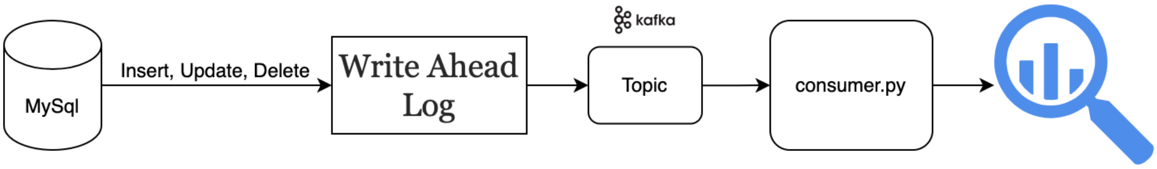
a) The idea to sync these data is to build some infrastructure CDC based. In this example, I will not be using a CDC tool (such as Debezium or AWS DMS, for example).
Instead, I will create a simple dataflow using Kafka and Python.

Kafka is the queue where we will manage to listen to any changes in the database, which includes inserts, updates, and so on. Whenever the Write Ahead Log is updated, a message will be sent to a Kafka topic.



b. i) I've never worked with this concept.
But I did some research and find some way to deal with this in the solution purpose of adjusting the *isolation.level* parameter. Unfortunately, I would have to study more to give a full answer to this.

b. ii) With the consumer.py, we can handle that easily. If we are making transformations on a determined table and we need to order these transformations, we can create logic in the consumer.py with the proper order to execute, with that, we won't face some problems like deleting an ingredient that has a relationship with recipes.

c. i) I would propose the creation of a control script, where we can trigger a job to alarm us (via slack or e-mail, for example), to perform some data quality checks.

For example, does the number of rows in MySql match the BigQuery-referenced table?

c. ii) We should create a code based on idempotency, so, we shall have a script that will perform something like (select * from tables) and make a big insert as the first time. This should work in a batch process since is not something we will ever do.