

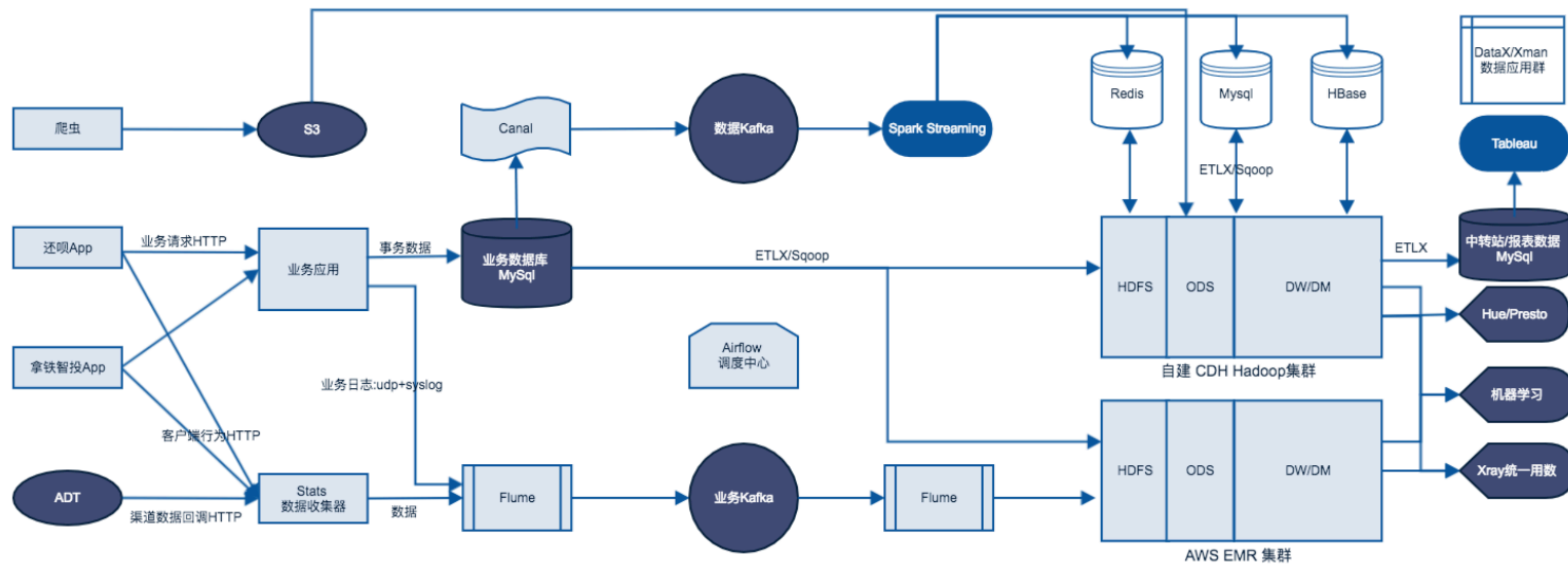
LatteData

AWS宁夏迁移方案

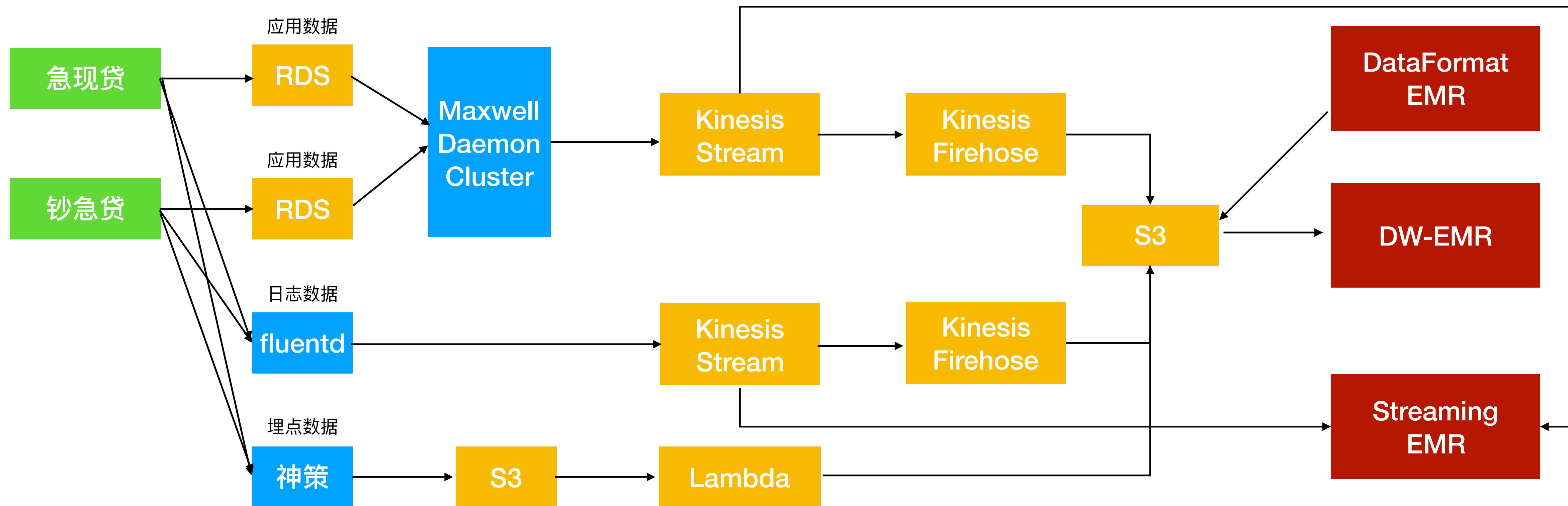
大数据部

现状

LatteData数据全流程



急现贷&钞急贷 - 实验



迁移前后

cn-north

cn-northwest

Airflow

统一用数

跳板机
Hue+Presto

北斗&
驾驶舱

Airflow

跳板机
Hue+Presto

统一用数

北斗&
驾驶舱

RDS

Kafka

CDH

JobEMR

Streaming
EMR

EtIxEMR

eTunnel
EMR

HDFS

S3

Mysql

HBase

eFuel

xman

CallBox

Tableau

eTunnel

RDS

Kinesis

DW-EMR

JobEMR

Streaming
EMR

EtIxEMR

eTunnel
EMR

S3

Mysql

HBase

eFuel

xman

CallBox

Tableau

eTunnel

迁移步骤

	迁移对象	迁移任务描述	实施部门	依赖项目
1	RDS	还呗核心业务数据库	运维部	需要数仓圈定RDS范围
2	StreamingEMR	基于RDS binlog的streaming应用迁移：电销、酷屏	大数据部	
3	Openshift环境	提供pod部署环境，CI/CD环境	运维部	
4	摩卡应用集群	摩卡前后端、账号体系、应用管理体系	技术保障部	3
5	ETLX	在宁夏区部署ETLX应用	大数据部	4
6	Airflow	在宁夏区部署Airflow	大数据部	
7	ETLX作业	基于新的RDS，配置全量的ETLX作业，验证ETLX全量抽数据的实效	大数据部	要区分哪些表历史上有过删除，然后评估是否需要将历史数据搬移过来
8	部分数仓以及作业	挑选某个数仓库，比如dbank.db, 通过distcp将数据同步到宁夏EMR，并配置数仓作业做数据持续运算	大数据部	验证数据仓数据完整性以及作业执行效率
9	全量数仓以及作业			
10	统一用数			
11	实时流			

迁移前需要弄清楚

- 宁夏区具体的价格优势是什么？能持续多久？
- 下面两个方案那个费用更低？
 - 宁夏区：EC2+EMR服务费 + S3 + 跨Region的网络费用
 - 北京区：纯EC2（去掉SSD）+S3+部分EMR