

面向电信领域的 海量实时数据技术与实践

 杨光明子

FOR BDCT PRESENTATION, 2016

目录

CONTENT

场景

架构

性能

演进

01

CONTENT

场景

架构

性能

演进



1

每天净增用户16.6万

2

每秒无线上网流量33G

3

每天话单数据10T

4

每日信令数据100T



2014

简单实时营销

2015

+实时位置运营

2016

+复杂实时营销
+内容业务处理

02

CONTENT

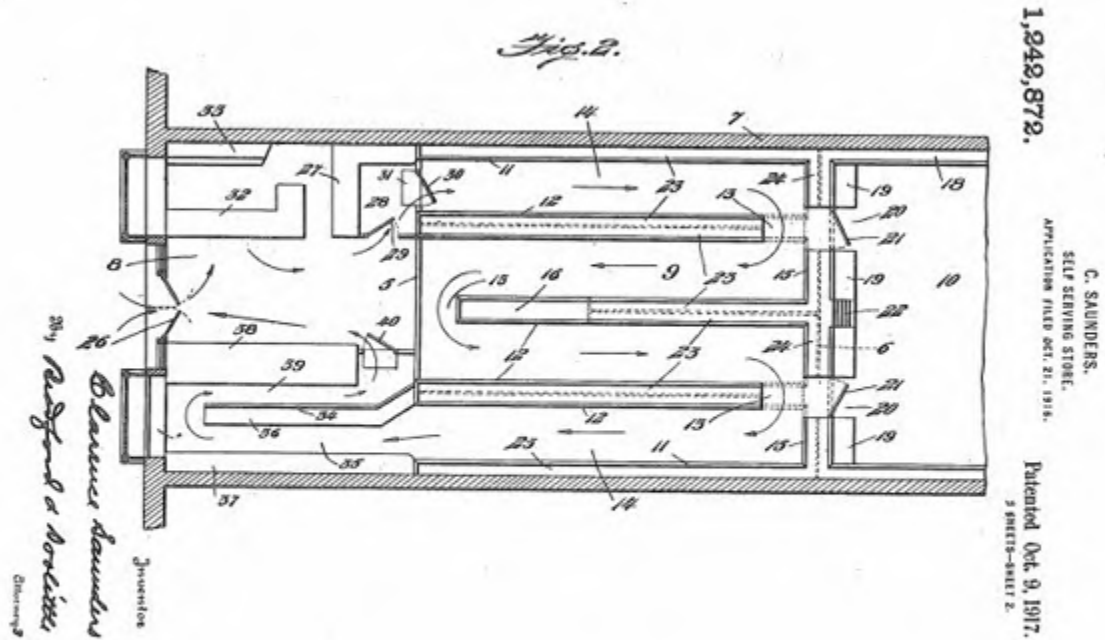
场景

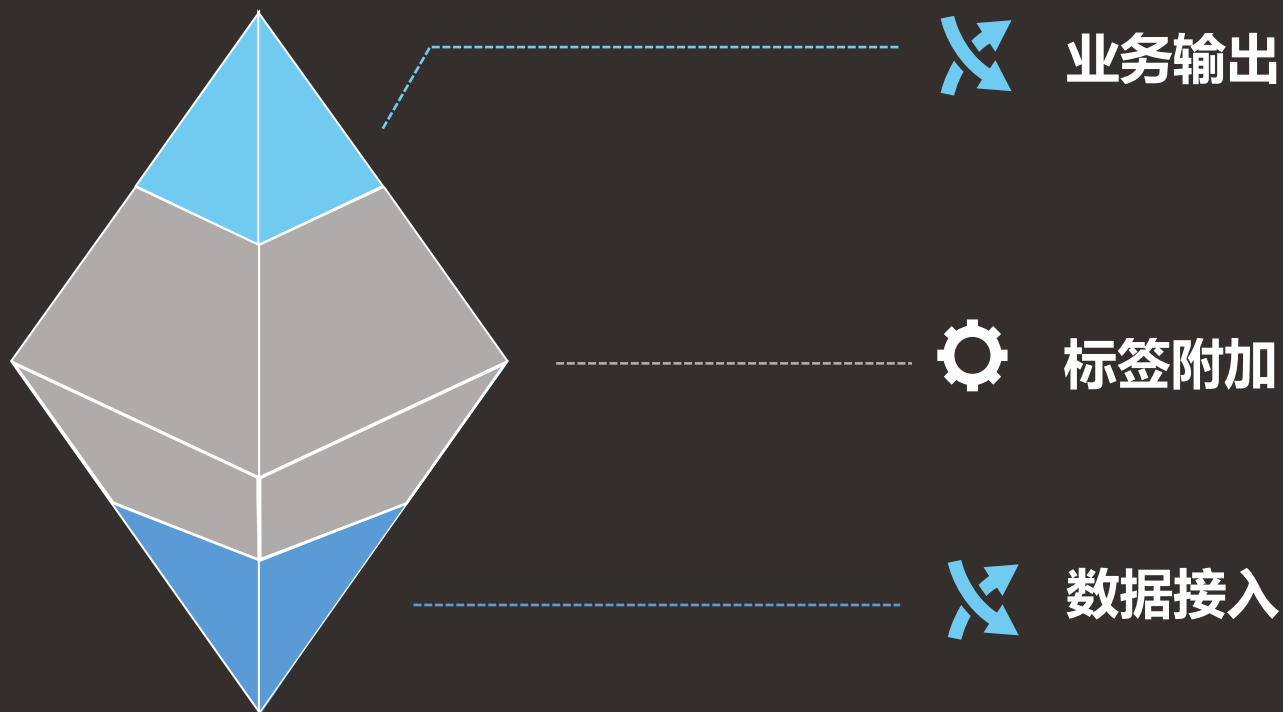
架构

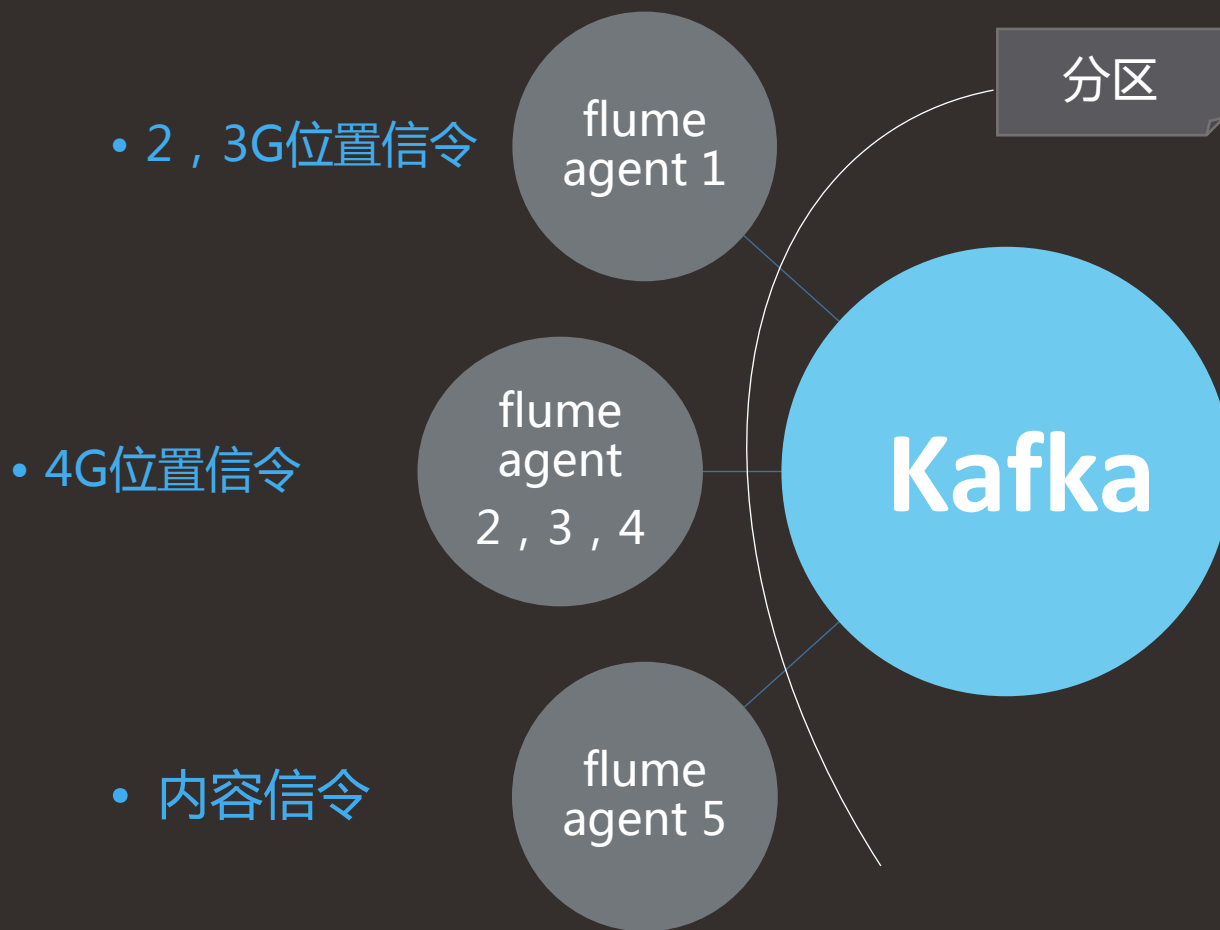
性能

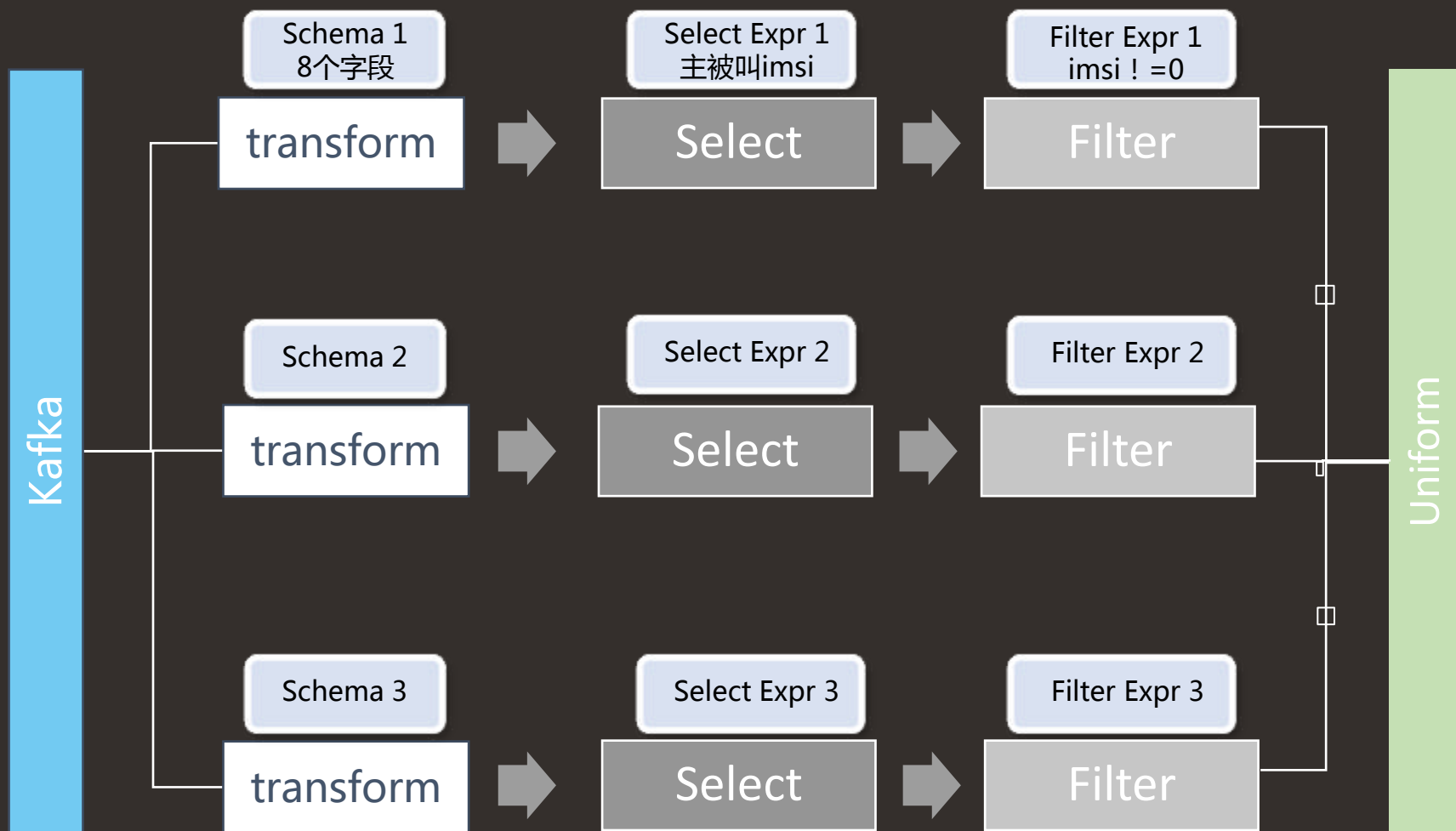
演进

Patent US 1242872



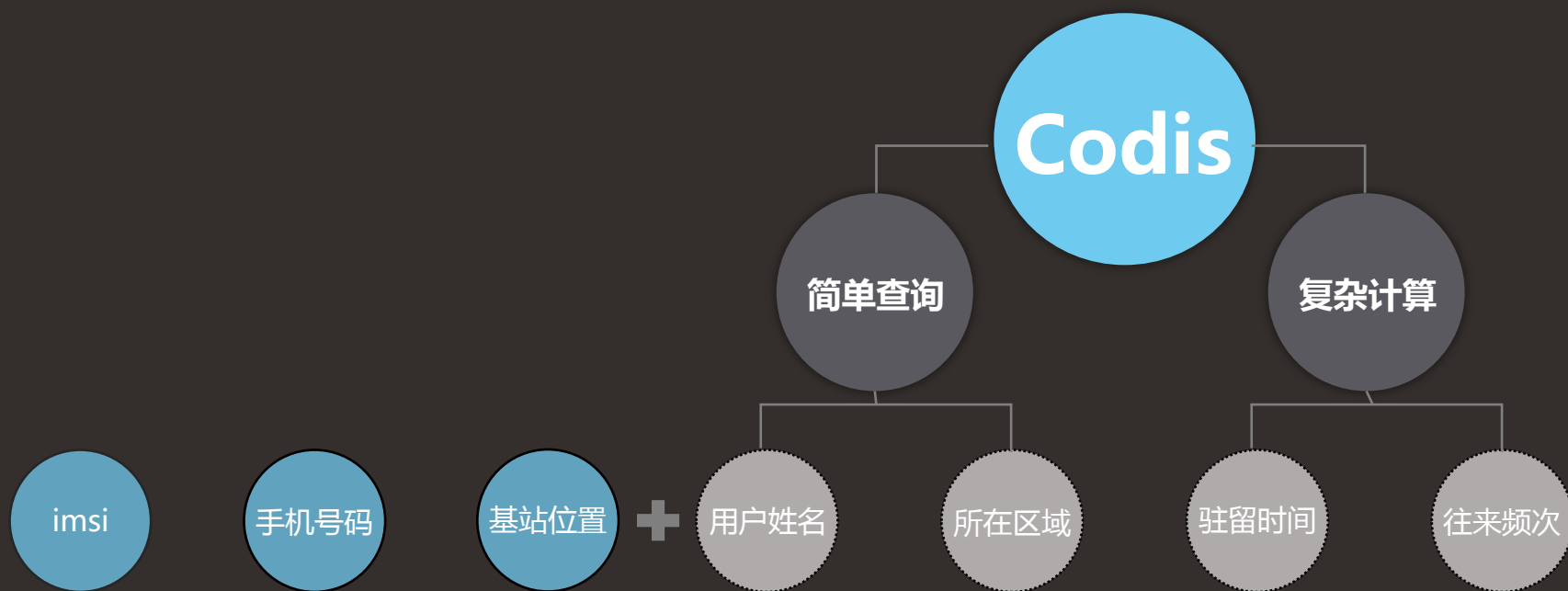


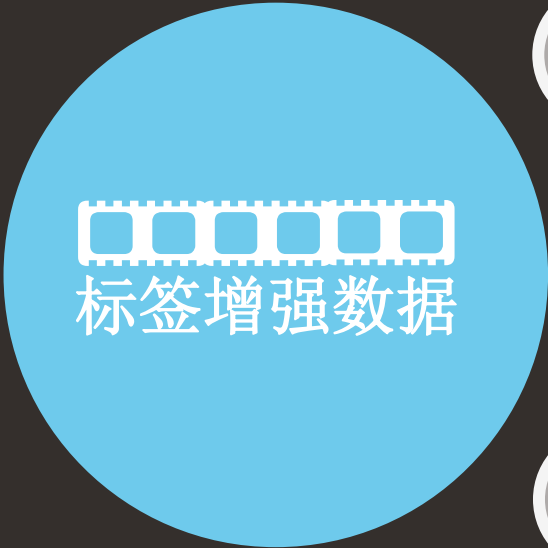


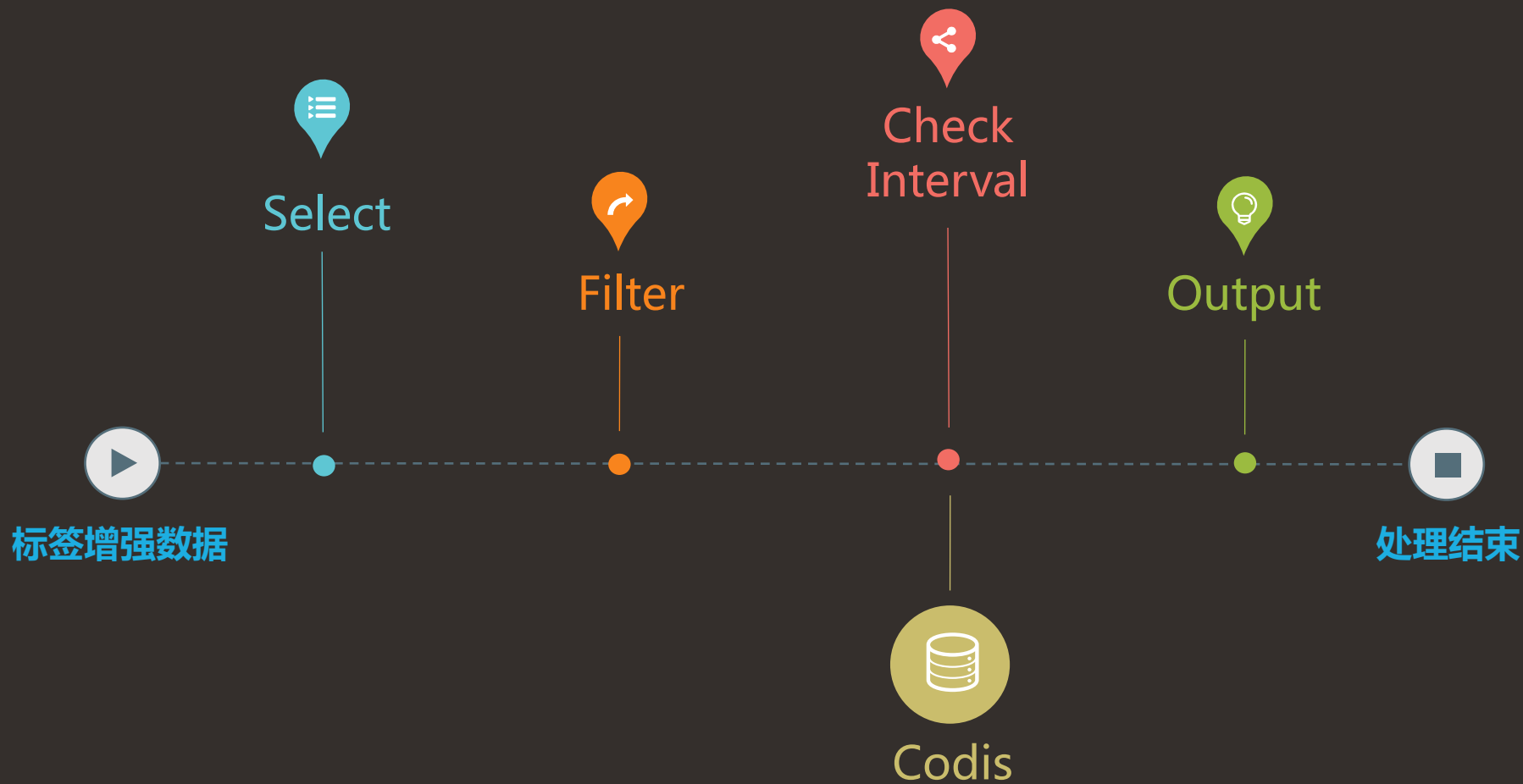


什么是标签？







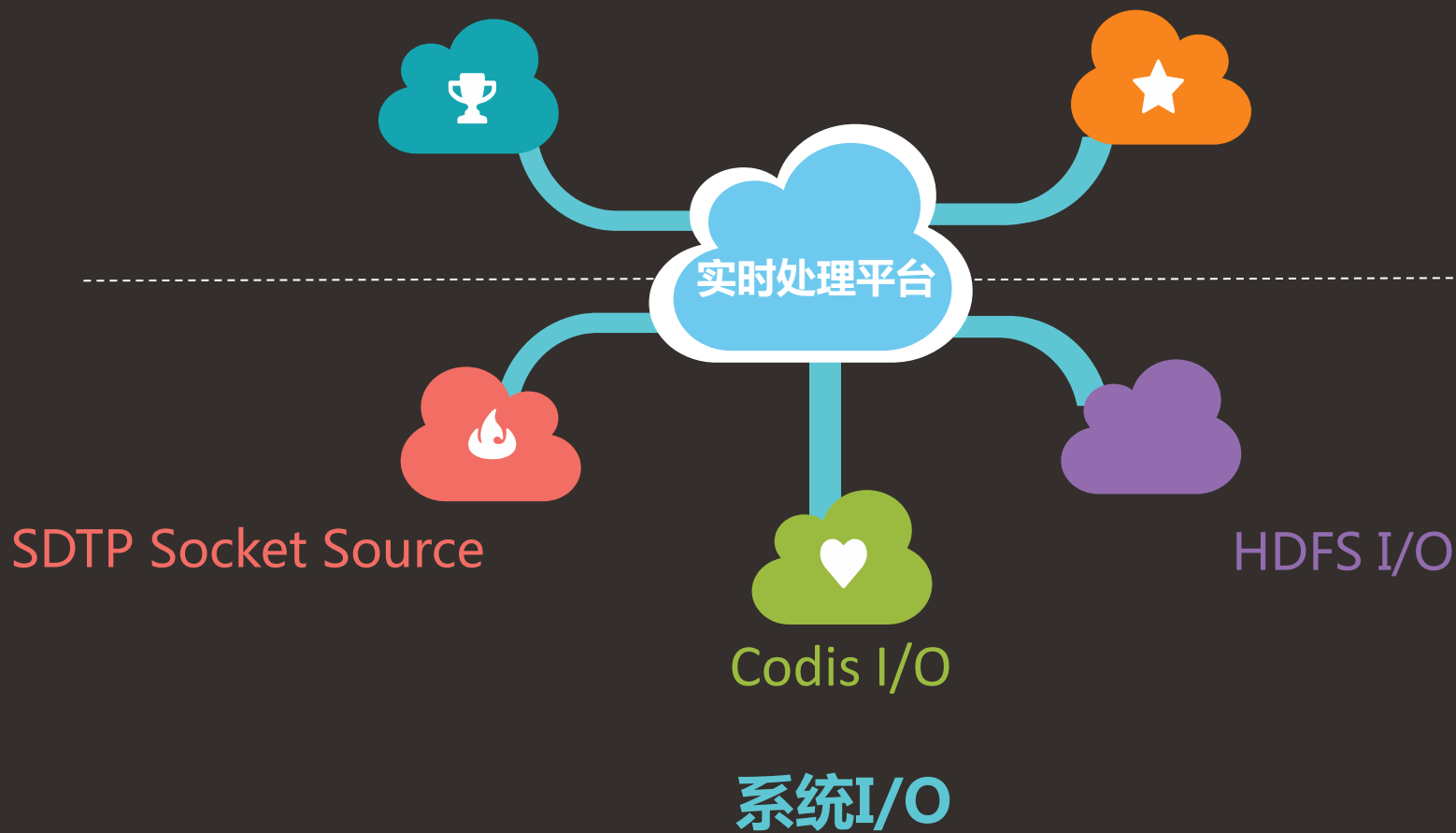




部署/配置

可视化配置界面

自动化部署



03

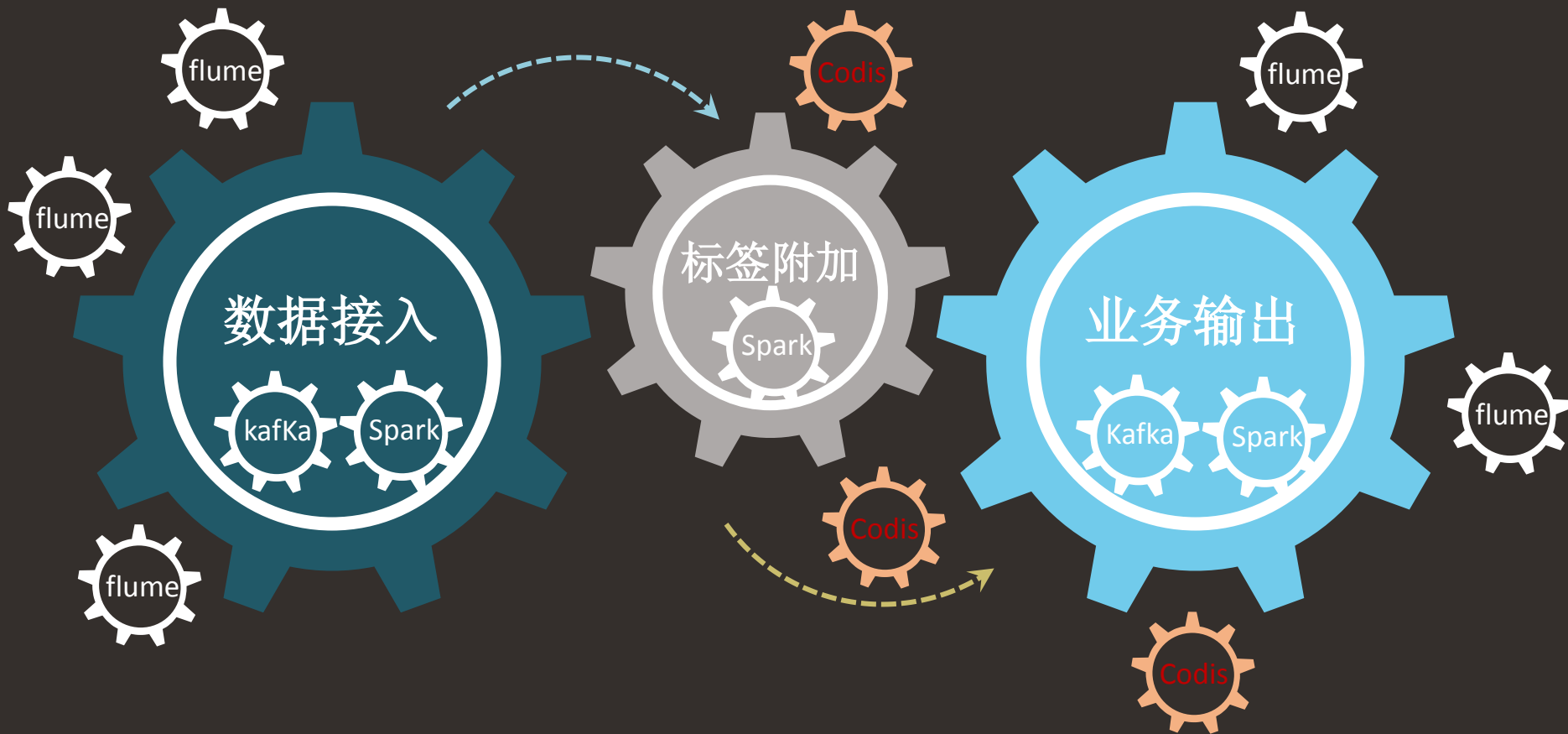
CONTENT

场景

架构

性能

演进





图一 测试环境Spark Streaming执行时间

Completed Batches (last 9 out of 9)

Batch Time	Input Size	Scheduling Delay ^(?)	Processing Time ^(?)	Total Delay ^(?)	Output Ops: Succeeded/Total
2016/12/06 15:16:00	47600 events	1 ms	5 s	5 s	1/1
2016/12/06 15:15:30	42000 events	1 ms	6 s	6 s	1/1
2016/12/06 15:15:00	48600 events	1 ms	5 s	5 s	1/1
2016/12/06 15:14:30	44800 events	1 ms	6 s	6 s	1/1

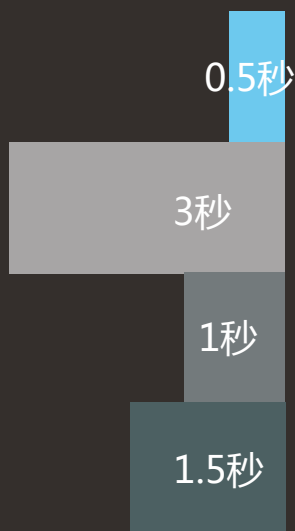
图二 任务详细执行时间

Output Op Id	Description	Duration	Status	Job Id	Duration	Stages: Succeeded/Total	Tasks (for all stages): Succeeded/Total	Error
0	foreachRDD at DataInterfaceTask.scala:69	+details 5 s	Succeeded	32	2 s	2/2	110/110	
				33	3 s	1/1	10/10	
				34	2 s	1/1	10/10	
				35	3 s	1/1	100/100	

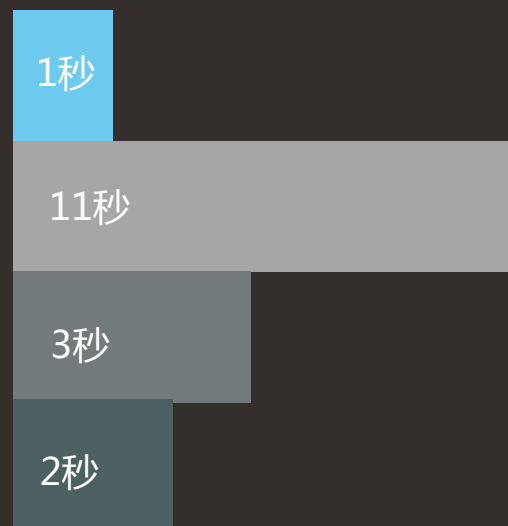


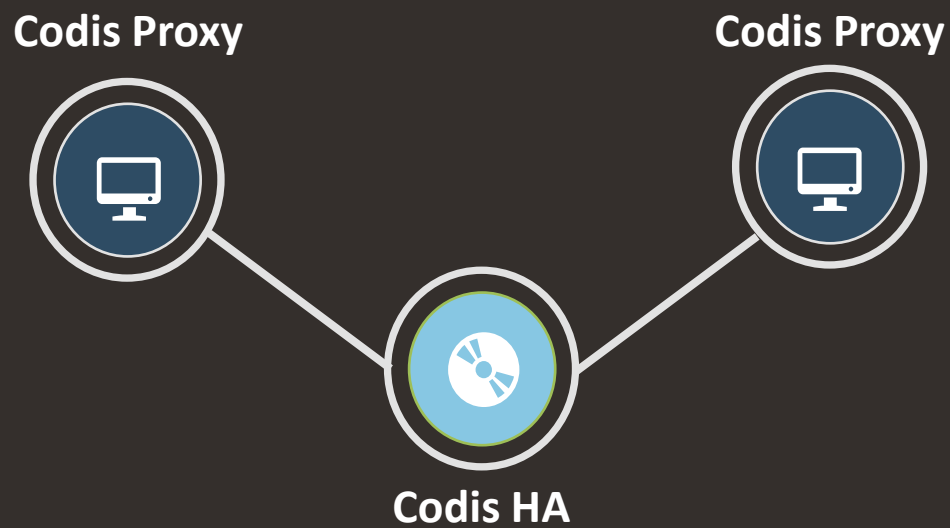
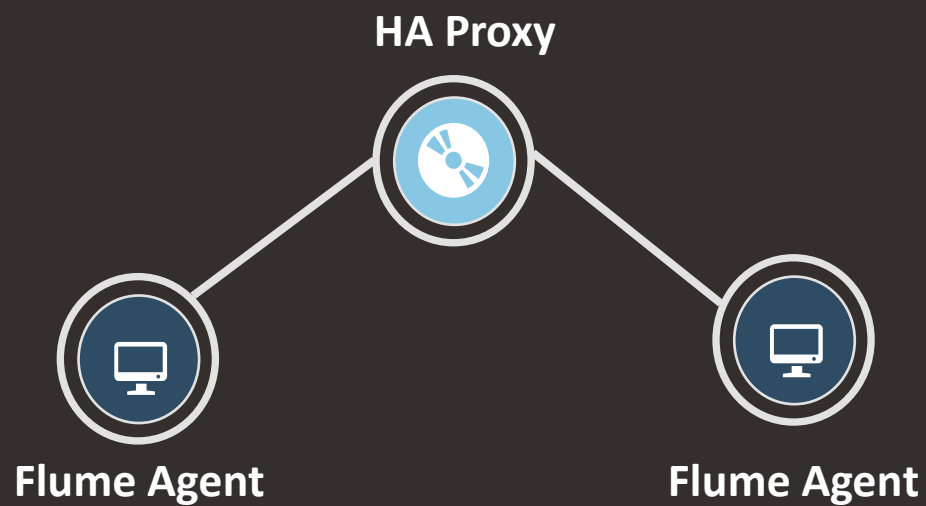
场景	30s数据量	Spark集群	Codis集群	Kafka Partition	业务数量
案例1	60万条	20/128G/32 core	10/128G	200	3
案例2	1000万条	28/512G/64 core	10/512G	1200	11

案例一 5秒



案例二 17秒





04

CONTENT

场景

架构

性能

演进



更多元化的场景

同时支持Spark streaming, Storm等主流计算引擎以支持不同的需求
支持启动多个实时处理作业, 将每个作业的结果纳入事件中心为复杂业务需求服务



更快的速度

支持Spark 2.0
支持原生redis-cluster



更健壮的架构

整合系统各组件HA

完全开源

<https://github.com/OCSP>

Thanks

杨光明子

18629652168

yangguangmz@asiainfo.com