

Planning Your Data Pipeline: Moovio+

The raw data

- Multi-line JSON files
- Resemble the strings passed by Kafka
- Each file consists of five users:
 - Heart rate measured each hour, 24 hours a day, every day

Health tracker data sample

```
{"device_id":0,"heartrate":52.8139067501,"name":"Deborah Powell","time":1.5778368E9}  
{"device_id":0,"heartrate":53.9078900098,"name":"Deborah Powell","time":1.5778404E9}  
{"device_id":0,"heartrate":52.7129593616,"name":"Deborah Powell","time":1.577844E9}  
{"device_id":0,"heartrate":52.2880422685,"name":"Deborah Powell","time":1.5778476E9}  
{"device_id":0,"heartrate":52.5156095386,"name":"Deborah Powell","time":1.5778512E9}  
{"device_id":0,"heartrate":53.6280743846,"name":"Deborah Powell","time":1.5778548E9}
```

Note that each line is a valid JSON object.

Health Tracker Data Schema

name: string

heartrate: double

device_id: long

time: float

Planning your data pipeline



- Where is the data coming from?
- How much data exists?
- What is this type of data?
- What are the SLA requirements around how it will be used?
- How frequently is it updated?
- What kind of inconsistencies or uncertainties might you anticipate?
- What might the raw → bronze → silver → gold levels look like?

The operations File



- The operations file is used to define and source the following composable pipeline functions:
 - `create_stream_writer`
 - `read_stream_delta`
 - `read_stream_raw`
 - `update_silver_table`
 - `transform_bronze`
 - `transform_raw`
 - `transform_silver_mean_agg_last_thirty`

```
|— plus
|   |— 00_ingest_raw
|   ...
|       |— includes
|           |— configuration
|               |— main
|                   |— python
|                       |— operations
```

Breakout session



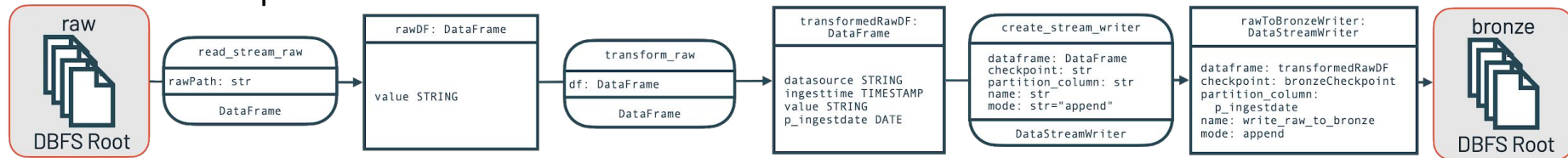
Planning your data pipeline

1. Considering the functions given in the operations file, fill in the function names, the arguments they will take, and the data types returned.
2. Fill in the schemas for the data frames.
3. Fill in the appropriate values for the datastream writer arguments.

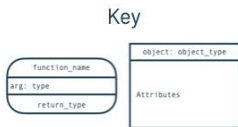
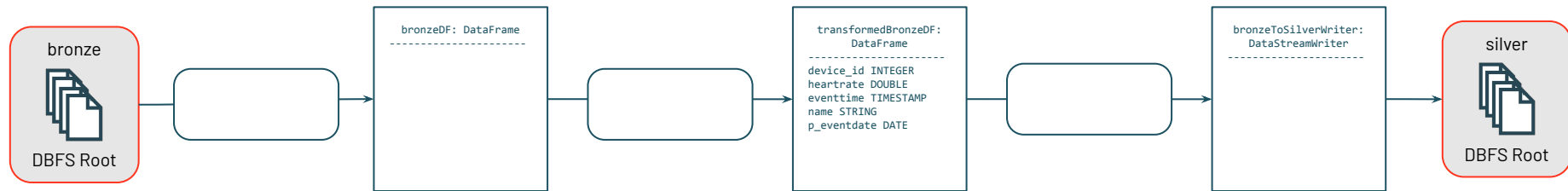
Group 1



Raw to Bronze Pipeline



Bronze to Silver Pipeline

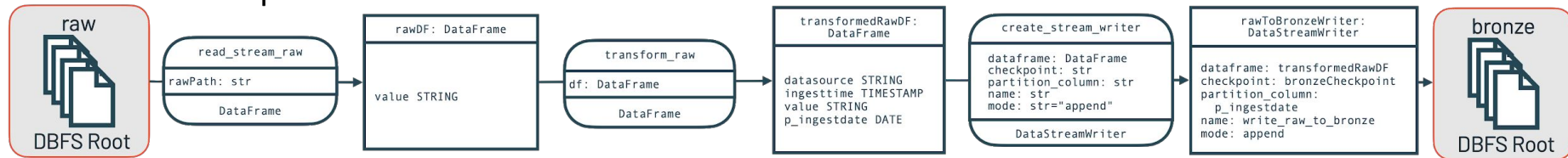


Moovio Plus Delta Architecture

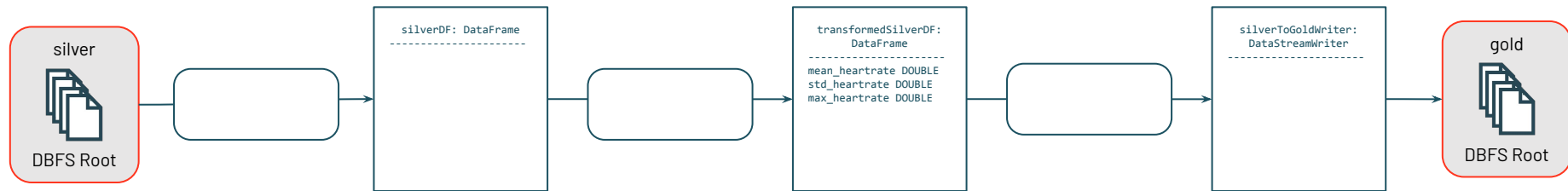
Group 2



Raw to Bronze Pipeline



Silver to Gold Pipeline



Key



Moovio Plus Delta Architecture

