



Yamibuy 迁移案例分享

刘恒涛

AWS 解决方案架构师



议程

- 客户背景
- 面临的挑战
- 迁移方案
- 迁移效果
- 经验总结

- 客户背景

客户背景

亚米网 (Yamibuy.com) 是北美最大的亚洲商品购物平台，主营最具特色的亚洲商品，涵盖食品、美妆、养生、电器、图书等生活类别。经过近几年的飞速发展，亚米网在北美华人圈建立了良好的品牌形象，成为北美华人线上购物的首选。

亚米网的创办起源于一瓶老干妈。创始人兼CEO周游曾在位于美国中西部的堪萨斯州留学，为买一瓶老干妈，要开车2个小时去最近的华人超市。想到还有更多人正在经历着和他相同的不便，本着服务在美华人的宗旨，2013年周游创办了亚米网。

如今，亚米网在美国洛杉矶、中国杭州、中国西安三地拥有超过三百人的团队，自主运营20万平方英尺的仓库以及自动化物流系统，自主销售商品万余种，并且与上千个知名品牌建立了深度合作关系、签署正品保障协议，为品牌和商家在北美的销售和品牌推广起到了关键性推动作用。



Yamibuy
亚米网

- 面临的挑战

客户面临的挑战

阿里服务带来的挑战：

阿里多次停机事故（RDS，SLB）

阿里服务支持不足

客户业务面临的挑战：

智能推荐需求

业务系统架构设计
-K8S需求

大数据平台架构升级优化需求

迁移面临的挑战

1. **工作负载涉及多个领域：**包括机器学习，多媒体，大数据，容器，关系数据库，Nosql数据库等
2. **系统复杂：**涉及20多个子系统，包括大数据平台，业务系统平台，后台任务平台
3. **涉及20+的AWS服务：**Route53, Cloudfront, EC2, ELB, EKS, ECR, APIGateway, Lambda, S3, Personalize, Athena, EMR, Redshift, Glue, Aurora, ElastiCache, DMS, Kinesis, Kinesis Firehose, Kinesis Analytics, EBS, Media Live, Media Package, Media Store, Media Convert。

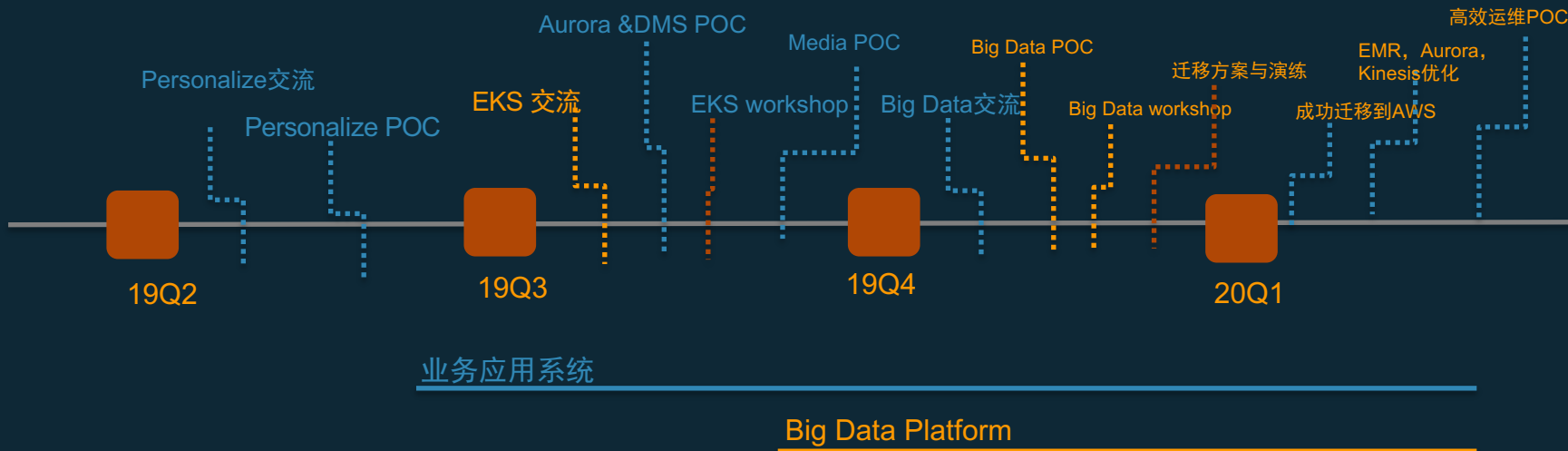
- **迁移历程与方案**

迁移历程

Activity

Coverage

Result



痛点

目标

微服务治理

Swarm 编排已无法满足需求

迁移swarm到EKS, 并协助客户落地EKS改造

大数据

自建CDH运维成本高, 无法满足需求

迁移CDH到EMR S3 Athena, Kinesis

机器学习

当前商品推荐等电商行业AI能力弱

引入托管服务和行业经验赋能客户, 已引入personalize

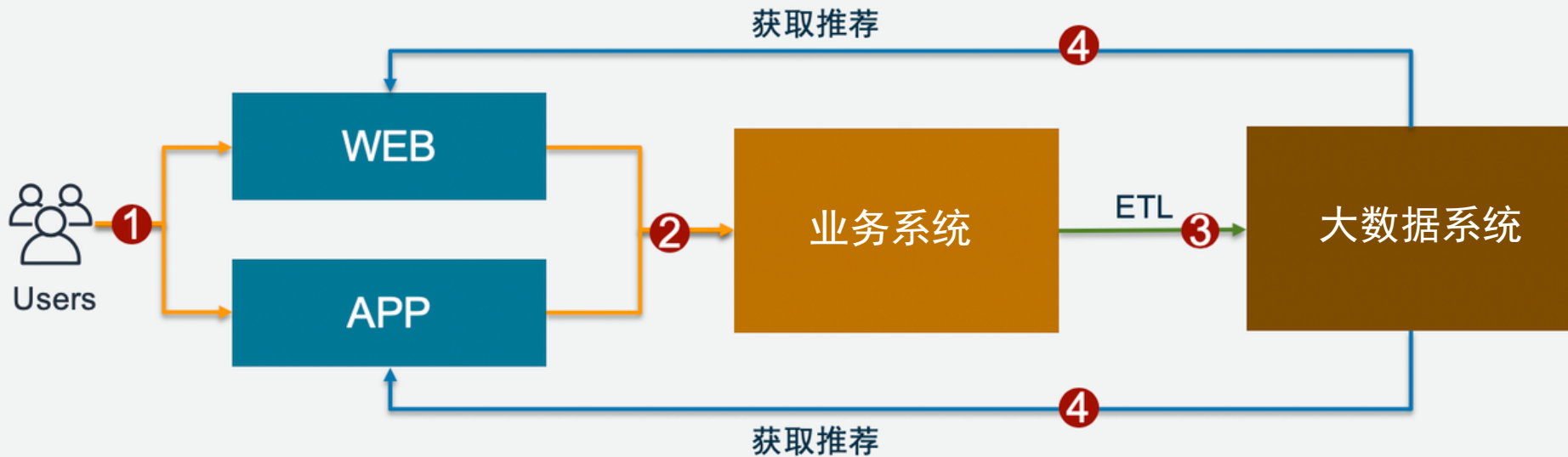
竞争对手

Aliyun

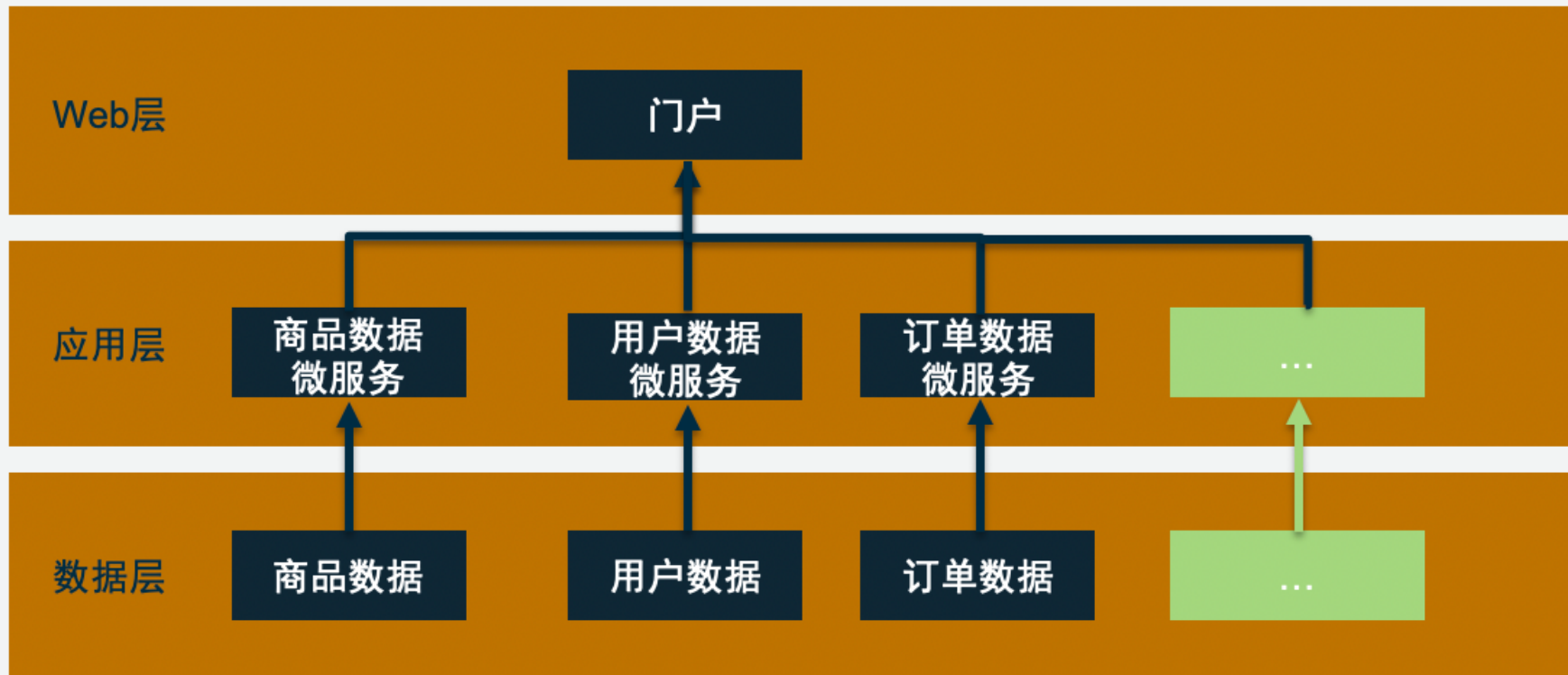
AWS的产品质量, 更好, 更及时的服务



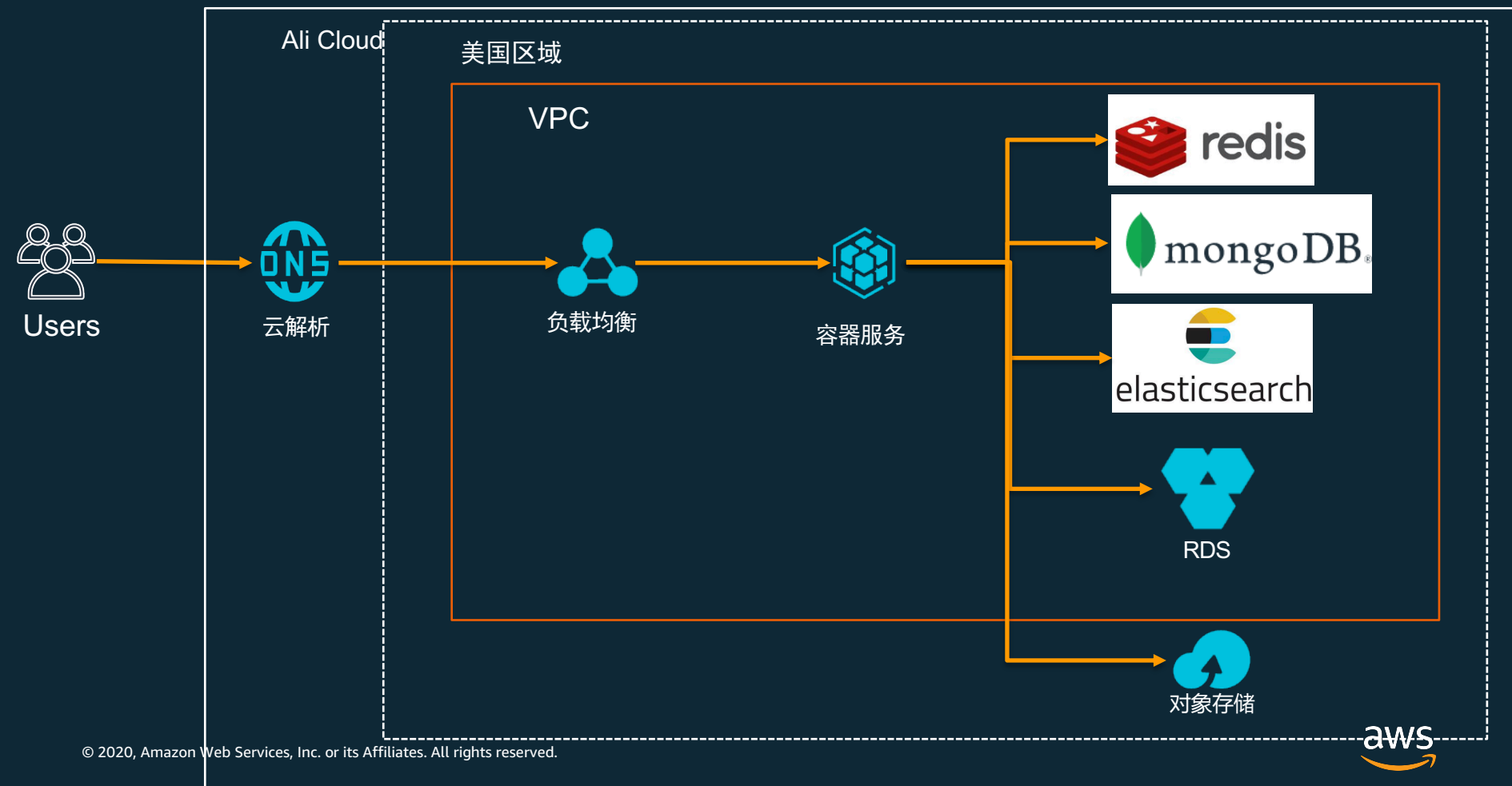
业务系统架构-Level 0



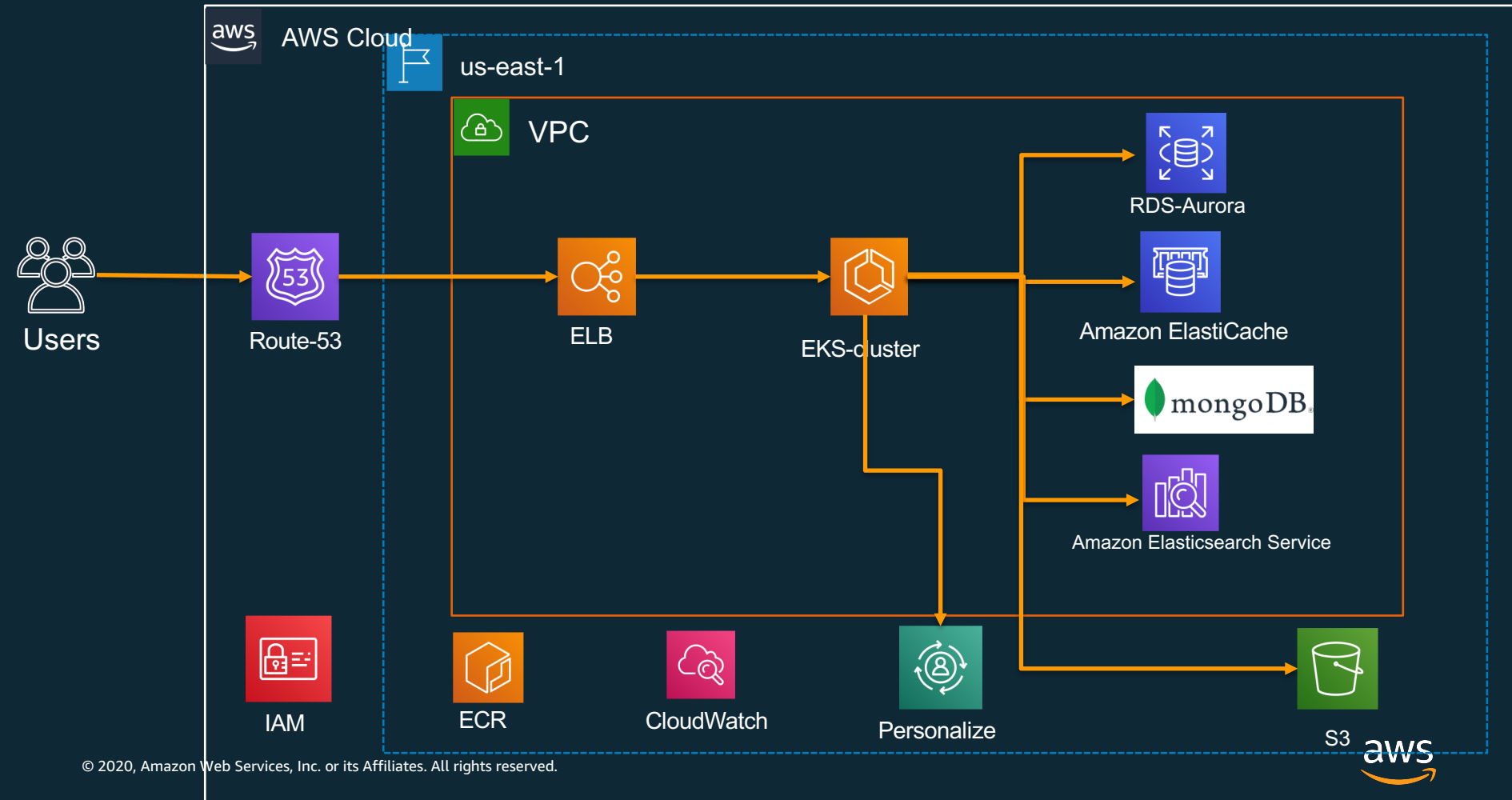
业务系统架构- level 1



业务系统架构（迁移前）



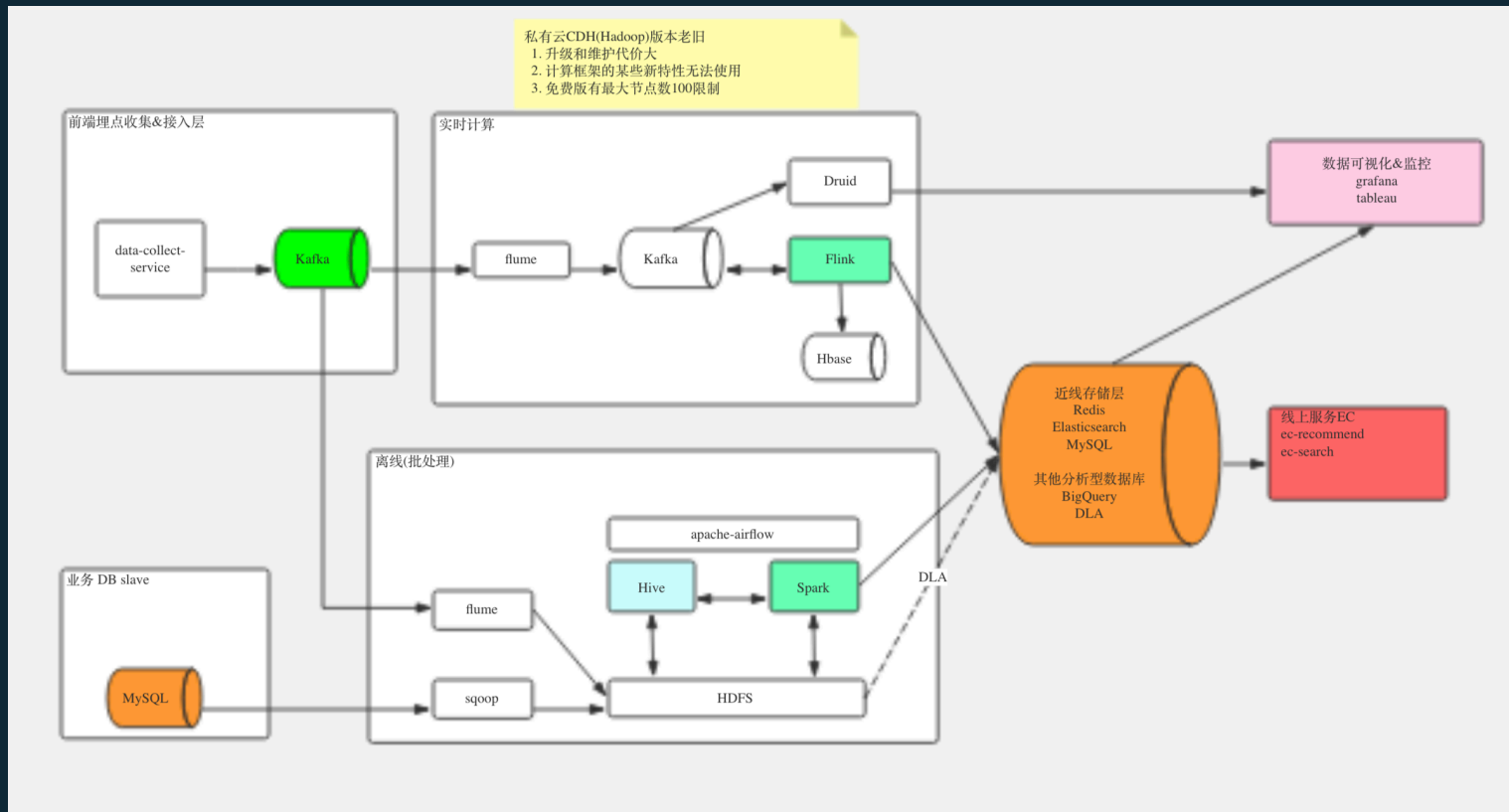
业务系统（迁移后）



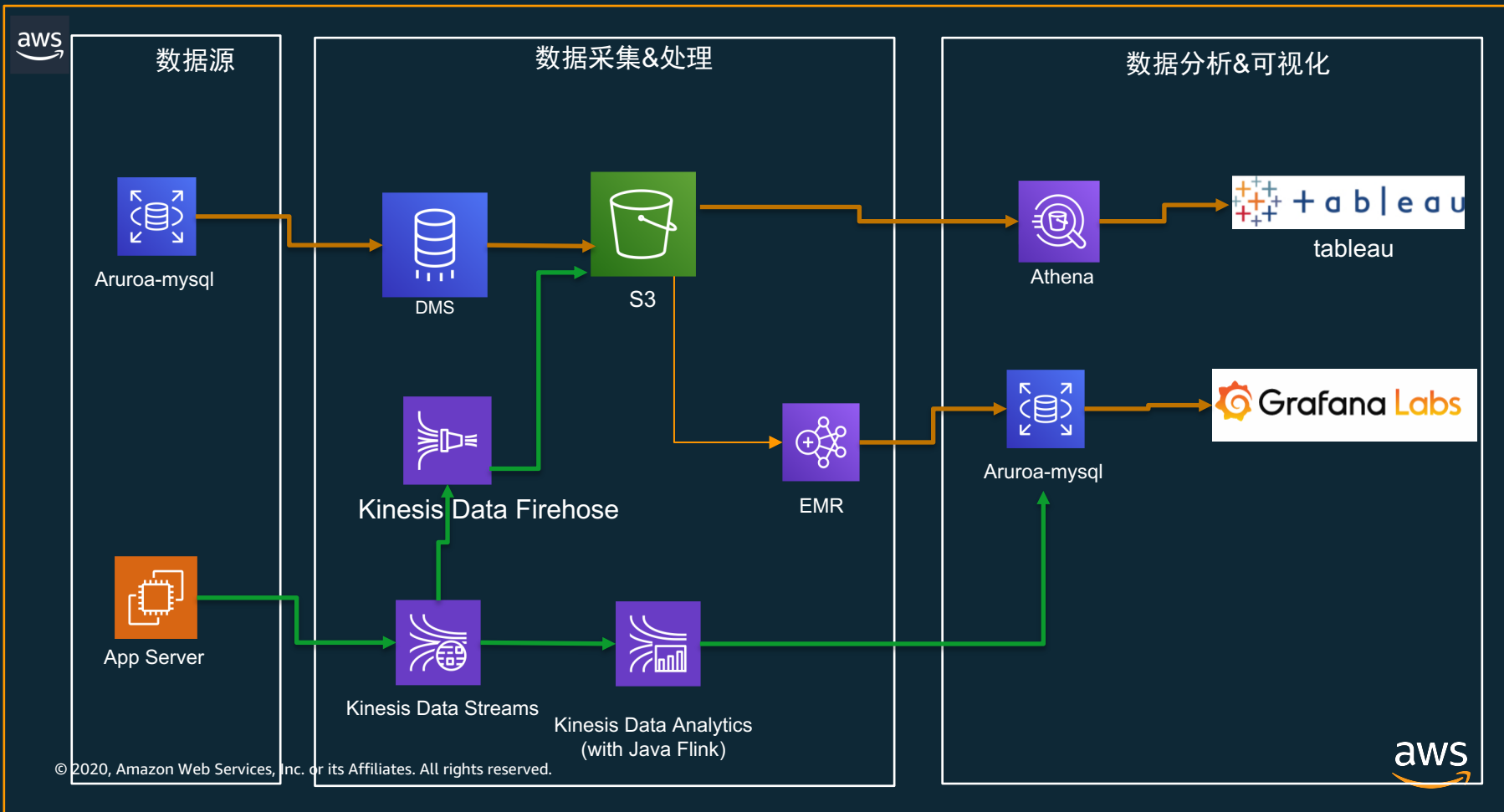
迁移前后架构对比

1. RDS 升级为**Aurora** 集群：相同配置下，性能更高，价格更优
2. 自建Redis升级为托管的**Amazon ElastiCache-Redis**集群：降低了运维成本
3. 自建ES升级为托管的**Amazon Elasticsearch**集群：降低了运维成本
4. Swarm集群升级为AWS托管的**EKS**集群：实现了客户架构升级的预期，同时降低了客户学习K8S的曲线
5. 通过Personalize实现智能推荐，解决了当前缺乏算法工程师，以及推荐效果差的问题。

大数据系统-level 1



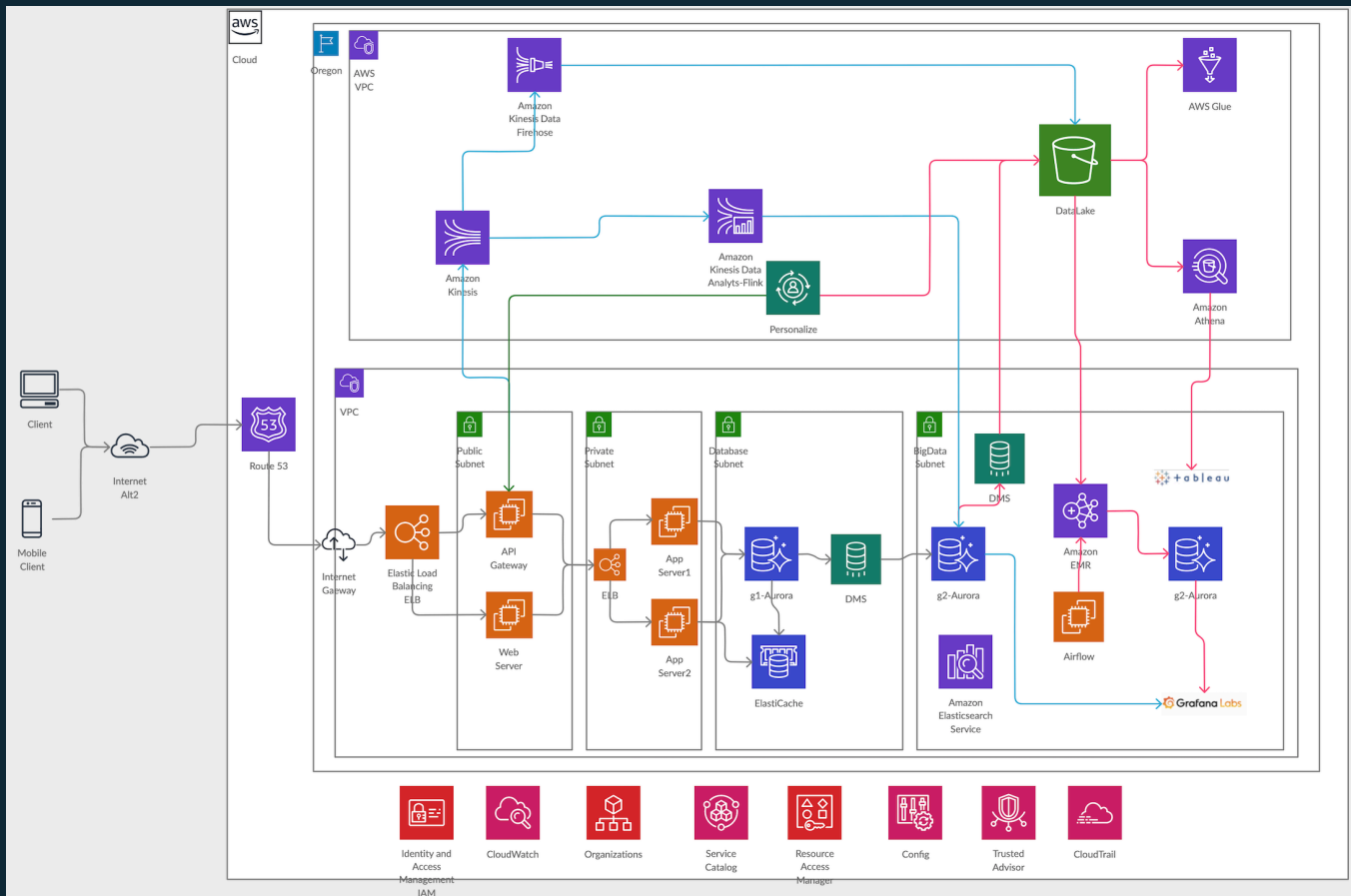
大数据系统架构-迁移后



迁移前后架构对比

1. 数据湖架构：让客户原有多点的数据集中存储和管控，同时能够更安全，更可靠的存储企业数据资产，通过Glue轻松构建企业数据资产目录。
2. 自建CDH 升级为**托管的EMR**：性能提升（通过多轮的性能优化原有4小时的任务优化到30分钟）
3. HDFS数据存储升级为**S3存储**：成本更低
4. 自建kafka集群升级为托管的Kinesis Data Stream/firehose：降低了运维成本
5. DLA,BigQuery 替换为**Athena**：更好的与AWS服务集成

客户当前总体架构（持续改进中）



- 迁移效果

客户收益

1. 客户当前使用4个Aurora集群（6个实例），包括生产和测试的在线电子商务平台，大数据分析平台。Aurora每月花费\$ 3.5K，客户反馈的价格比Ali RDS便宜了近2/3（\$ 12每月K）
2. 数据存储在基于S3的数据湖中，与客户之前使用的HDFS存储方法相比，成本降低了60%。
3. 通过Personalize实现智能推荐，解决了当前缺乏算法工程师，以及推荐效果差的问题。
4. 通过EKS的升级，满足了客户业务系统架构升级需求，同时降低了客户学习K8S的曲线

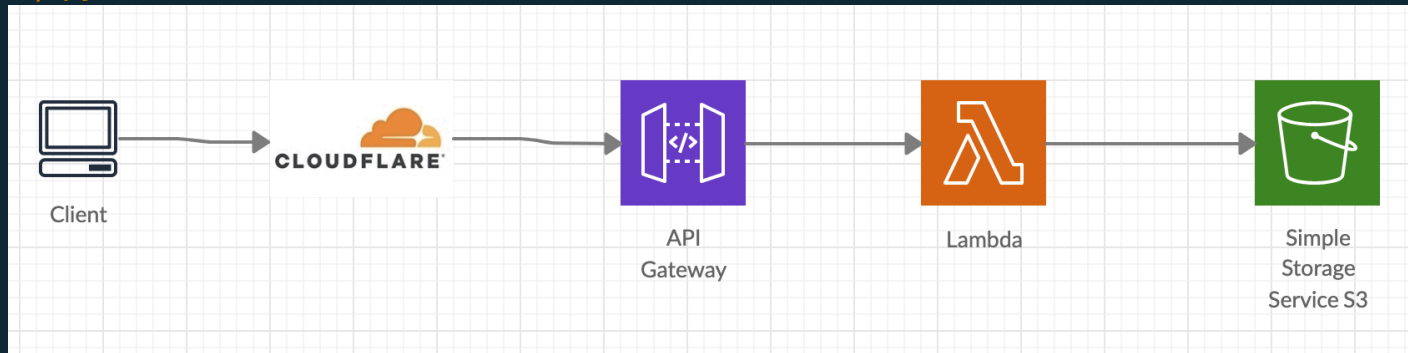
- **经验总结**

选择AWS的原因

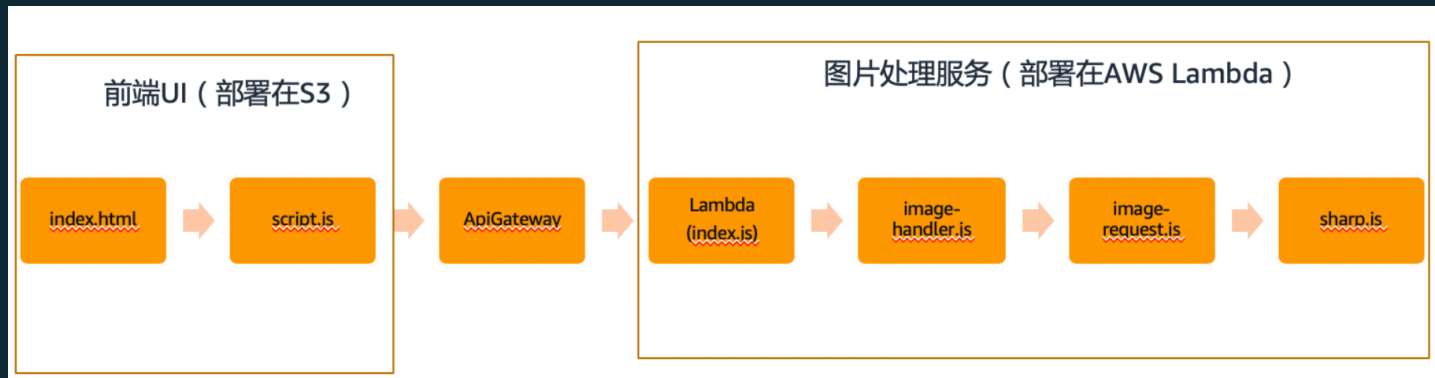
- **技术支持：** AWS提供了优质，快速的支持服务，通过微信，电话，定期拜访的方式快速支持客户需求与问题
- **解决方案支持：** 通过定期的客户拜访获取客户需求并持续跟进落地方案，提供符合行业特性的，符合AWS最佳实践的，经过POC验证的，可落地的解决方案。
- **性能：** 通过方案验证阶段的对比测试，上线后的实际运行效果体验，性能方面得到了客户的认可
- **成本：** 通过AWS的最佳实践提供成本最优的方案。如Aurora，EMR+S3，后续Spot的使用等

如何在不修改代码的情况下迁移Ali图片处理服务

架构:



代码处理过程:



Q: 如何选择DMS的复制实例？

实例的选择可以逐级提升，注意监控复制实例的 CPU，内存，磁盘的使用情况，如果异常了看具体哪个指标过高然后适当调整。CDC方式主要和数据的更新频率，涉及的表数量，以及当前数据量有关系的。

https://docs.aws.amazon.com/zh_cn/dms/latest/userguide/CHAP_ReplicationInstance.html#CHAP_ReplicationInstance.InDepth

Q: DMS如何打开详细日志进行问题核查?

aws dms modify-replication-task --replication-task-arn arn:aws:dms:us-east-1:517141035927:task:TAL3REOMJSYU2MAV4XS3WRH7PE --replication-task-settings file:///Users/lht/bigdata/DMS/settings.json

```
{
  "Logging": {
    "EnableLogging": true,
    "LogComponents": [
      {
        "Id": "SOURCE_UNLOAD",
        "Severity": "LOGGER_SEVERITY_DEFAULT"
      },
      {
        "Id": "TARGET_LOAD",
        "Severity": "LOGGER_SEVERITY_DEBUG"
      },
      {
        "Id": "SOURCE_CAPTURE",
        "Severity": "LOGGER_SEVERITY_DEFAULT"
      },
      {
        "Id": "TARGET_APPLY",
        "Severity": "LOGGER_SEVERITY_DETAILED_DEBUG"
      },
      {
        "Id": "TASK_MANAGER",
        "Severity": "LOGGER_SEVERITY_DEFAULT"
      }
    ],
    "CloudWatchLogGroup": "dms-tasks-xxxx",
    "CloudWatchLogStream": "dms-task-xxxxx"
  }
}
```

Q: DMS如何导出RDS数据到S3为parquet?

1. 控制台, Extra connection attributes。设置 dataFormat=parquet;
2. 指定parquet 格式后, parquet文件会自动保存列明, 然后通过glue 爬取创建表, 通过athena 或者redshift 访问就会有列名了。也可以导入后通过glue 为数据增加列名

Q: DMS同步到S3的增量数据处理？

DMS同步RDS数据到S3后的增量数据会多一列OP，I代表插入，D代表删除，U代表更新，通过Athena 查询会多出这一列，同时数据会有多条。

可以通过glue来处理将数据进行清洗保留最终状态，或者直接将数据同步到Redshift

<https://amazonaws-china.com/blogs/database/python-code-to-download-dms-task-logs-using-the-aws-dms-task-id/>

DMS问题整理

<https://quip-amazon.com/BMBhAwaMDlfV>

经验总结

1. 迁移是一个漫长的过程，逐步获得客户信任的过程
2. 想客户之所想，虑客户之未虑，客户的问题往往不仅限于我们的服务
3. 迁移工具是迁移成功的充分条件，并非必要条件
4. 迁移成功是新的开始

特别鸣谢：刘磊，孙华，白鹤，蒋力

Q&A

谢谢！