

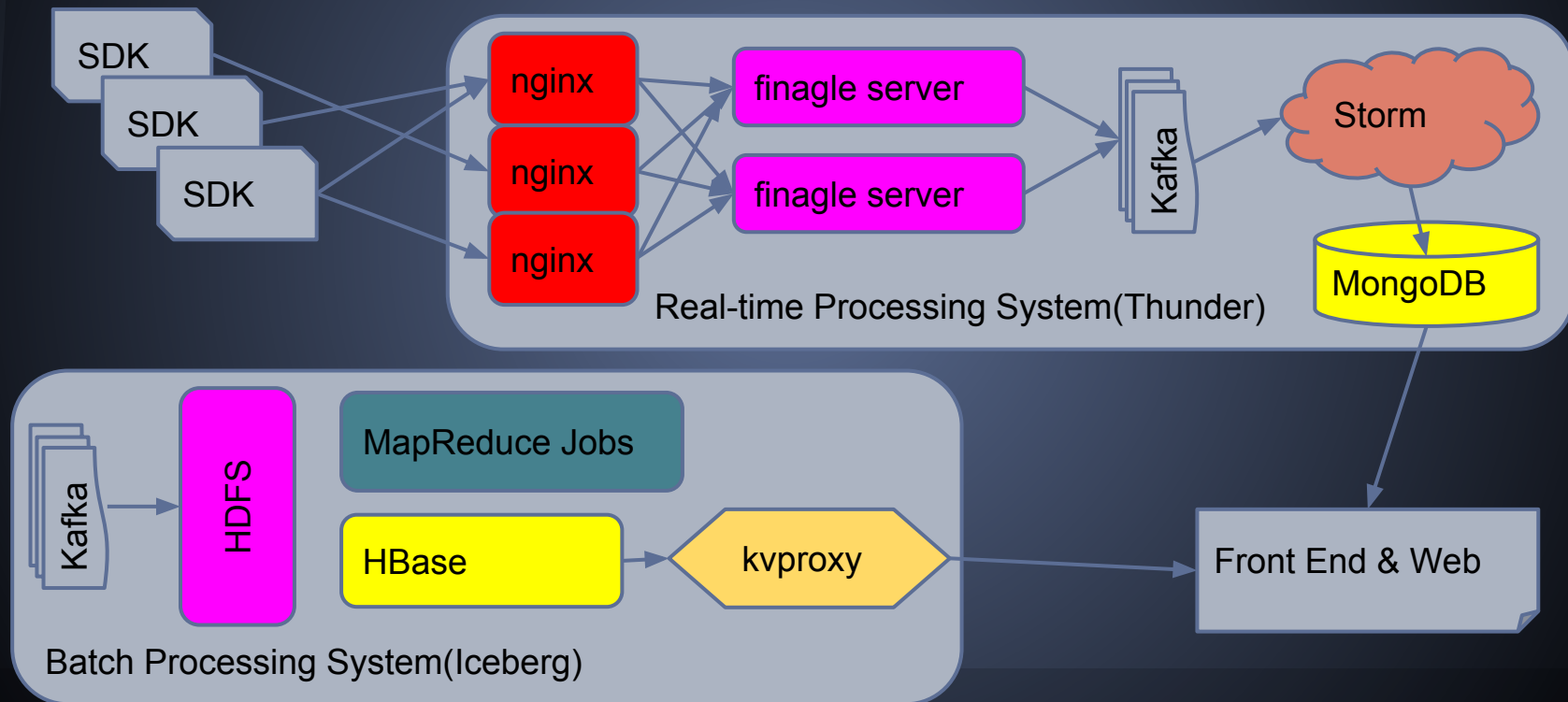
Umeng Analytical Arch

Zhang Yan 2014.10

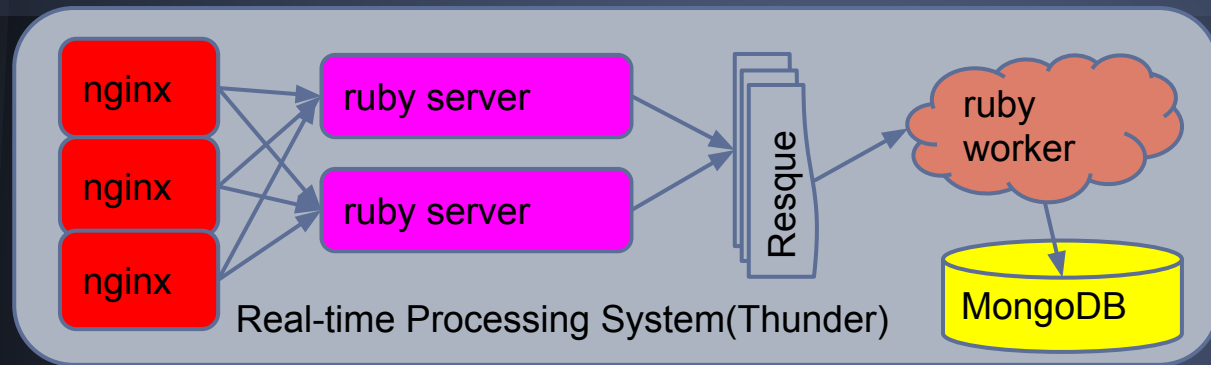
Outlines

- how are we doing
- real-time system
- batch system
- lambda arch
- future & challenge

How are we doing

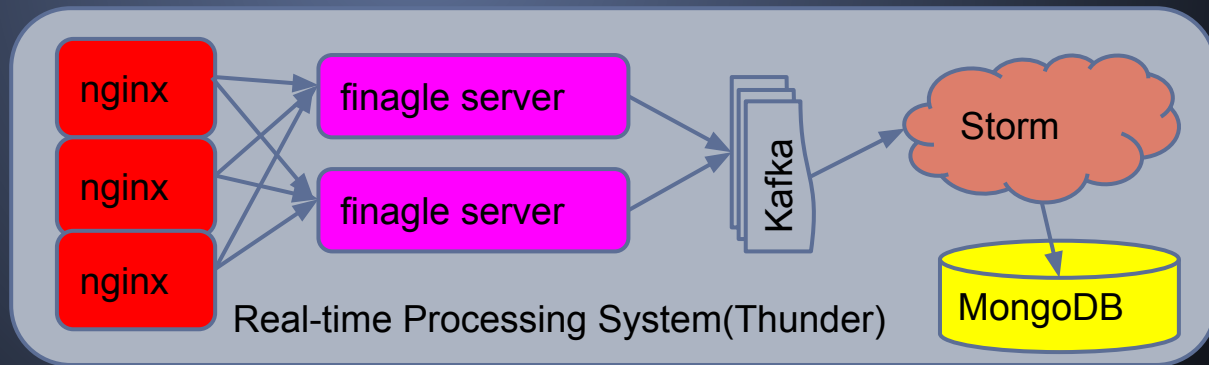


Real-time system



Evolve ...

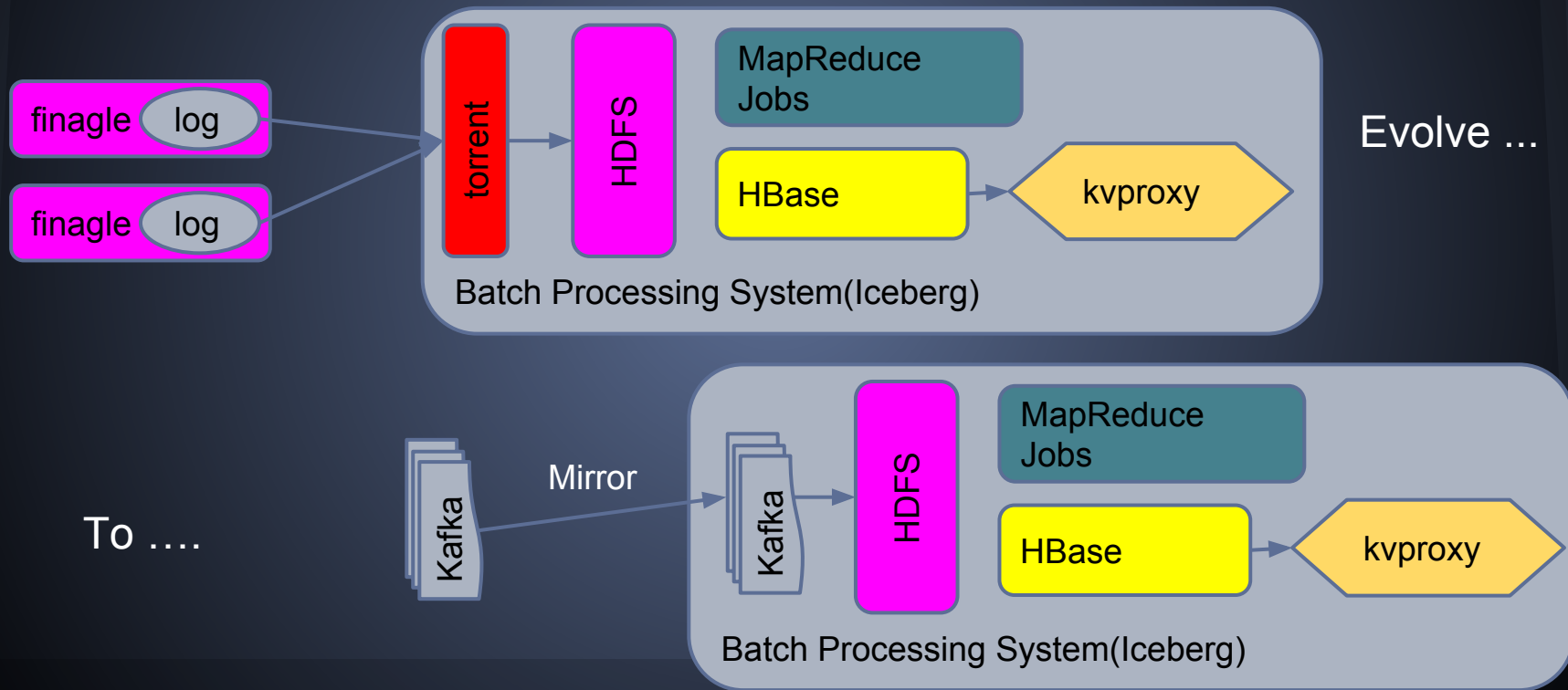
To



Real-time system(cont.)

- technology we use
 - java, scala, ruby
 - nginx, resque, mongodb
 - finagle, kafka, storm
- challenge we face
 - data skew
 - mongodb
 - global lock is bad.
 - table lock, but still good not enough.

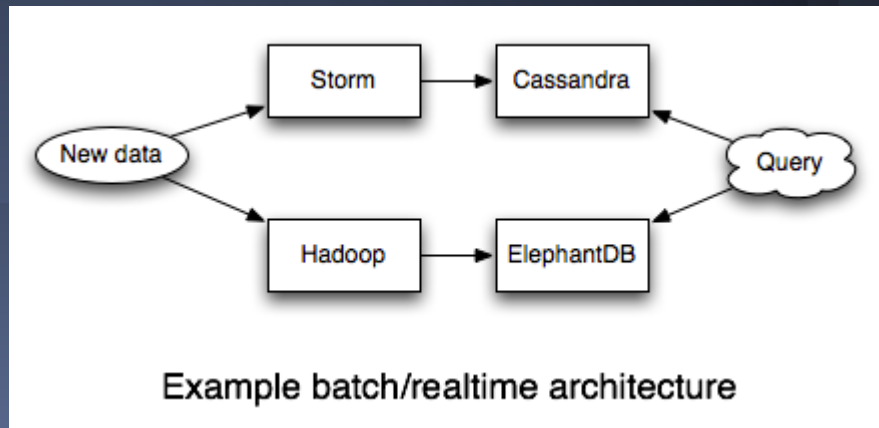
Batch system



Batch system(cont.)

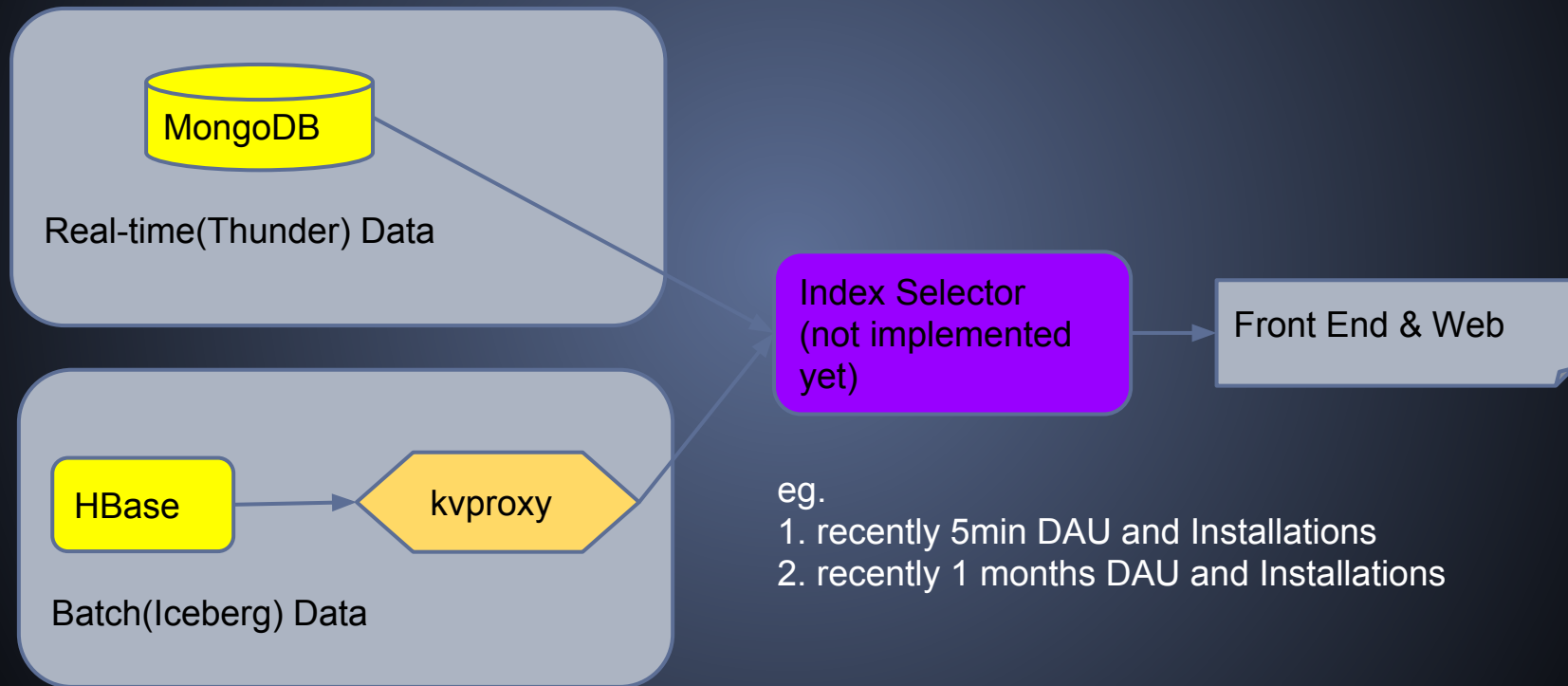
- technology we use
 - java, scala, python, c/c++
 - hadoop, hdfs, hbase, pig, hive, opentsdb
- challenge we face
 - data processing in a uniform way
 - optimize hbase usage(hash prefixed key, bulk load)
 - big amount of precomputation(HyperLogLog)
 - debugging, deployment, measurement etc.

Lambda arch



- merge real-time and batch
- complete each other
- real-time
 - fast, imprecise, narrow scope
- batch
 - slow, accurate, wide scope

Lambda Arch(cont.)



Future & Challenge

- just-in-time computation
 - druid, impala, spark etc.
- unify real-time & batch processing in one piece of code.
- data availability(internal & external)