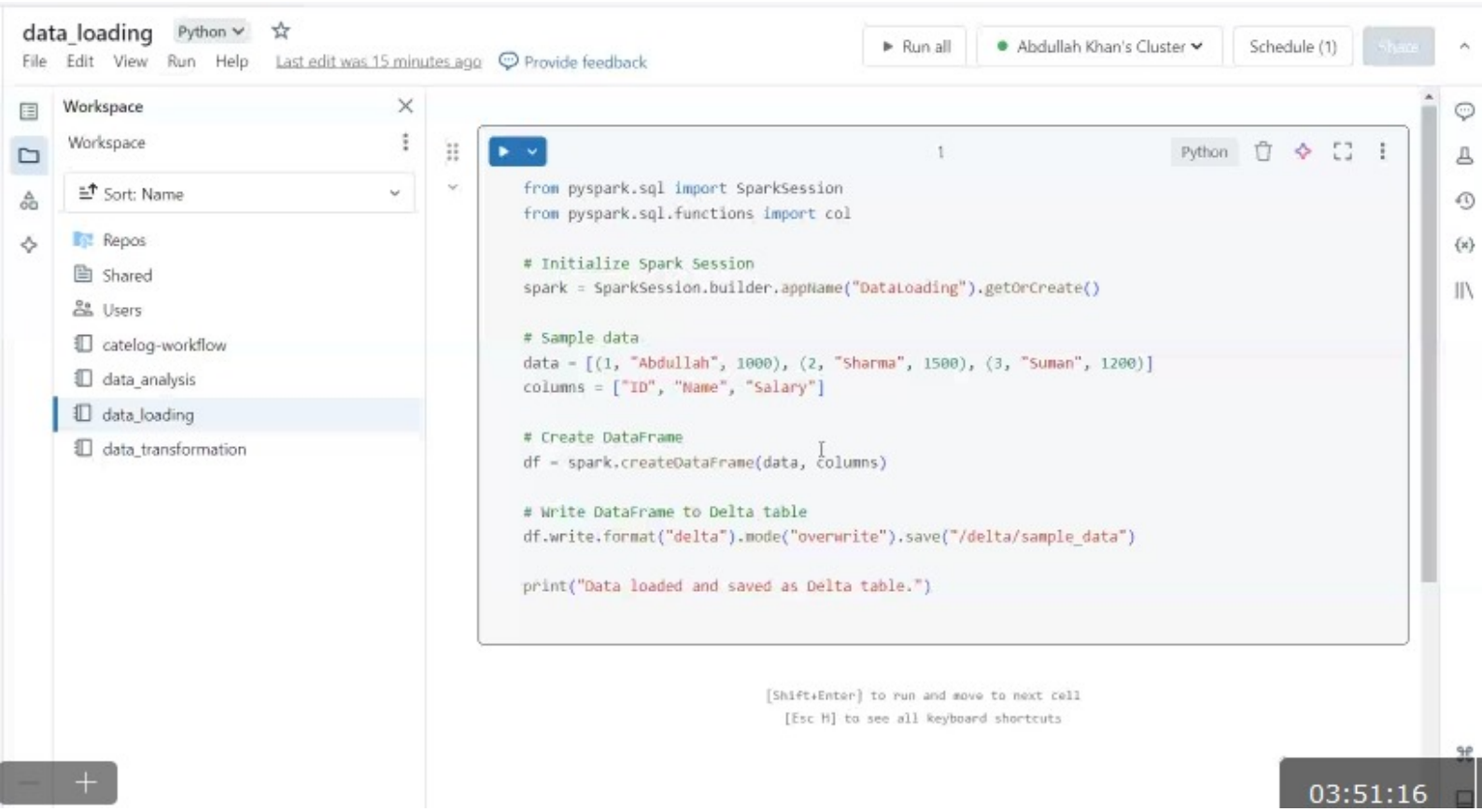
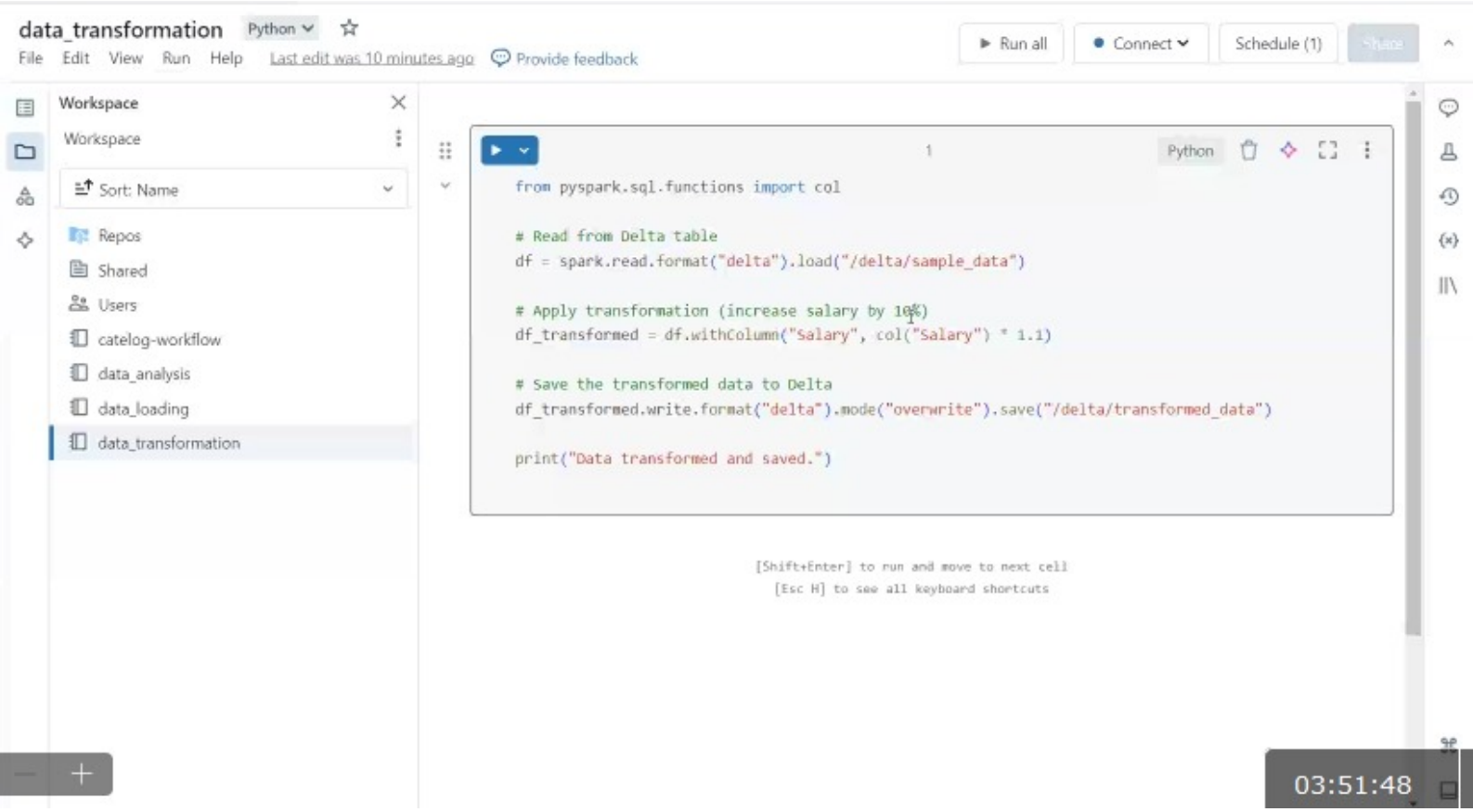
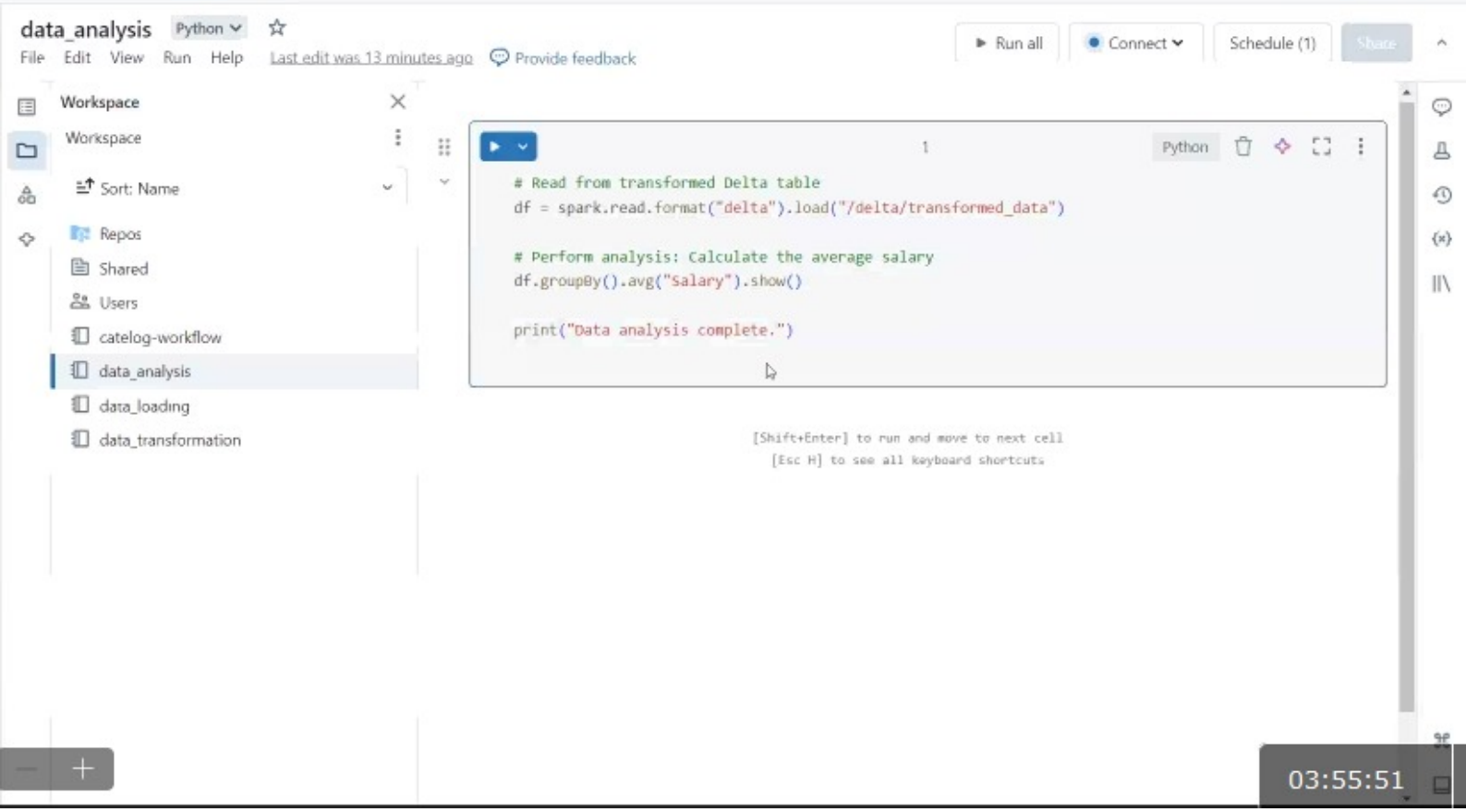

data_loading
Python
File Edit View Run Help
Last edit was 15 minutes ago
Provide feedback
Run all
Abdullah Khan's Cluster
Schedule (1)
Workspace
Workspace
Sort: Name
Repos
Shared
Users
catelog-workflow
data_analysis
data_loading
data_transformation
1
Python
from pyspark.sql import SparkSession
from pyspark.sql.functions import col
# Initialize Spark Session
spark = SparkSession.builder.appName("DataLoading").getOrCreate()
# Sample data
data = [(1, "Abdullah", 1000), (2, "Sharma", 1500), (3, "Suman", 1200)]
columns = ["ID", "Name", "Salary"]
# Create DataFrame
df = spark.createDataFrame(data, columns)
# Write DataFrame to Delta table
df.write.format("delta").mode("overwrite").save("/delta/sample_data")
print("Data loaded and saved as Delta table.")
[Shift+Enter] to run and move to next cell
[Esc H] to see all keyboard shortcuts
03:51:16


data_transformation
Python
File Edit View Run Help
Last edit was 10 minutes ago
Provide feedback
Run all
Connect
Schedule (1)
Workspace
Workspace
Sort: Name
Repos
Shared
Users
catelog-workflow
data_analysis
data_loading
data_transformation
1
Python
from pyspark.sql.functions import col
# Read from Delta table
df = spark.read.format("delta").load("/delta/sample_data")
# Apply transformation (increase salary by 10%)
df_transformed = df.withColumn("Salary", col("Salary") * 1.1)
# Save the transformed data to Delta
df_transformed.write.format("delta").mode("overwrite").save("/delta/transformed_data")
print("Data transformed and saved.")
[Shift+Enter] to run and move to next cell
[Esc H] to see all keyboard shortcuts
03:51:48


data_analysis
Python
File Edit View Run Help
Last edit was 13 minutes ago
Provide feedback
Run all
Connect
Schedule (1)
Share
Workspace
Workspace
Sort: Name
Repos
Shared
Users
catelog-workflow
data_analysis
data_loading
data_transformation
1
Python
# Read from transformed Delta table
df = spark.read.format("delta").load("/delta/transformed_data")
# Perform analysis: Calculate the average salary
df.groupBy().avg("Salary").show()
print("Data analysis complete.")
[Shift+Enter] to run and move to next cell
[Esc H] to see all keyboard shortcuts
03:55:51