Presentation Title: Presentation template v0.1

Daniel W. Engels, PhD

First Slide

- Transition to this slide on first mouse click
- From here until end of presentation all slides are on a 15 second timer
- At the end of 15 seconds, the next slide will automatically appear

20 slides total

- A total of 20 slides are allowed for content for a total presentation time of 5 minutes

Slide Format

- This is a generic slide format:
  - White background
  - DataScience@SMU in bottom left corner
  - Page number in bottom right corner

Hacking the Slide Format

- Hacking the slide format is allowed
- However,
  - The total number of slides
    - must be 21
  - The slides must transition automatically every 15 seconds
  - The aspect must be the standard format (not 16:9)
  - The footer information must be present and in the same position as this template

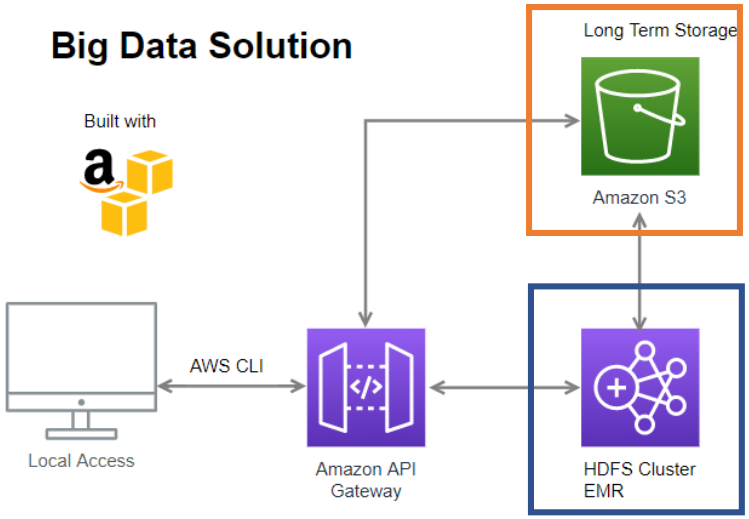
AWS Based System

AWS Elastic MapReduce Cluster

Pre-configured Hadoop System

- Hadoop
- Hive
- Hue

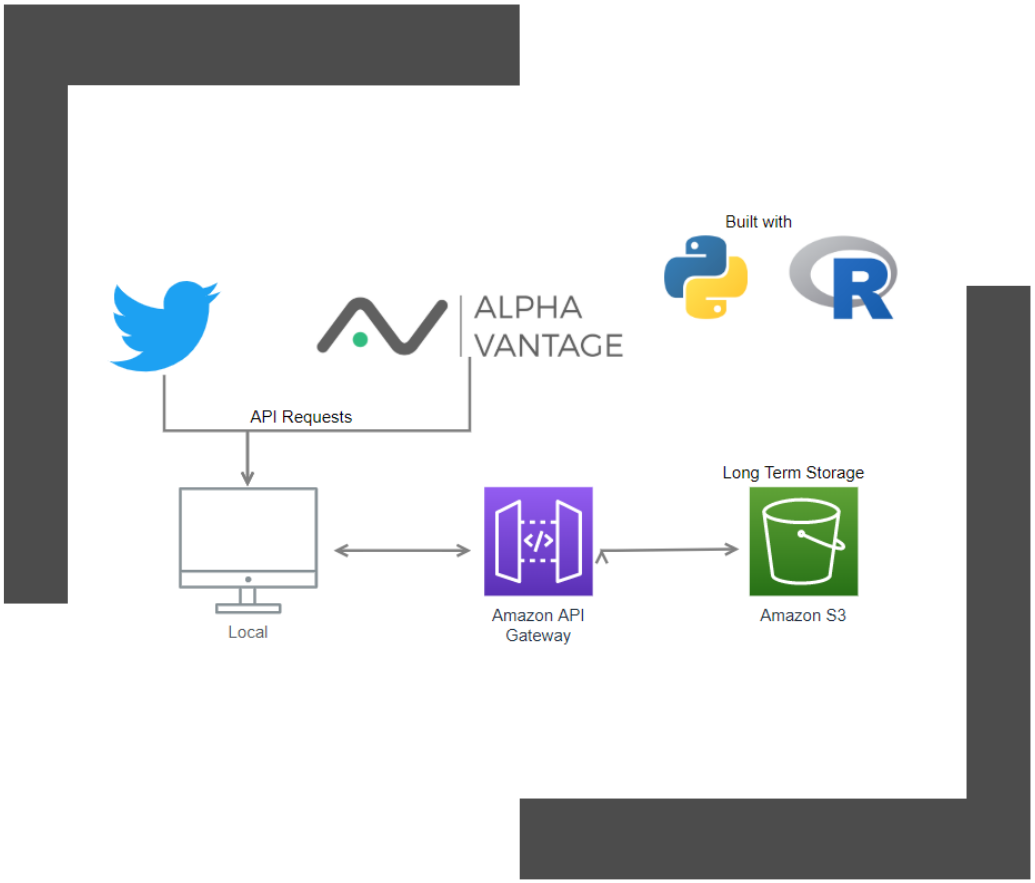
S3 Storage



How it Works

- Cluster created on demand
- Data loaded from S3 bucket
- Runs Hive scripts to create database
- Interact with database with
  - Hue Interface
  - Command Line

Data Collection



Intra-day Stock Prices

Tweets

Tweet Data

Features

- Tweet Text
- Hashtags
- Urls
- Mentions
- Post time

300,000 json files

Last Slide

- Fade to black after 15 seconds