



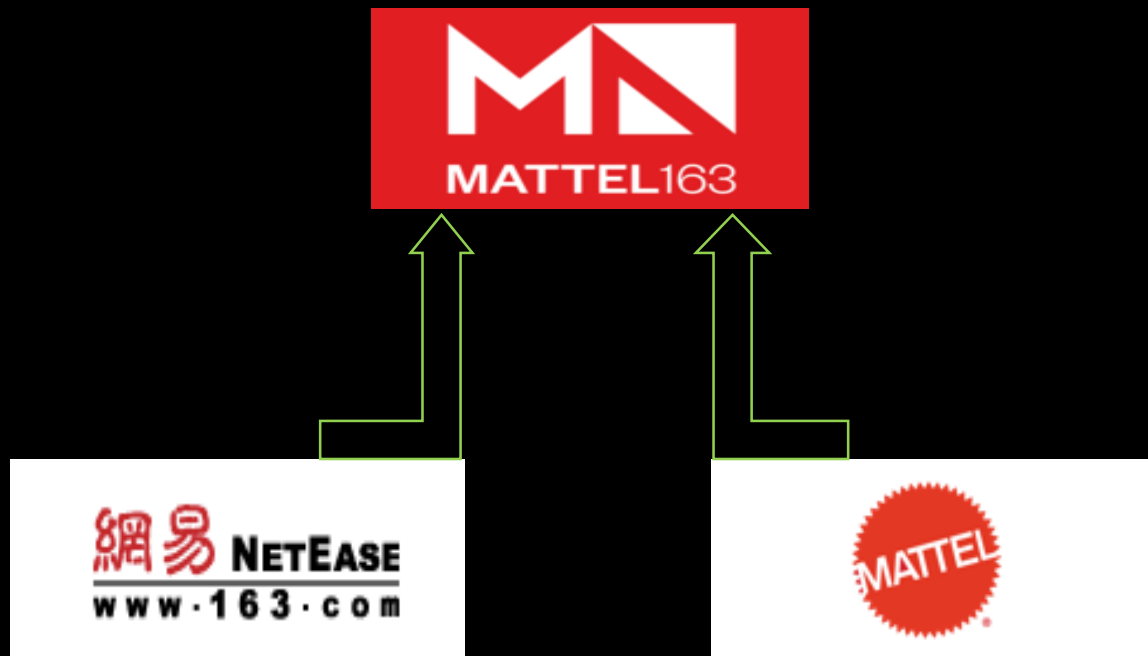
# 云上的游戏 数据仓库实践

陈炜斌 Mattel163 数据开发工程师

张孝峰 AWS 资深解决方案架构师



# 关于Mattel163



# UNO手游——Android iOS版全球发布

出海手游下载榜第11名 (部分区域除外)

- 用户数: 五千万以上





# Mattel163使用AWS

- 遍布全球
- 经济实惠
- 自动化
- 简单易用
- 使用事件驱动开发



# Mattel163使用AWS



# Mattel163使用AWS



MATTEL163

# 云端的数据湖架构



以存储更多种类数据  
和更多样化的处理  
手段为目标



以更强的业务价值  
输出为前提



以Hadoop生态体系为主的大数据平台

崭新的数据湖架构

AWS 中国（宁夏）区域由西云数据运营  
AWS 中国（北京）区域由光环新网运营



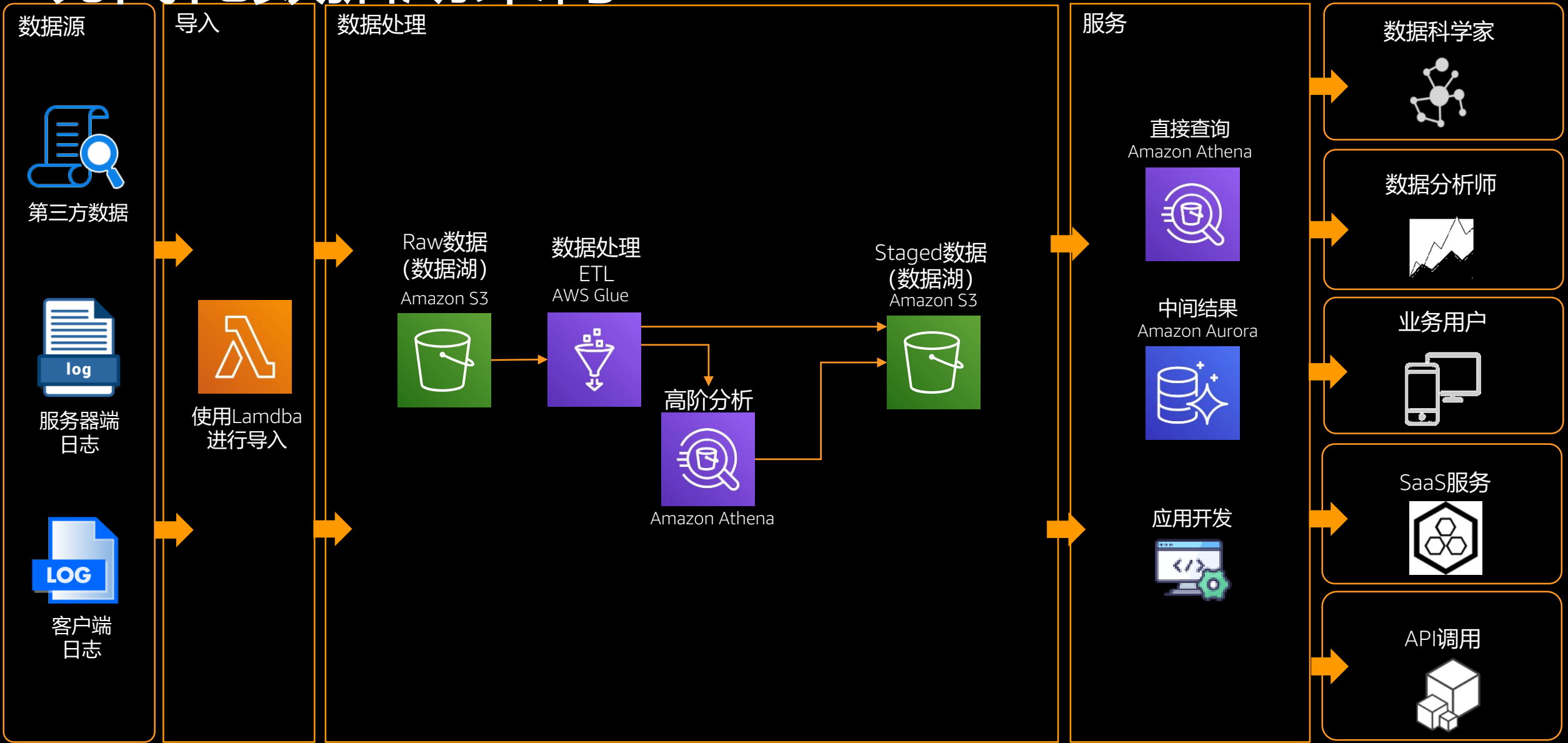


# Mattel163的日志数据处理流程

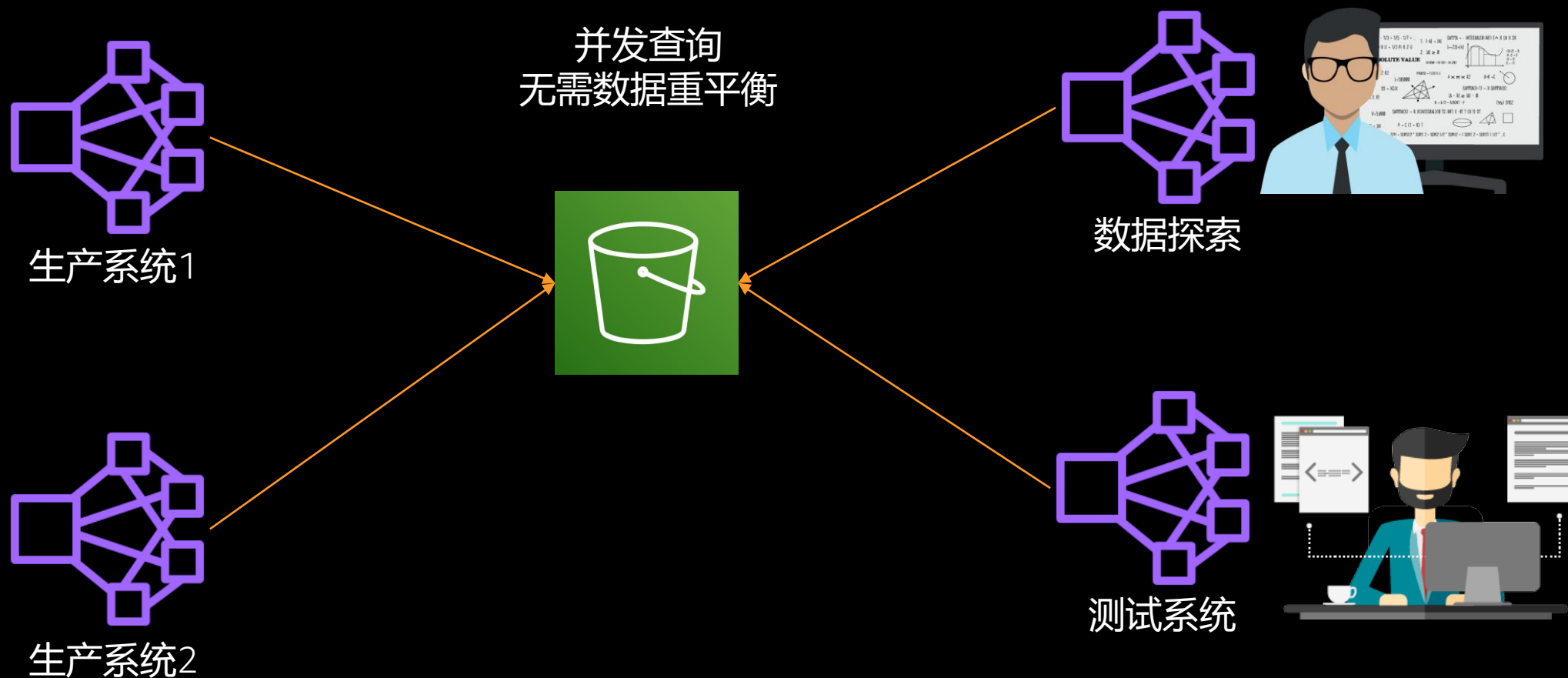




# 现代化数据湖架构



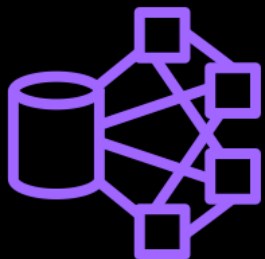
# S3支持多个EMR集群同时查询同一批数据



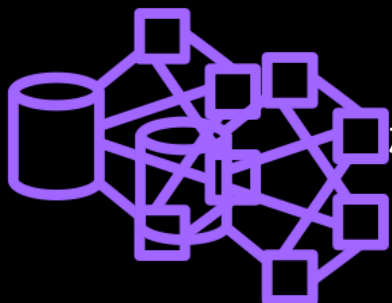
存放在HDFS

1PB数据

存放在S3



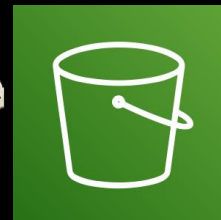
复制因子 = 3  
3 PB



磁盘预留空间 25%  
4 PB



节省接近  
九成的费用!



1 PB  
(按实际使用计算)

单价  
ST1: \$0.045 / 每月GB

总价  
\$ 188,743.68 / 月

单价  
S3: 约 \$0.02155 / 每月GB

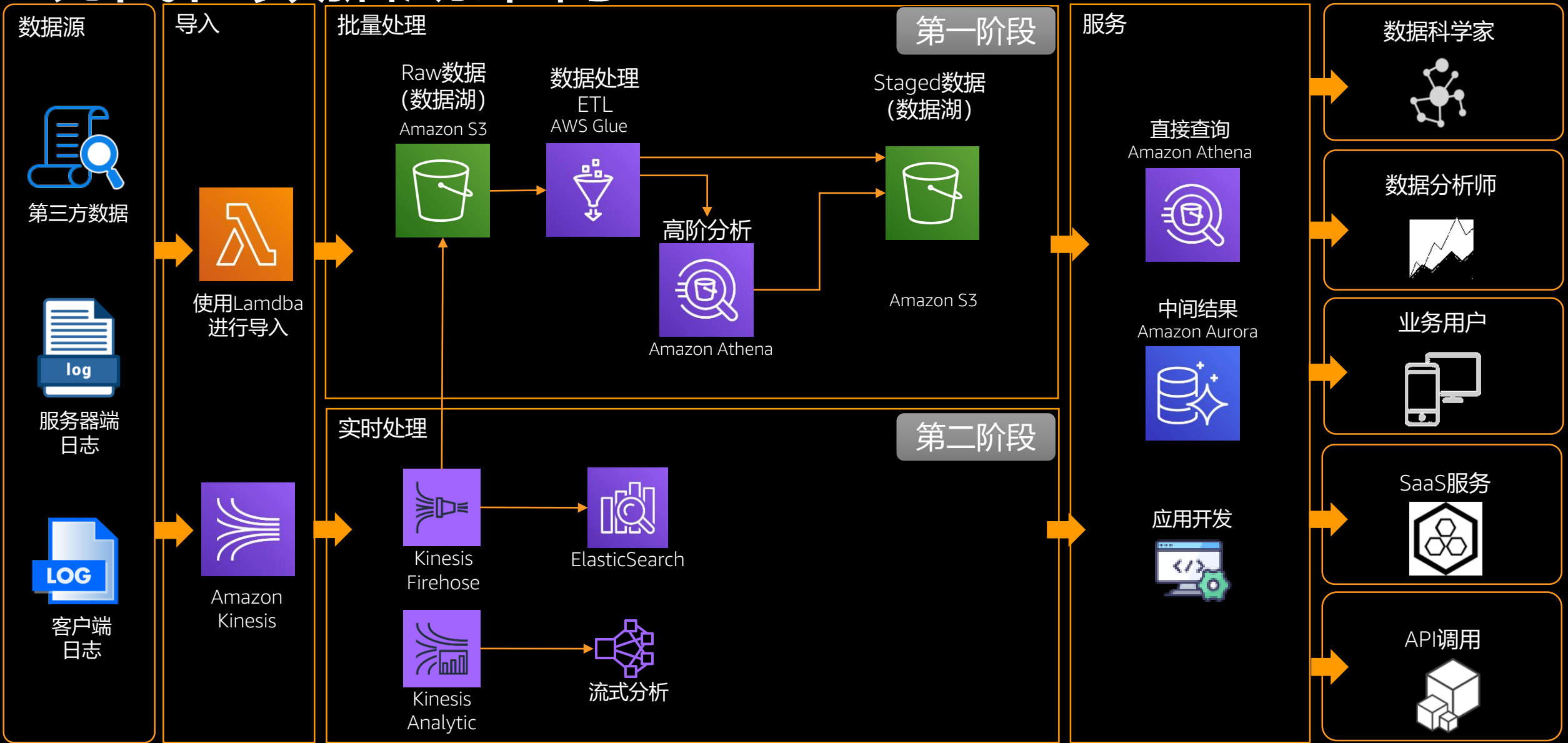
总价  
\$ 22,067.2 / 月

AWS 中国 (宁夏) 区域由西云数据运营  
AWS 中国 (北京) 区域由光环新网运营

价格以美国东部 (弗吉尼亚北部) 为例 <https://aws.amazon.com/s3/pricing/>



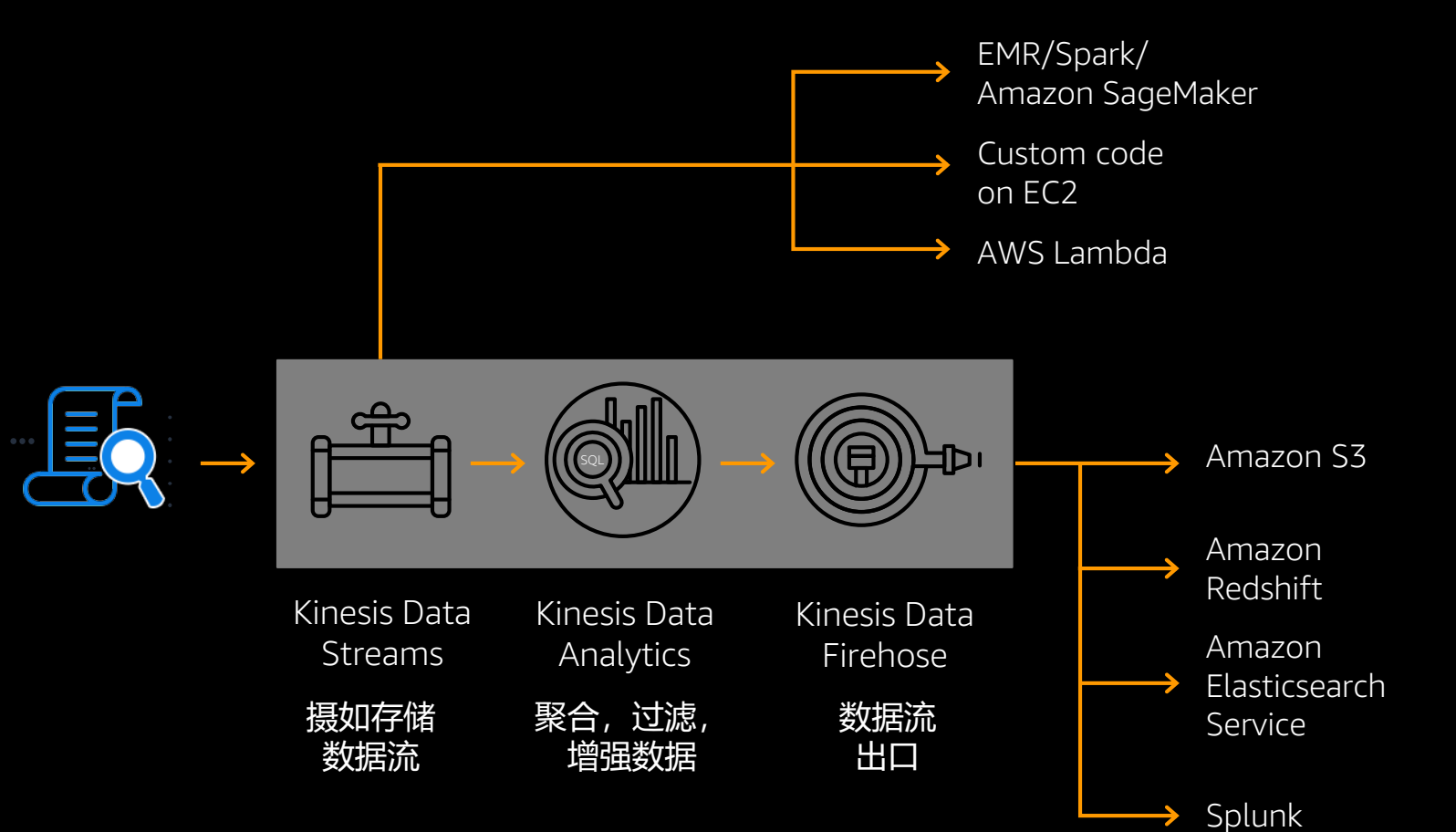
# 现代化数据湖架构





# Amazon Kinesis Data Streaming

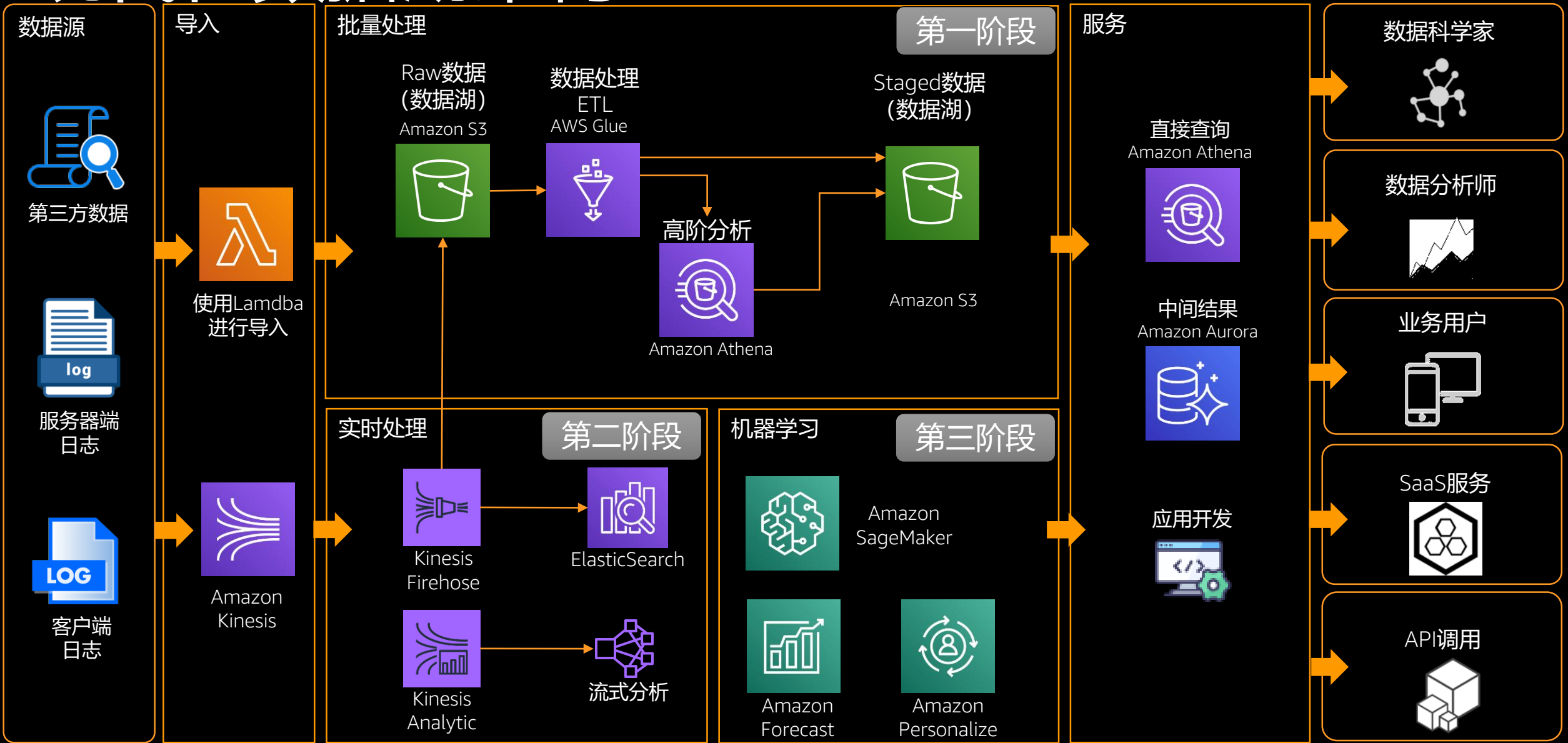
实时收集，处理和分析数据流



- 实时
- 全托管
- 可扩展
- 安全
- 成本可控











AWS 中国 (宁夏) 区域由西云数据运营  
AWS 中国 (北京) 区域由光环新网运营

# 现代化数据湖架构



# Amazon ML 堆栈：最广泛、最深入的功能

## AI 服务

视觉			语音		语言		聊天机器人	预测	推荐
									
REKOGNITION IMAGE	REKOGNITION VIDEO	TEXTRACT	POLLY	TRANSCRIBE	TRANSLATE	COMPREHEND	LEX	FORECAST	PERSONALIZE

## ML 服务



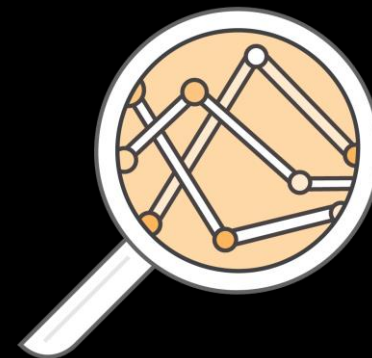
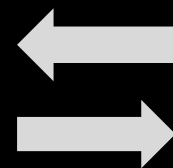
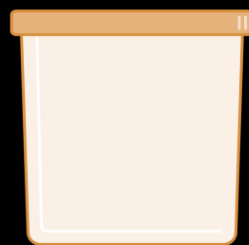
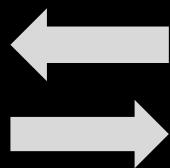
## ML 框架 + 基础设施

框架	接口	基础设施
 TensorFlow	 mxnet	 PYTORCH
	 GLUON	 Keras
		 EC2 P3 与 P3dn
		 EC2 G4
		 EC2 C5
		 FPGA
		 Greengrass
		 Elastic Inference
		 Inferentia

AWS 中国 (宁夏) 区域由西云数据运营  
AWS 中国 (北京) 区域由光环新网运营



# AWS完整的云端数据湖方案



## 面向未来 新数据类型的支持

吸纳新的数据类型,  
构建差异化的体验:  
图片/视频, 社交媒体, IoT传感器及  
更多数据

## 单一数据源 ("data lake")

适合于任何数据和任何分析  
高度安全, 极度弹性  
再无数据孤岛

## 面向各种任务类型的分 析平台

具备最新的分析能力  
机器学习, 自然语言处理, 托管  
Hadoop服务等



# Thank you!



MATTEL163

