



数禾应用分析

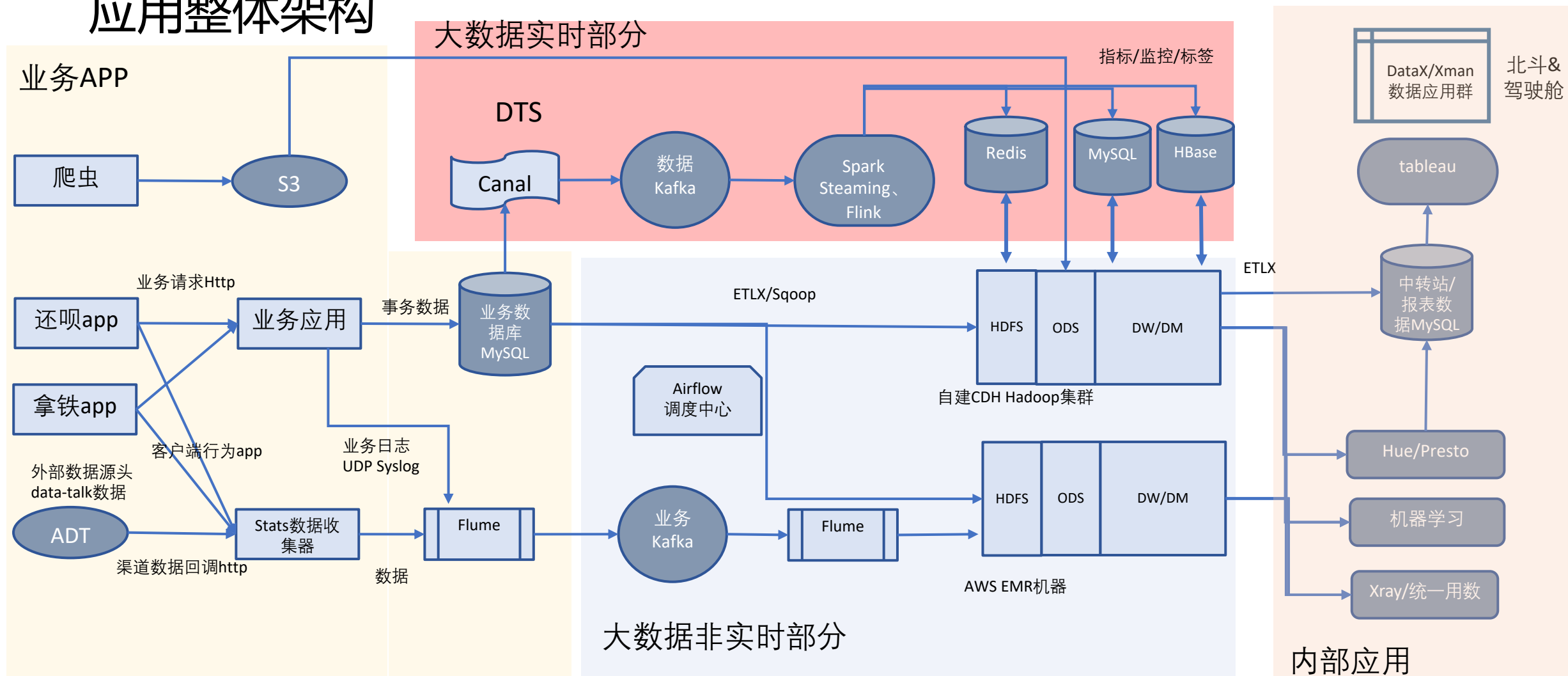
SA.尹广东



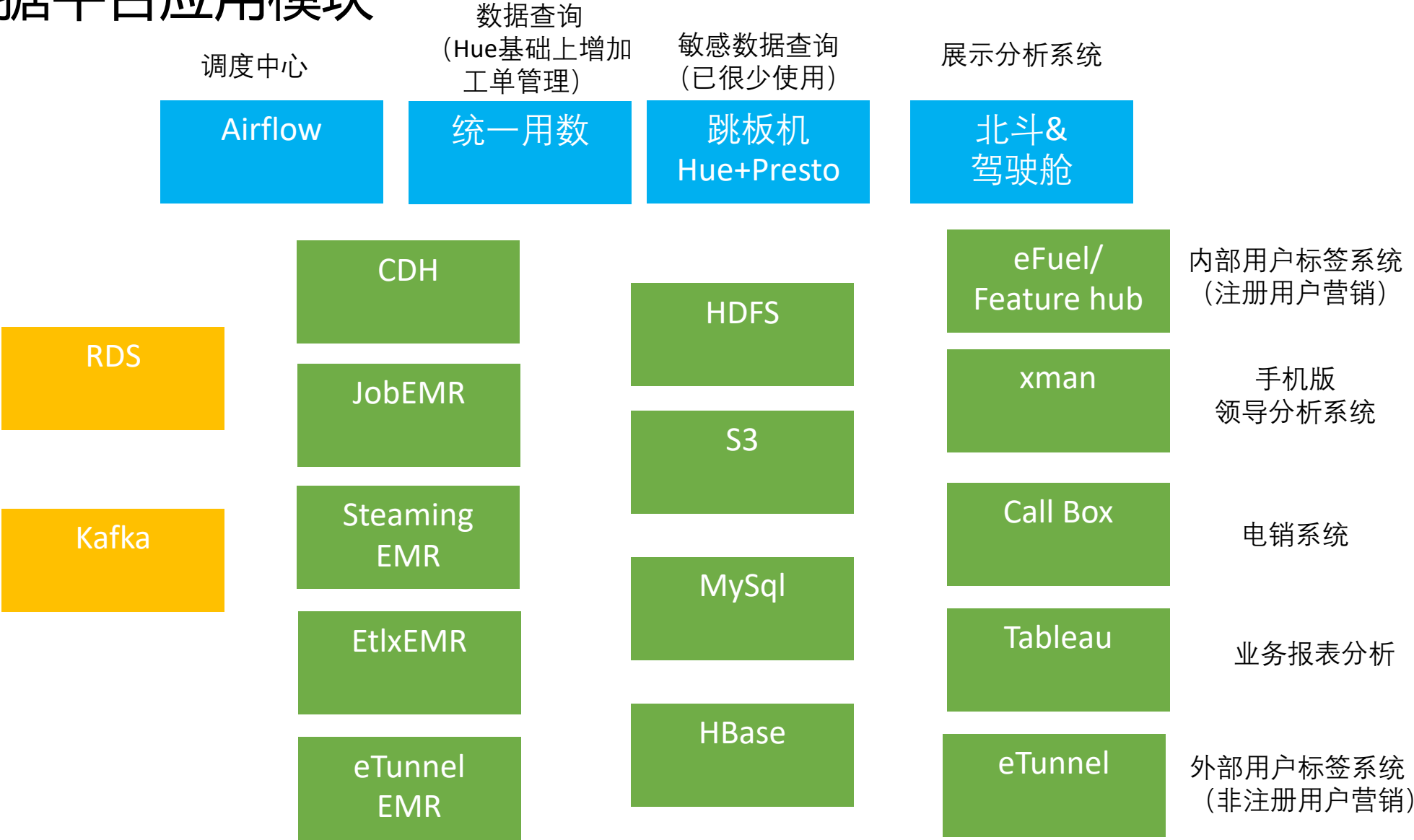
目录

- 数禾系统架构
- AWS大数据平台优点
- 用户下一步工作反馈
- AWS总体优势总结

应用整体架构



大数据平台应用模块



数禾使用AWS 情况 (2020年4月)

month	prod	total_cost
2020/04	Amazon Elastic Compute Cloud	2096461
2020/04	Amazon Relational Database Service	1105273
2020/04	Amazon Simple Storage Service	350869
2020/04	Tax	248129
2020/04	DataTransfer	146847
2020/04	Amazon ElastiCache	88793
2020/04	AmazonCloudWatch	78399
2020/04	Amazon Elastic MapReduce	68186
2020/04	AWS Direct Connect	44852
2020/04	Amazon EC2 Container Service	39237
2020/04	Amazon DynamoDB	32460
2020/04	Amazon Simple Queue Service	30294
2020/04	Amazon Kinesis Firehose	24842
2020/04	Amazon Kinesis	13498
2020/04	AWS Config	5806
2020/04	Amazon Elastic File System	3545
2020/04	AWS Database Migration Service	3429
2020/04	AWS Lambda	1864
2020/04	Amazon Elastic Container Service for Kubernetes	372
2020/04	Amazon Simple Notification Service	333
2020/04	Amazon Virtual Private Cloud	121
2020/04	AWS Key Management Service	5
2020/04	Amazon EC2 Container Registry (ECR)	1
2020/04	AWS CloudTrail	0

目录

- 数禾系统架构
- **AWS大数据平台优点**
- 用户下一步工作反馈
- AWS总体优势总结

AWS的大数据平台优点

AWS	阿里
能够满足数禾的所有要求	部分功能不满足，根据数禾的需求进行后台开发
平台稳定性高	由于定制开发造成程序冲突引起系统重启
S3的吞吐量可达到260GB/s以上。	单账号的吞吐量为12GB/s（可以开Case提升Limit），数禾的大数据目前10GB/s（从EMR到S3）
数禾大数据系统大量使用了Spot实例和RI，节约了大量成本	阿里 EMR 对 Spot 的支持弱，也不支持加密
AWS有SLA保障（99.9%）	无SLA保障
AWS EMR和开源社区主版本一致性很高。	阿里对开源 Hadoop 有进行较多定制化，而且对社区版本跟进速度不快。未来有平台绑定风险
Apache Flink实时计算是AWS EMR大数据组件形式存在的，使用功能上不会区别太大，主要可能还是成本上的差异，不过我们的EMR是Stateless而	阿里那边是Stateful的，相比较阿里云这块服务弹性和灵活性较差。

目录

- 数禾系统架构
- AWS大数据平台优点
- 用户下一步工作反馈
- AWS总体优势总结

用户反馈大数据下一步计划

- 大数据迁移阿里云的计划执行中。
- 大数据完成阿里云迁移后，考虑对AWS上CDH Hadoop进行重建（使用AWS EMR），并按照系统逻辑考虑部分系统回迁（迁移范围待定，口头可能在50%）
- 大数据考虑多云架构，防止阿里云的绑定
- 基础架构组估算迁移后的用量每月100万左右（不含大数据）
- 数禾希望AWS的EDP有阶梯方式供数禾参考选择

目录

- 数禾系统架构
- AWS大数据平台优点
- 用户下一步工作反馈
- **AWS总体优势总结**

AWS优势总结

- AWS平台优势：
 - AWS平台稳定性与成熟度大于阿里（阿里大数据平台根据数禾需求多次进行后台调整，造成程序的运行问题）
 - AWS开源程度高于阿里（阿里服务定制较多例如EMR，后续会造成平台绑定）
- AWS的服务优势
 - AWS作为全球云计算的领导，有更好的全球技术储备与资源
 - 随着AWS各服务在中国的持续落地，可在图数据库、人工智能等领域提供数禾使用
- 数禾获得AWS人力投入支持
 - AWS有针对数禾BD CSM SA TAM专职团队，人员保障，迅速响应数禾当前与未来的发展
 - AWS在上海有AI研究院，可以更好的与数禾在人工智能、风控等领域进行投入与合作
- 数禾整体拥有成本优势
 - AWS价格高于阿里，目前努力通过EDP的模式减少数禾的整体成本
 - 综合考虑到服务成熟度、技术稳定性、开放性、专职人员支持等方面的隐形因素可为数禾提供更好的整体拥有成本

Thank you!

还呗app



还呗 信用卡账单分期新选择

还呗是基于生活消费多场景的分期服务平台，是由重庆市分众小额贷款有限公司版权所有，并由上海数禾信息科技有限公司提供技术服务，面向年轻人提供账单分期和商品分期服务。用户可在购物、还款场景下，享受更灵活、更便捷、更高效的分期生活服务。

立即下载

了解详情



拿铁智投App



拿铁智投 智能投顾开拓者之一



拿铁智投是国内智能投顾领域早期开拓型产品之一。拿铁智投不断进行技术创新，深化大数据与人工智能技术在产品中的运用，持续优化产品及体验。拿铁智投亦不断丰富业务场景，为白领提供多样的智能化理财方案。白领投资的每一笔钱都可以通过拿铁智投快速定制投资方案。

立即下载

了解详情

AWS EMR VS Ali EMR

□ 优势：

- 1. 完全支持spot实例，阿里云只支持集群扩展过程中，任务实例使用竞价实例。
- 2. 支持fleet模式启动集群，阿里云不支持
- 3. 弹性伸缩支持更多种的监控指标
- 4. 理论上EMR+S3模式性能优于阿里云EMR+OSS效率（OSS带宽有限制）
- 5. EMR节点带宽优势（EC2带宽优于ECS）
- 6. 支持本地、S3静态数据加密和传输中数据加密，阿里云不支持存储数据加密
- 7. 支持用户自定义镜像，阿里云不支持
- 8. AWS EMR提供99%SLA，阿里云EMR不提供SLA

□ 劣势：

- 1. 阿里云EMR管理使用更方便：
 - a) 支持Knox、ranger、openLDAP等组件，用户非常方便配置EMR上的权限管理。
 - b) 提供托管Gateway，用户可以安全方便的使用各种UI：hdfsui、yarnui、hueui
- 2. 阿里云提供EMR开发运维工作台：EMR Workspace，任务 workflow 界面，用户开发和维护job非常方便、用户体验好。
- 3. 阿里云EMR部分集群类型支持ESSD IO性能高