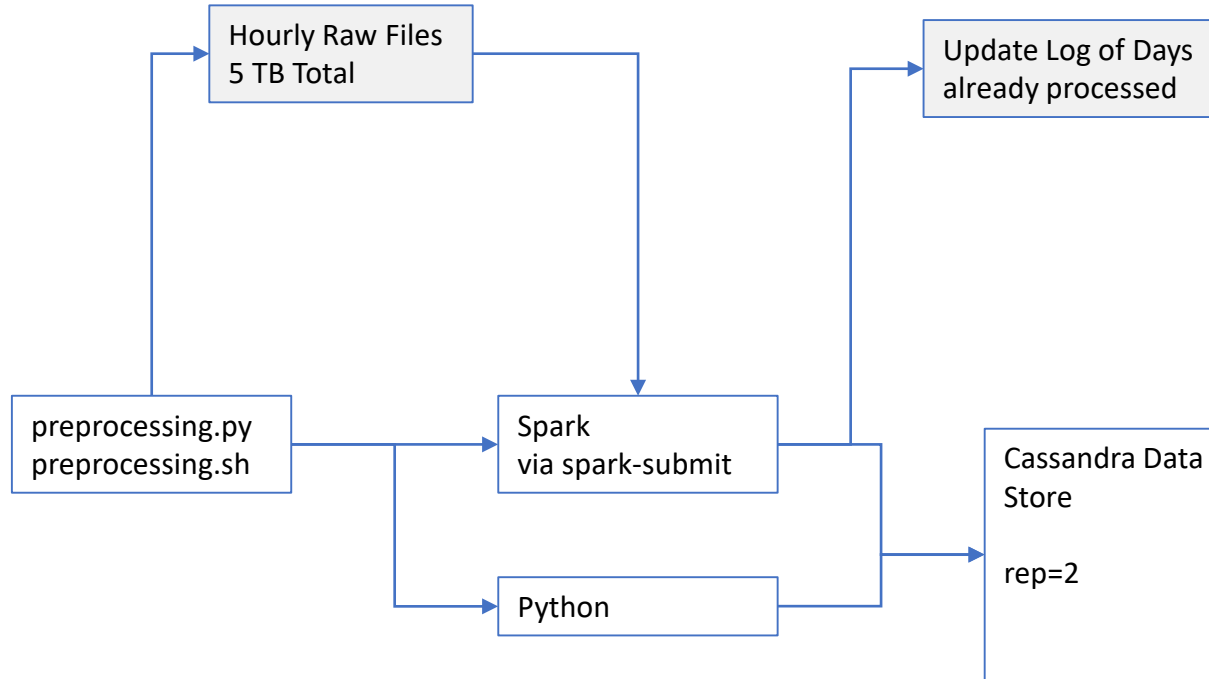
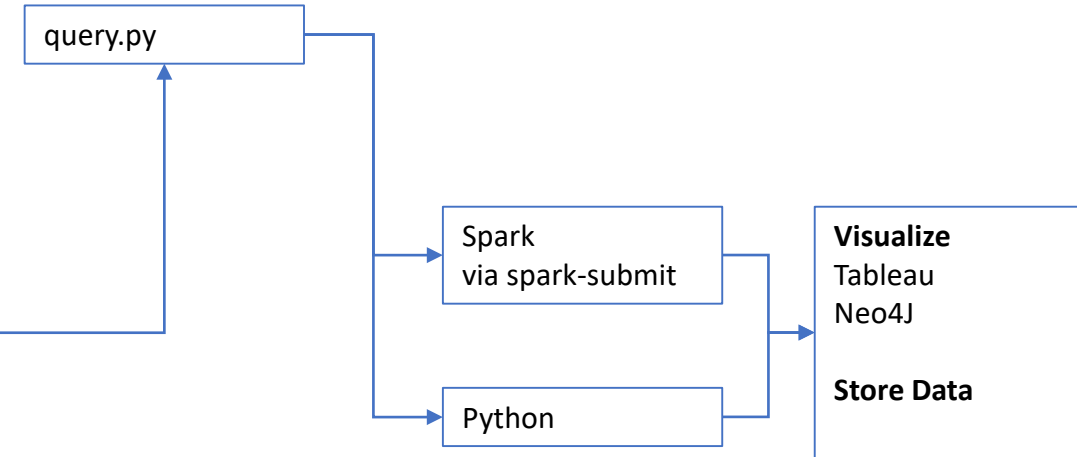


## Preprocessing



## Querying



Preprocessing:

Aggregate Raw Hourly Files by **Day**

Daily\_Aggregated\_Files.json

### example

Barack\_Obama,\_Sr\_20170808.json

date: 2017-08-08

language: en

topic: Barack\_Obama,\_Sr

daily\_views: 200

## To Do

- Operationalize a Preprocessing Script
- Preprocess all Raw Data (limit to 1TB to start?)
- Write Query Document Structure
- Setup Tableau or something else?