Note: This assignment is intentionally open ended so please clearly state any assumptions

Imagine that you are working on a data platform solution for a retail channel engagement customer, and your team is looking for the next important metric to deliver to end enterprise (the merchants/retailers).

The customer provided an insight that the most frequently requested metric by all retailers is an average of how long users spend on their flyers each day. This would allow them to measure the effectiveness of the flyer, as well as understanding what resonates more with their consumers.

Background information for this exercise:
- Assume that the channel engagement platform is a mobile or tablet application.
- Flyers have unique identifiers
- Merchants have unique identifiers
- Merchants have multiple flyers

You have been given a sample dataset with the following schema:
timestamp, user_id, event, flyer_id, merchant_id

There are a number of events which the mobile app collects as the shoppers the app. These are the 'event" types in the preceding schema.

- flyer_open - Whenever a flyer is opened by a user, to see the flyer contents
- item_open - Whenever an item is opened by the user, to see the item details
- list_flyers - Whenever a listing of the flyers is shown
- shopping_list_open - Whenever a user opens their shopping list
- favorite - Whenever a user adds a merchant to their favourites.

Note: The user is considered to close the old flyer on opening a new flyer.

Your task is to:
1. Compute the unique users in the dataset.
2. Compute the average time on flyer per user.
3. Compute the average time users spent on **flyer_id = 2020004**
4. Explain how your algorithm scales for:
   a) 1 Million Events (~10 MB of data)
   b) 1 Trillion Events (~10 TB of data)

You need to write your program code in Python3 or Scala.