

CLOUDBERA

Educational Services

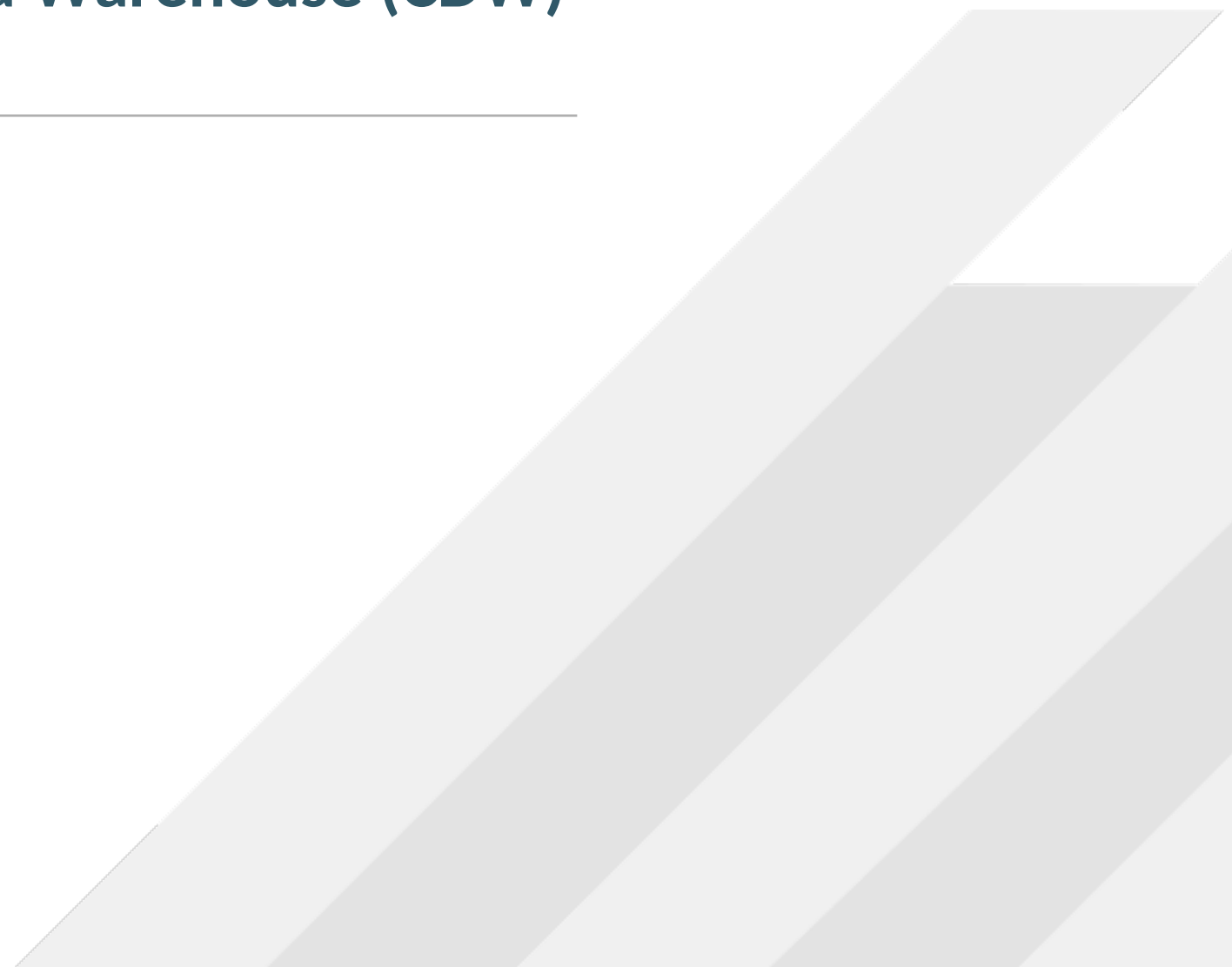
Introduction to Cloudera Data Warehouse

Self-Service Analytics in the Cloud
with CDP



Cloudera Data Warehouse (CDW) Overview

Chapter 1G



Course Chapters

- **Cloudera Data Warehouse (CDW) Overview**
- The CDW Web Interface
- Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)
- Querying Data from CDW Web Interface (Data Analyst Track)
- Managing Virtual Warehouses (Data Engineering Track)
- Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

Chapter Topics

Cloudera Data Warehouse (CDW) Overview

- **Introduction to This Course**
- What Is CDW?
- Benefits of CDW

Course Learning Objectives

- Identify the primary purpose and benefits of CDW
- Access CDW and navigate different pages within it

Data Engineer

- Create a Database Catalog and a Virtual Warehouse in CDW
- Describe how CDW scales automatically
- Change settings for Hive and Impala Virtual Warehouses

Data Analyst

- Run a query using Hue or DAS with a CDW Virtual Warehouse
- Connect to a Virtual Warehouse from the command line
- Connect a third-party tool to a CDW Virtual Warehouse

Chapter Topics

Cloudera Data Warehouse (CDW) Overview

- Introduction to This Course
- **What Is CDW?**
- Benefits of CDW

Learning Objective

- Identify the primary purpose and components of CDW

What Is CDW



- **CDW enables creation of data warehouses and data marts for analysts**
- **CDW has two components**
 - Database Catalogs
 - Virtual Warehouses

Chapter Topics

Cloudera Data Warehouse (CDW) Overview

- Introduction to This Course
- What Is CDW?
- **Benefits of CDW**

Learning Objective

- Identify five benefits that CDW provides

Benefits

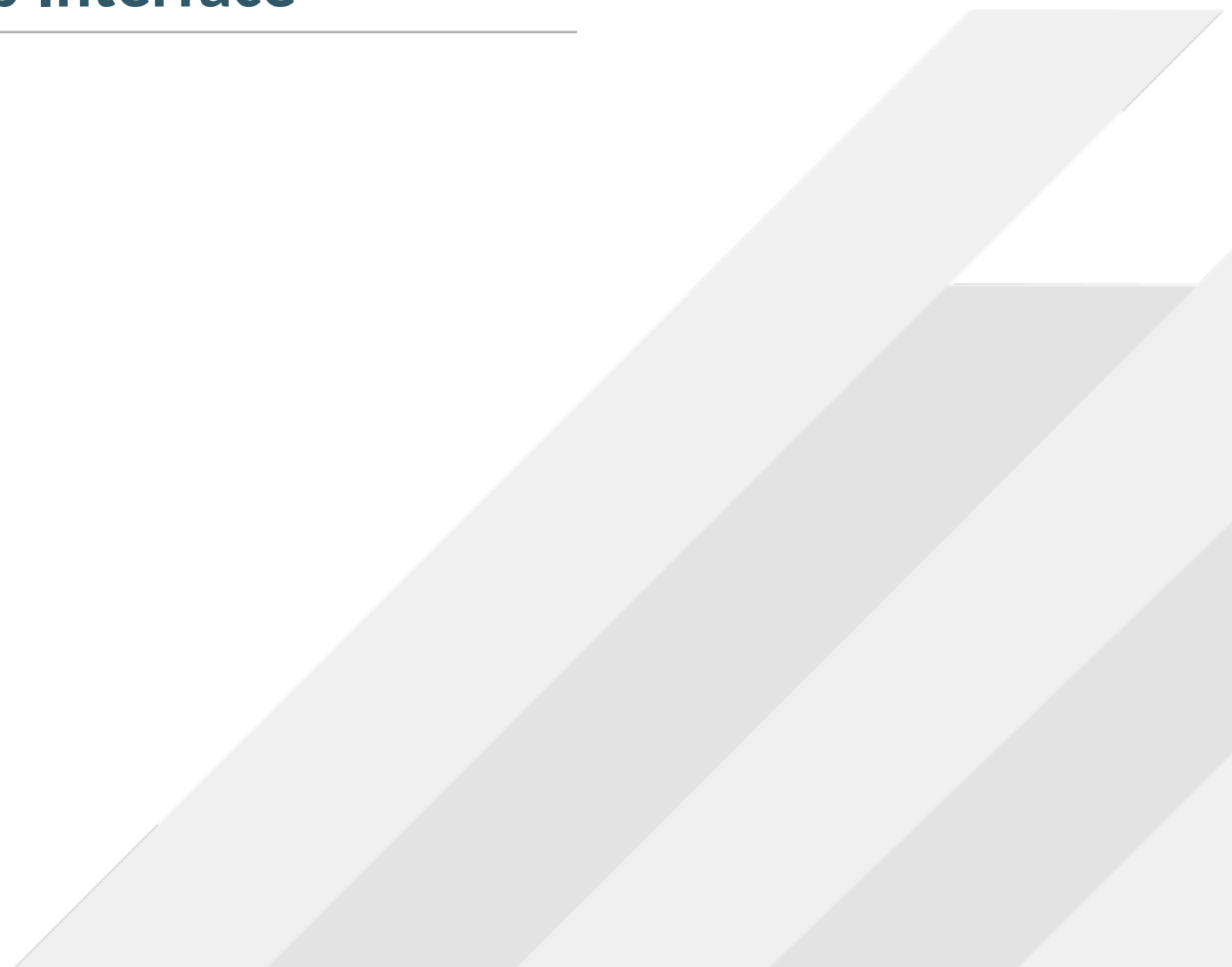
Data warehouses and data marts are

- **Automatically configured and isolated**
- **Optimized for existing workloads when moved to the cloud**
- **Auto-scaled to meet varying demands**
- **Auto-suspended to save costs**
- **Compliant with security controls**



The CDW Web Interface

Chapter 2G



Course Chapters

- Cloudera Data Warehouse (CDW) Overview
- **The CDW Web Interface**
- Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)
- Querying Data from CDW Web Interface (Data Analyst Track)
- Managing Virtual Warehouses (Data Engineering Track)
- Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

Chapter Topics

The CDW Web Interface

- **How to Access CDW**
- CDW Orientation

Learning Objective

- Access CDW using Single Sign-On

Exercise

- Access your CDP home page
- Click on the Data Warehouse icon to access CDW

Chapter Topics

The CDW Web Interface

- How to Access CDW
- **CDW Orientation**

Learning Objective

- **Navigate the different pages within CDW**

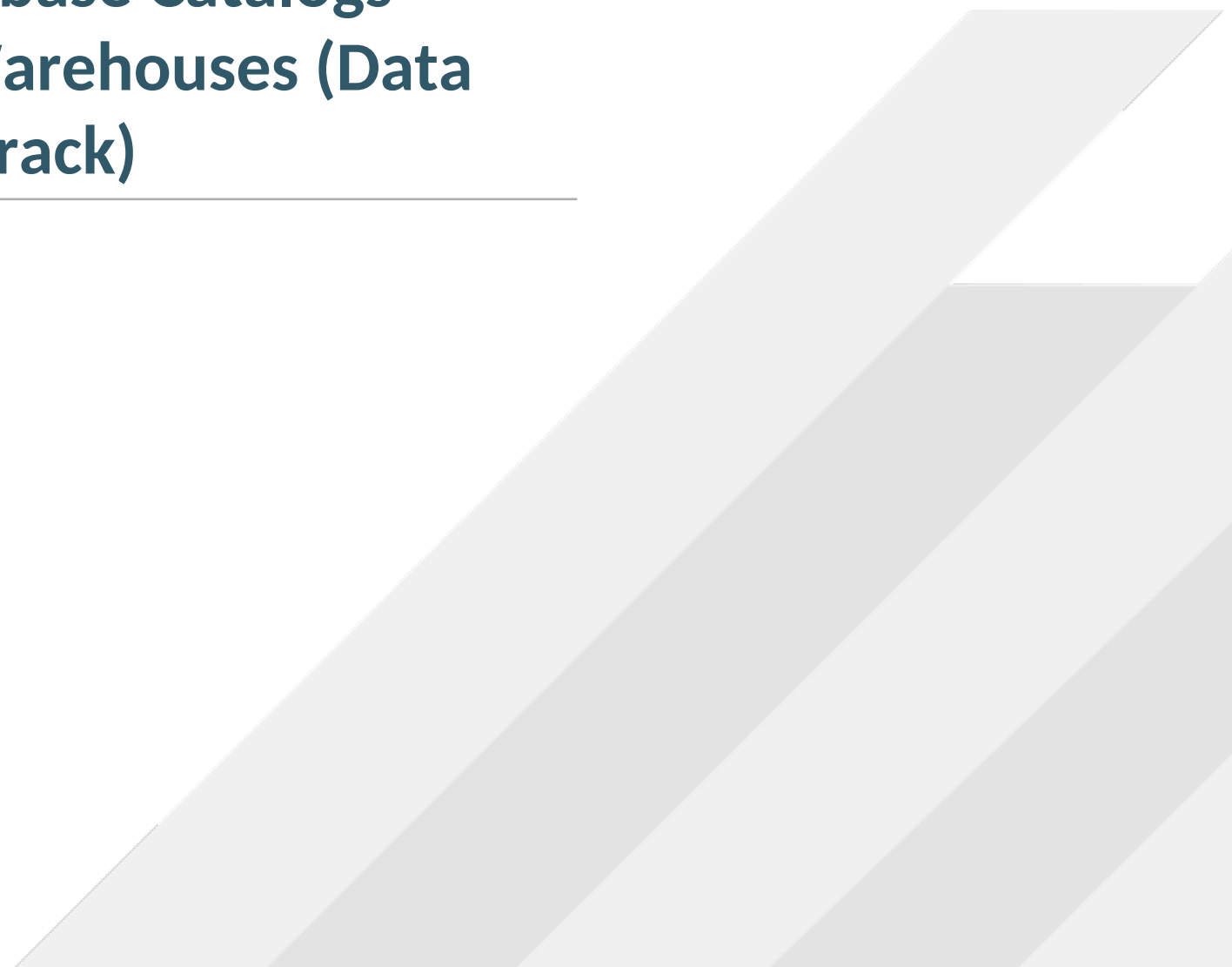
Exercise

- 1. Explore the sidebar: Click the grid to show the list of CDP applications, collapse and expand the sidebar, and view the Help**
- 2. Find the Overview page and expand the Environments section**
- 3. Click on an entity and look for the entities that are associated with it (note the Database Catalog and Virtual Warehouses)**
- 4. Filter the Database Catalog or Virtual Warehouses section to show only a few, including the Database Catalog or Virtual Warehouse you just noted**
- 5. Go to the Database Catalogs page and find the same Database Catalog that you just noted, using the filter or page navigation if necessary**
- 6. Go to the Virtual Warehouses page and find the same Virtual Warehouses you just noted, using the filter or page navigation if necessary**



Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)

Chapter 3E



Course Chapters

- Cloudera Data Warehouse (CDW) Overview
- The CDW Web Interface
- **Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)**
- Querying Data from CDW Web Interface (Data Analyst Track)
- Managing Virtual Warehouses (Data Engineering Track)
- Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

Chapter Topics

Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)

- **Creating a Database Catalog**
- Creating a Virtual Warehouse

Learning Objective

- Create a Database Catalog in CDW

Exercise

- **Create a Database Catalog for testing**

Chapter Topics

Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)

- Creating a Database Catalog
- **Creating a Virtual Warehouse**

Learning Objective

- Create a Virtual Warehouse in CDW

Decision Factors (Type)

- **Use case**
 - What features are needed
- **File format**
- **Personal preferences**

Decision Factors (Size)

- Concurrent queries or users
- Query complexity
- Data set size

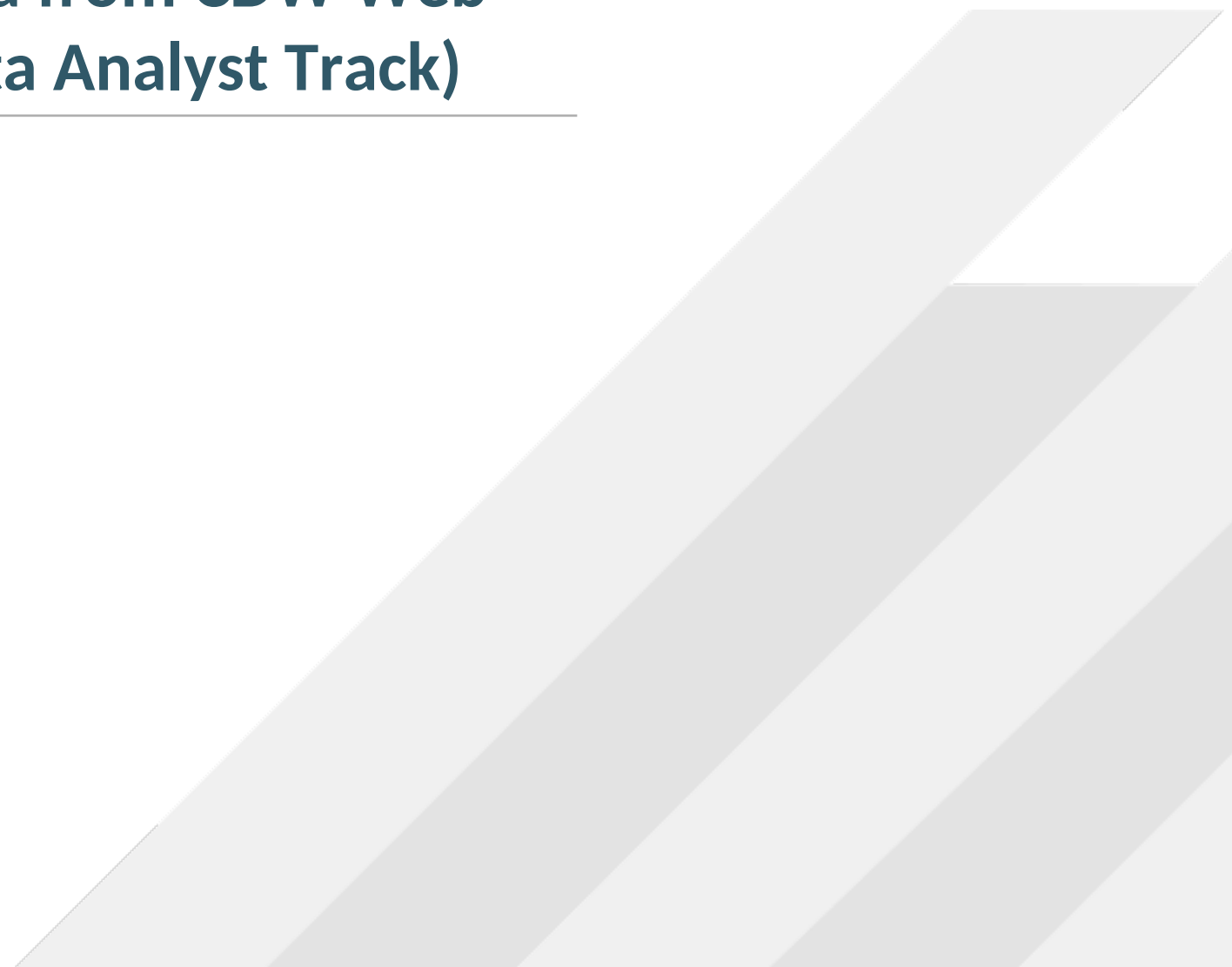
Exercise

- **Create a Virtual Warehouse associated with your test Database Catalog**
 - Use Hive
 - Use the smallest size
- **Suspend it and the test Database Catalog after it's been created**



Querying Data from CDW Web Interface (Data Analyst Track)

Chapter 3A



Course Chapters

- Cloudera Data Warehouse (CDW) Overview
- The CDW Web Interface
- Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)
- **Querying Data from CDW Web Interface (Data Analyst Track)**
- Managing Virtual Warehouses (Data Engineering Track)
- Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

Chapter Topics

Querying Data from CDW Web Interface (Data Analyst Track)

- **Accessing Query Editors**
- Setting Workload Password
- Querying with Hue
- Querying with DAS

Learning Objective

- Access the query editors for Hue and DAS from a CDW Virtual Warehouse

Exercise

- **Open a query editor**
 - If necessary, continue to next video to set up your workload password

Chapter Topics

Querying Data from CDW Web Interface (Data Analyst Track)

- Accessing Query Editors
- **Setting Workload Password**
- Querying with Hue
- Querying with DAS

Learning Objective

- **Set a workload password to allow access to Hue and DAS**

Exercise

- Create your workload password
- Test it by accessing Hue or DAS

Chapter Topics

Querying Data from CDW Web Interface (Data Analyst Track)

- Accessing Query Editors
- Setting Workload Password
- **Querying with Hue**
- Querying with DAS

Learning Objective

- Run a query on a table using Hue

Exercise

- In Hue, complete a simple `SELECT *` query on a table you can access

Chapter Topics

Querying Data from CDW Web Interface (Data Analyst Track)

- Accessing Query Editors
- Setting Workload Password
- Querying with Hue
- **Querying with DAS**

Learning Objective

- Run a query on a table using DAS

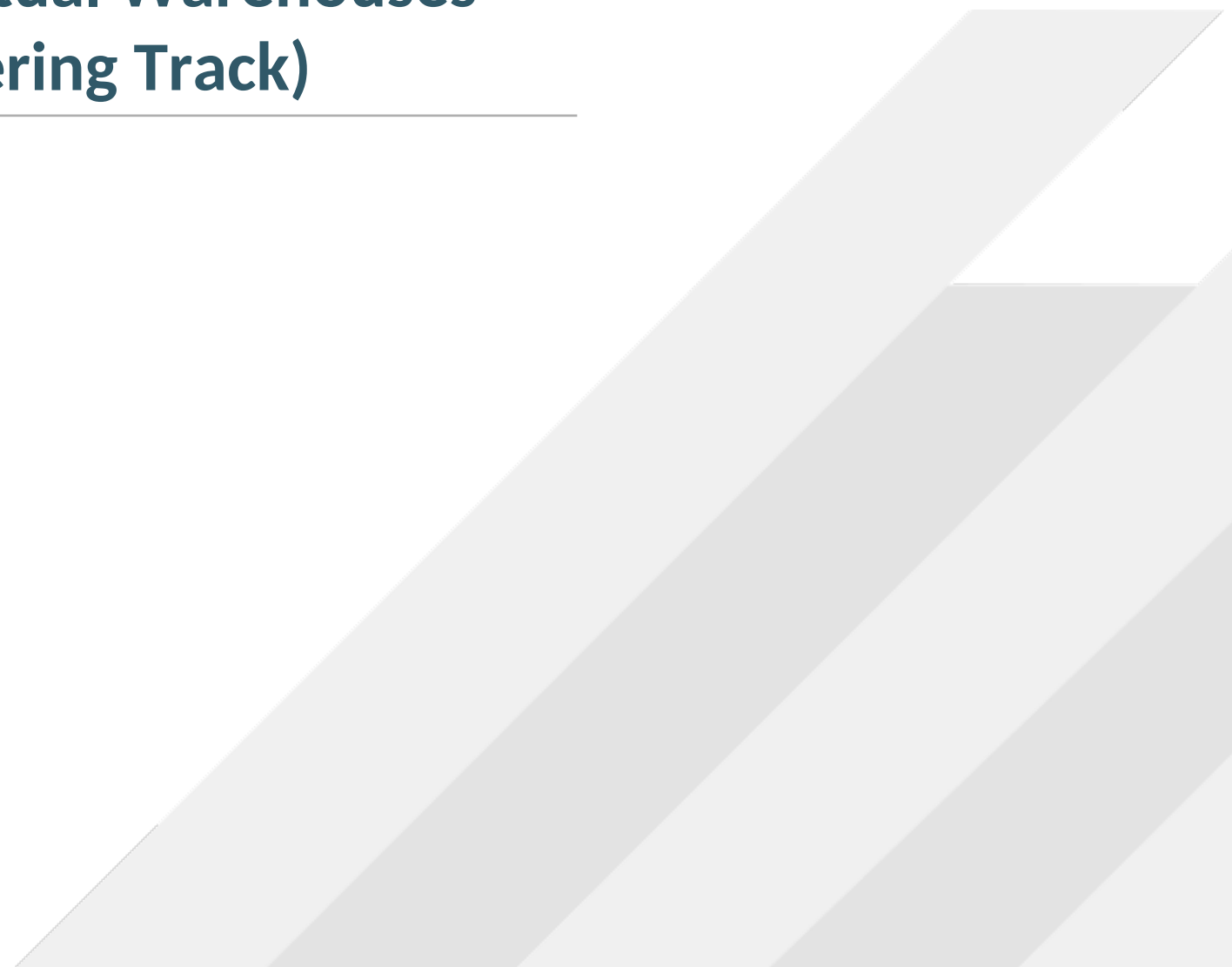
Exercise

- In DAS, complete a simple `SELECT *` query on a table you can access



Managing Virtual Warehouses (Data Engineering Track)

Chapter 4E



Course Chapters

- Cloudera Data Warehouse (CDW) Overview
- The CDW Web Interface
- Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)
- Querying Data from CDW Web Interface (Data Analyst Track)
- **Managing Virtual Warehouses (Data Engineering Track)**
- Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

Chapter Topics

Managing Virtual Warehouses (Data Engineering Track)

- **Auto-Scaling**
- Additional Tuning for Hive
- Additional Tuning for Impala

Learning Objective

- Describe how CDW auto-scales Virtual Warehouses
- Set the auto-scale range for a Virtual Warehouse

Exercise

- **Change the auto-scale settings**
 - For a new Virtual Warehouse, on creation
 - For an existing Virtual Warehouse

Pay attention to what values are available for different initial sizes

Chapter Topics

Managing Virtual Warehouses (Data Engineering Track)

- Auto-Scaling
- **Additional Tuning for Hive**
- Additional Tuning for Impala

Learning Objective

- **Set auto-scale settings unique to Hive Virtual Warehouses**

Hive Settings

Headroom = Number of nodes to keep free

Wait Time = Amount of time query is in queue

Exercise

- **Change the Headroom/Wait Time settings for a new or existing Hive Virtual Warehouse**

Chapter Topics

Managing Virtual Warehouses (Data Engineering Track)

- Auto-Scaling
- Additional Tuning for Hive
- **Additional Tuning for Impala**

Learning Objective

- **Set auto-scale settings unique to Impala Virtual Warehouses**

Autoscale Mode

- **Conservative**
 - Auto-scale up 60 seconds after maximum utilization
 - Immediately auto-scale down when possible
- **Aggressive**
 - Immediately auto-scale up at maximum utilization
 - Auto-scale down 60 seconds after demand drops
- **Balanced**
 - Auto-scale up 30 seconds after maximum utilization
 - Auto-scale down 30 seconds after demand drops

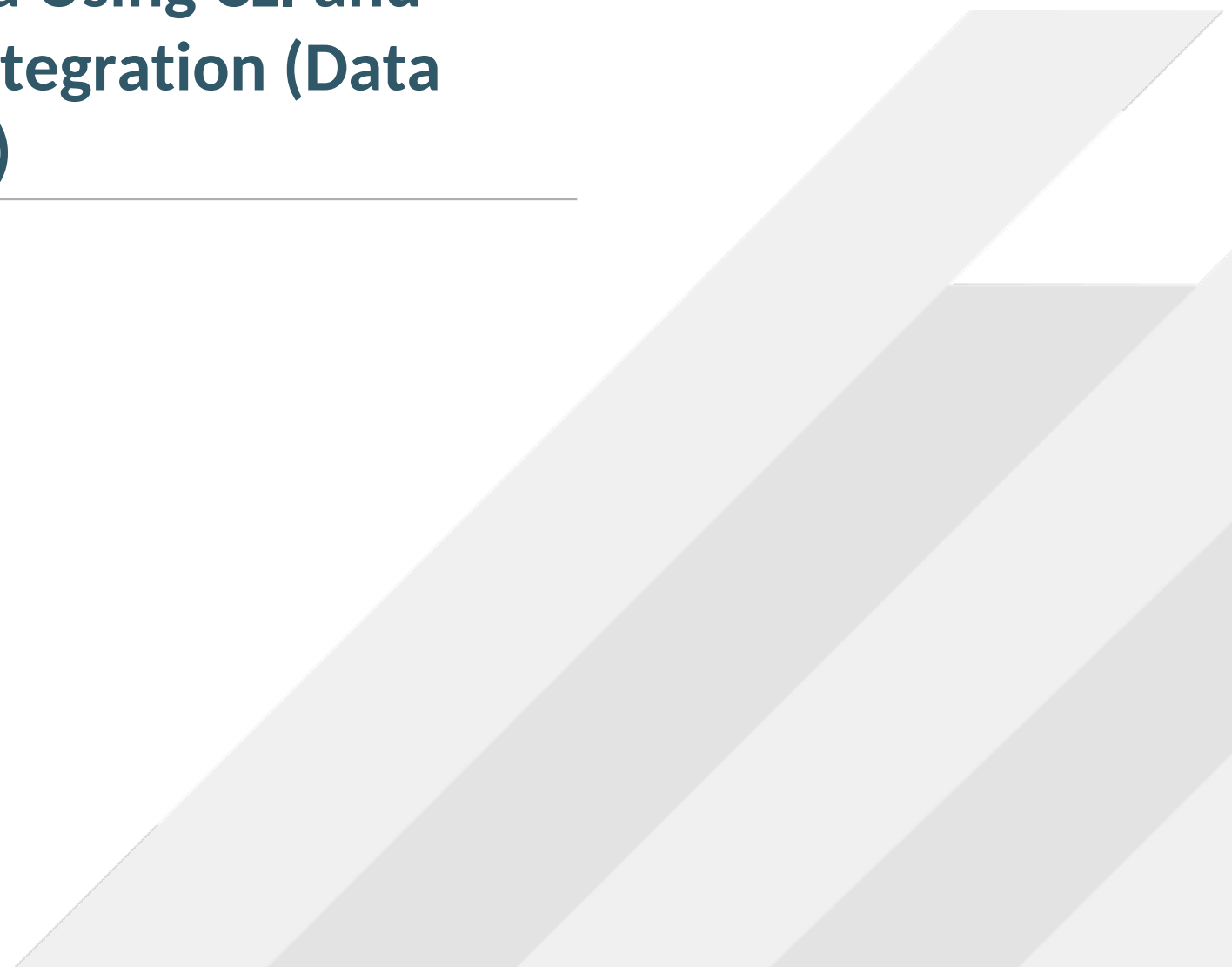
Exercise

- **Change the Autoscale Mode setting for a new or existing Impala Virtual Warehouse**



Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

Chapter 4A



Course Chapters

- Cloudera Data Warehouse (CDW) Overview
- The CDW Web Interface
- Creating Database Catalogs and Virtual Warehouses (Data Engineering Track)
- Querying Data from CDW Web Interface (Data Analyst Track)
- Managing Virtual Warehouses (Data Engineering Track)
- **Querying Data Using CLI and Third-Party Integration (Data Analyst Track)**

Chapter Topics

Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

- **Using Impala Shell**
- Using Third-Party Tools

Learning Objective

- **Connect to Impala Shell from the command line**

Installing Impala Shell

- **If connecting from a CDP node, skip installation**
- **Install requires:**
 - Non-Windows computer
 - Python 2.7
 - A pip installer

Exercise

- **Install Impala Shell**
 - Unless using a cluster node
- **Use Impala Shell to query a table**

Chapter Topics

Querying Data Using CLI and Third-Party Integration (Data Analyst Track)

- Using Impala Shell
- **Using Third-Party Tools**

Learning Objective

- **Connect a third-party tool to Hive**

CDW and Third-Party Tools

- **Integration with third-party tools using**
 - ODBC
 - JDBC
- **Example: Tableau with ODBC**

Exercise

- **Connect a business intelligence (BI) application to Hive**
 - Download a driver
 - Download a BI if necessary