BEATING ONE BUS AWAY AT ITS OWN GAME

MICHAEL MAKRIS

Project Question:
Can I improve mass transit rider experience by providing accurate predictions of bus arrivals and trip time based on past performance?

Second, is it possible is to help riders decide which bus to take, all factors considered?

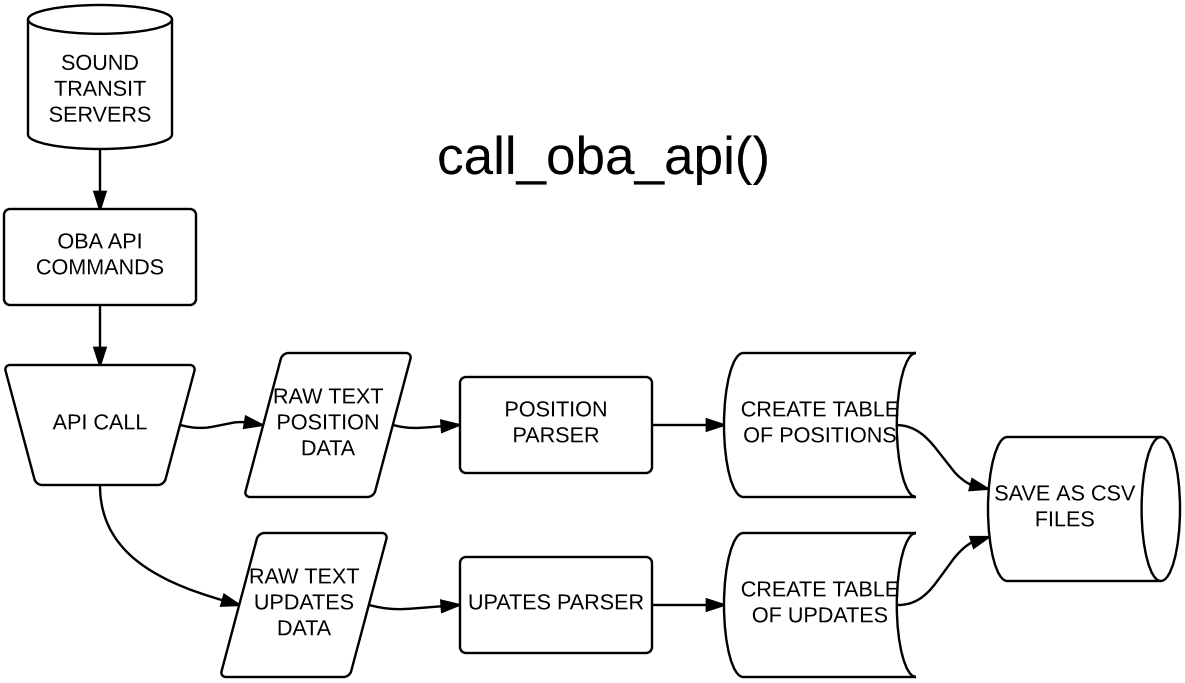
DATA GATHERED:

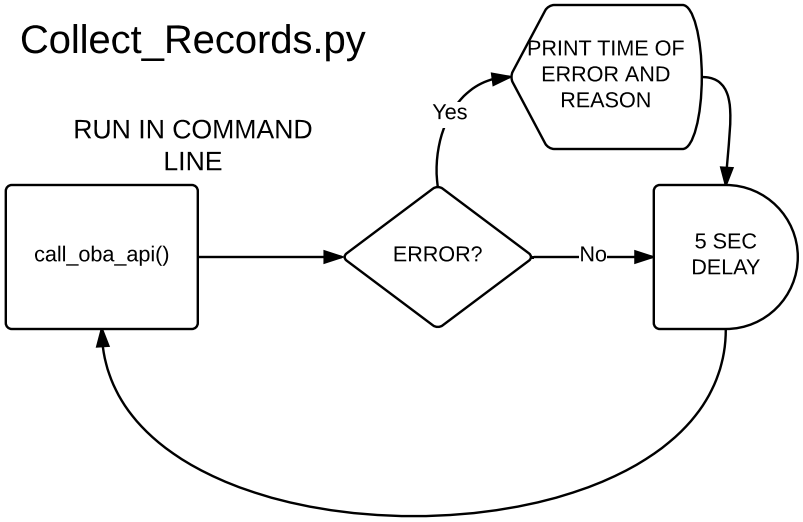
STATIC AGENCY DATA:

SCHEDULES
STOPS
STOP TIMES
ROUTES
TRIPS

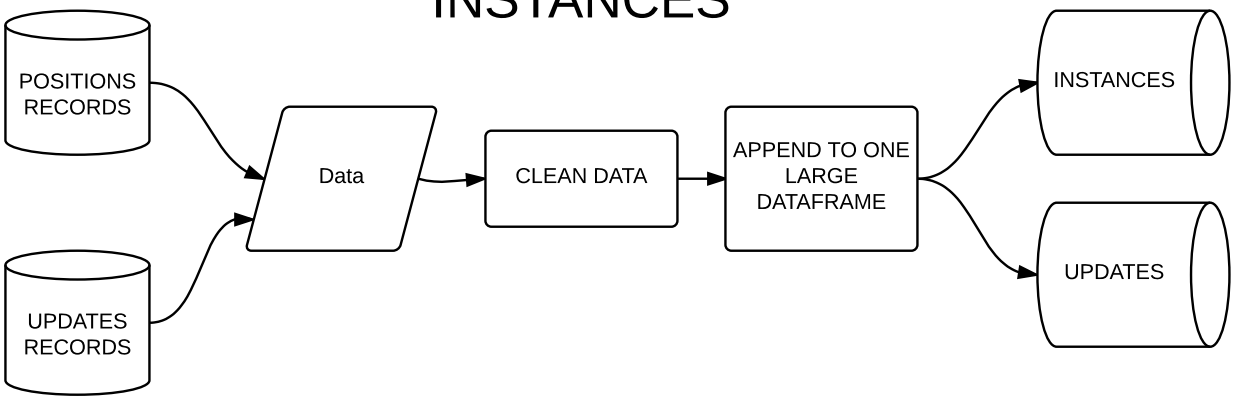
REAL TIME DATA:

BUS POSITIONS
AGENCY UPDATES

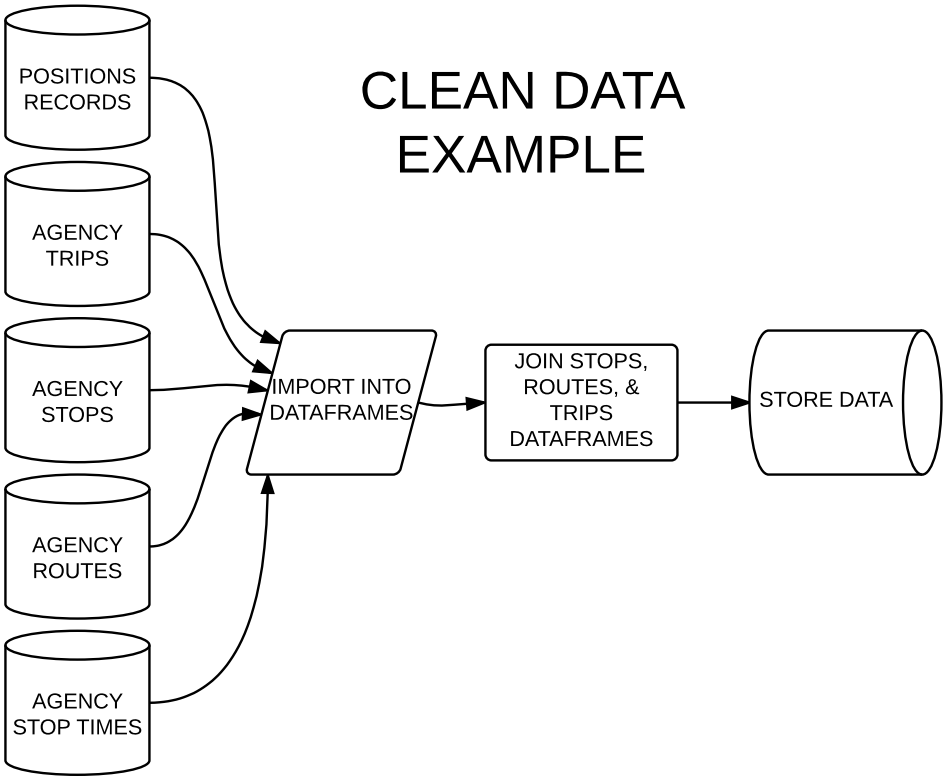

call_oba_api()
SOUND TRANSIT SERVERS
OBA API COMMANDS
API CALL
RAW TEXT POSITION DATA
POSITION PARSER
CREATE TABLE OF POSITIONS
RAW TEXT UPDATES DATA
UPATES PARSER
CREATE TABLE OF UPDATES
SAVE AS CSV FILES


Collect_Records.py
RUN IN COMMAND LINE
call_oba_api()
ERROR?
Yes
No
PRINT TIME OF ERROR AND REASON
5 SEC DELAY

CREATE INSTANCES



CLEAN DATA EXAMPLE



INSIGHTS

1. Bus Data Frequency Issues:

5 Seconds too Often

2. Large Amounts of Data:

Expected ~ 4 GB Raw

3. Basic Dataframe Commands Fail with Large Datasets:

Learning to Use the "Proper" DataFrame Commands

CONTINUING WORK

COLLATING COMPLETE 2-WEEK HISTORY ON COMMAND

INPUTS:
Agency ID
Bus Route ID

OPTIONAL INPUT:
Stop ID

USE REGRESSION TO PREDICT CURRENT DELAY AND ARRIVAL TIME BASED ON CURRENT CONDITIONS