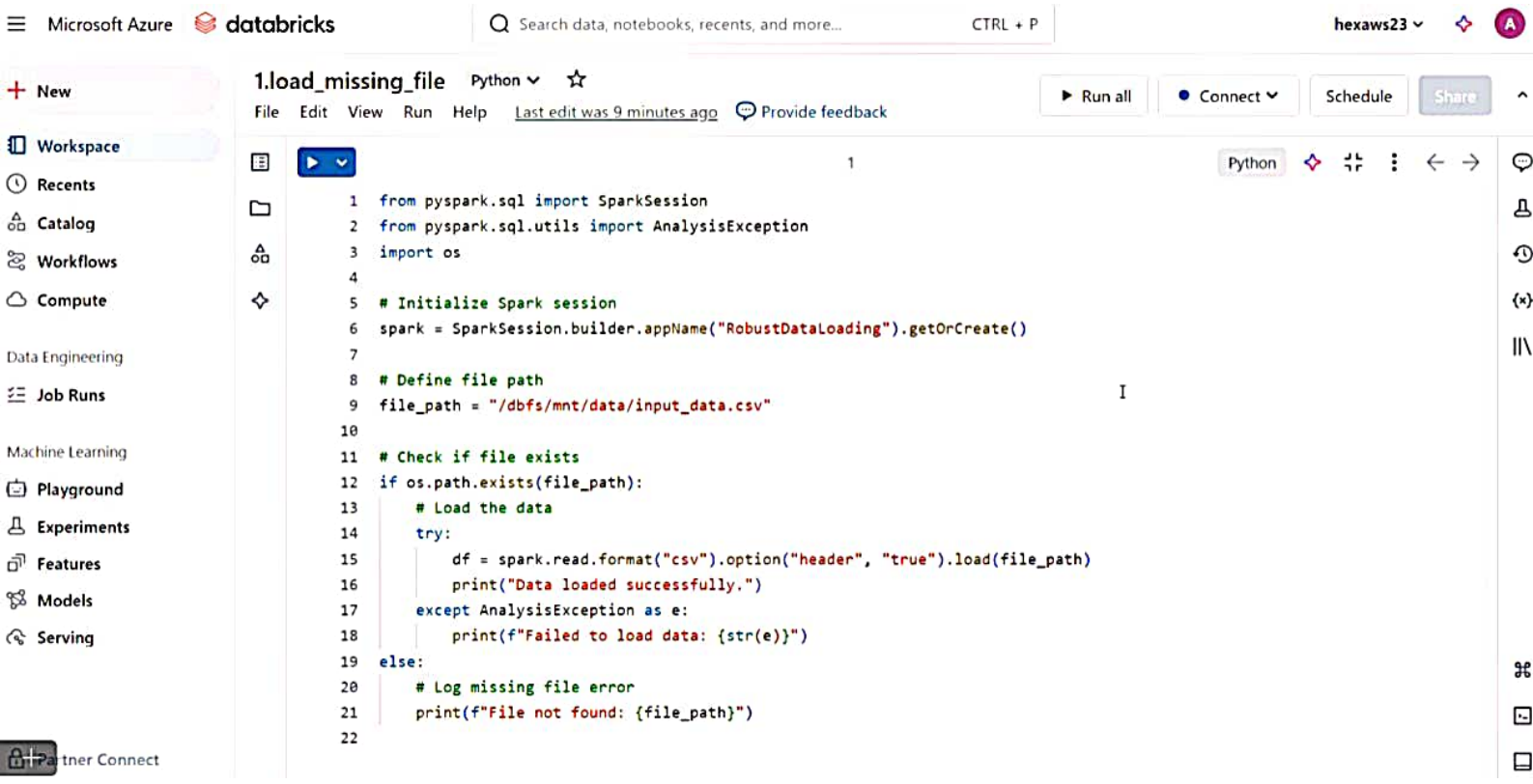
Microsoft Azure
databricks
Search data, notebooks, recents, and more...
CTRL + P
hexaws23
New
Workspace
Recents
Catalog
Workflows
Compute
Data Engineering
Job Runs
Machine Learning
Playground
Experiments
Features
Models
Serving
Partner Connect
1.load_missing_file
Python
File Edit View Run Help
Last edit was 9 minutes ago
Provide feedback
Run all
Connect
Schedule
Share
1
Python
from pyspark.sql import SparkSession
from pyspark.sql.utils import AnalysisException
import os
# Initialize Spark session
spark = SparkSession.builder.appName("RobustDataLoading").getOrCreate()
# Define file path
file_path = "/dbfs/mnt/data/input_data.csv"
# Check if file exists
if os.path.exists(file_path):
# Load the data
try:
df = spark.read.format("csv").option("header", "true").load(file_path)
print("Data loaded successfully.")
except AnalysisException as e:
print(f"Failed to load data: {str(e)}")
else:
# Log missing file error
print(f"File not found: {file_path}")

2.retry_file_load Python

File Edit View Run Help Last edit was 8 minutes ago Provide feedback

Run all Connect Schedule

1 Python

```python
from pyspark.sql.functions import col
import time

# Number of retries
max_retries = 3
retries = 0
success = False

# Define the transformation logic
def transform_data():
    df = spark.read.format("csv").option("header", "true").load("/dbfs/mnt/data/input_data.csv")
    df_transformed = df.withColumn("Salary", col("Salary").cast("double") * 1.1)
    df_transformed.write.format("delta").mode("overwrite").save("/delta/transformed_data")
    print("Data transformed successfully.")

# Retry loop
while retries < max_retries and not success:
    try:
        transform_data()
        success = True
    except Exception as e:
        retries += 1
        print(f"Transformation failed. Retry {retries}/{max_retries}")
```

databricks

Search data, notebooks, recents, and more... CTRL + P

hexaws23

2.retry_file_load Python

File Edit View Run Help Last edit was now Provide feedback

Run all Connect Schedule Share

1 Python

```python
# Number of retries
max_retries = 3
retries = 0
success = False
# Define the transformation logic
def transform_data():
    df = spark.read.format("csv").option("header", "true").load("/dbfs/mnt/data/input_data.csv")
    df_transformed = df.withColumn("Salary", col("Salary").cast("double") * 1.1)
    df_transformed.write.format("delta").mode("overwrite").save("/delta/transformed_data")
    print("Data transformed successfully.")

# Retry loop
while retries < max_retries and not success:
    try:
        transform_data()
        success = True
    except Exception as e:
        retries += 1
        print(f"Transformation failed. Retry {retries}/{max_retries}")
        time.sleep(5)
if not success:
    print("Transformation failed after max retries.")
```

3.data_analysis Python

File Edit View Run Help Last edit was 6 minutes ago Provide feedback

Run all | Connect | Schedule | Share

1 Python

```
# Define Delta table path
delta_table_path = "/delta/transformed_data"

# Check if the Delta table exists
try:
    spark.sql(f"DESCRIBE HISTORY delta.`{delta_table_path}`").show()
    # If table exists, perform analysis
    df = spark.read.format("delta").load(delta_table_path)
    df.groupBy("Department").avg("Salary").show()
    print("Data analysis completed successfully.")
except AnalysisException as e:
    print(f"Data not found or incomplete: {str(e)}. Skipping analysis.")
```

[Shift+Enter] to run and move to next cell

[Esc H] to see all keyboard shortcuts