苏宁OLAP 引擎发展之路

赵云





Geekbang》. TGO 銀鹏會

全球技术领导力峰会

500+高端科技领导者与你一起探讨技术、管理与商业那些事儿



⑤ 2019年6月14-15日 | ⑥上海圣诺亚皇冠假日酒店



扫码了解更多信息

苏宁OLAP引擎负责人 苏宁大数据平台负责人







统一报表门户



统一指标口径



统一模型视图



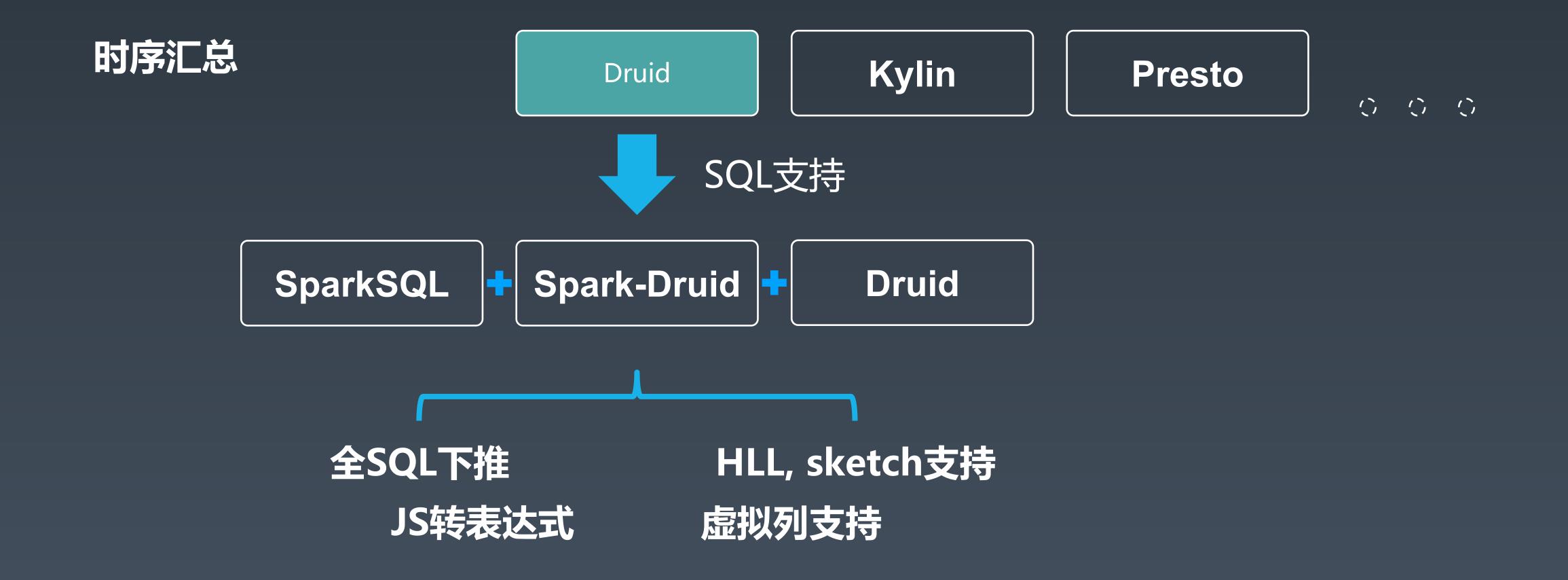


统一SQL引擎

- 河 演进和优化
- 回整体架构



Druid



Druid优化



实时peon OOM优化



动态查询优先级



大查询限制



明细



PG

非时序汇总

PG Table

PG PG PG Table B

SQL
LogicalPlan



PG Optimizer

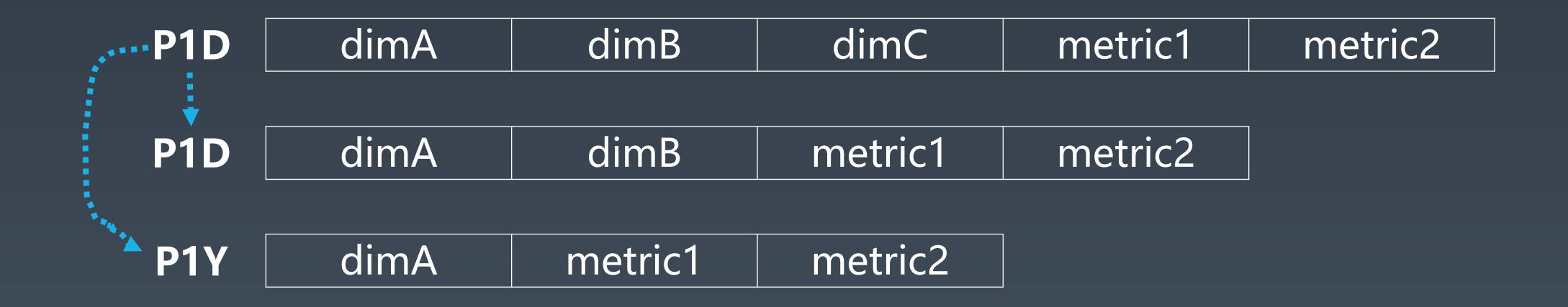


SQL

日期格式转换 字段类型转换 AB表替换

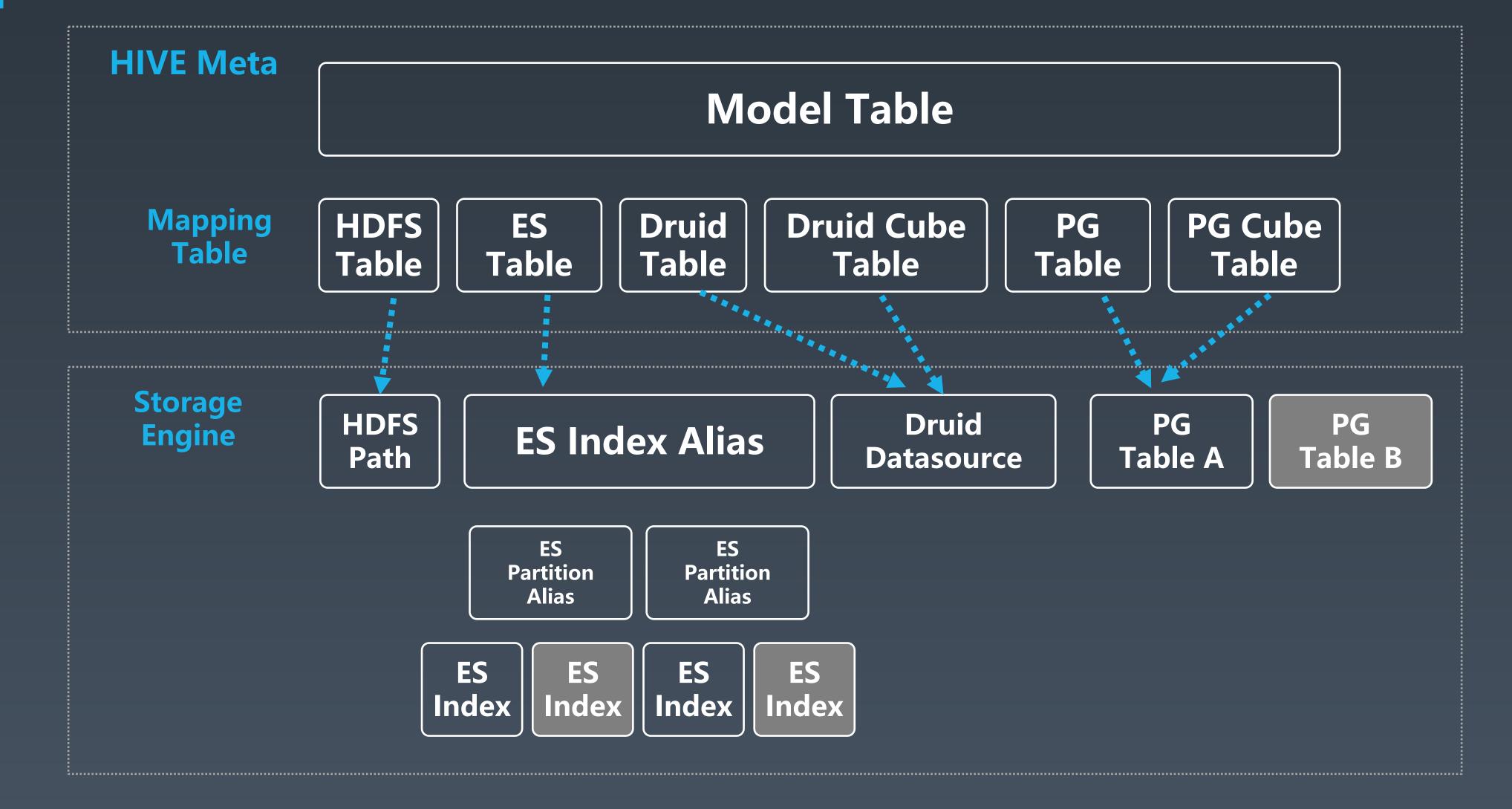
CUBE

在汇总模型上建立cube

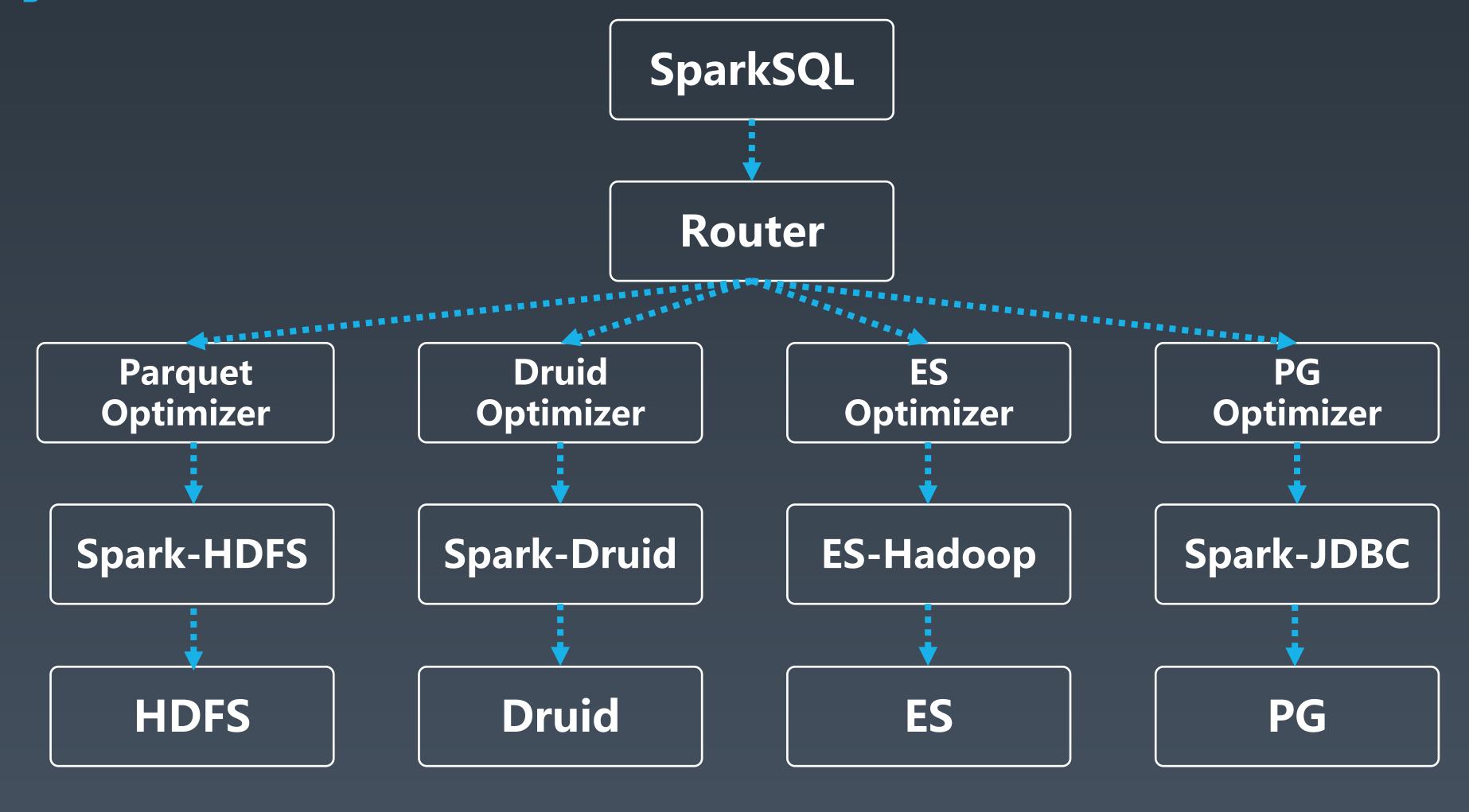


根据粒度、维度等路由

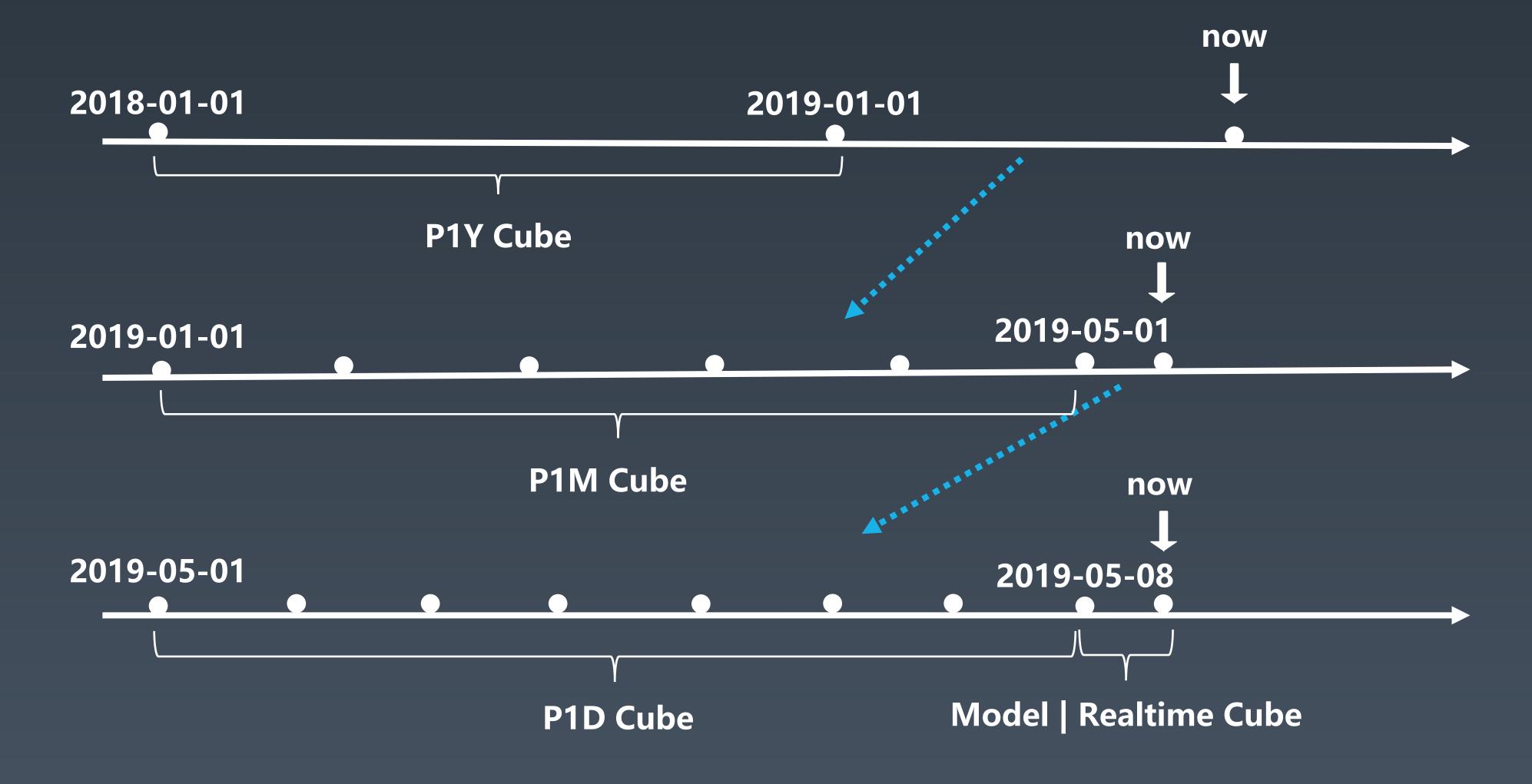
元数据



查询引擎



分段汇总





预计算

复杂查询

dimA, dimB, dimC	111
dimC	001
dimA, dimC	101

预定义SQL

select
from ...

•••

group by ??

......

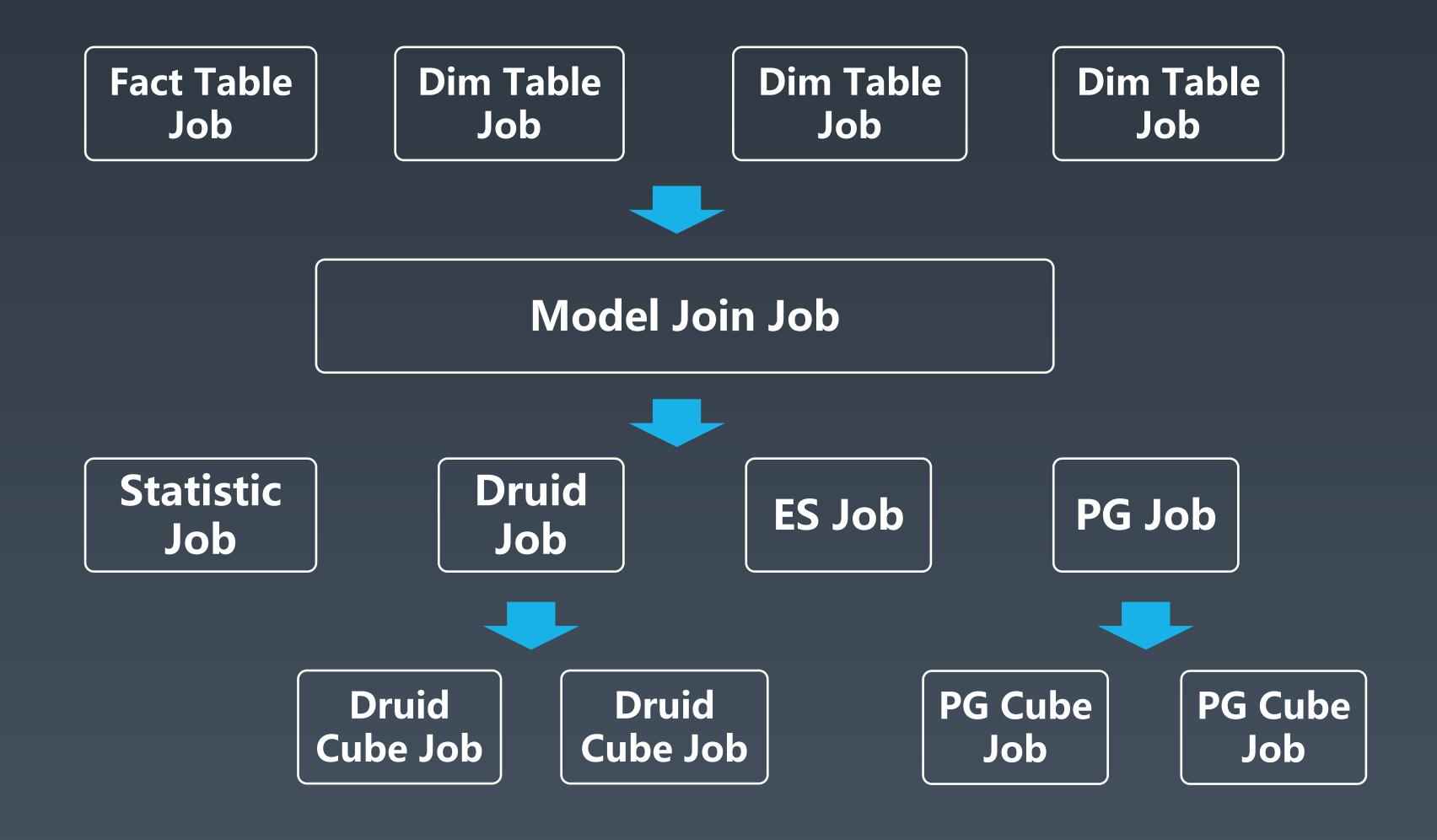
....

cuboid	dimA	dimB	dimC	metric1	metric2
111	XX	XX	XX	XX	XX
111	XX	XX	XX	XX	XX
001			XX	XX	XX
001			XX	XX	XX
101	XX		XX	XX	XX

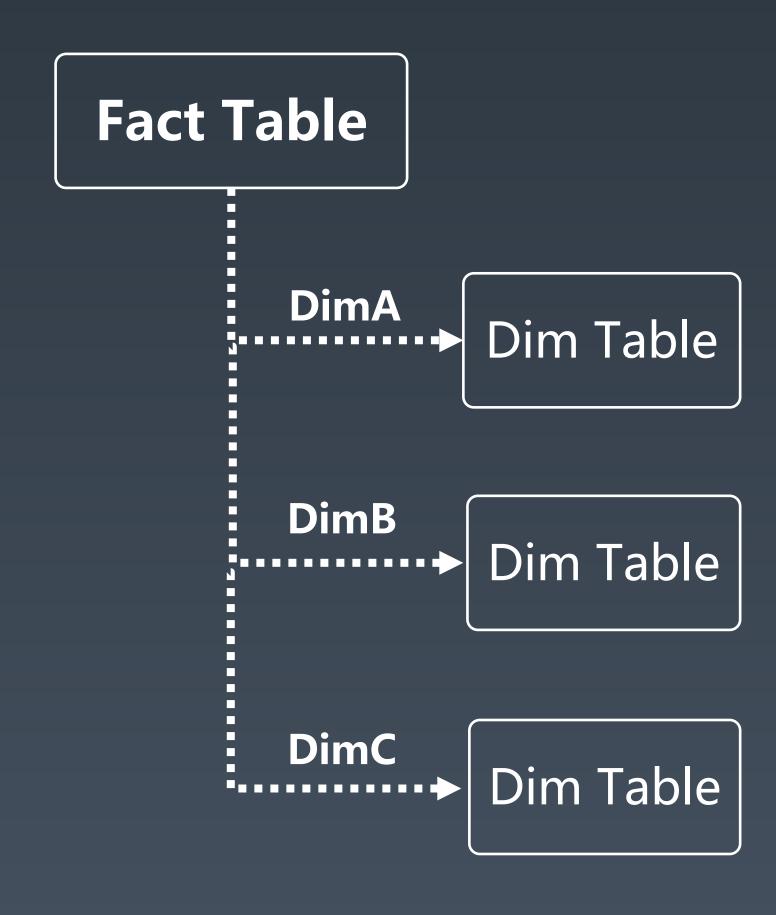
COST

采集 Histogram **Count Distinct(HLL)** Count where group by having order by select 计算 应用 报表开发 熔断 路由 隔离

离线任务



数据倾斜优化



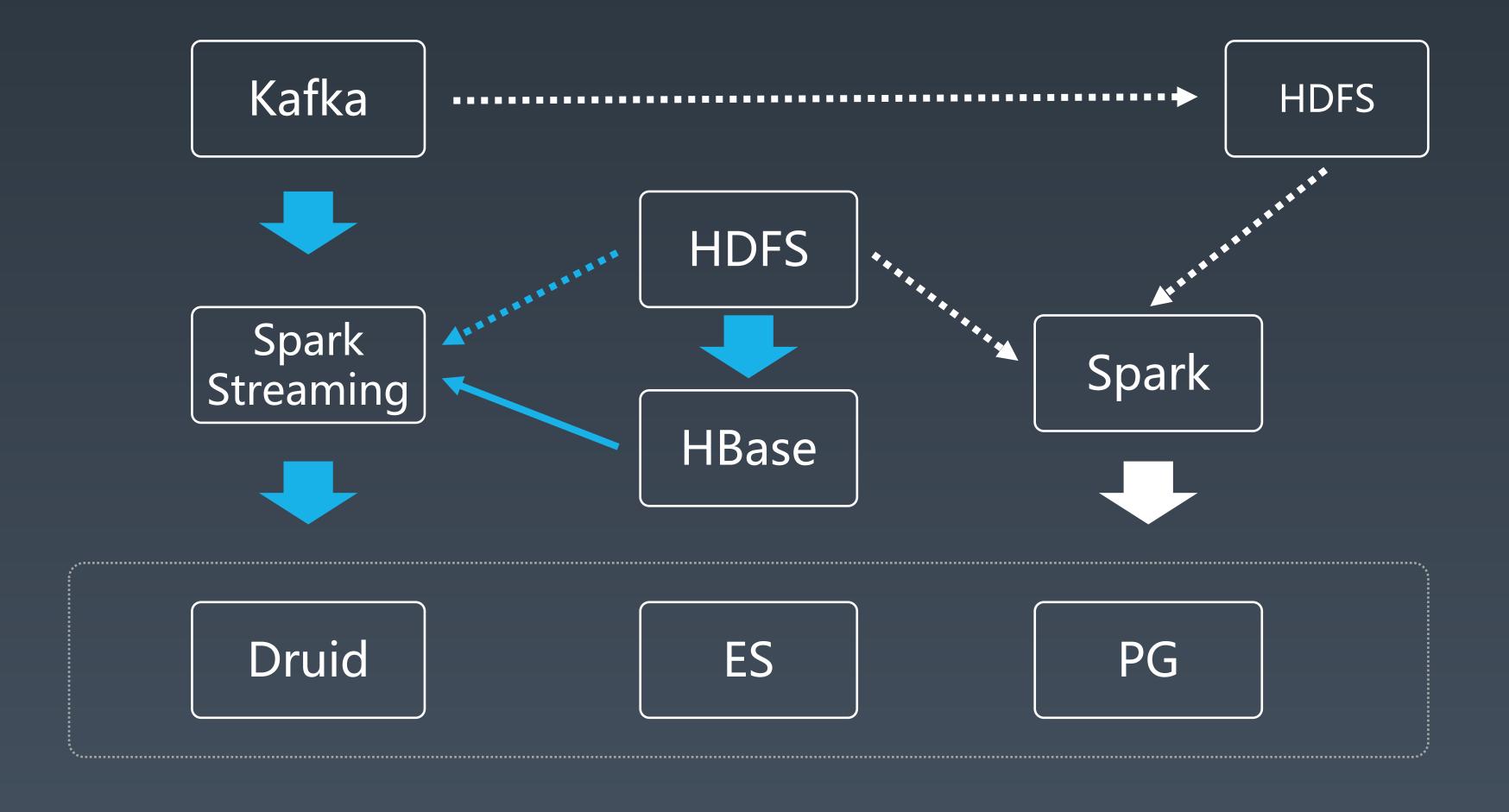
统一关联字段采样

小维表 Broadcast

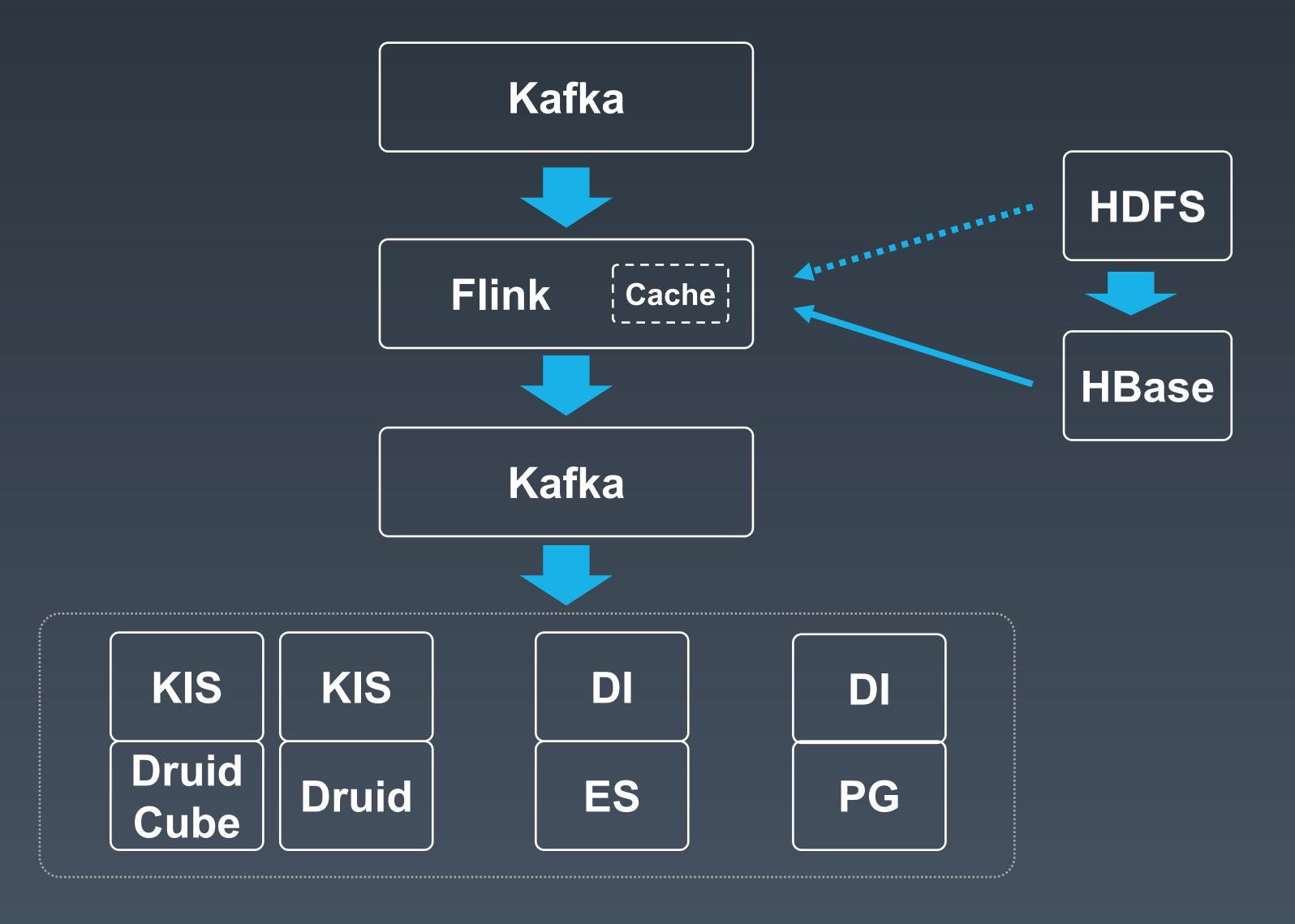
大维表分2段处理



实时任务



实时任务演进



OLAP整体架构

离线任务

实时任务

统计任务

元数据 服务 异步接口 同步接口 系统管理 数据集 数据流 调度 CUBE 模型 资源隔离 扩缩容 HA管理 资源调度 Cost 任务 查询 解析 路由 引擎

优化

文件存储

存储引擎

缓存

汇总存储

Cost服务 基数查询

RDB

明细存储

流控

服务 监控

监控

APM

日志

监控

任务

监控

苏宁大数据

数据应用产品 自主分析 数据预测 业务数据产品 报表门户 数据中台 开发工具 标签开发 报表开发 数据质 数 数据安全管 指标 标签 据 维度开发 指标开发 资 量管理 7 实时任务 离线任务 管 数据仓库 理 理 数据采集 管理平台 离线计算 实时计算 存储 搜索 OLAP 大数据平台 引擎 引擎 引擎



极客邦科技 会议推荐2019

ArchSummit

深圳

全球架构师峰会

大会: 7月12-13日 培训: 7月14-15日 **Arch**Summit

北京

全球架构师峰会

大会: 12月6-7日 培训: 12月8-9日

5月

6月

7月

10月

11月

12月

QCon

北京

广州

全球软件开发大会

大会: 5月6-8日 培训: 5月9-10日

QCon

全球软件开发大会

培训: 5月25-26日 大会: 5月27-28日 GTLC GLOBAL TECH LEADERSHIP CONFERENCE

上海

技术领导力峰会

时间: 6月14-15日

GMITC

北京

全球大前端技术大会

大会: 6月20-21日 培训: 6月22-23日 QCon

上海

全球软件开发大会

大会: 10月17-19日培训: 10月20-21日

GMITC

深圳

全球大前端技术大会

大会: 11月8-9日 培训: 11月10-11日

AiCon

北京

全球人工智能与机器学习大会

大会: 11月21-22日 培训: 11月23-24日

THANKS! QCon O