



Design Pattern of Hippo Service

Roger

What is this?

Borrow the concept from

Microservice

????????????????

單一大架構

Step 1: get data



Step 2: process data



Step 3: dump data



Step 4: get other data



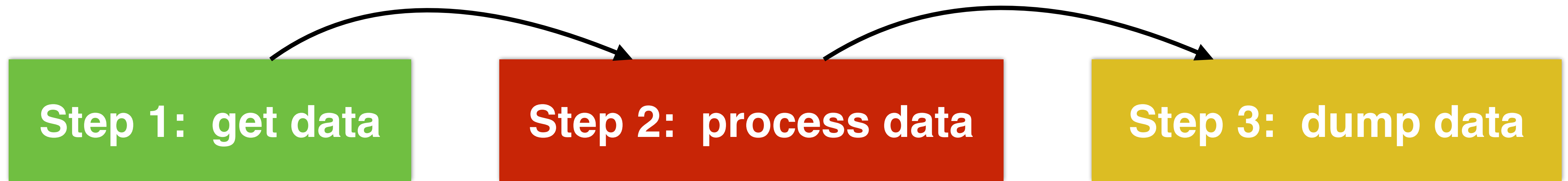
Step N:

Problem1: App太大包

Problem 2: 一步驟掛, 全掛

Problem 3: 每一步驟不能用不同技術

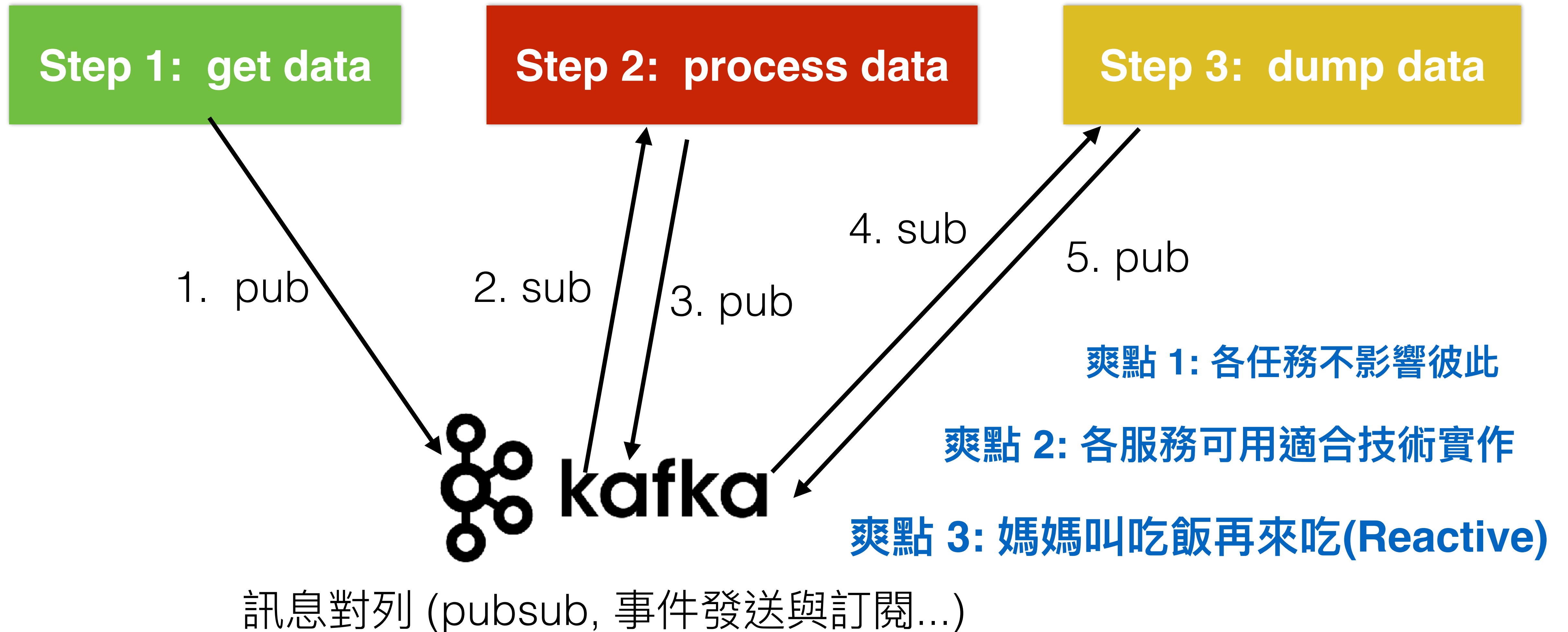
排程架構



Problem 1: 每一步驟該設幾點?

Problem 2: 一步驟掛, 還是接著跑!

微服務架構



HIPPO types by purpose

ETL

- **Frontier**
- **Batch-ETL**
- **DB-ETL**
- **Stream-ETL**

Science

- **ML Service**

Application

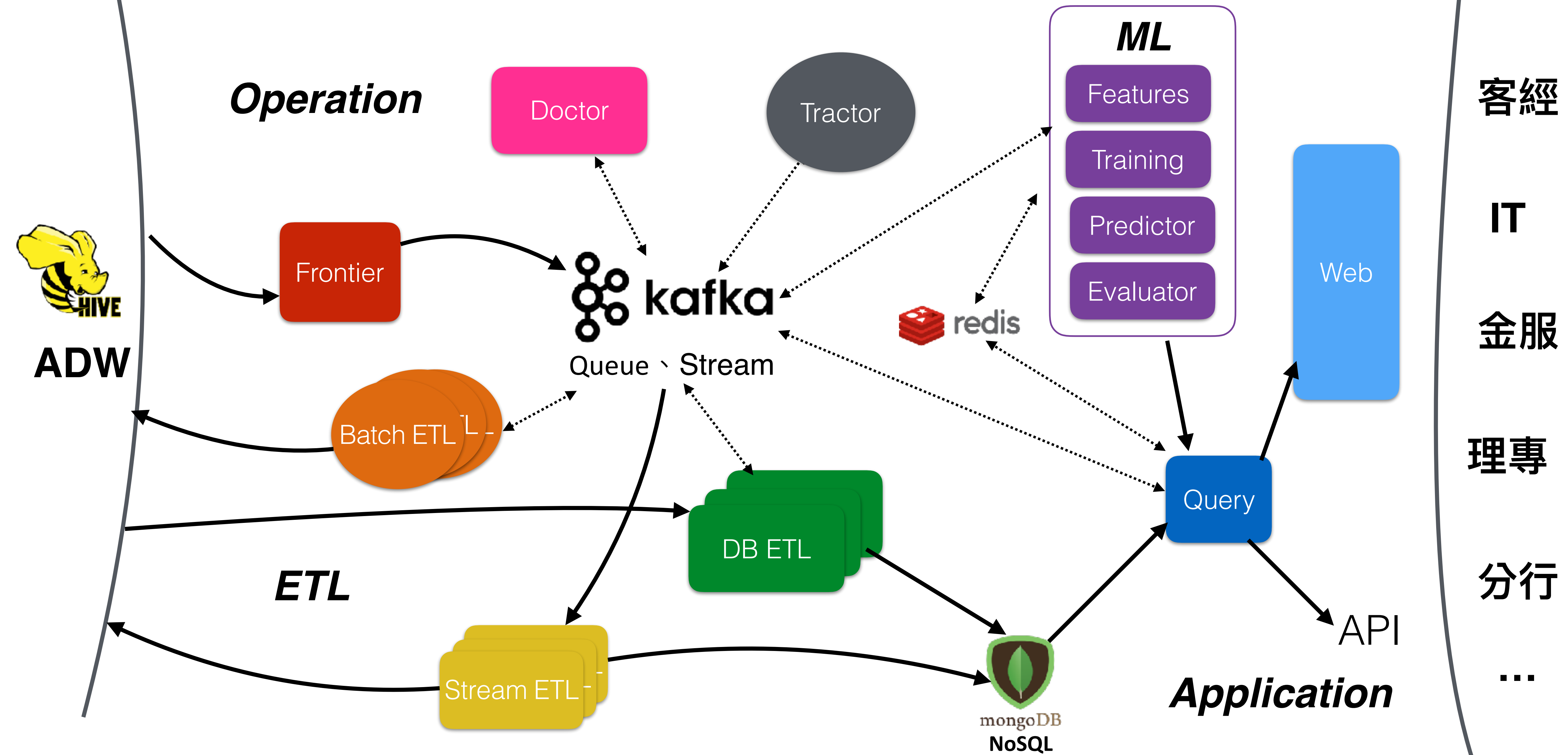
- **Query Service**
- **Web Service**

Operation

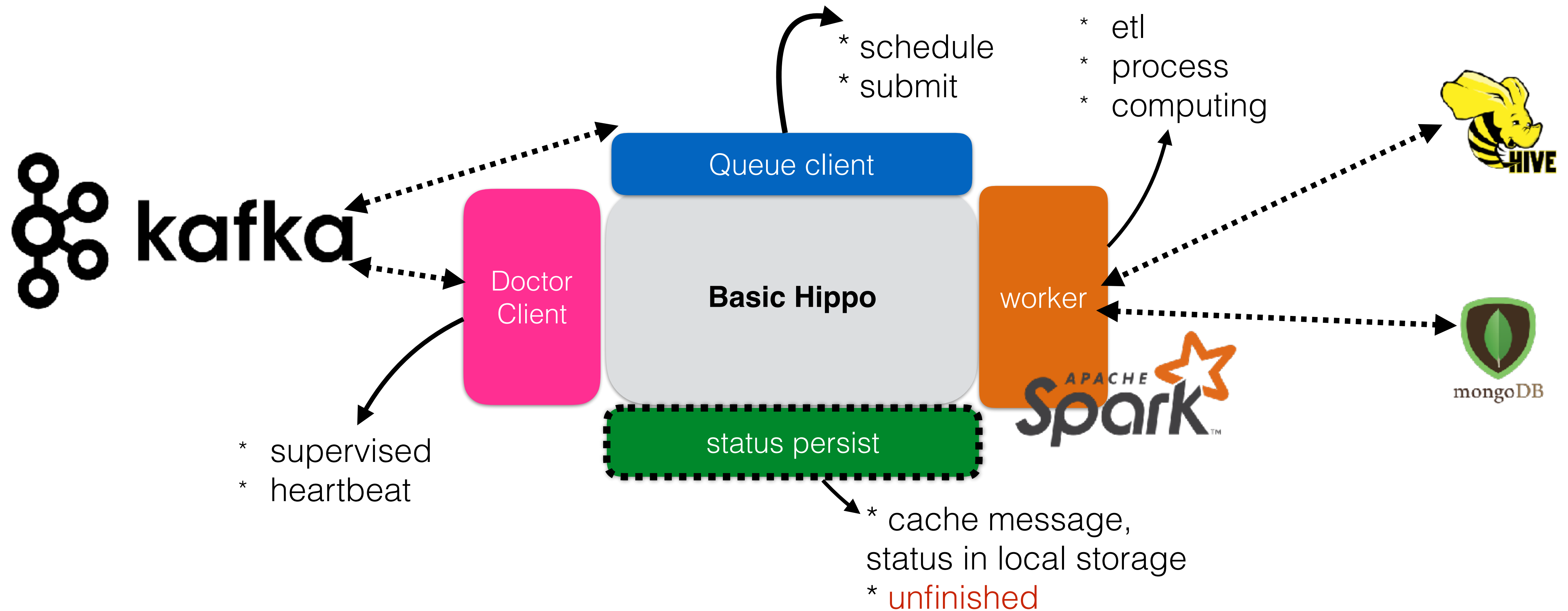
- **Doctor**
- **Tractor**

One service may comprises mixed types

Future HIPPO Views



Basic hippo structure



 python™

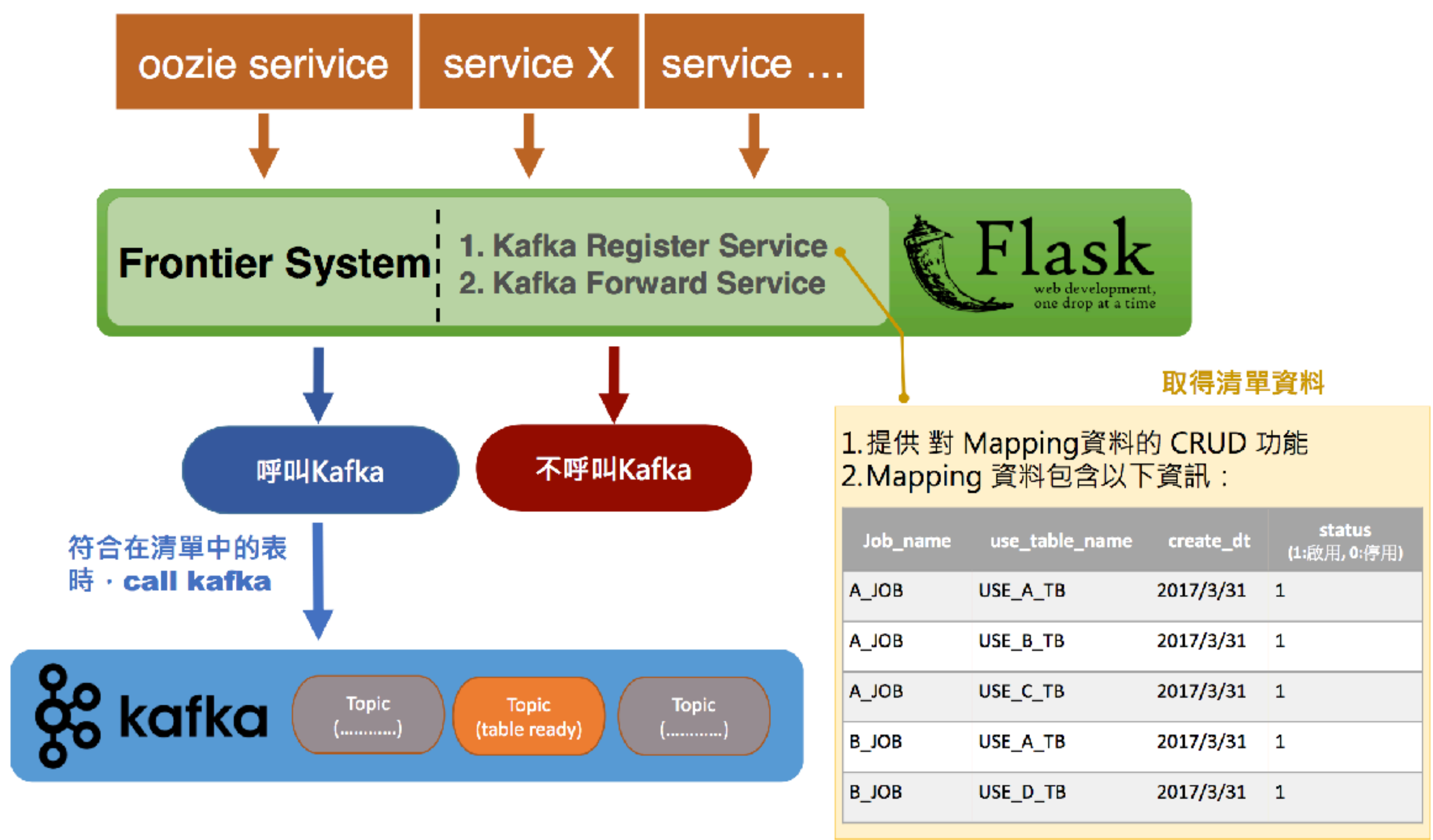
language

 **Scala**

ETL

Frontier

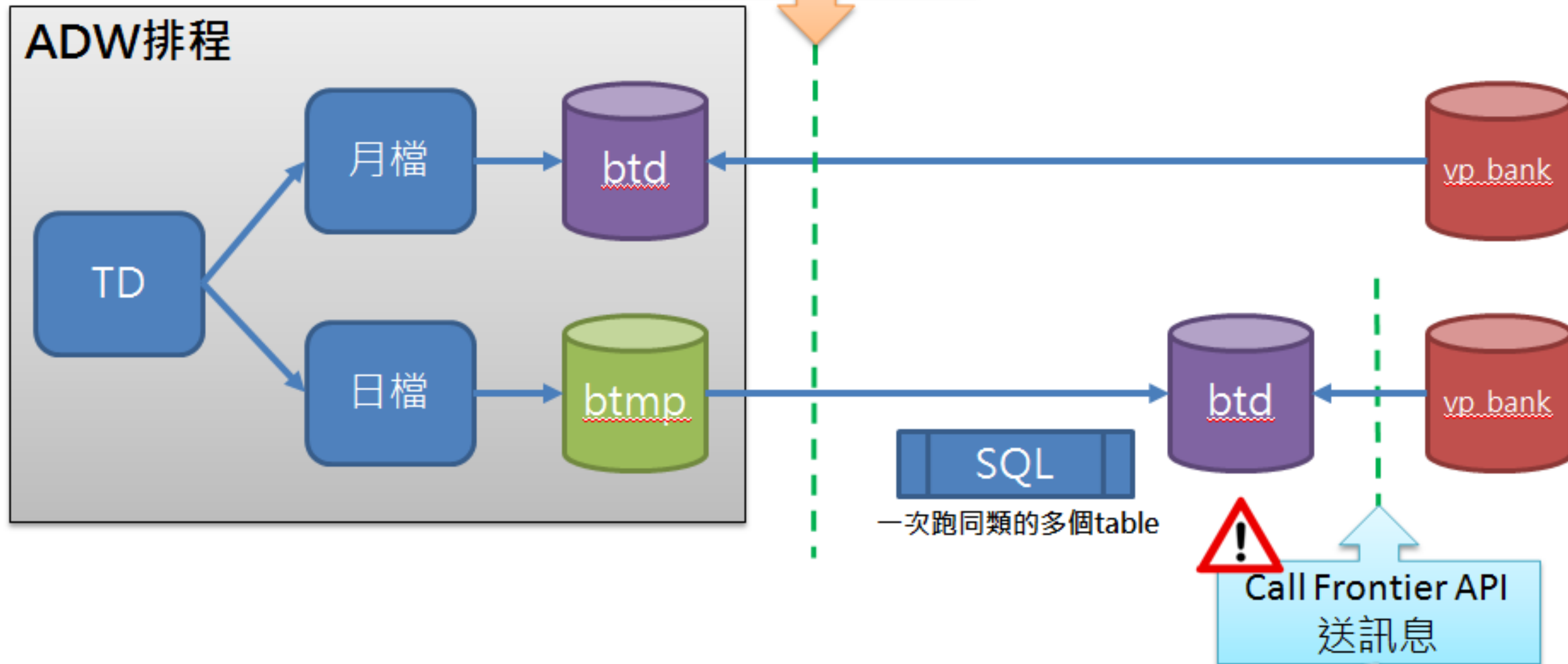
架構圖



【傳送到kafka的訊息內容】

月檔: {"table": "PARTY_TRU_ACCUM_BAL_M", "partition": [{"name": "YYYYMM", "value": "201703"}],
"db": "btd", "method": "overwrite_partition", "exec_date": "1493394946"}
日檔: {"table": "PARTY_EDW_CUST_INS_SALES", "partition": [{"name": "", "value": ""}], "db":
"btmp", "method": "partition", "exec_date": "1493394946"}

Call Frontier API
送訊息



【傳送到kafka的訊息內容】

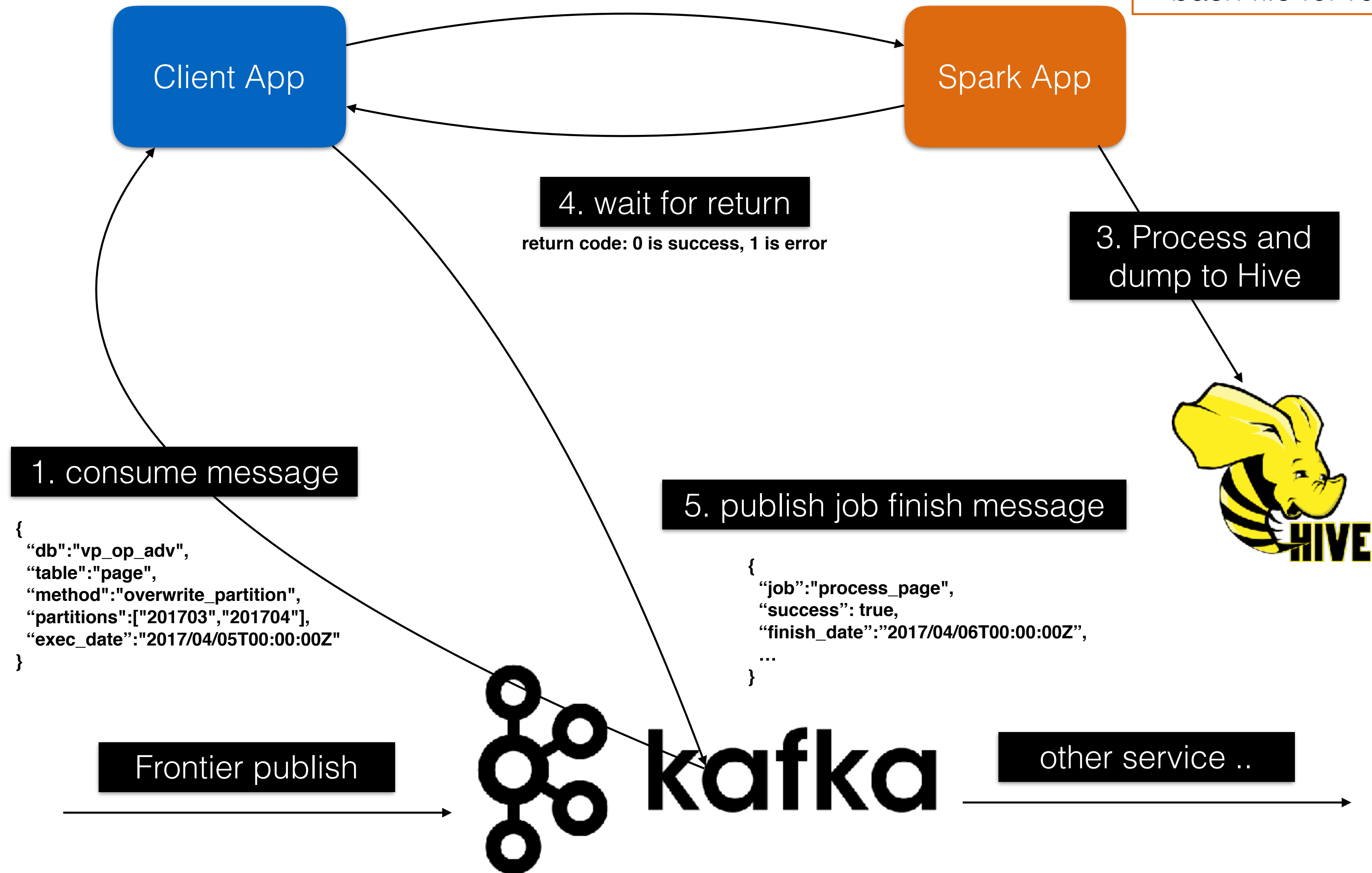
日檔: {"table": "PARTY_EDW_CUST_INS_SALES", "partition": [{"name": "", "value": ""}],
"db": "btd", "method": "partition", "exec_date": "1493394946"}

Batch-ETL

- * Kafka client
- * Process lib
- * Akka or Thread

2. spark-submit
service call "bash" or "command"

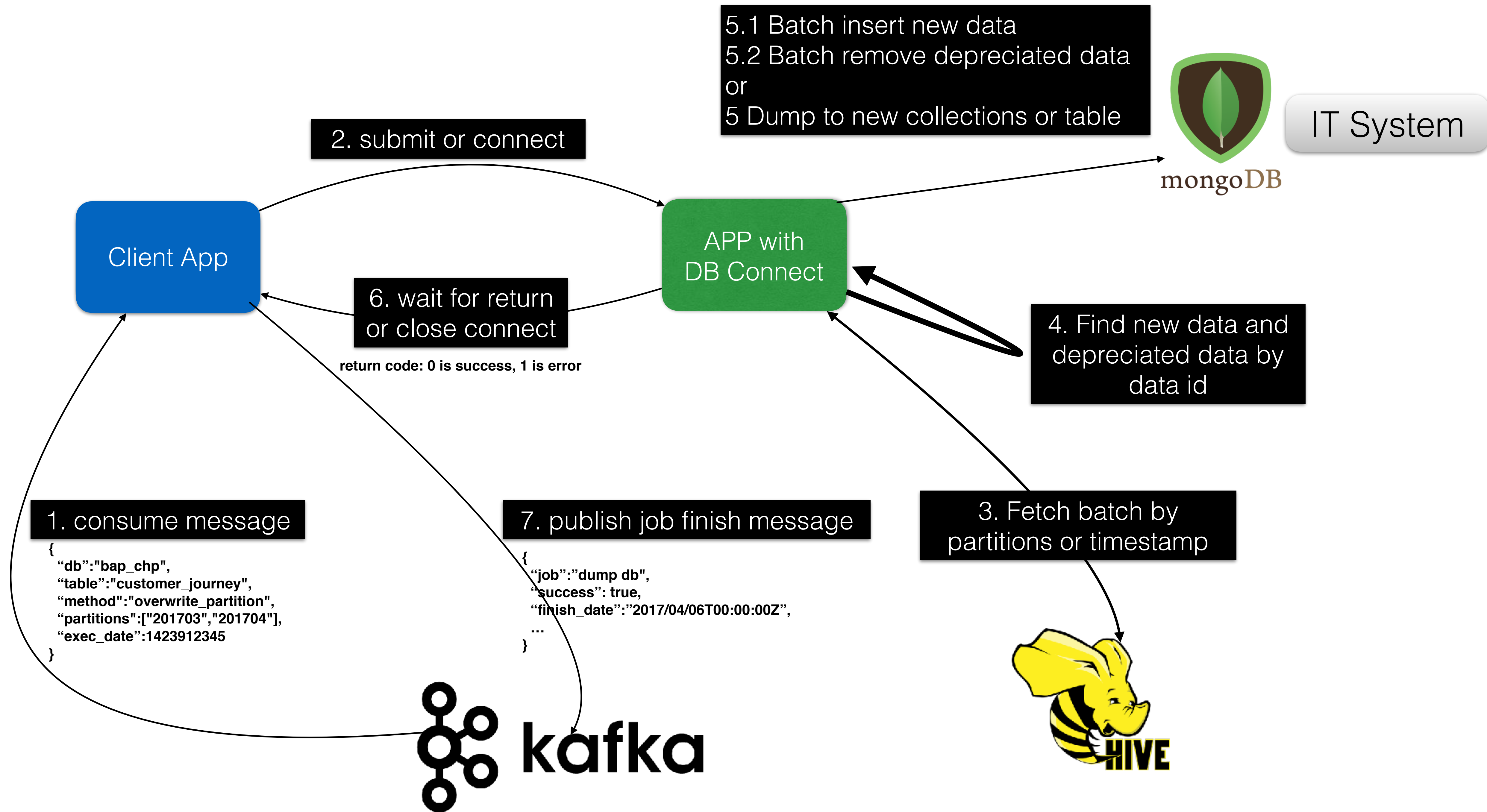
- * Yarn-client
- * Hive SQL
- * jar or py
- * bash file for running app



Structure

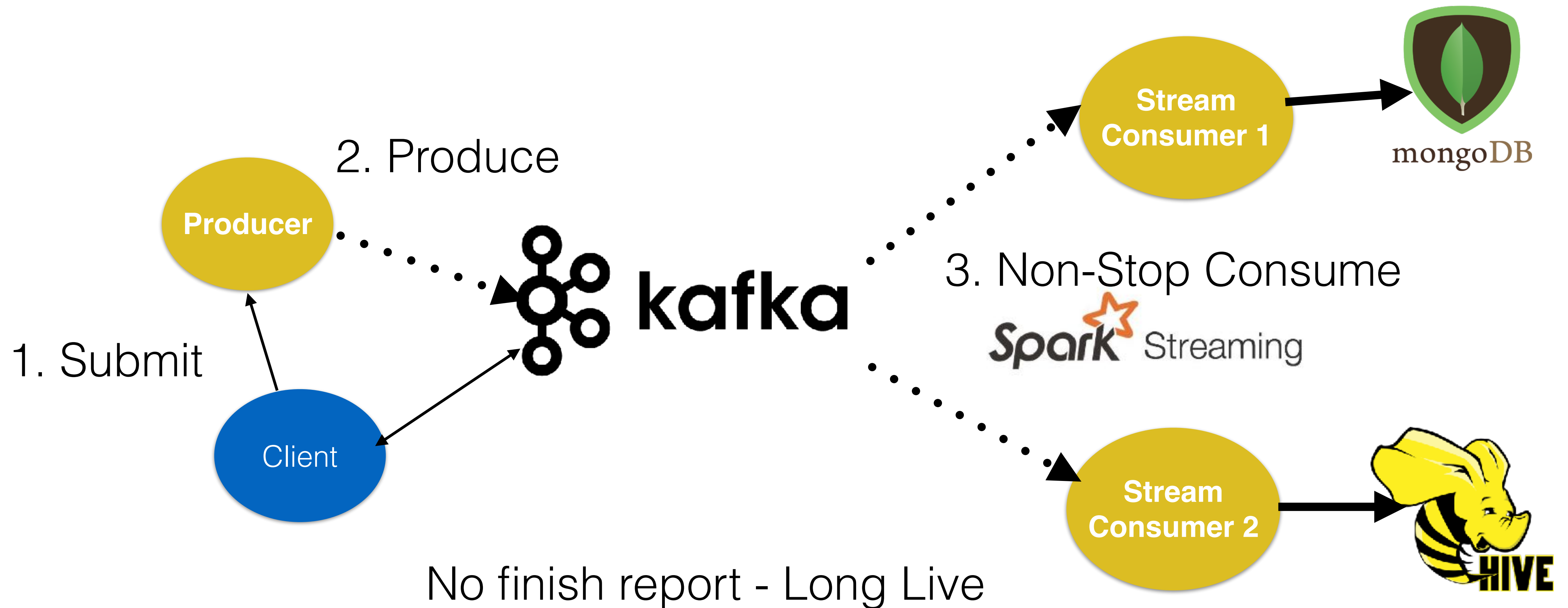
- start-client.sh
- start-spark.sh
- client app folder
- spark app folder

DB-ETL



Stream-ETL

Same source, but different consume purposes



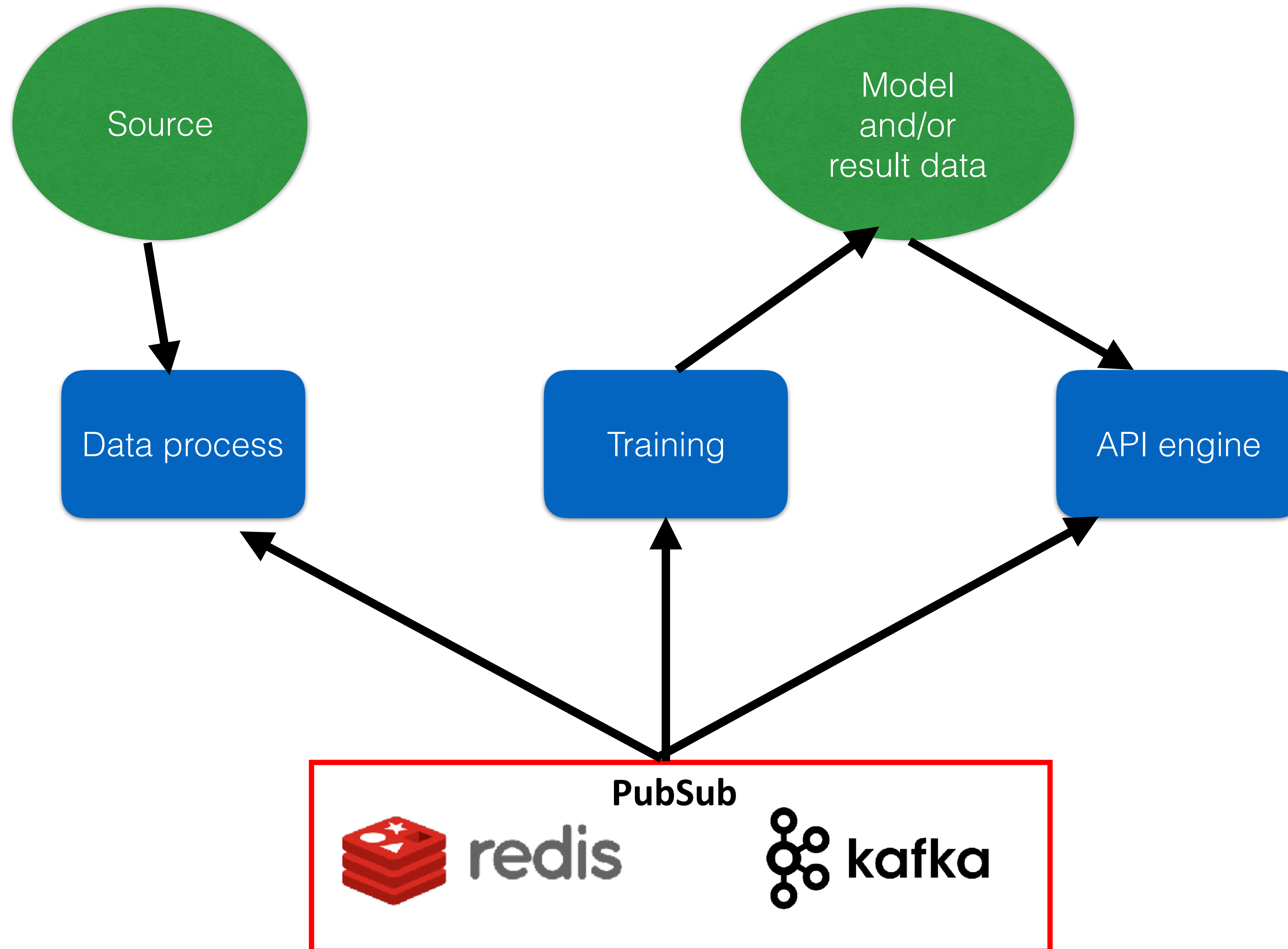
Neha Narkhede, 2016

ETL Is Dead, Long Live Streams: real-time streams w/ Apache Kafka

https://www.youtube.com/watch?v=I32hmY4diFY&list=PLVeYbWw30yOKWUZOrLqeVjWf_Vfvy3AGf

Science

ML Service

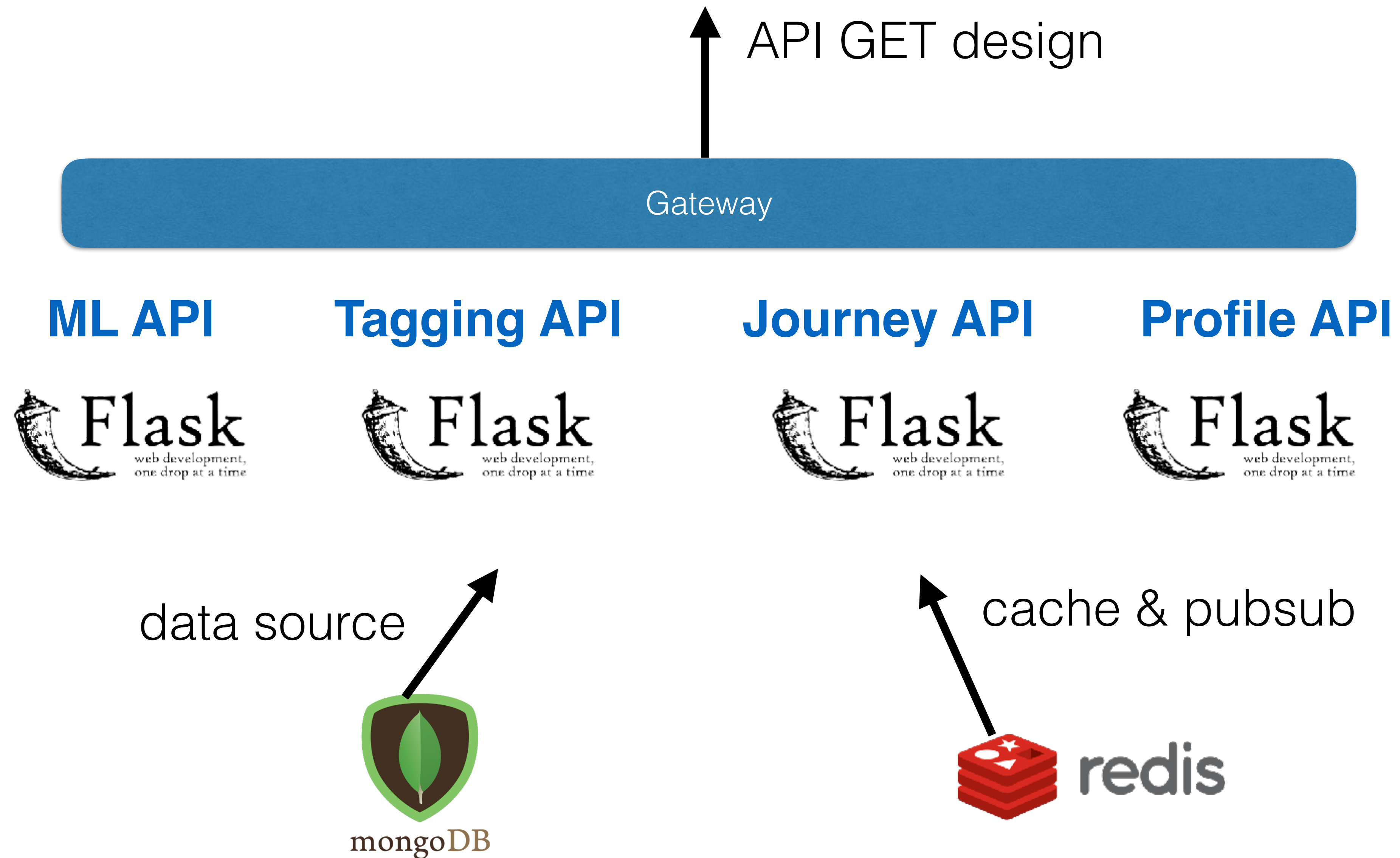


Repository

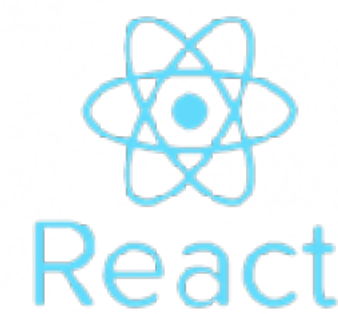
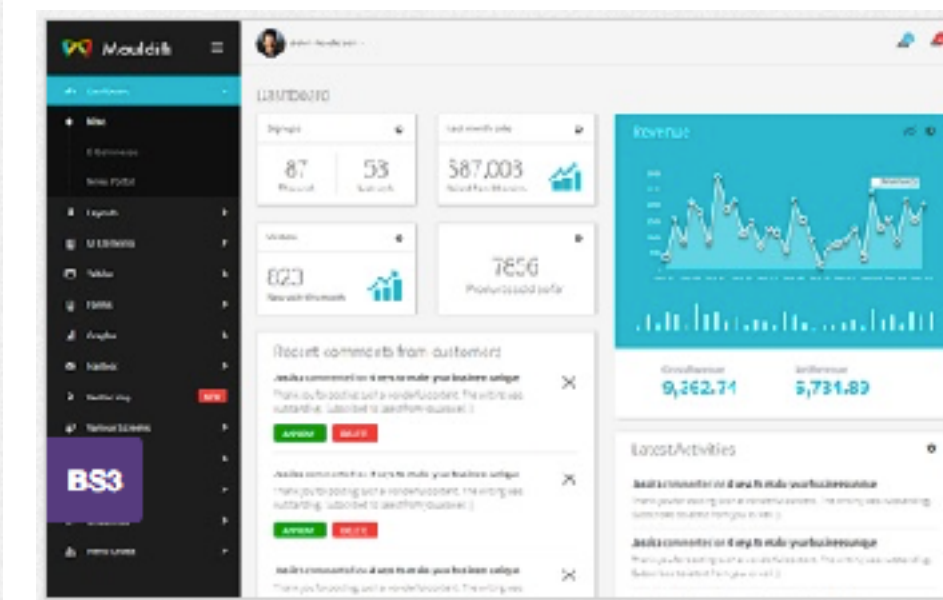
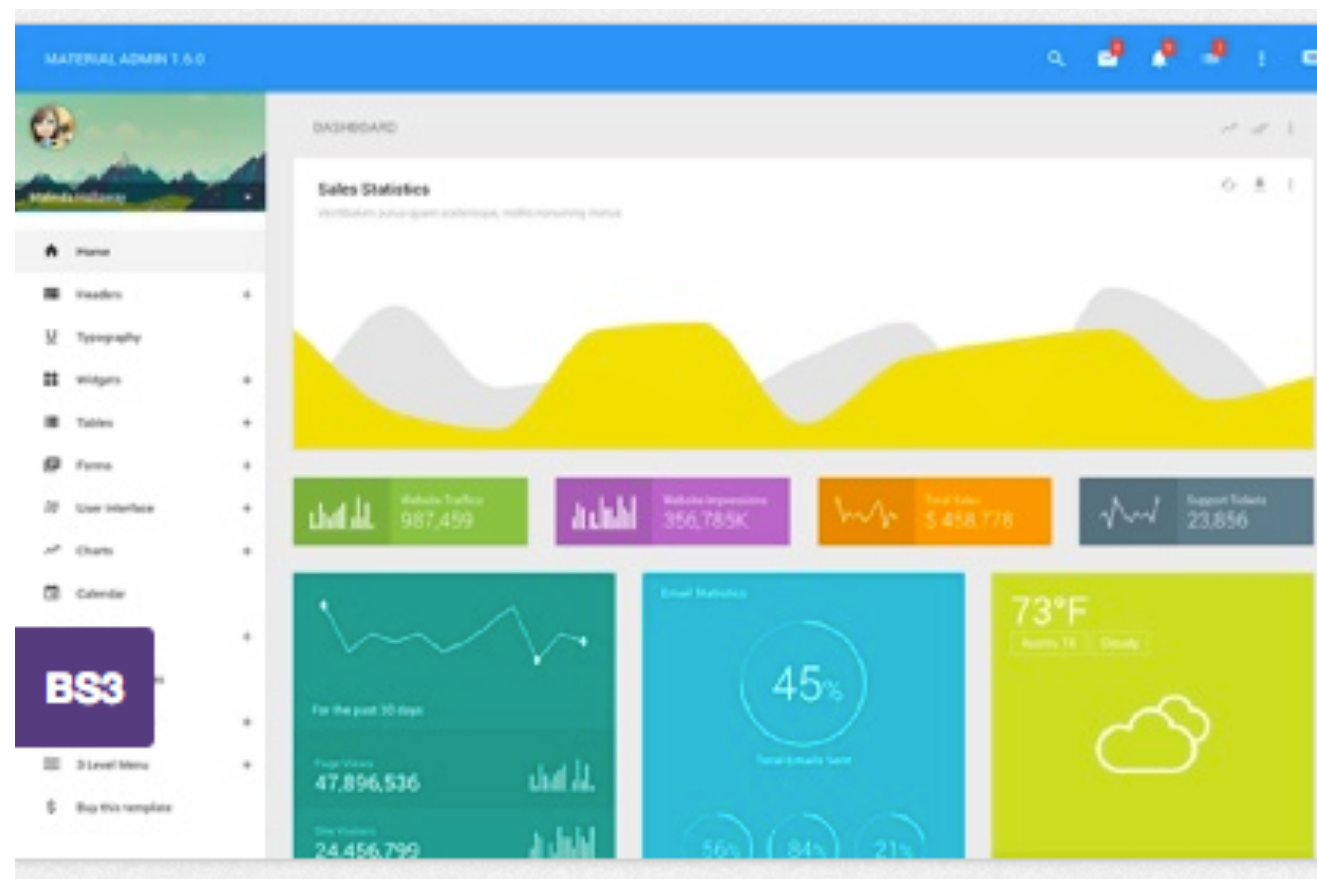
github.com/b967050008/MLServices

Application

Query Service



Web Service

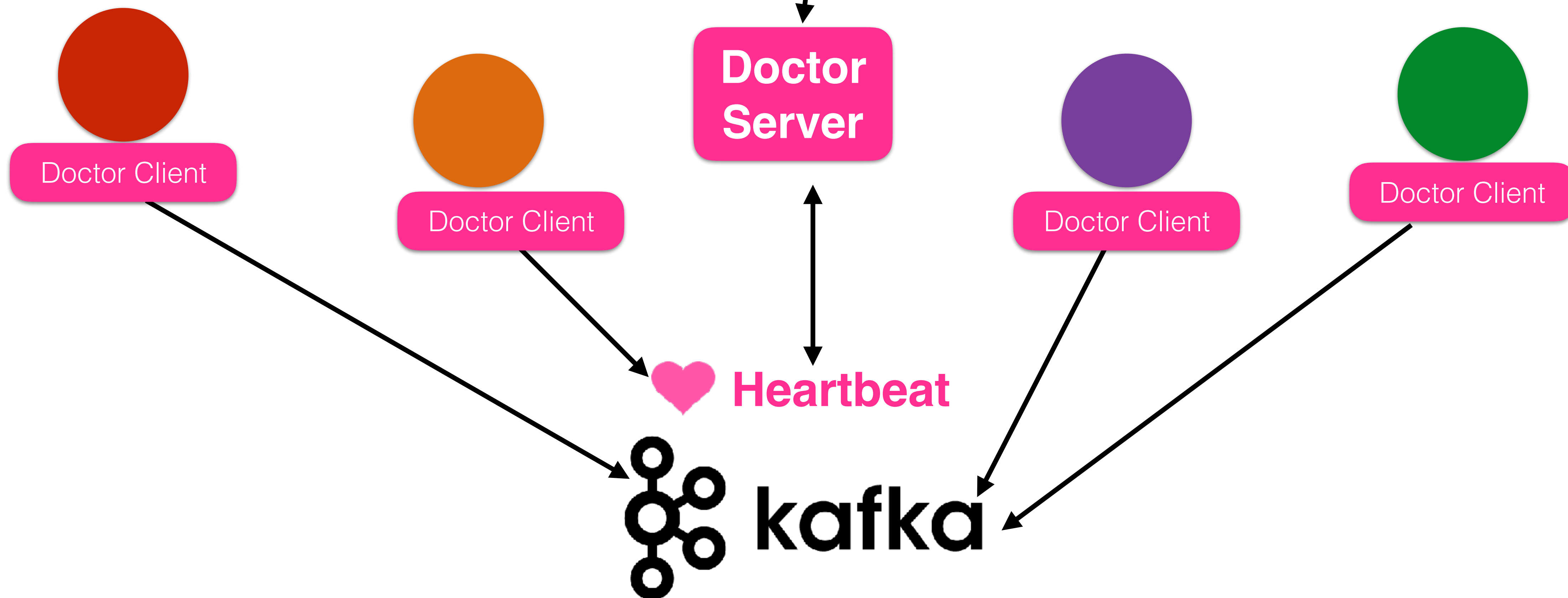
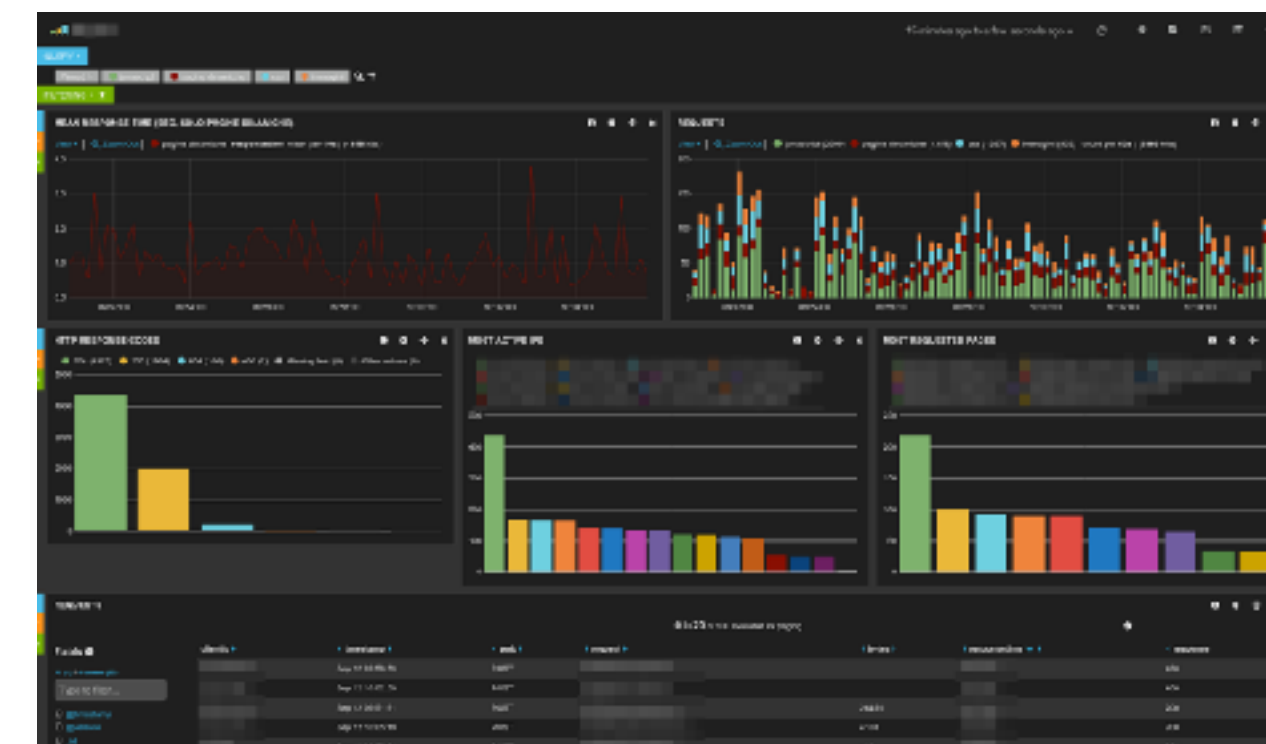


Need help from Mandy?

Operation

Doctor

Scheduler
Status persistence
Supervise



Tractor



kafka

topic: hippo-finish



save message



mongoDB

track hippo
mission status

