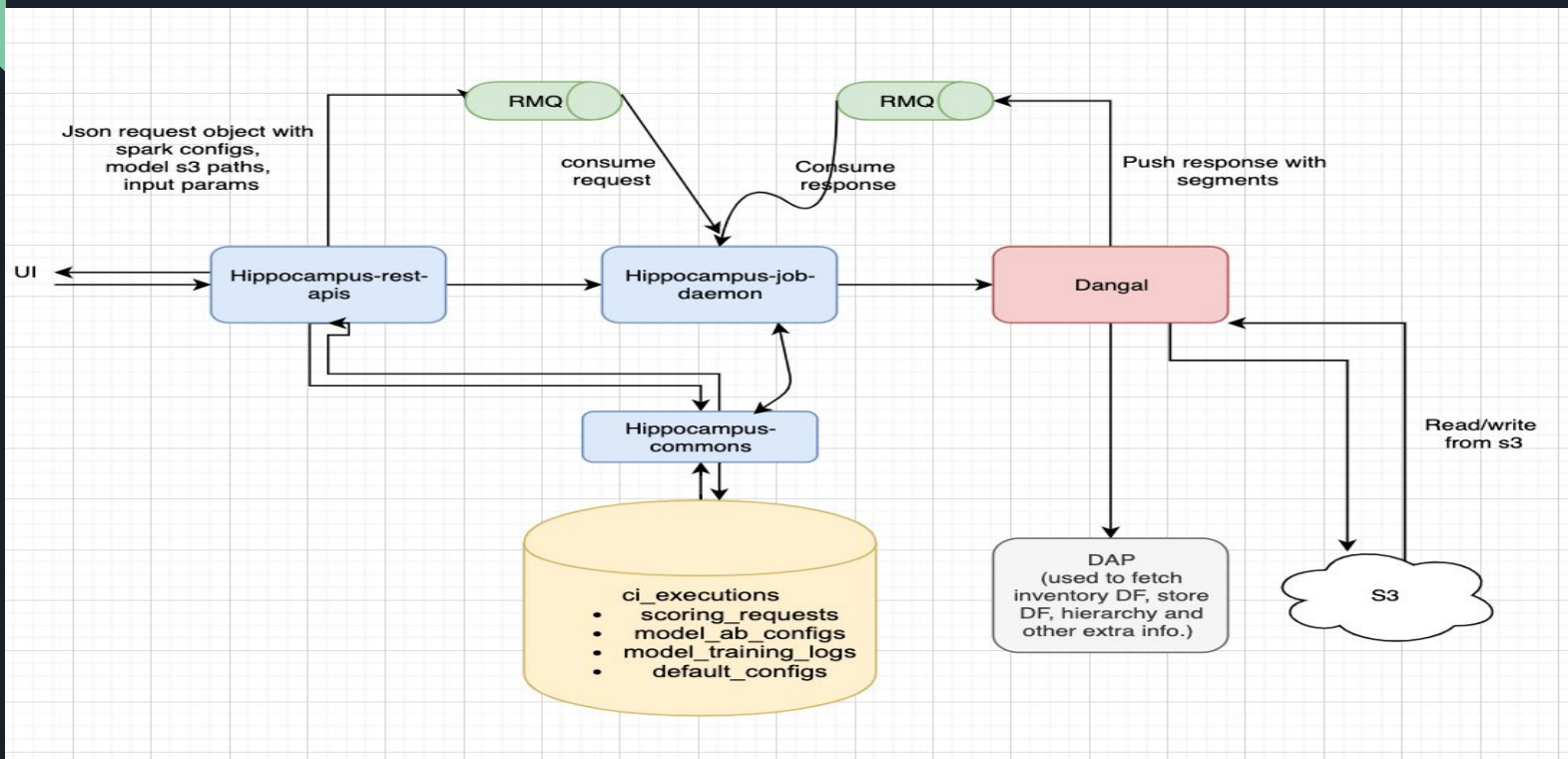




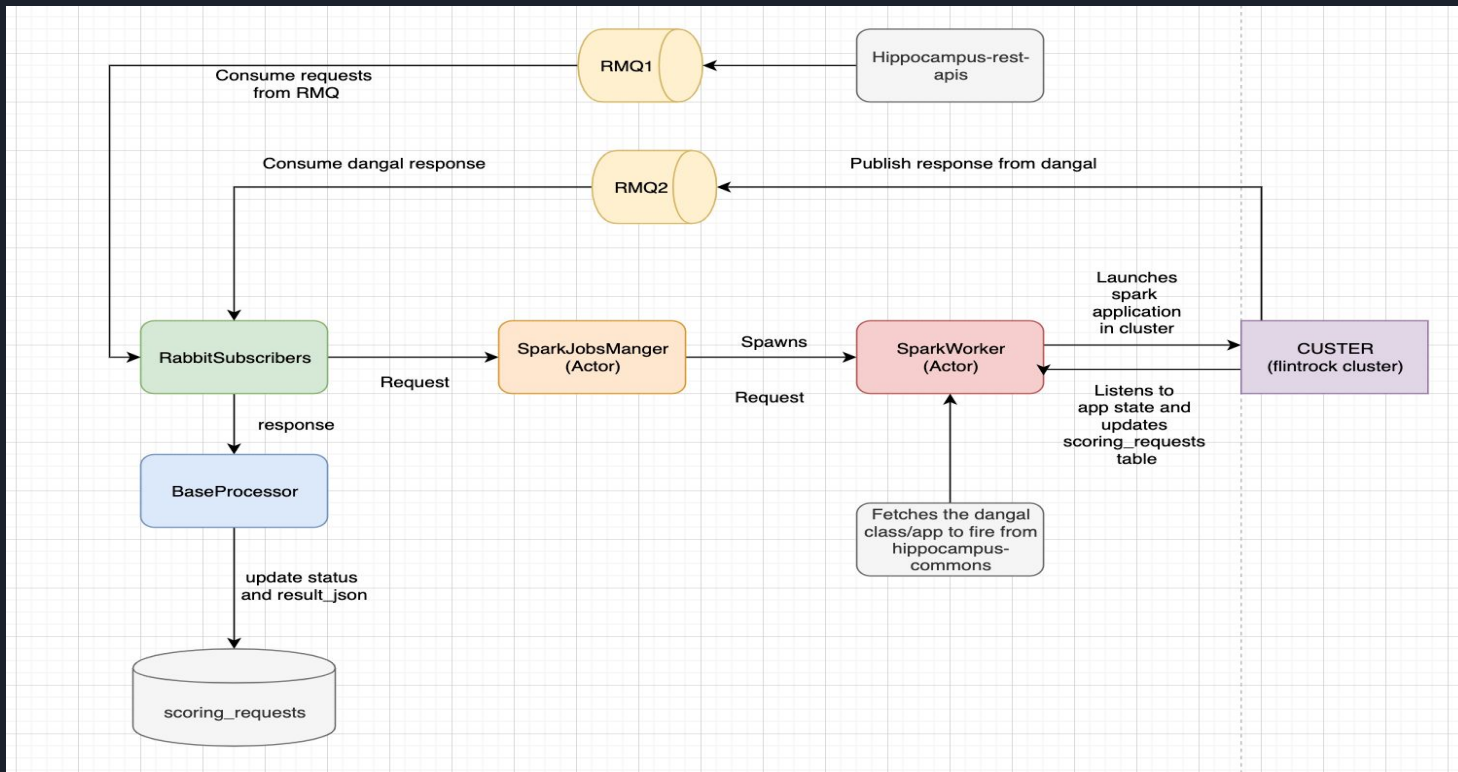
SPARK SQL INFERENCE

Data Science

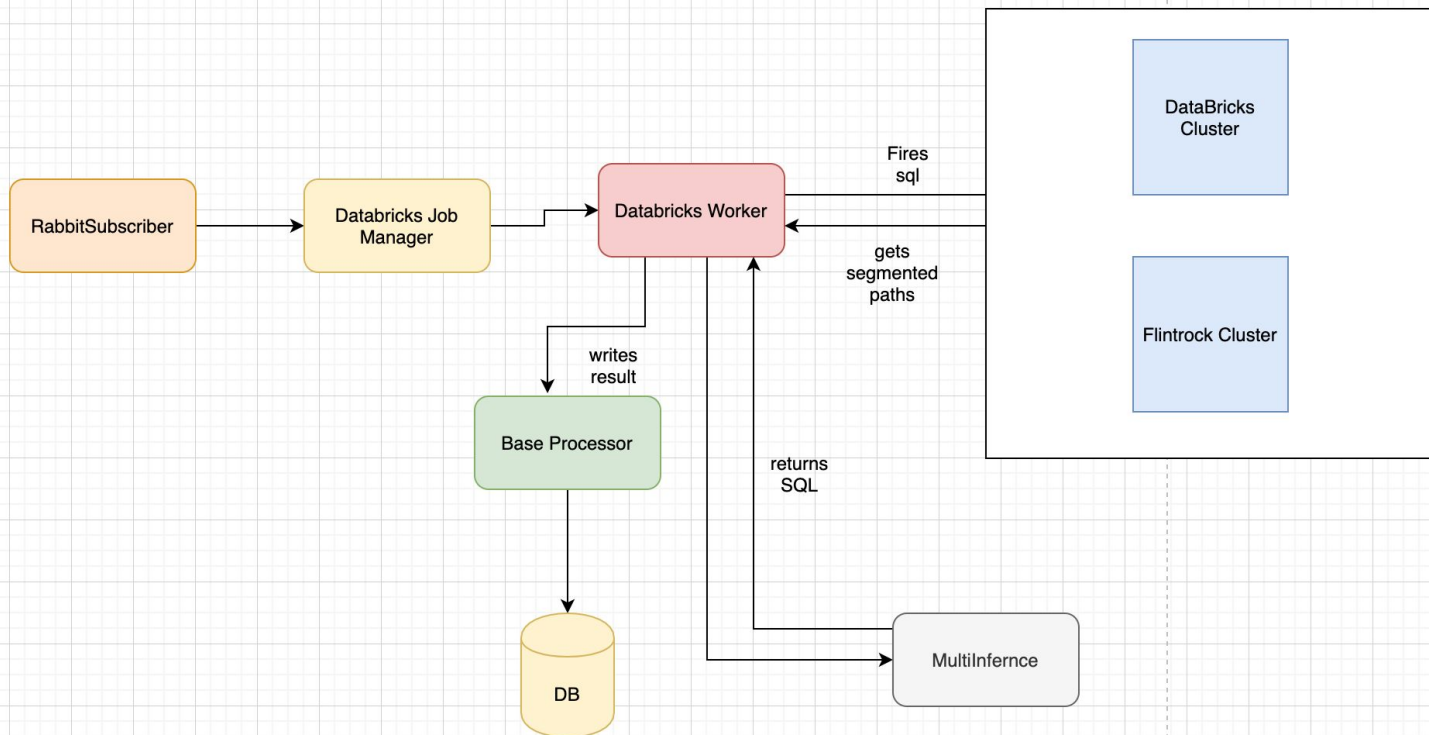
Old flow



Old Flow



New Flow



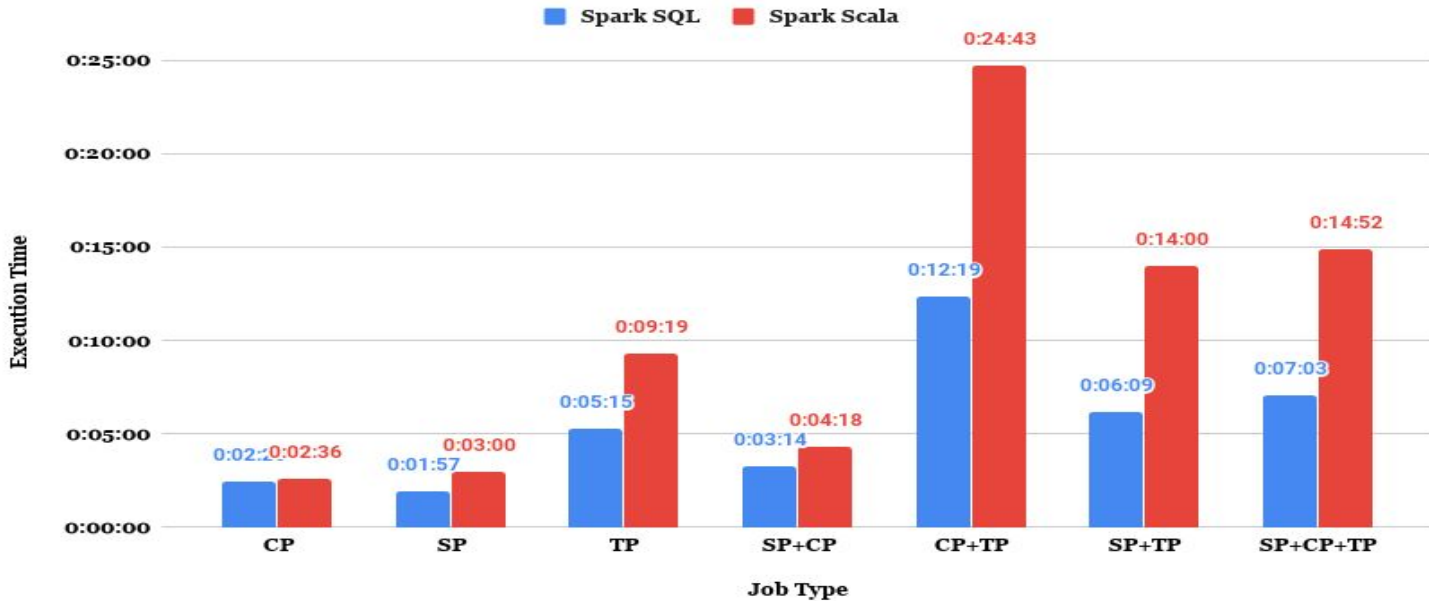


New Design

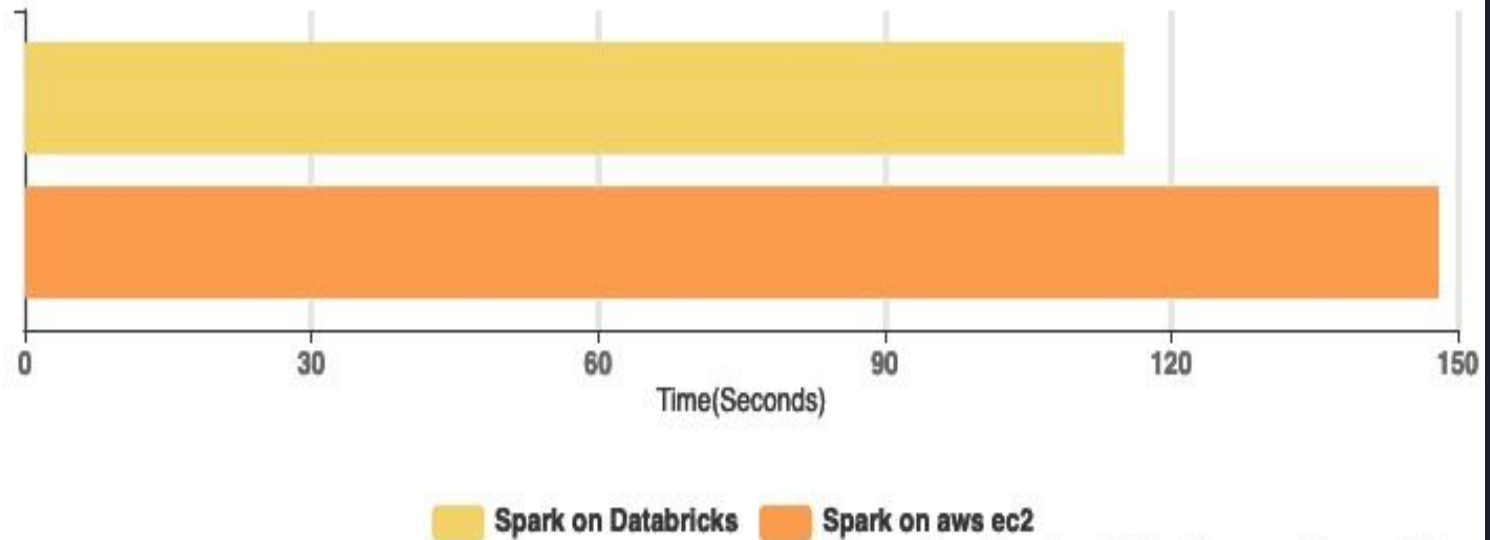
- Ingest request from different queue
- Whole sql is created for the combination by multi inference class
- Hippocampus-job-daemon creates the sql and fires on databricks cluster
- The response is collected as usual and written in `scoring_requests` DB.

Comparison

Performance - Spark SQL & Spark Scala



Dangal - Databricks Run time VS aws ec2 Run time





Comparison

- Region: Nightly
- Org id: 50241
- Userbase: 407715
- For Multi-inference at most 60% improvement and minimum 25% time improvement.
- For single inference we see the slight improvement in CP, 25% improvement in SP and almost 50% improvement in TP.
- On average, spark on Databricks performed 20-25% better than spark on AWS.

Thank you...

