Predict Bart Station Capacity

Jonathan Hilgart

What if you wanted to know how full the bart station would be before you arrived?

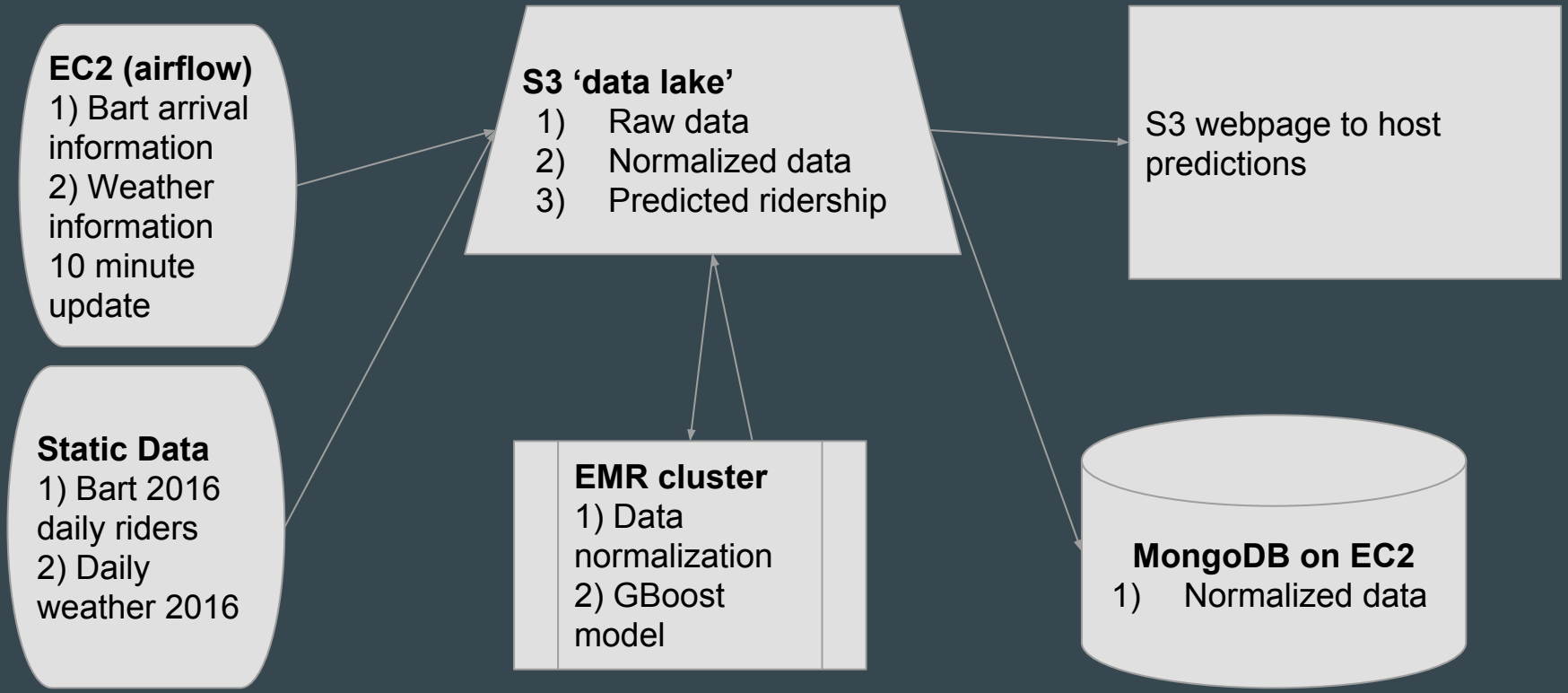
Data Sources - Historical

- 2016 historic daily rider exits per BART station
  - Origin Station
  - Destination Station
  - Number of Exits
  - Hour of the day
- 2016 historic daily weather for San Francisco (zip 94105)
  - Wind speed
  - Humudity
  - Minimum temp.
  - Max temp
  - Avg. temp
  - Sea pressure

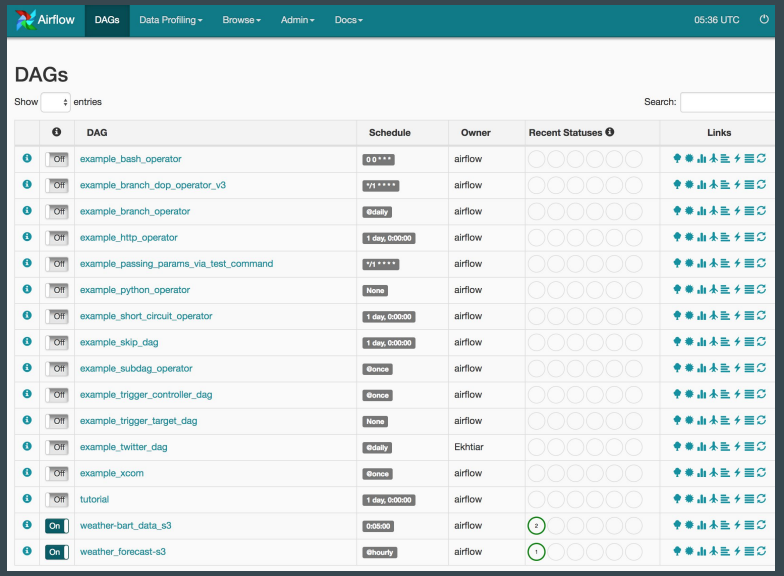
Data Sources - Live

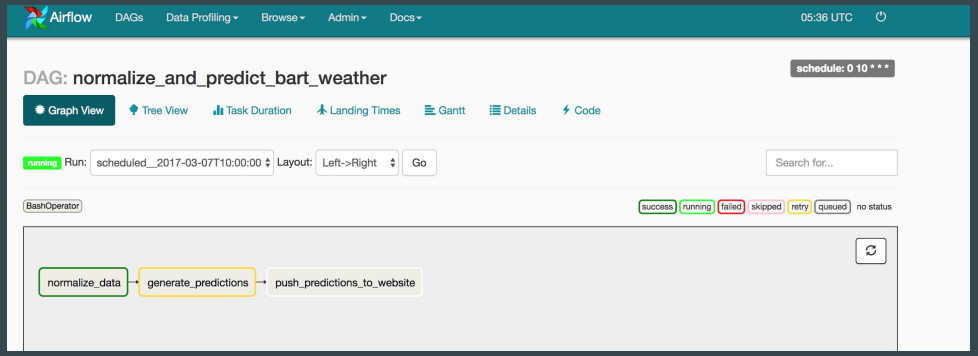
- Live BART information pulled every ten minutes
  - Origin station
  - Destination station
  - Number of trains
  - Bike flag
  - Minutes until arrival
- Live weather information pulled every ten minutes
  - Wind
  - Min. temp
  - Max. temp
  - Avg. temp
  - Humidity
  - Sea pressure

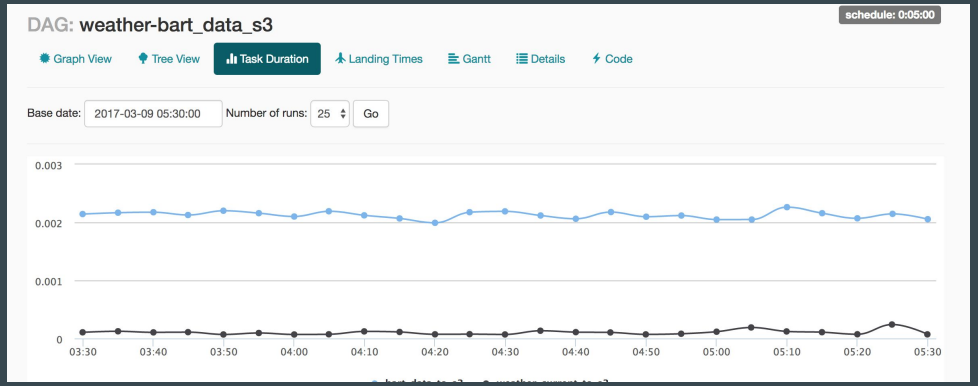
Architecture



Airflow


Airflow
DAGs
Data Profiling
Browse
Admin
Docs
05:36 UTC
DAGs
Show entries
Search:
DAG
Schedule
Owner
Recent Statuses
Links
example_bash_operator
example_branch_dop_operator_v3
example_branch_operator
example_http_operator
example_passing_params_via_test_command
example_python_operator
example_short_circuit_operator
example_skip_dag
example_subdag_operator
example_trigger_controller_dag
example_trigger_target_dag
example_twitter_dag
example_xcom
tutorial
weather-bart_data_s3
weather_forecast-s3
airflow
Ekhtiar


Airflow
DAGs
Data Profiling
Browse
Admin
Docs
05:36 UTC
DAG: normalize_and_predict_bart_weather
schedule: 0 10 * * *
Graph View
Tree View
Task Duration
Landing Times
Gantt
Details
Code
running
Run: scheduled__2017-03-07T10:00:00
Layout: Left->Right
Go
Search for...
BashOperator
success
running
failed
skipped
retry
queued
no status
normalize_data
generate_predictions
push_predictions_to_website


DAG: weather-bart_data_s3
schedule: 0:05:00
Graph View
Tree View
Task Duration
Landing Times
Gantt
Details
Code
Base date: 2017-03-09 05:30:00
Number of runs: 25
Go
0.003
0.002
0.001
0
03:30
03:40
03:50
04:00
04:10
04:20
04:30
04:40
04:50
05:00
05:10
05:20
05:30

Next Steps & Challenges

- Challenges
  - BART data in XML format. Needed to convert to JSON
  - Sparse documentation for Airflow
  - Differences between spark interactive and spark-submit
  - Unix time
- Next Steps
  - Change daily prediction to hourly prediction
  - Incorporate direction of BART train (North or South)
  - Incorporate localized weather data for each BART station
  - Add visualizations
  - Incorporate MUNI data
  - Lambda architecture with ‘speed’ layer

Results

http://bart-capacity-predictions.com.s3-website-us-east-1.amazonaws.com/