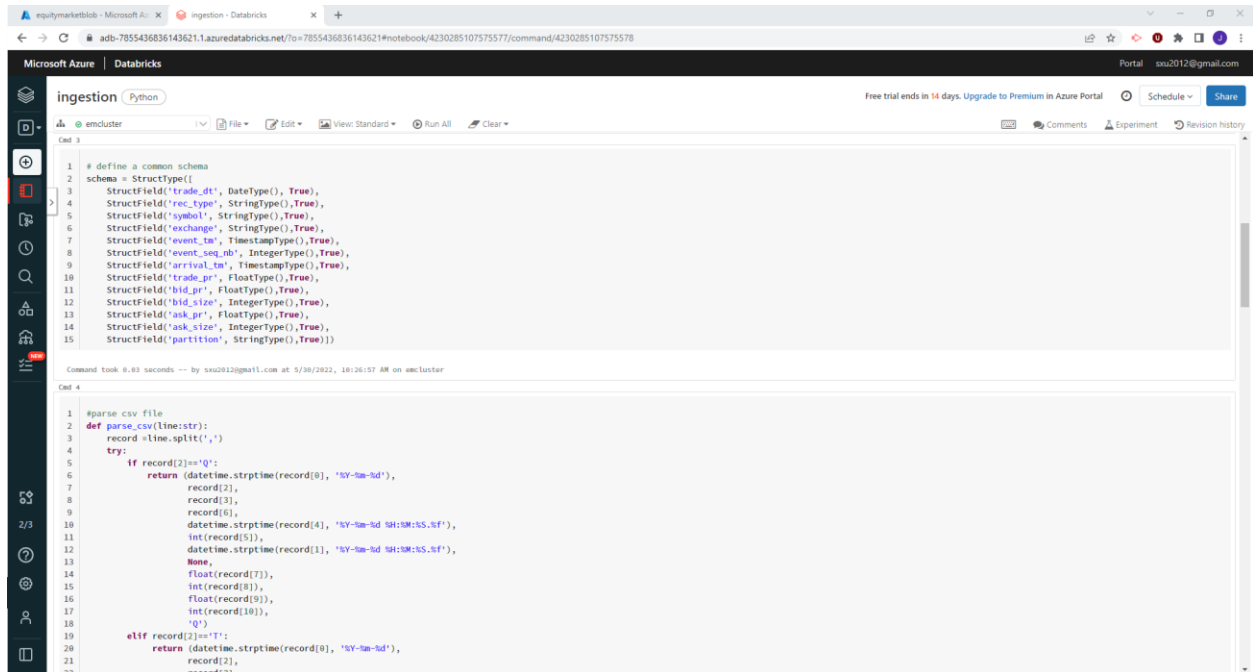


Azure Databricks:

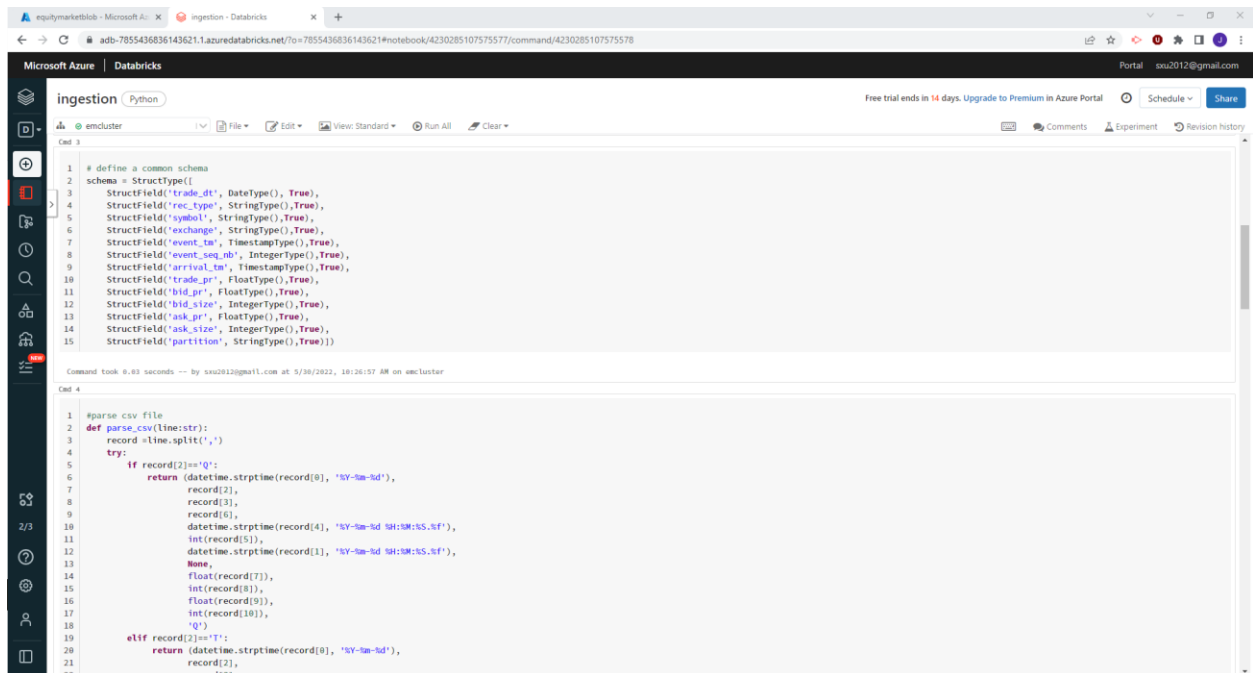


The screenshot shows the Azure Databricks web interface. The notebook is titled "ingestion" and is written in Python. The interface includes a sidebar with navigation icons (Home, Recent, Workspace, Clusters, Jobs, Recent Clusters, Recent Jobs, Recent Clusters, Recent Jobs, Recent Clusters, Recent Jobs) and a top bar with the user's name "sou2012@gmail.com" and a trial status "Free trial ends in 14 days. Upgrade to Premium in Azure Portal".

```
1 # define a common schema
2 schema = StructType([
3     StructField('trade_dt', DateType(), True),
4     StructField('rec_type', StringType(), True),
5     StructField('symbol', StringType(), True),
6     StructField('exchange', StringType(), True),
7     StructField('event_tm', TimestampType(), True),
8     StructField('event_seq_nb', IntegerType(), True),
9     StructField('arrival_tm', TimestampType(), True),
10    StructField('trade_pr', FloatType(), True),
11    StructField('bid_pr', FloatType(), True),
12    StructField('bid_size', IntegerType(), True),
13    StructField('ask_pr', FloatType(), True),
14    StructField('ask_size', IntegerType(), True),
15    StructField('partition', StringType(), True)])

Command took 0.03 seconds -- by sou2012@gmail.com at 5/30/2022, 10:26:57 AM on emc1cluster

1 #parse csv file
2 def parse_csv(line:str):
3     record = line.split(',')
4     try:
5         if record[2]=='Q':
6             return (datetime.strptime(record[0], '%Y-%m-%d'),
7                     record[2],
8                     record[3],
9                     record[6],
10                    datetime.strptime(record[4], '%Y-%m-%d %H:%M:%S.%f'),
11                    int(record[5]),
12                    datetime.strptime(record[1], '%Y-%m-%d %H:%M:%S.%f'),
13                    None,
14                    float(record[7]),
15                    int(record[8]),
16                    float(record[9]),
17                    int(record[10]),
18                    'Q')
19         elif record[2]=='T':
20             return (datetime.strptime(record[0], '%Y-%m-%d'),
21                     record[2],
22                     record[3],
```



This screenshot is identical to the one above, showing the same Azure Databricks notebook with the same code in two cells. The interface elements, including the sidebar and top bar, are also identical.

```
1 # define a common schema
2 schema = StructType([
3     StructField('trade_dt', DateType(), True),
4     StructField('rec_type', StringType(), True),
5     StructField('symbol', StringType(), True),
6     StructField('exchange', StringType(), True),
7     StructField('event_tm', TimestampType(), True),
8     StructField('event_seq_nb', IntegerType(), True),
9     StructField('arrival_tm', TimestampType(), True),
10    StructField('trade_pr', FloatType(), True),
11    StructField('bid_pr', FloatType(), True),
12    StructField('bid_size', IntegerType(), True),
13    StructField('ask_pr', FloatType(), True),
14    StructField('ask_size', IntegerType(), True),
15    StructField('partition', StringType(), True)])

Command took 0.03 seconds -- by sou2012@gmail.com at 5/30/2022, 10:26:57 AM on emc1cluster

1 #parse csv file
2 def parse_csv(line:str):
3     record = line.split(',')
4     try:
5         if record[2]=='Q':
6             return (datetime.strptime(record[0], '%Y-%m-%d'),
7                     record[2],
8                     record[3],
9                     record[6],
10                    datetime.strptime(record[4], '%Y-%m-%d %H:%M:%S.%f'),
11                    int(record[5]),
12                    datetime.strptime(record[1], '%Y-%m-%d %H:%M:%S.%f'),
13                    None,
14                    float(record[7]),
15                    int(record[8]),
16                    float(record[9]),
17                    int(record[10]),
18                    'Q')
19         elif record[2]=='T':
20             return (datetime.strptime(record[0], '%Y-%m-%d'),
21                     record[2],
22                     record[3],
```


Microsoft Azure | Databricks

equity-market-data-analysis

Runs Tasks

Task name * equity-market-data-analysis

Type * Notebook Source * Local

Path * /Users/sxu2012@gmail.com/equity-market-analysis-daily

Cluster * emcluster (8.00 GB | 4 Cores | DBR 10.4 LTS | Spark 3.2.1 | Scala 2.12)

Parameters * UI | JSON

Alerts * xsu2012@gmail.com Start Success Failure X

Timeout in Seconds * 180

Advanced options

Job details

Job ID 143767312279023

Creator xsu2012@gmail.com

Run as xsu2012@gmail.com

Tags + Tag

Git Add Git settings

Schedule None Edit schedule

Clusters emcluster Driver: Standard_F4, Workers: Standard_F4, 0 workers, 10.4 LTS (includes Apache Spark 3.2.1, Scala 2.12)

View cluster Swap Spark UI Logs Metrics

Email alerts No alerts Edit alerts

Permissions

Microsoft Azure

sbcap

Overview

Subscription (mvp) Azure subscription 1

Subscription ID 88045111-50ef-4121-b158-4449e5b40272

Location West US 3

Deployments 3 Succeeded

Resources Recommendations

Filter for any field... Type == all Location == all Add filter

Showing 1 to 2 of 2 records Show hidden types

Name	Type	Location
equitymarketblob	Storage account	West US
equitymarketdb	Azure Databricks Service	West US

Page 1 of 1