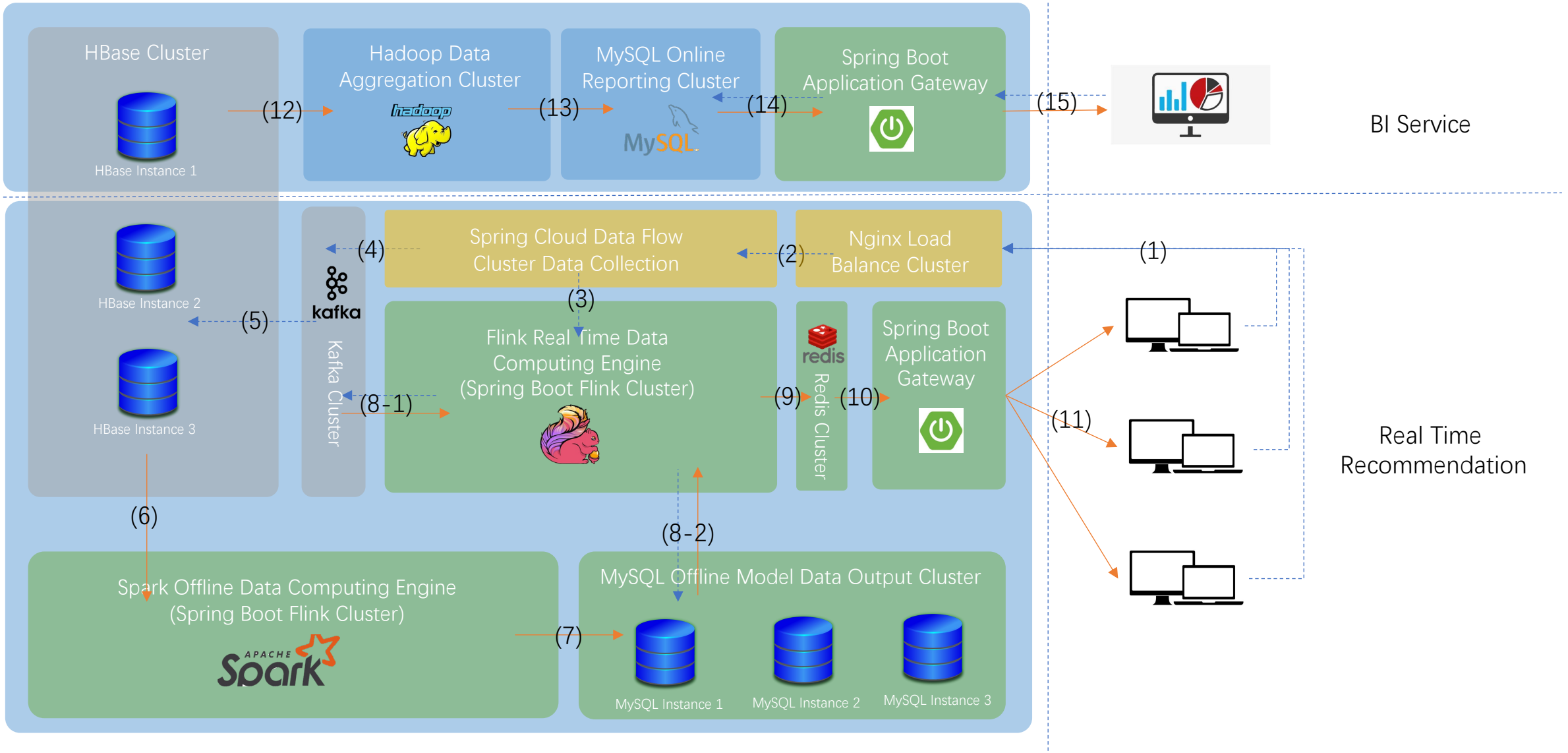
Real-time Recommendation System (RRS) & BI Platform Data Flow Chart



RRS Feature & Advantages

1. Millisecond-level response
2. Visit lifecycle based real-time computing
3. High concurrency, high tolerance
4. TB-level Data Storage
5. Customized Online Modelling Engine Base on TensorFlow
6. Real time computing allow some computing latency. But responding data to user should be fast as possible.
7. Usually within 15s after a user action a time, a result should be calculated out and presented to user when a user does something next time.
8. Scalable

Data Flow Description

1. Use datacloudanalytics.js to collect user behavior data from online shop site. Then send data to Nginx Load Balance Cluster(NLBC) for data request forwarding
2. Nginx Cluster(NLBC) will send off equivalently to each instance in Spring Cloud Data Flow Cluster(SCDFC).

3. Data Flow Cluster(SCDFC) is clustered of spring boot application instance, for matching data from history data model result. It first send data off to Flink Real-time Data Computing Engine(FRDCE) ready for low cost computing and low latency response to user end.
4. Data Flow Cluster(SCDFC) will then send data off to Kafka Cluster(KC)
5. KC will then send data asynchronously to Hbase for history data storage.
6. Spring Offline Data Computing Engine (SODCE) will extract data from Hbase to modelling on several dimensions including but not only users, products, events, consumer amount, which is clustered of Spring boot Application Instances.
7. SODCE will send off post process modelling result data to MySQL Offline Modelling Data Output Cluster for data storage and query request from Flink (FRDCE), usually this part of data should only be computed on once a day and should be during a low traffic period, due to system risk avoiding policy.
8. Computing Process has two steps not in order:
    1. Flink Cluster(FRDCE) consumes the history data within kafka data remaining duration e.g. 72 days for user behavior correlation analysis and product analytics for getting a recommending result.
    2. Flink Cluster(FRDCE) consumes the modeling data which define user group, user feature, user’ s latest attention points and are related to relevant products to score the product and get a top ranking product list for presenting user end.
9. After result calculated out, Flink Cluster will send result to redis for high concurrency query
10. When user get next page or js proactively request result data with a short interval, Spring Boot Application Gateway(SBA) will fetch the result data and respond to user of low latency.