Data Engineer Challenge

We have multiple apps emitting different kinds of events, such as when a push notification has been received, when a certain screen was viewed, or the app has been updated. We also get an event called “user_engagement” for each session of use, with the time the user was active in the app.

We would like to be able to plot the number of active users per day. Our analytics database is an SQL database, so the data from the event should be loaded in the following table:

active_user_table

Field name	Type	Mode	Description
date	DATE	REQUIRED	
active_user_count	INTEGER	REQUIRED	

1. Using the data in the attached json file, create a script to load the data from the events into the table above. You can use python, or any language you feel familiar with. Please provide instructions on how to run the script.
2. How would you design the ETL process for it to automatically update daily?
   How would you scale this process if we got tens or hundreds of millions of events per day?
3. Suggest any target architecture to cater for this growth.

We expect the result as a github repository. Please leave the commit history in.

Notes:

1. event_date is UTC. Our users are based in the US.
2. We say a user is active if the engagement time is at least 3 seconds and any valuable events occured at least once.