Starburst

Trino/Starburst Training:
PyStarburst introduction/examples

Starburst Academy
v1.0.0

Session
5 of 5

Connection Before Content

Lester Martin - https://about.me/lestermartin

- Educational Engineer @ Starburst
  - Build the content
  - Teach the class
  - Repeat
- 30 years of technology experience
  - Started my journey on a TRS-80 Model III
  - Played most every role, but consider myself a programmer at my core
  - Half of career in transactional systems and the second half in analytical processing
  - A DECADE of “big data” experience to include
    - Trino/Starburst, Hadoop, Hive, Spark
    - NiFi, Kafka, Storm, Flink
    - HBase, MongoDB

Session #5 objectives

- Differentiate SQL-based data engineering from programming-based
- Understand how PyStarburst implements lazy execution with Starburst Galaxy
- Explore the DataFrame API
- Write Python code to perform analytical queries and transformation processing

Session #5 objectives

- Differentiate SQL-based data engineering from programming-based
- Understand how PyStarburst implements lazy execution with Starburst Galaxy
- Explore the DataFrame API
- Write Python code to perform analytical queries and transformation processing

YES... in 1 hour!!

Session #5 objectives

- Differentiate SQL-based data engineering from programming-based
- Understand how PyStarburst implements lazy execution with Starburst Galaxy
- Explore the DataFrame API
- Write Python code to perform analytical queries and transformation processing

YES… in 1 hour!!

SO… we'd better get moving!!

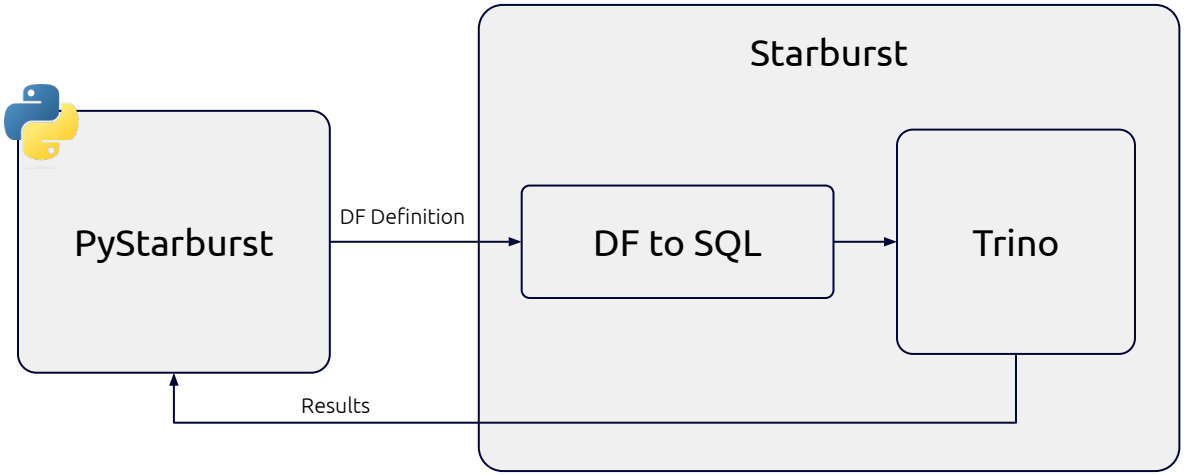
PyStarburst Overview

```
df_missions = df_missions.with_column("date", f.sql_expr("COALESCE(TRY(date_parse(\"date\", '%a %b %d, %Y %H:%i UTC')), NULL)"))

print(df_missions.schema)

df_missions = df_missions\
    .filter(col("date") > datetime(2000, 1, 1))\
    .sort(col("date"), ascending=True)

df_missions.show()
```



- PySpark-like Syntax
- Lazy execution
- Python gets converted to SQL
- Heavy lifting done by Trino

Links: API Docs & Example Code

Currently supported on Starburst Galaxy and Starburst Enterprise.

Starburst

POIs

- The lab guide has it ALL!
- WE MADE IT TO THE FINAL SESSION
  - ~~January 31st – *Creating & querying data lake tables*~~
  - ~~February 14th – *Modern table formats & Apache Iceberg (Happy Valentine's Day!!)*~~
  - ~~February 28th – *Materialized views & data pipelines*~~
  - ~~March 13th – *Experience Warp Speed in action*~~
  - March 27th – *Transformation processing with PyStarburst*
- All code is already present in the jupyter notebook

POIs

- The lab guide has it ALL!
- WE MADE IT TO THE FINAL SESSION
  - ~~January 31st – *Creating & querying data lake tables*~~
  - ~~February 14th – *Modern table formats & Apache Iceberg (Happy Valentine's Day!!)*~~
  - ~~February 28th – *Materialized views & data pipelines*~~
  - ~~March 13th – *Experience Warp Speed in action*~~
  - March 27th – *Transformation processing with PyStarburst*
- All code is already present in the jupyter notebook

Sessions 1-4 recordings available

Starburst

Today’s approach

- I’m going to perform the labs myself & you can…
  - do them along with me
  - or just watch (and maybe do later)
- We ALL come from DIFFERENT experiences & backgrounds, so…
  - **ASK QUESTIONS**
    - **FILL UP THE CHAT!**
    - **SHARP FOLKS THERE TO ADDRESS YOUR QUESTIONS!!**

Today’s approach

- I’m going to perform the labs myself & you can…
  - do them along with me
  - or just watch (and maybe do later)
- We ALL come from DIFFERENT experiences & backgrounds, so…
  - **ASK QUESTIONS**
    - **FILL UP THE CHAT!**
    - **SHARP FOLKS THERE TO ADDRESS YOUR QUESTIONS!!**

Starburst

HANDS-ON LABS!!

Starburst Academy