute Node
0110101010101010101011
0101011101010101010110

Compute Node
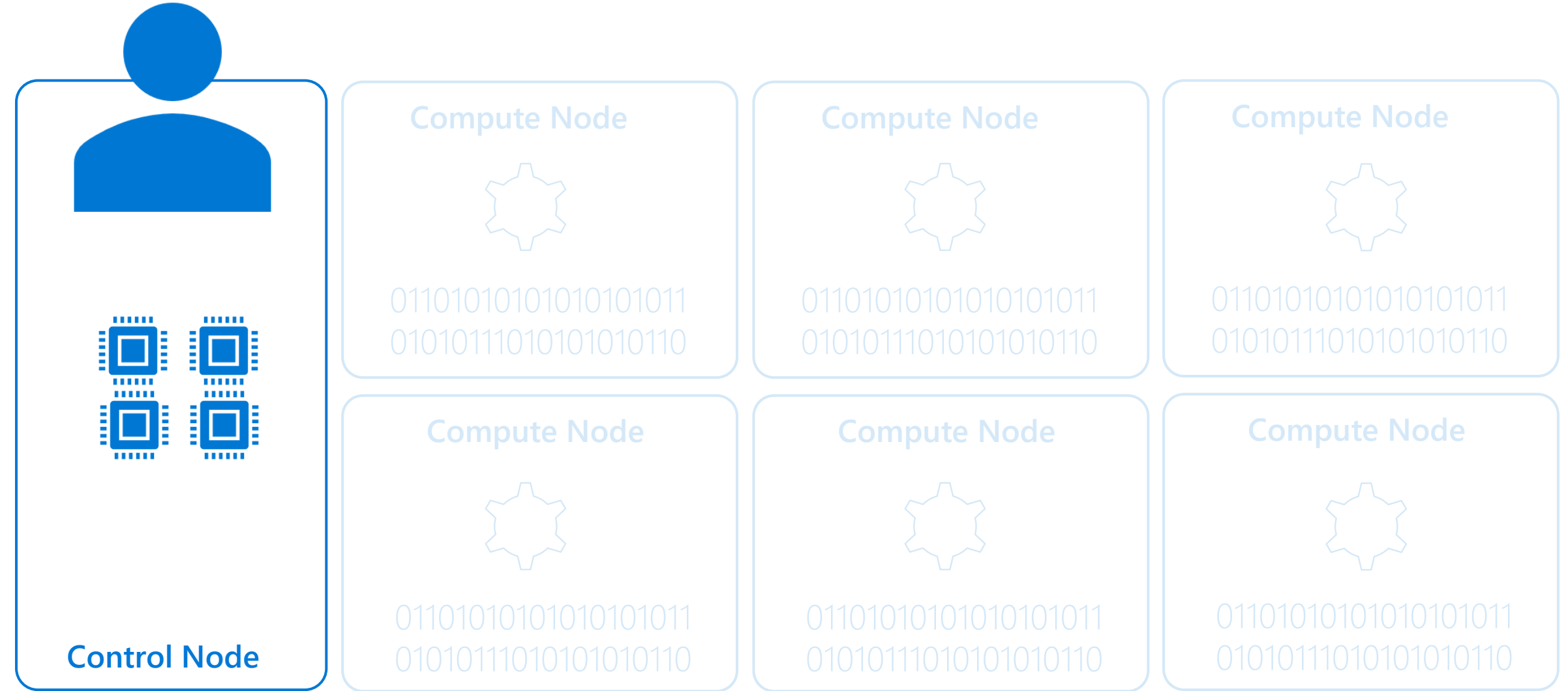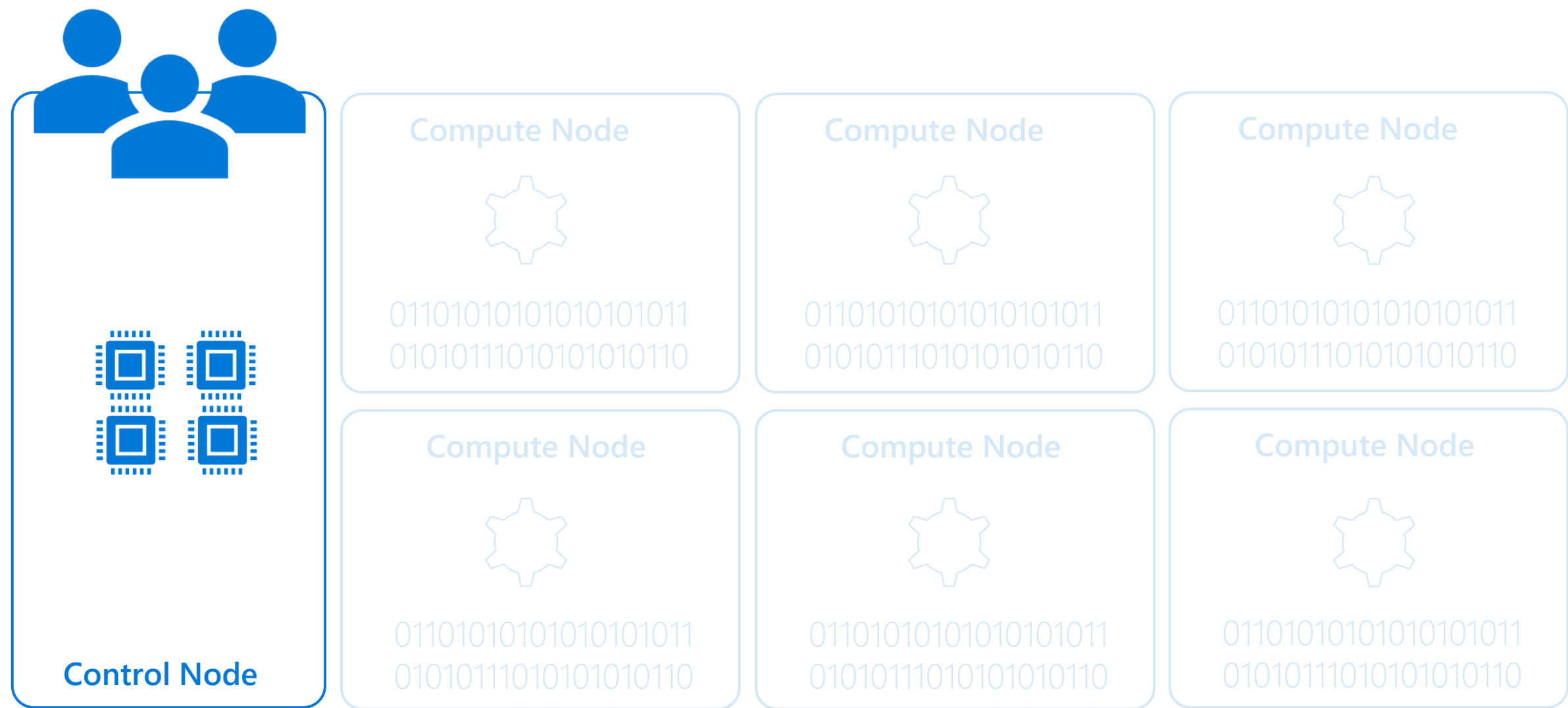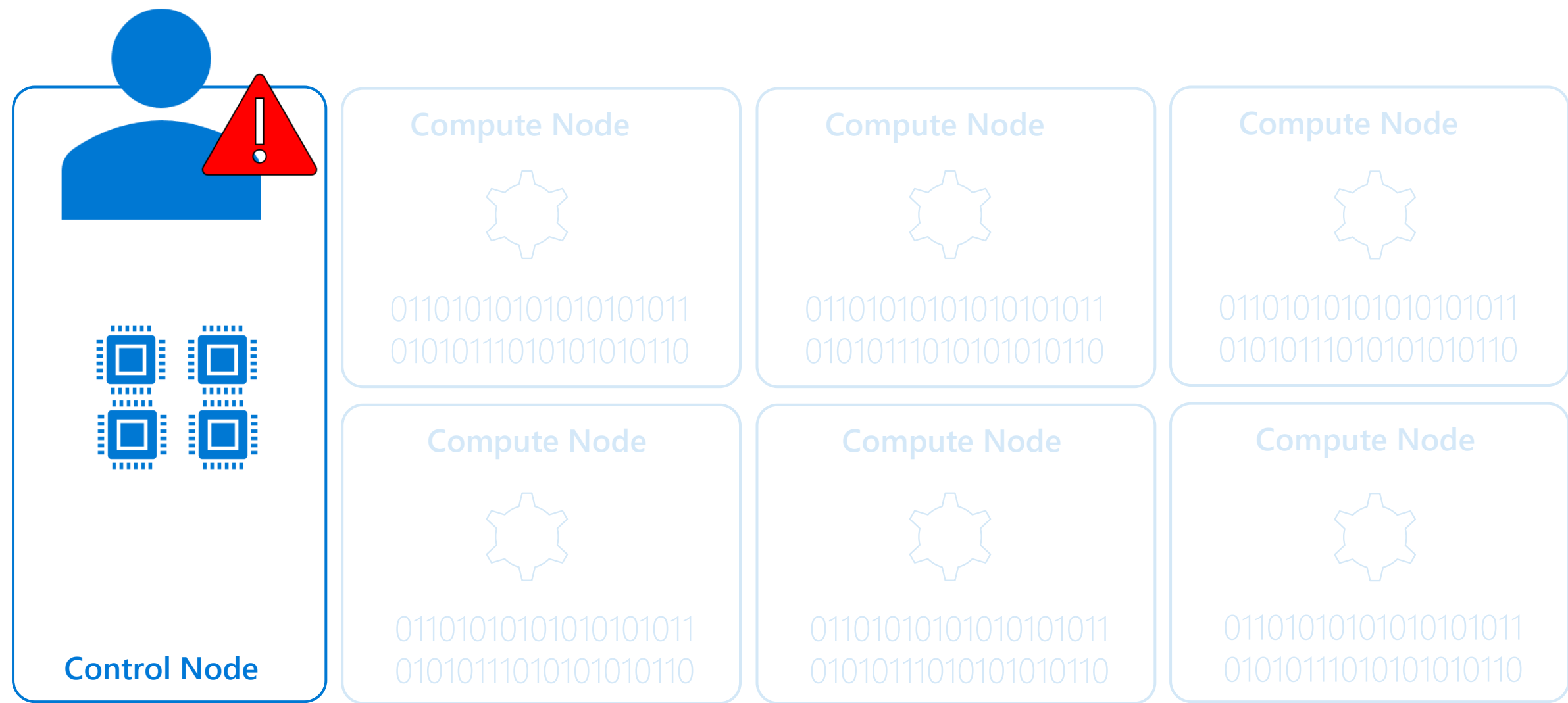0110101010101010101011
0101011101010101010110

Compute Node
0110101010101010101011
0101011101010101010110

Compute Node
0110101010101010101011
0101011101010101010110

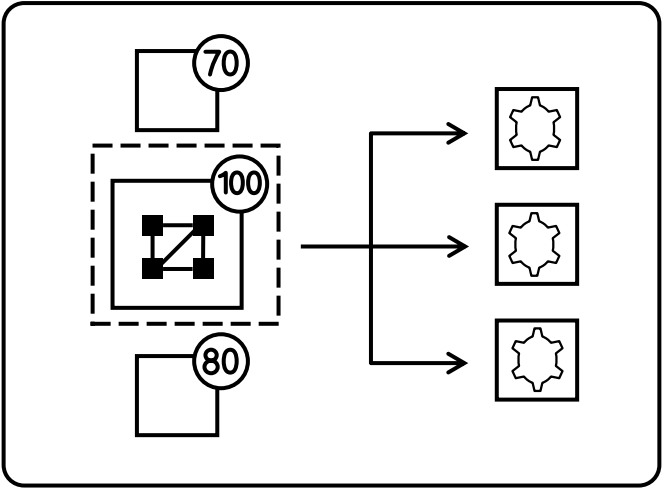Compute Node
0110101010101010101011
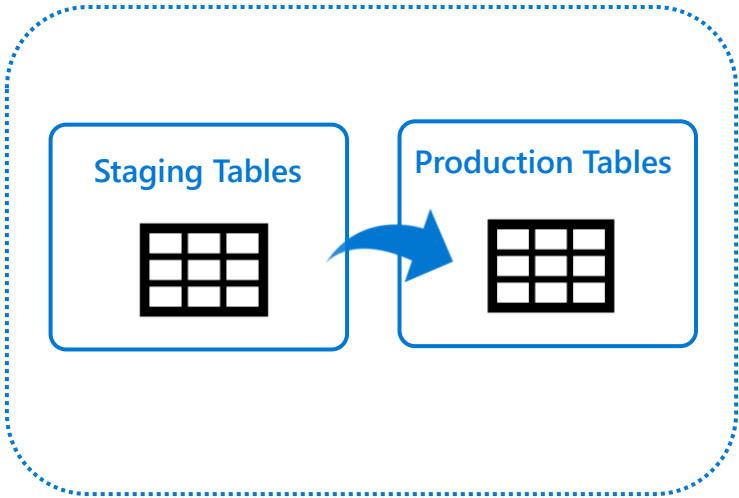0101011101010101010110

Implement Workload Management

Use PolyBase, the Copy command or the Copy Activity

PolyBase

Copy command

Staging Tables

Production Tables

Copy data activity
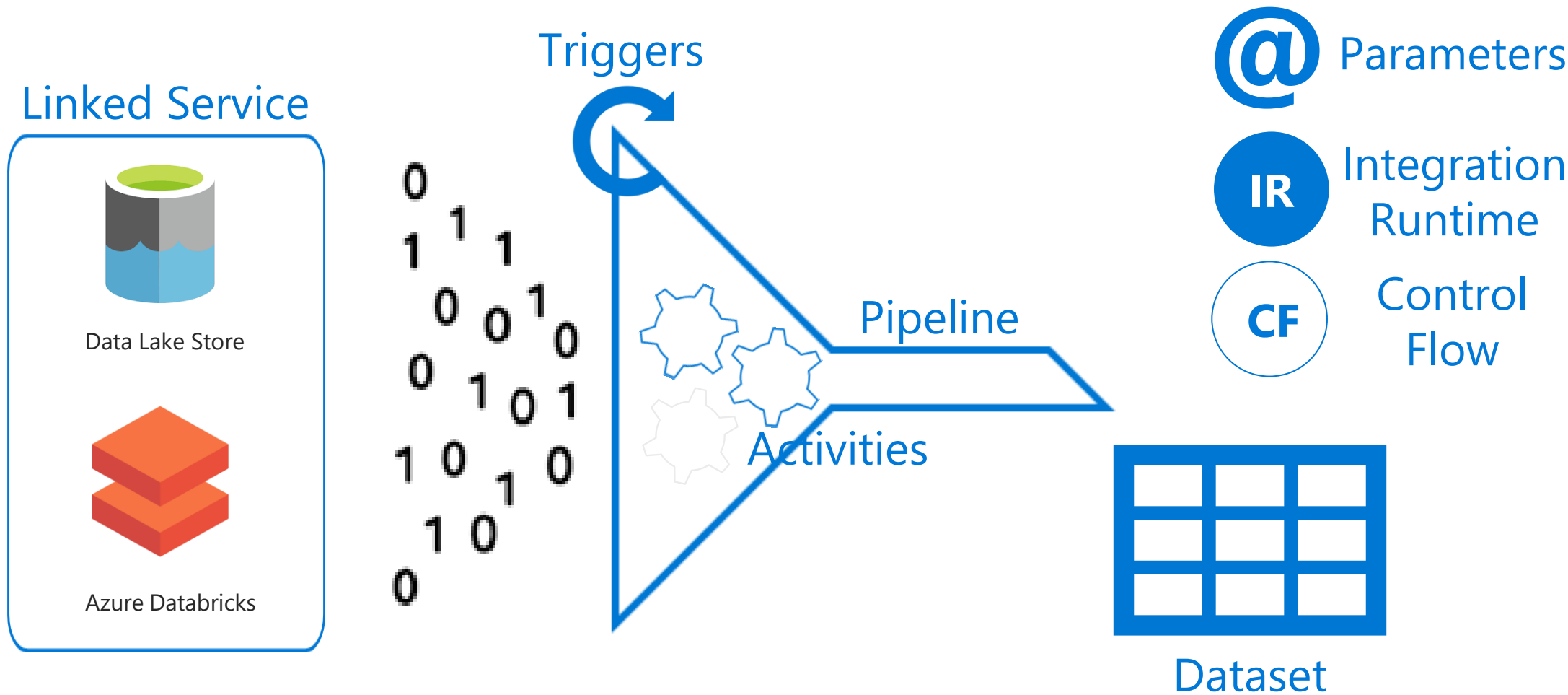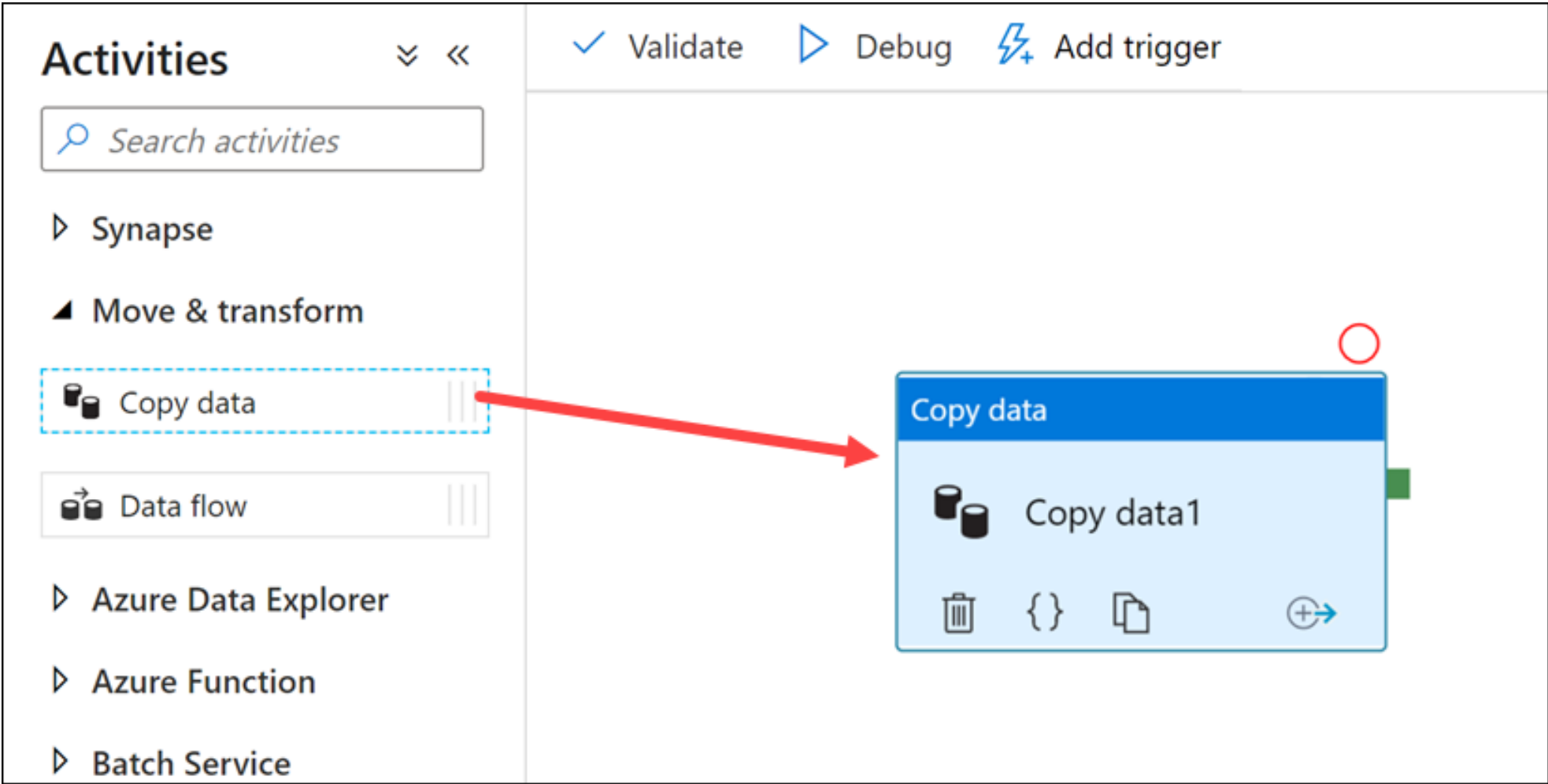
Copy data

Copy Sales
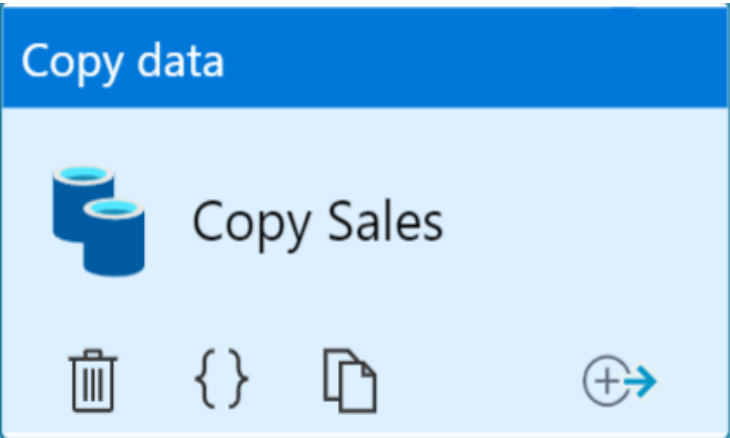
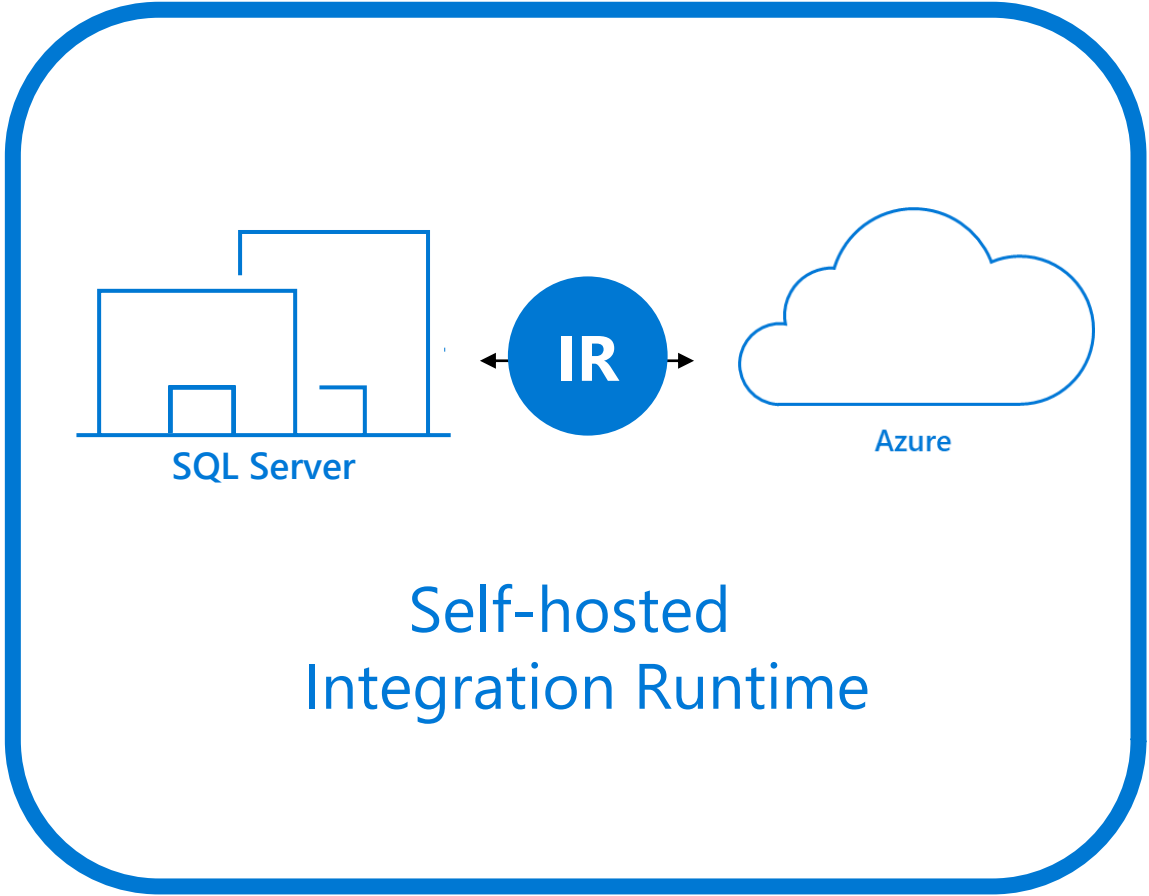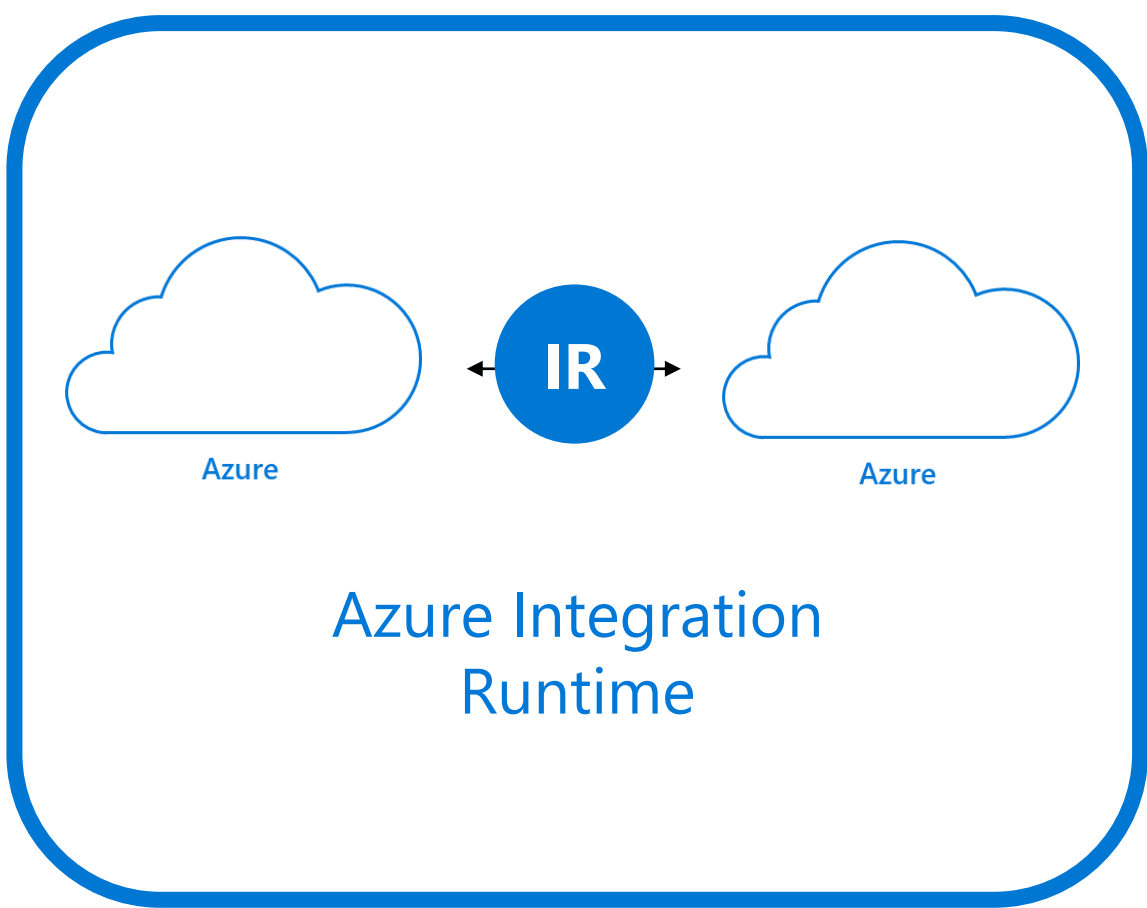# Lesson 02: Petabyte-scale ingestion with Azure Data Factory

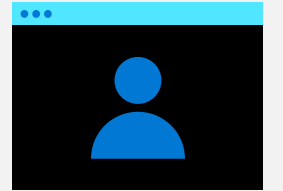# Azure Data Factory/Synapse pipeline revision

Petabyte-scale ingestion with Azure Data Factory

Understanding integration

Lab: Ingest and load data into the data warehouse

Lab overview

This lab teaches students how to ingest data into the data warehouse through T-SQL scripts and Synapse Analytics integration pipelines. The student will learn how to load data into Synapse dedicated SQL pools with PolyBase and COPY using T-SQL. The student will also learn how to use workload management along with a Copy activity in a Azure Synapse pipeline for petabyte-scale data ingestion.

Lab objectives

After completing this lab, you will be able to:

Use data loading best practices in Azure Synapse Analytics

Petabyte-scale ingestion with Azure Data Factory