

DAMIS

中国数据资产管理峰会

CHINA DATA ASSET MANAGEMENT SUMMIT

携程机票大数据架构最佳实践

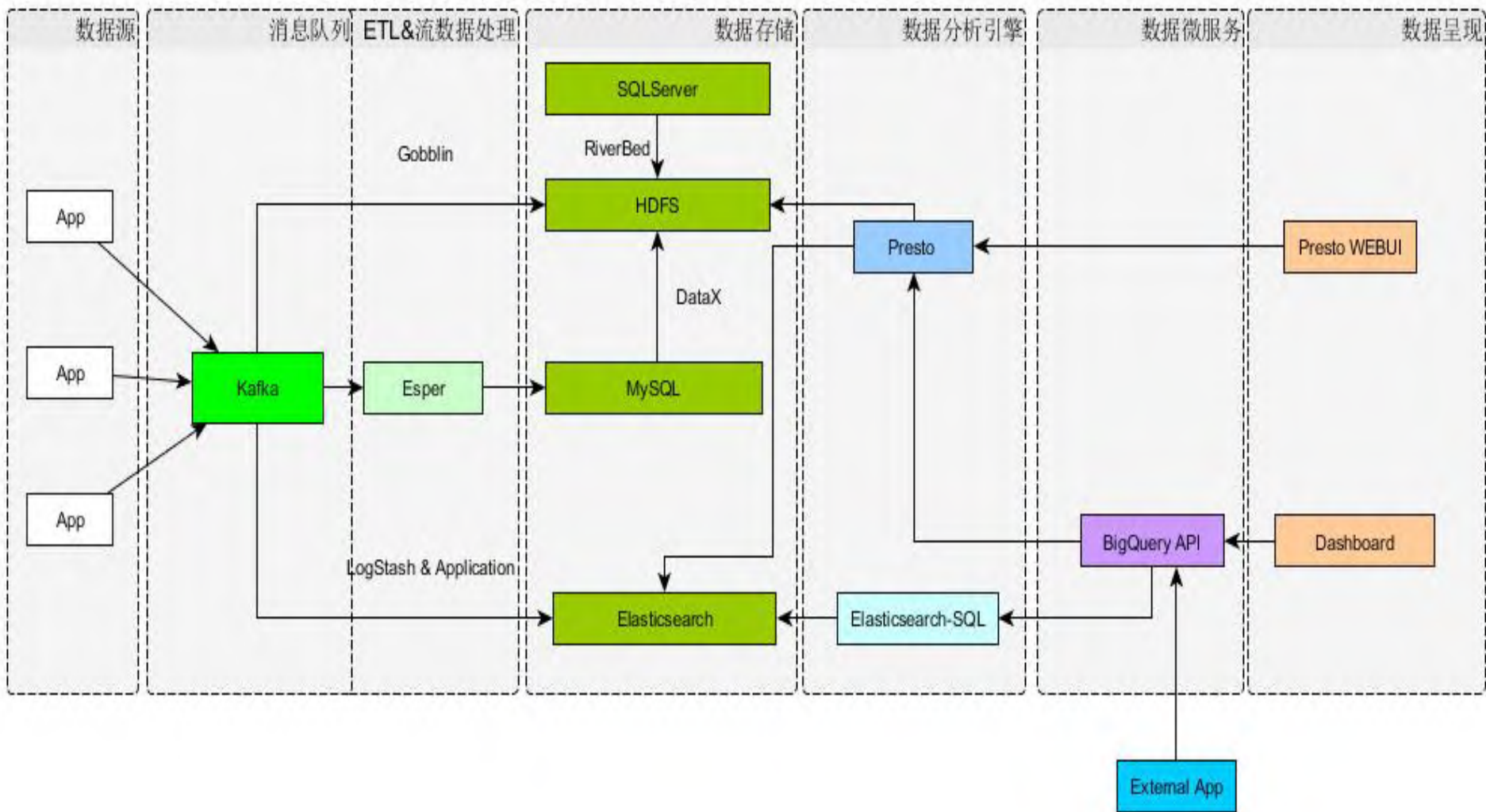
演讲人：许鹏

目录

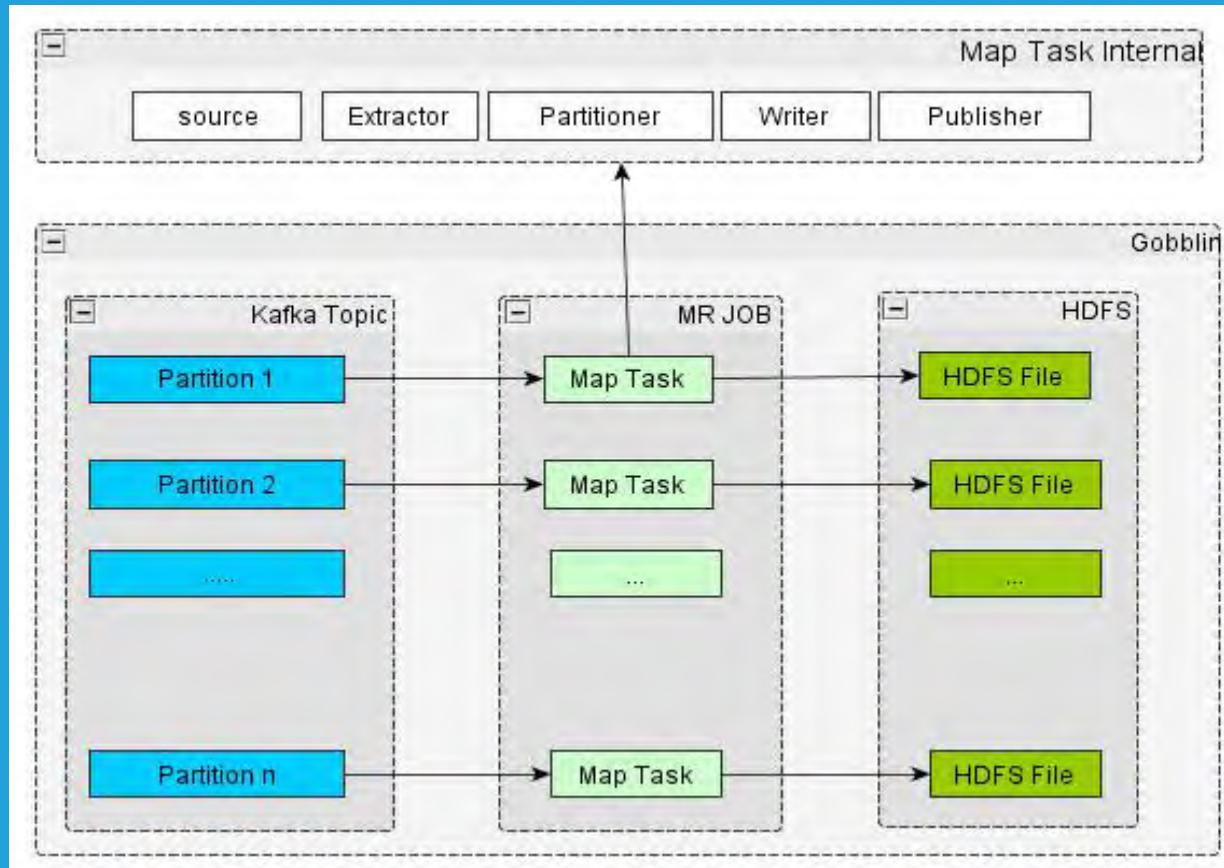
- 数据平台技术选型
- 用户交互
- 项目实践和运维



整体框架



ETL PipeLine -- Gobblin

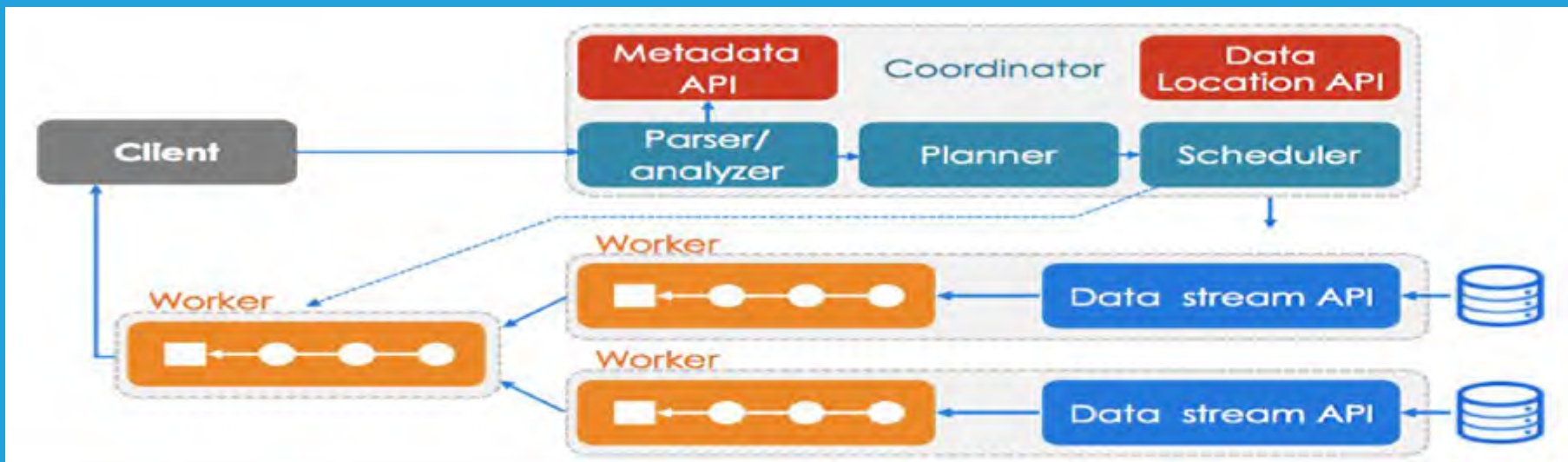


JSON -> Avro -> ORC



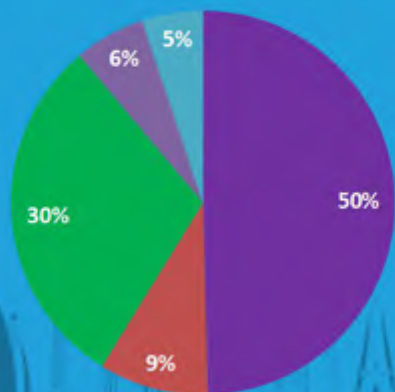
JSON -> ORC

分析引擎 - Presto



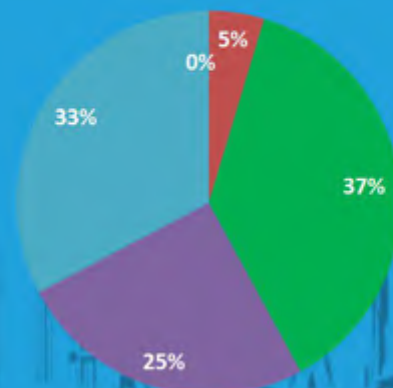
PRESTO 响应时间分布(毫秒)

■ <500 ■ 500~1000 ■ 1000~10000 ■ 10000~60000 ■ >60000



HIVE 响应时间分布(毫秒)

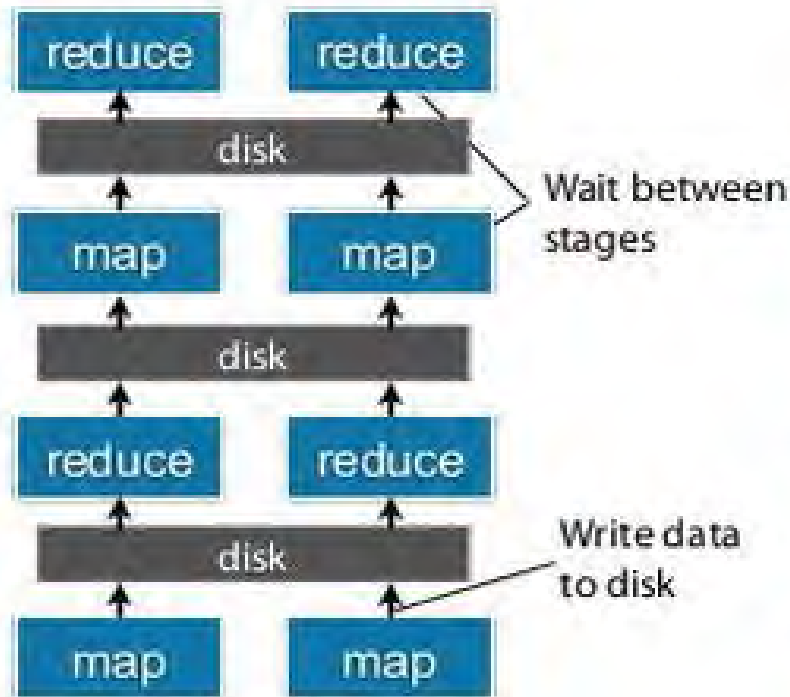
■ <500 ■ 500~1000 ■ 1000~10000 ■ 10000~60000 ■ >60000



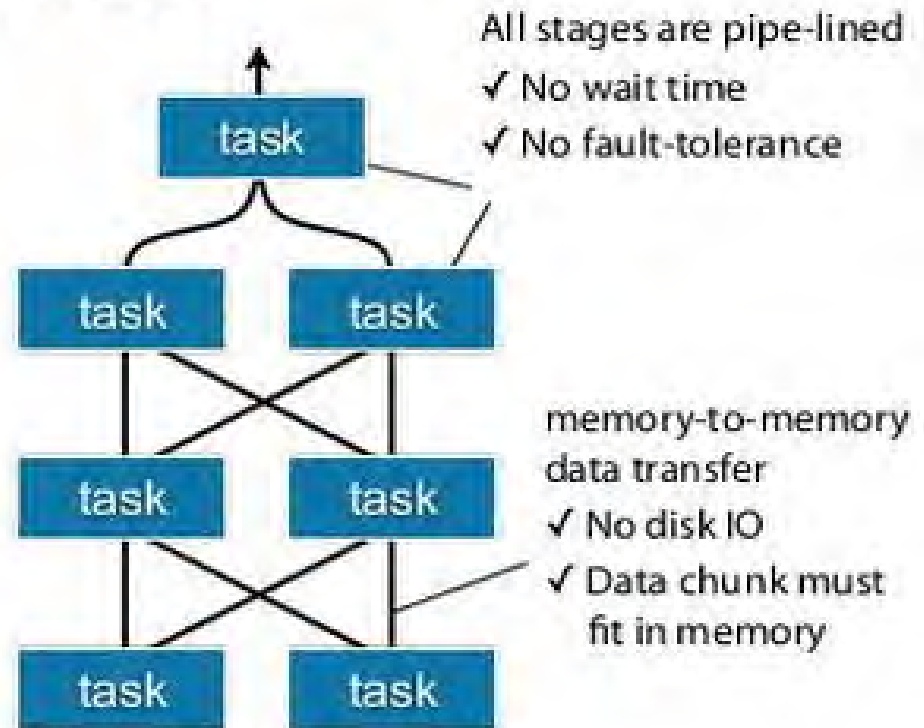
分析引擎比较 -- Presto

- 流水线式
- 操作基于内存
- 支持多租户模式，资源复用率高

MapReduce



Presto



近实时搜索 - Elasticsearch

- 近实时搜索引擎
- 支持横向扩展
- 高可用
- 易于管理
- 社群活跃，有专门的商业公司支持

Cluster



Legend



ES调优和运维

OS Level

- File Handler
 - ulimit
- Memory
 - turn off swappiness
- I/O Scheduler
 - CFQ
 - vm.dirty_ratio
 - vm.dirty_background_ratio

Elasticsearch

- Cluster Level
 - Shard allocation
 - Concurrent recovery
- Index Level
 - Shard
 - Replica
 - Refresh
 - Merge
 - Schema Design
 - Balance indexing and query

name	load_5m	heap.percent	cpu	indexing	search	ip	node.role	master	shard_num	disk.avail
	15.92	64	21	1575	82		d	-	477	14.6tb
	13.95	61	16	1812	71		d	-	477	19.6tb
	13.62	57	8	1724	74		d	-	476	10tb
	12.29	61	9	1473	69		d	-	477	18.4tb
	12.01	67	10	1853	79		d	-	478	11.3tb
	11.81	74	10	1961	70		d	-	477	15.3tb
	10.95	67	11	1737	69		d	-	477	43.7tb
	10.50	53	20	1602	81		d	-	478	16.7tb
	9.67	72	7	1803	76		mdi	-	475	9.3tb
	9.32	61	9	1602	72		d	-	478	16.8tb
	9.06	60	5	1726	71		d	-	477	17.8tb
	9.01	64	17	1646	80		d	-	477	11.5tb
	9.00	55	13	1607	70		di	-	478	12.8tb
	8.92	71	7	1720	57		d	-	477	14.5tb
	8.15	71	9	1883	50		md	-	478	7.9tb

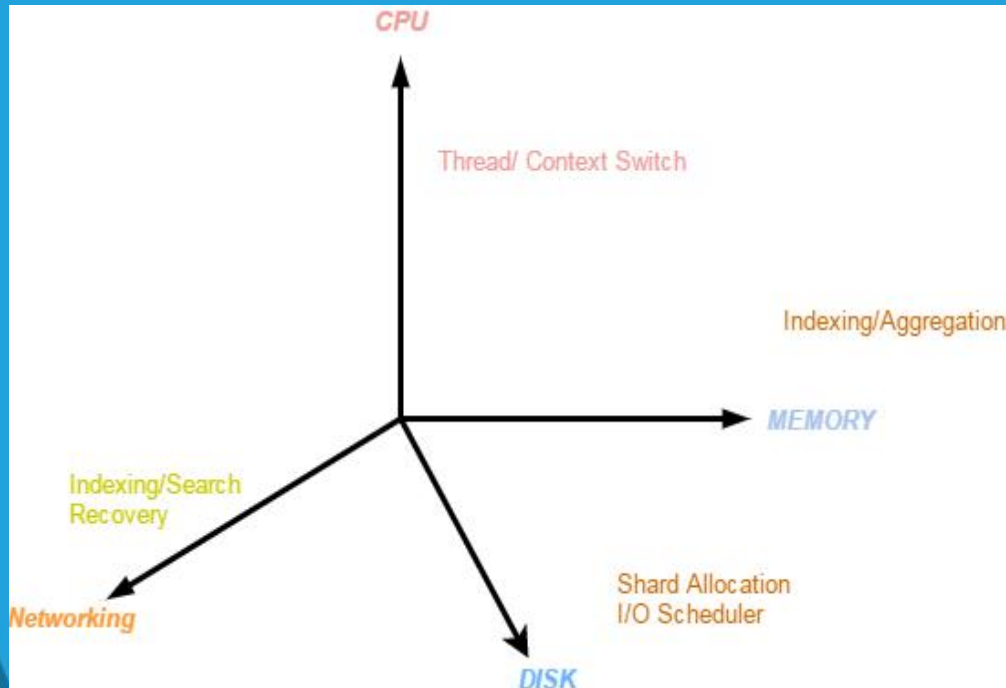
ES调优和运维 continuous

OS Level

- File Handler
 - ulimit
- Memory
 - turn off swappiness
- I/O Scheduler
 - CFQ
 - vm.dirty_ratio
 - vm.dirty_background_ratio

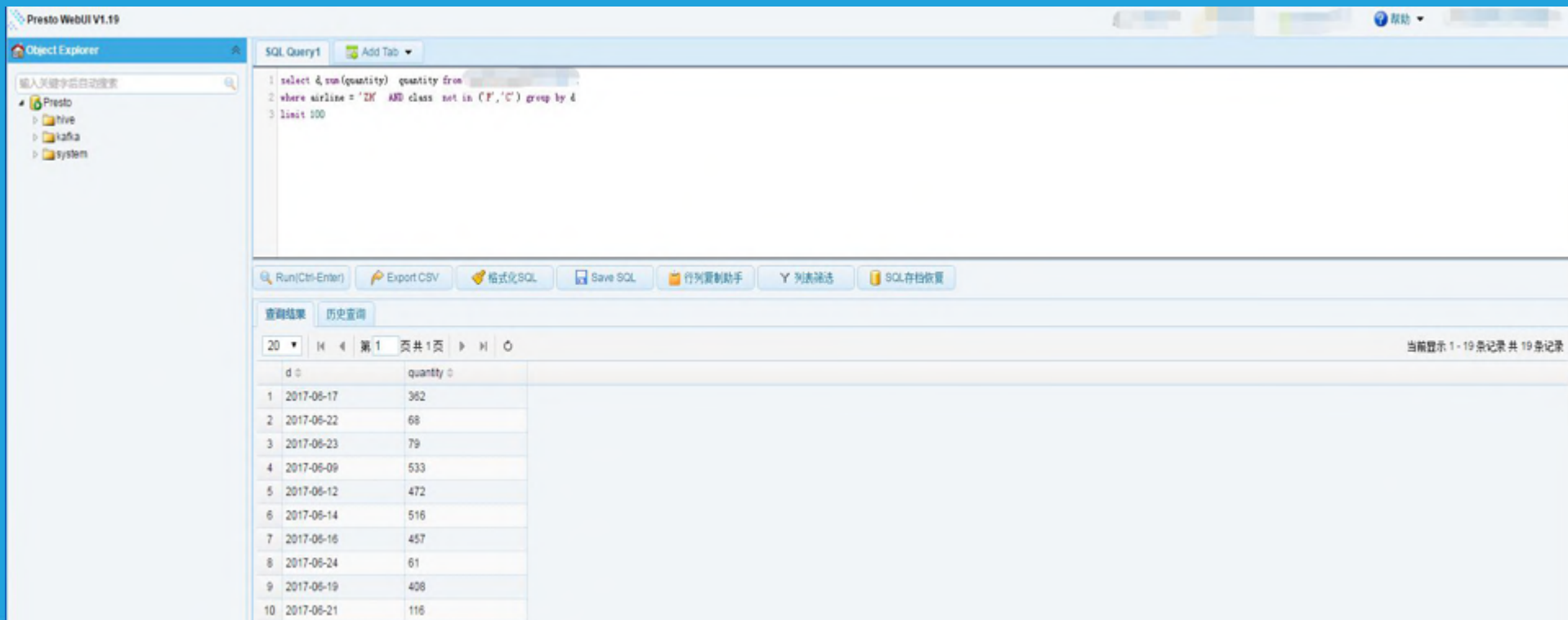
Elasticsearch

- Cluster Level
 - Shard allocation
 - Concurrent recovery
- Index Level
 - Shard
 - Replica
 - Refresh
 - Merge
 - Schema Design
 - Balance indexing and query



统一化的查询接口 - SQL

- 降低用户学习成本
- 缩短产品推广时间
- 引入Elasticsearch-SQL，支持使用SQL语法对ES进行查询
- ESPER/Meteo, 支持类SQL语句，对数据进行实时聚合，易于配置



The screenshot displays the Presto WebUI V1.19 interface. On the left, the 'Object Explorer' shows a tree view with 'Presto', 'hive', 'kafka', and 'system' folders. The main area shows a SQL query editor with the following text:

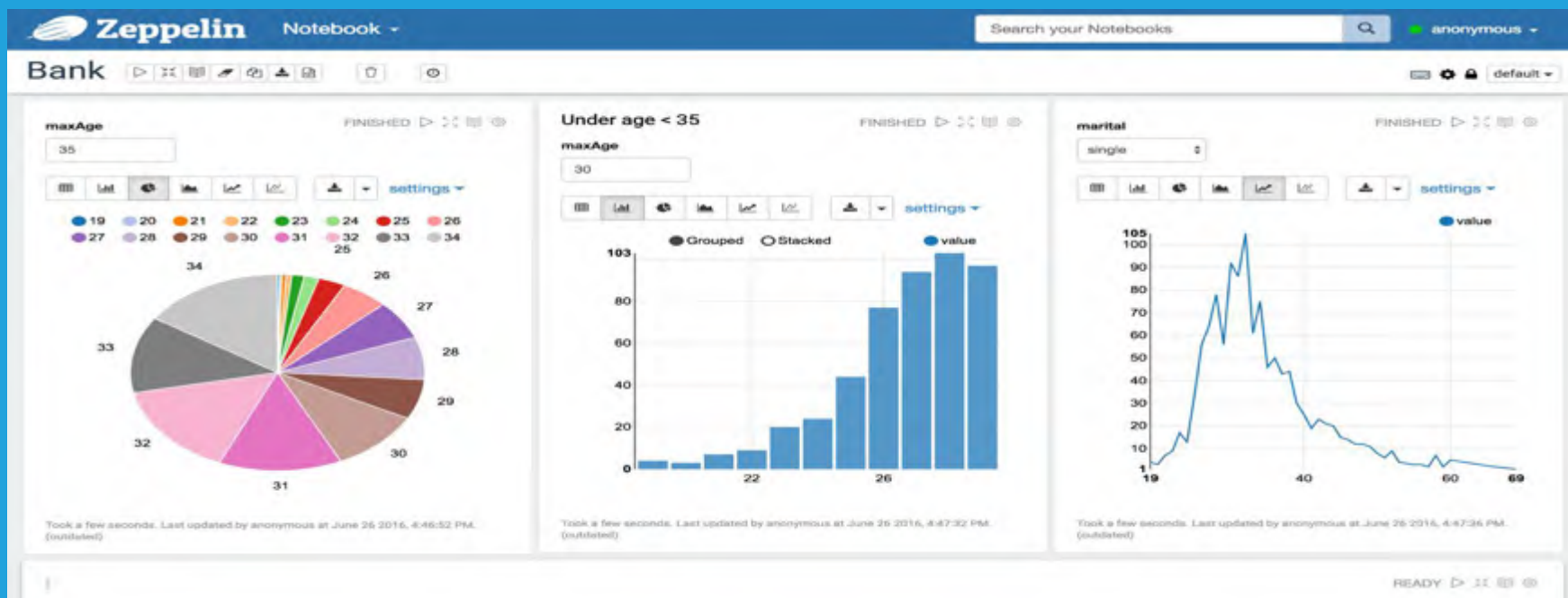
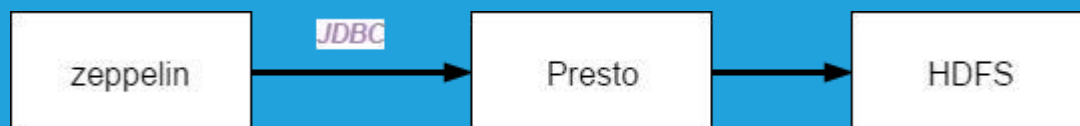
```
1 select & sum(quantity) quantity from   
2 where airline = 'ZH' AND class not in ('F','C') group by d   
3 limit 100
```

Below the editor, there are buttons for 'Run(Ctrl-Enter)', 'Export CSV', '格式化SQL', 'Save SQL', '行列限制助手', 'Y 列表筛选', and 'SQL存档恢复'. The '查询结果' (Query Results) tab is active, showing a table with 10 rows of data. The table has two columns: 'd' and 'quantity'. The data is as follows:

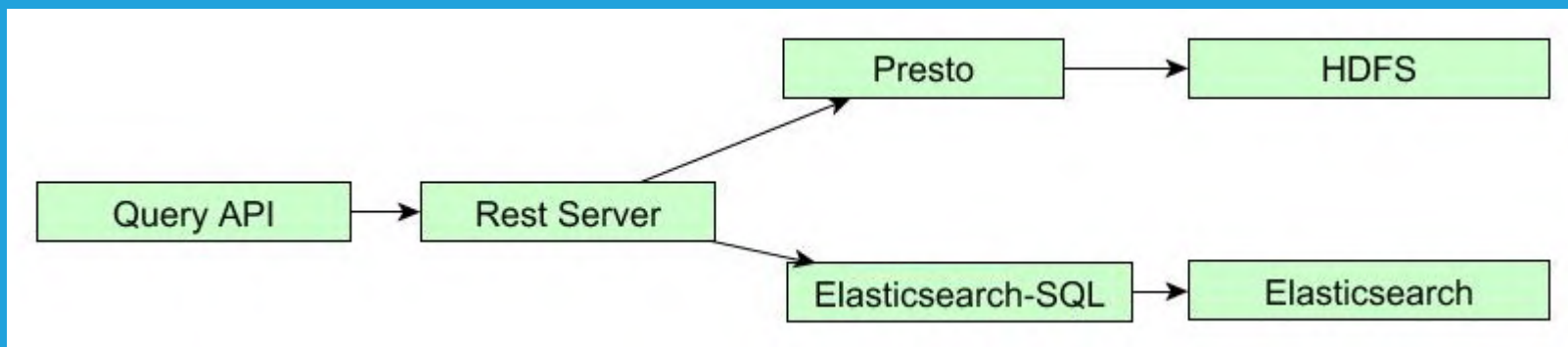
	d	quantity
1	2017-06-17	362
2	2017-06-22	68
3	2017-06-23	79
4	2017-06-09	533
5	2017-06-12	472
6	2017-06-14	516
7	2017-06-16	457
8	2017-06-24	61
9	2017-06-19	408
10	2017-06-21	116

At the bottom right of the results area, it says '当前显示 1 - 19 条记录 共 19 条记录'.

数据可视化 -- Zeppelin



数据微服务 – Rest查询接口



- 统一的查询入口
- 统一的权限管理
- 统一的查询语言

任务调度器 – Job Scheduler

- Zeus <https://github.com/ctripcorp/dataworks-zeus>

Scheduler调度内核

宙斯支持复杂依赖的任务调度
依赖任务零延时触发
失败任务连锁报警

以下为依赖任务举例：

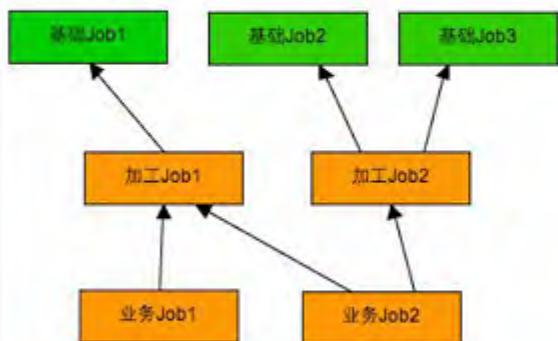
图案示意说明

独立Job：可以设定调度周期

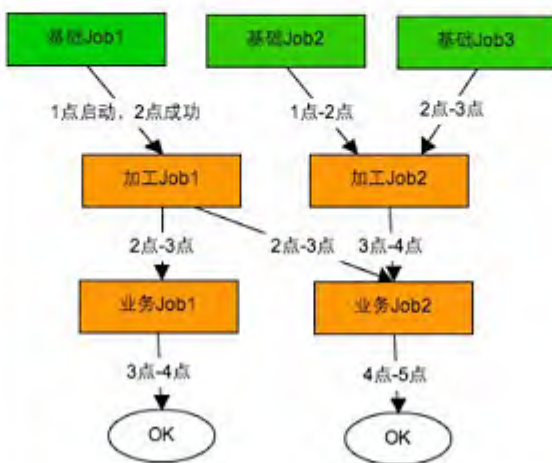
依赖Job：不能设定调度周期，依赖任务运行成功后自动触发调度

失败的Job，并且发出报警

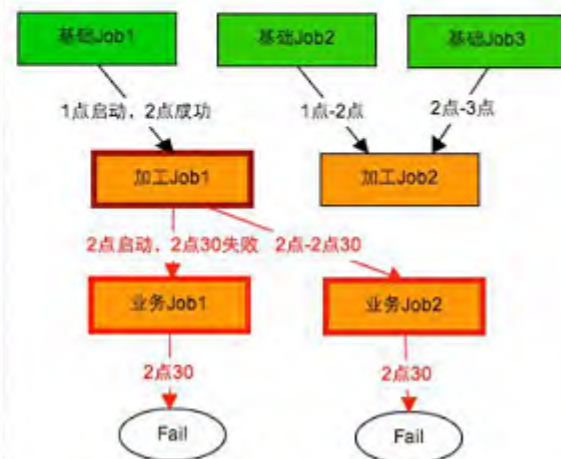
Job依赖关系图



成功的Job调度时序图



失败的Job调度时序图



数据团队能力建设

	后台技术	前端技术和框架	Linux知识
引擎开发	√		
交互界面设计		√	
平台运维	√		√
平台规划	√	√	√
开发语言	Java/Scala/Python	Javascript, React.js	



DAMIS

中国数据资产管理峰会

CHINA DATA ASSET MANAGEMENT SUMMIT

THANK YOU !

