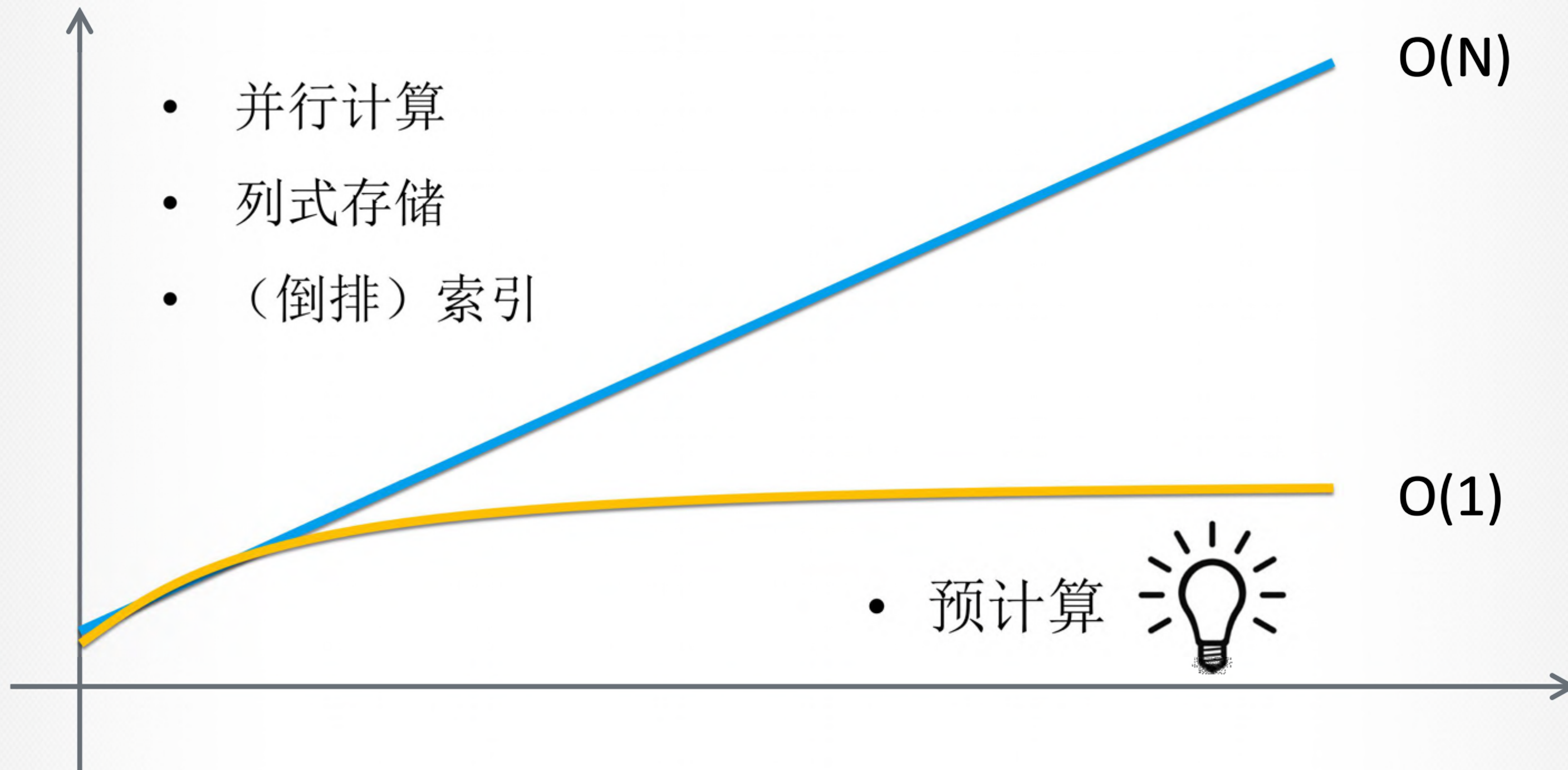


Apache Kylin v1.6 之 新Streaming OLAP实现

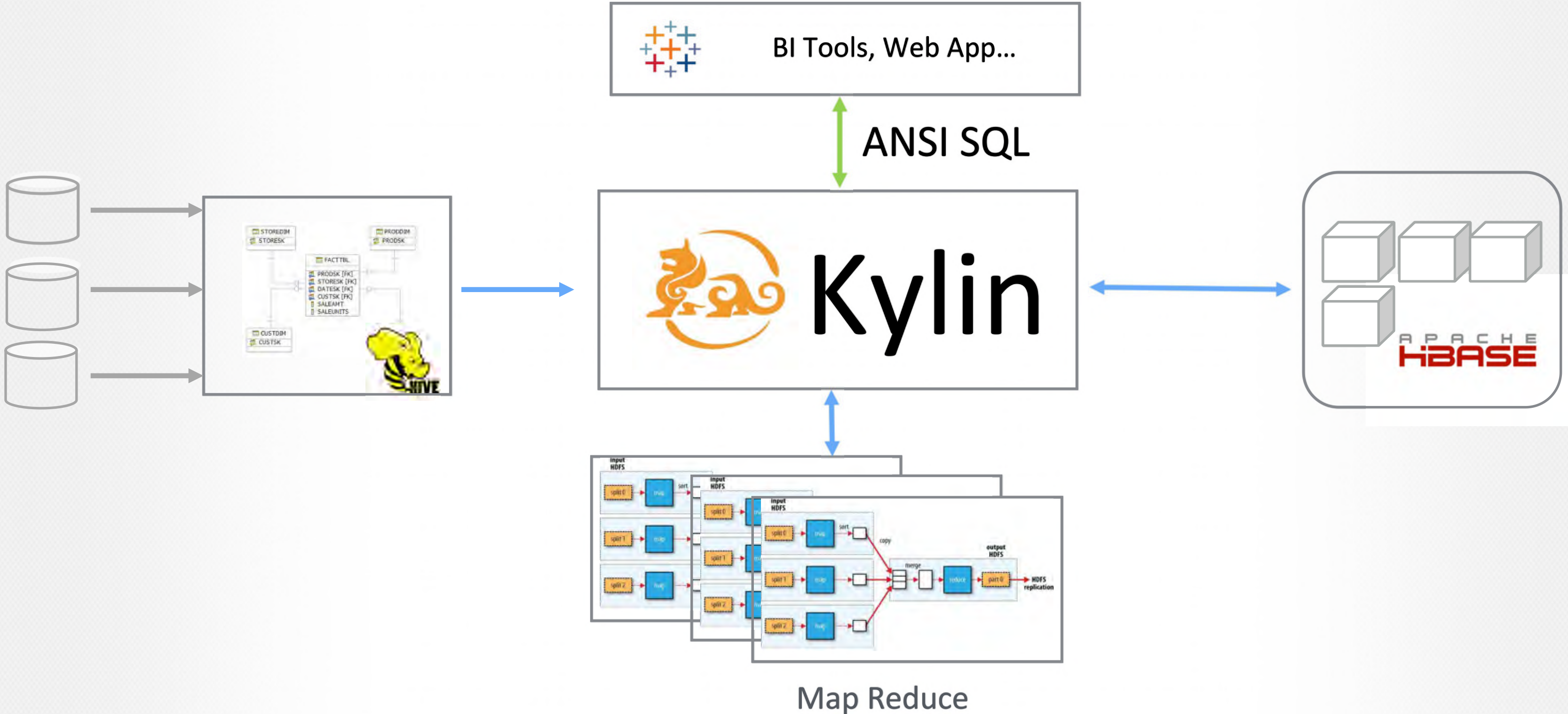
Agenda

- Apache Kylin架构回顾
- V1.5的Streaming尝试
- V1.6的新Streaming实现
- 未来展望

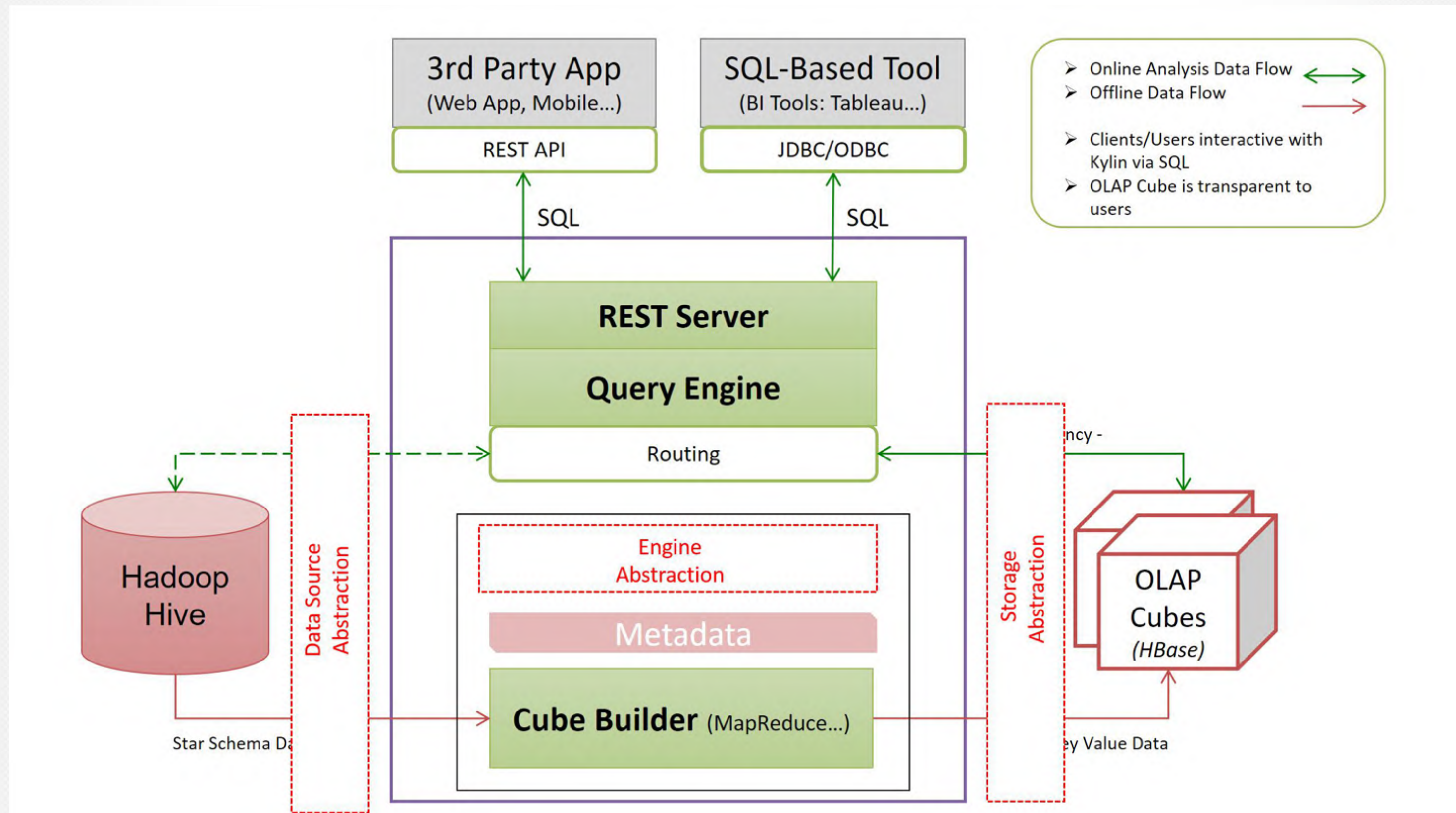
Kylin的目标



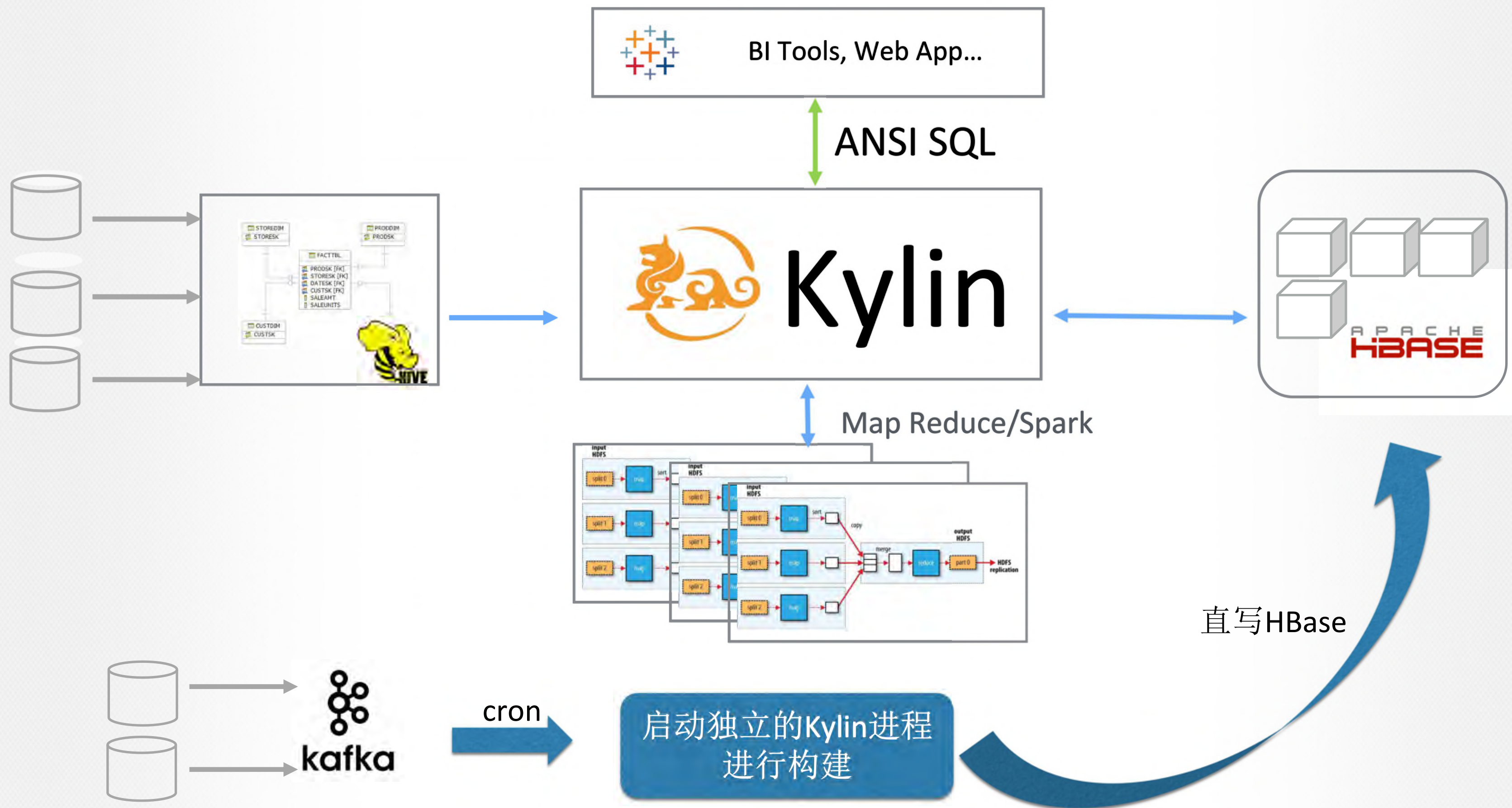
Kylin的架构



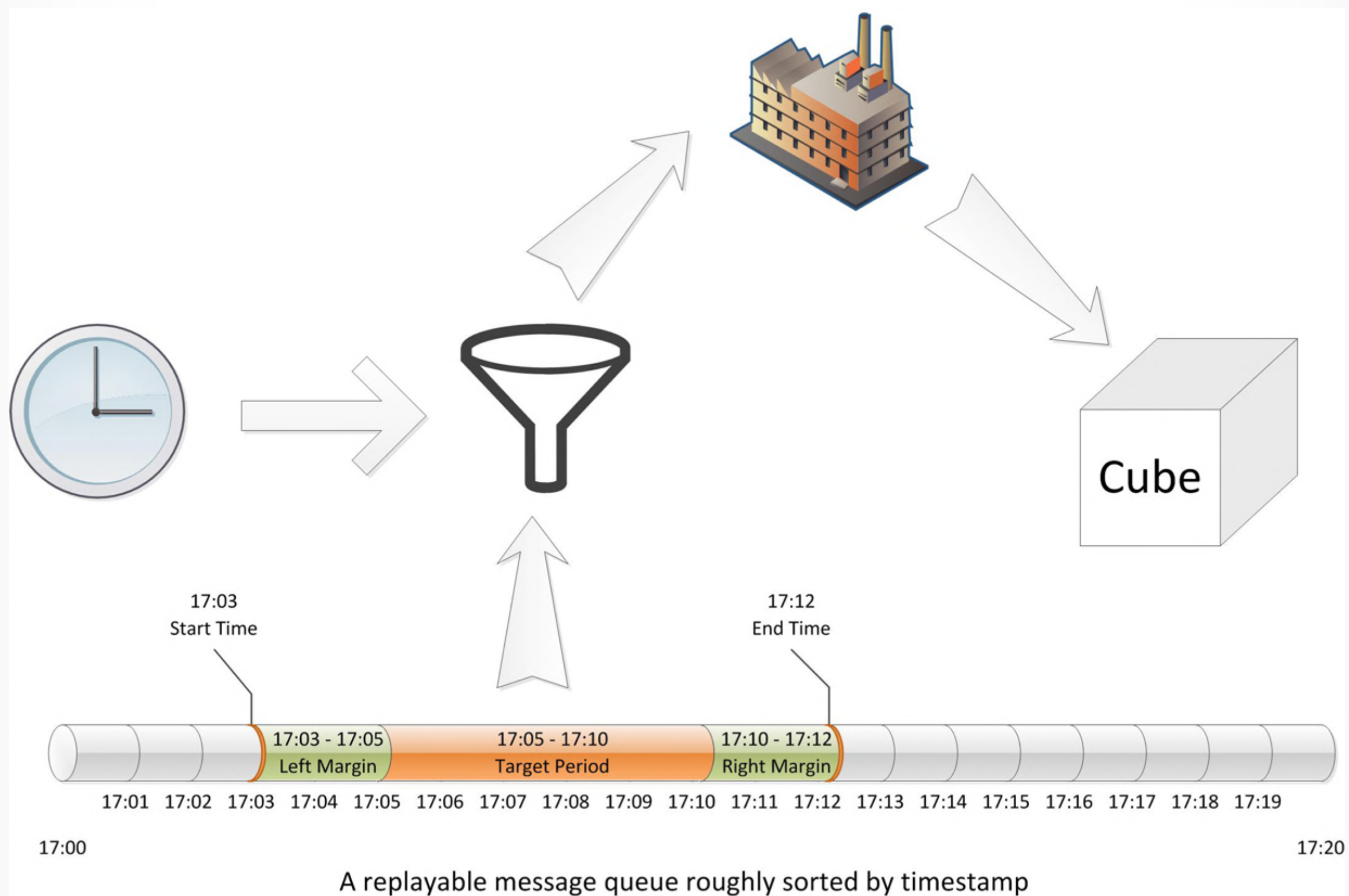
V1.5 核心组件解耦



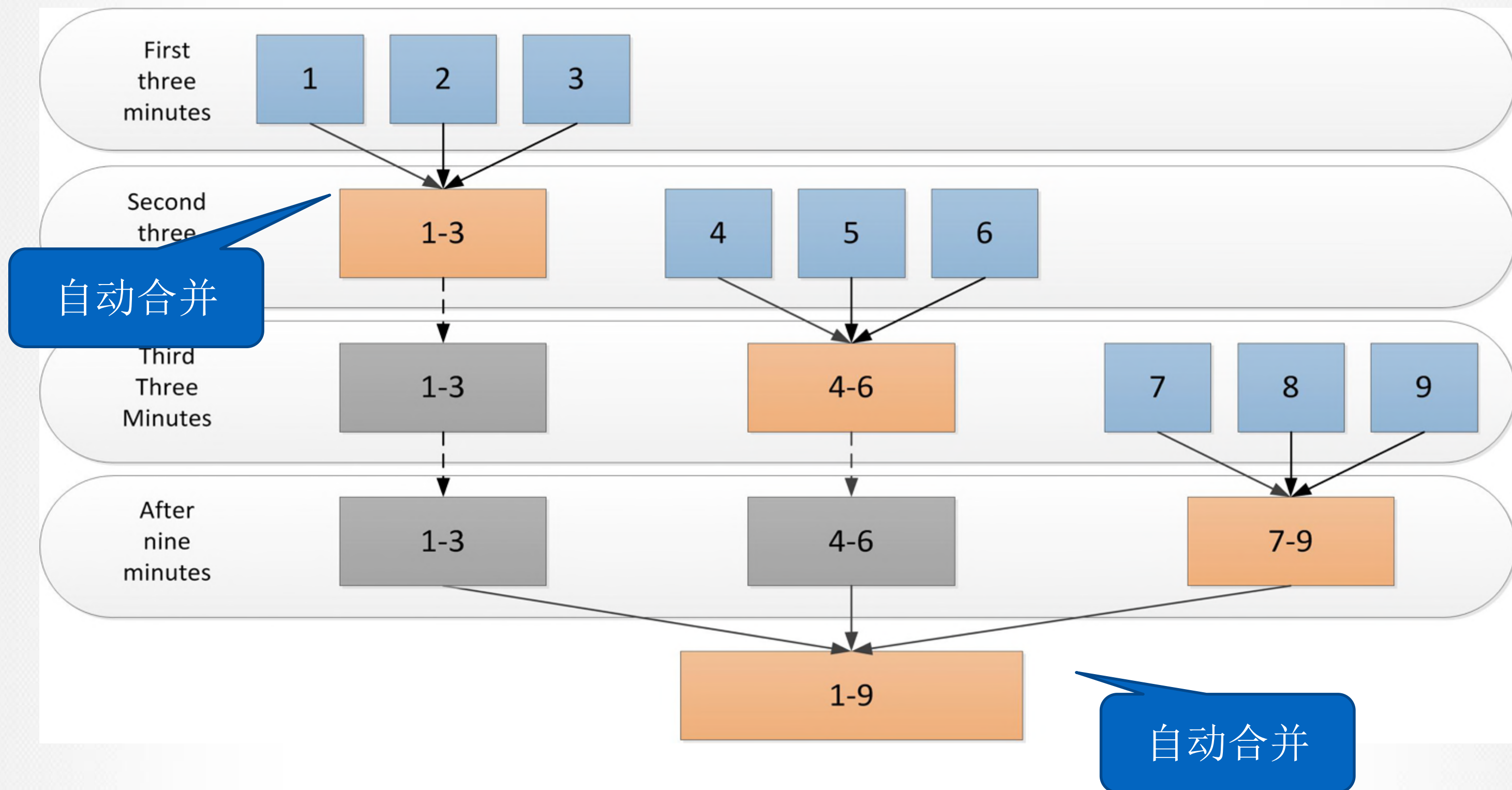
V1.5的Streaming



流数据分段 - 使用时间近似寻找



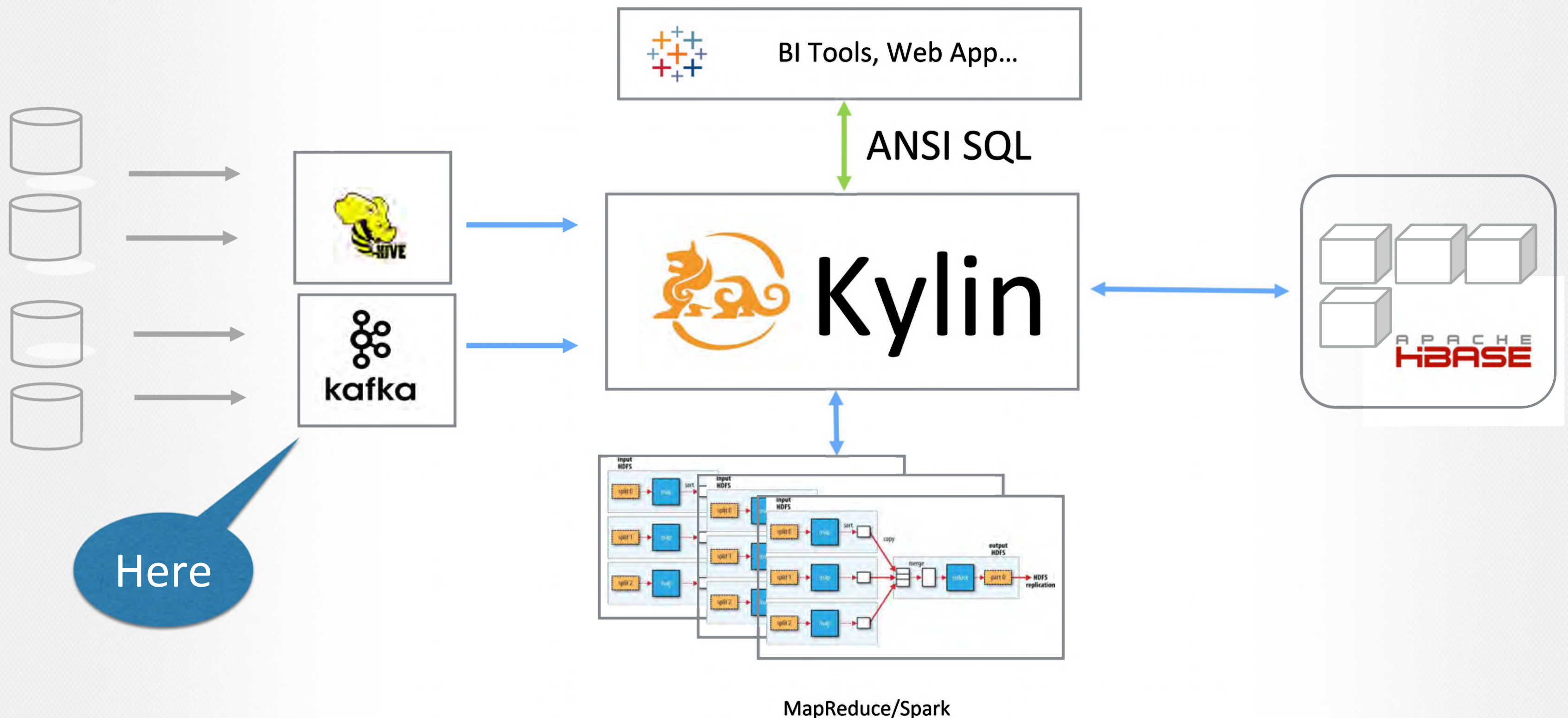
小碎片逐渐合并成大Segment



收获及缺陷

- 解决了从无到有的问题
- 构建不能自动伸缩
- 近似二分查找会丢失数据
- 构建任务难以监控
- 错误恢复困难
- 整体运维成本高

v1.6 Streaming: 可扩展架构下的完美实现

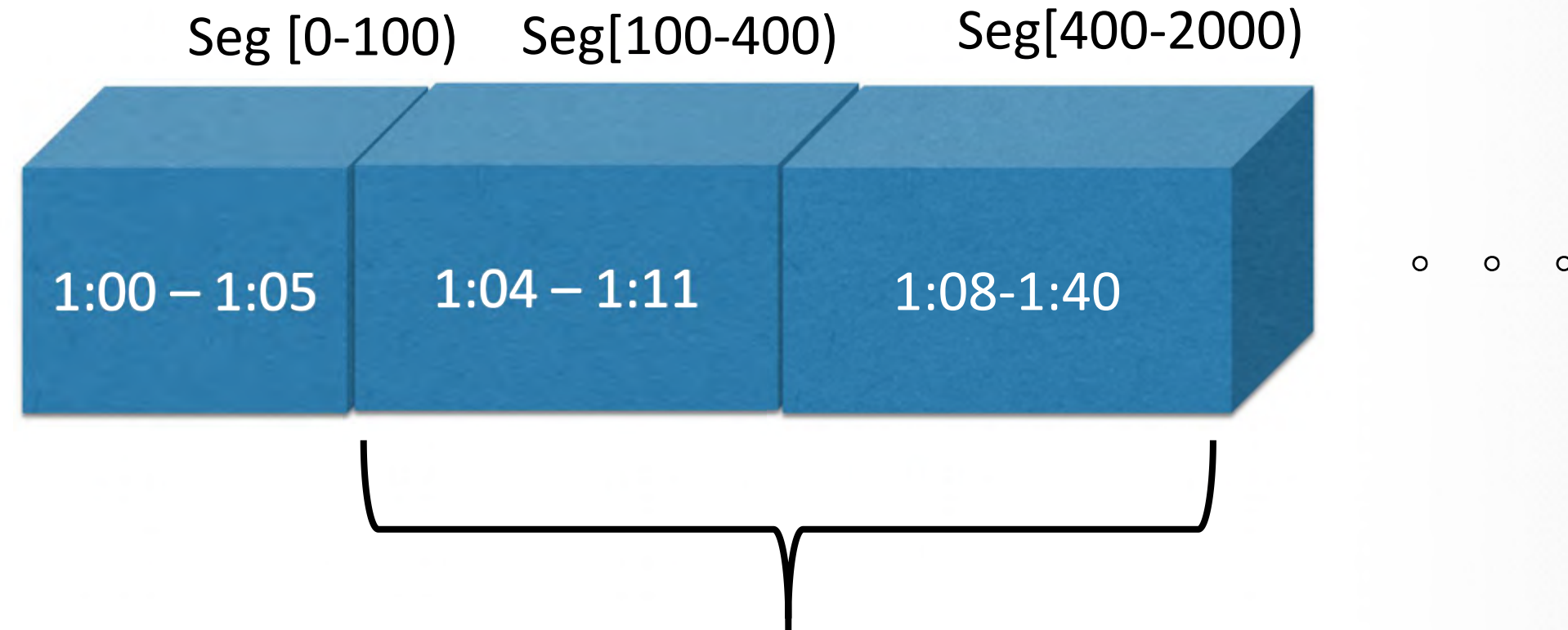


完全可伸缩

- 重用既有MR构建引擎
- 从几千条到几亿条数据，一次轻松构建
- 可随意暂停或更改构建频率

Segment按offset切分

- Segment按offset切分, 不能有重合
- Segment之间允许有时间值重合
- 确保了数据一致性和查询准确性



查询1:10的数据, 将扫描包含它的所有Segment

其它改进

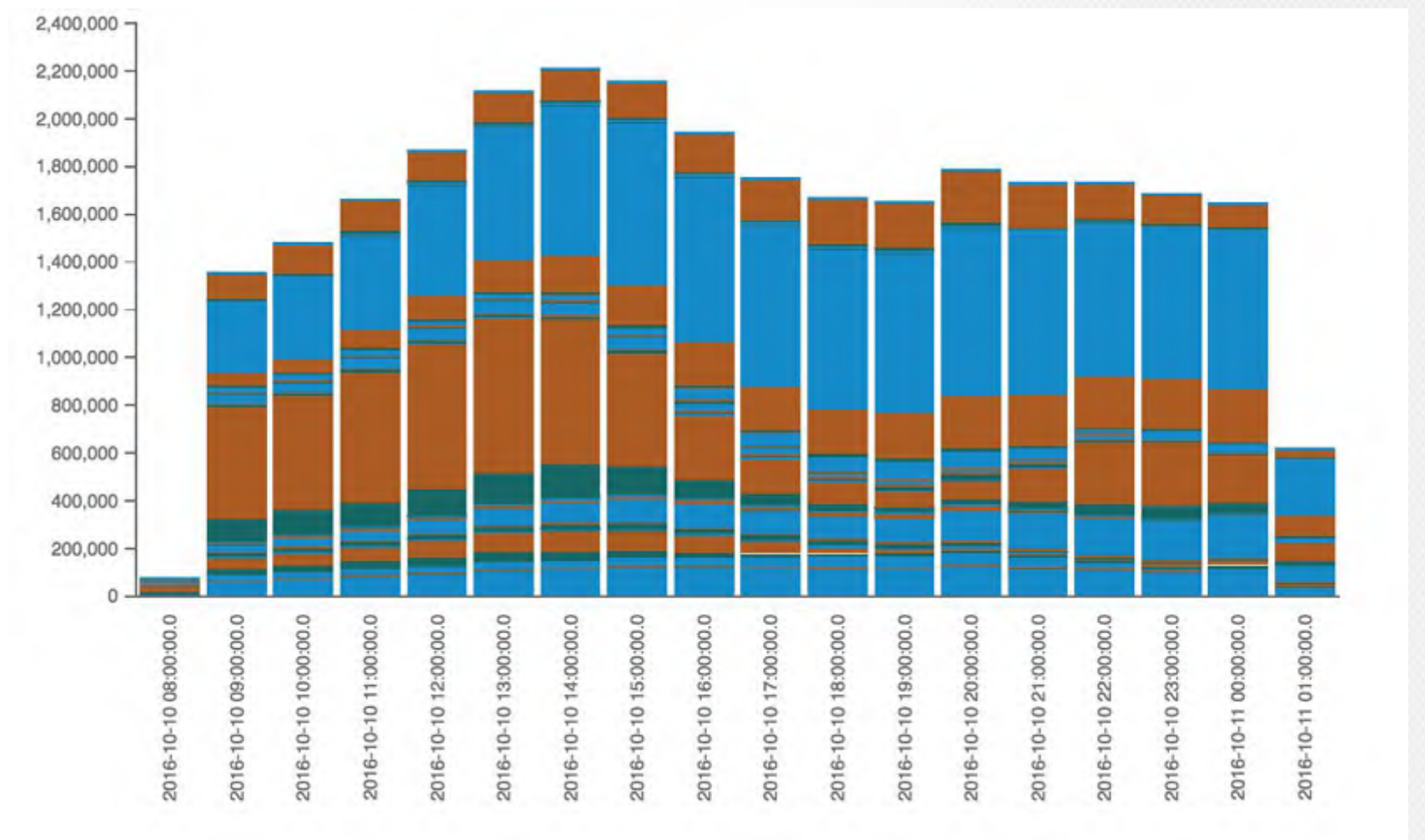
- 自动寻找开始和结束的offset
- 支持嵌套式JSON消息，支持自定义时间格式
- 允许多segment并行构建/合并
- Rest API触发

目前状态

- 公开测试进行中

Jobs in: LAST ONE WEEK NEW PENDING RUNNING FINISHED ERROR DISCARDED

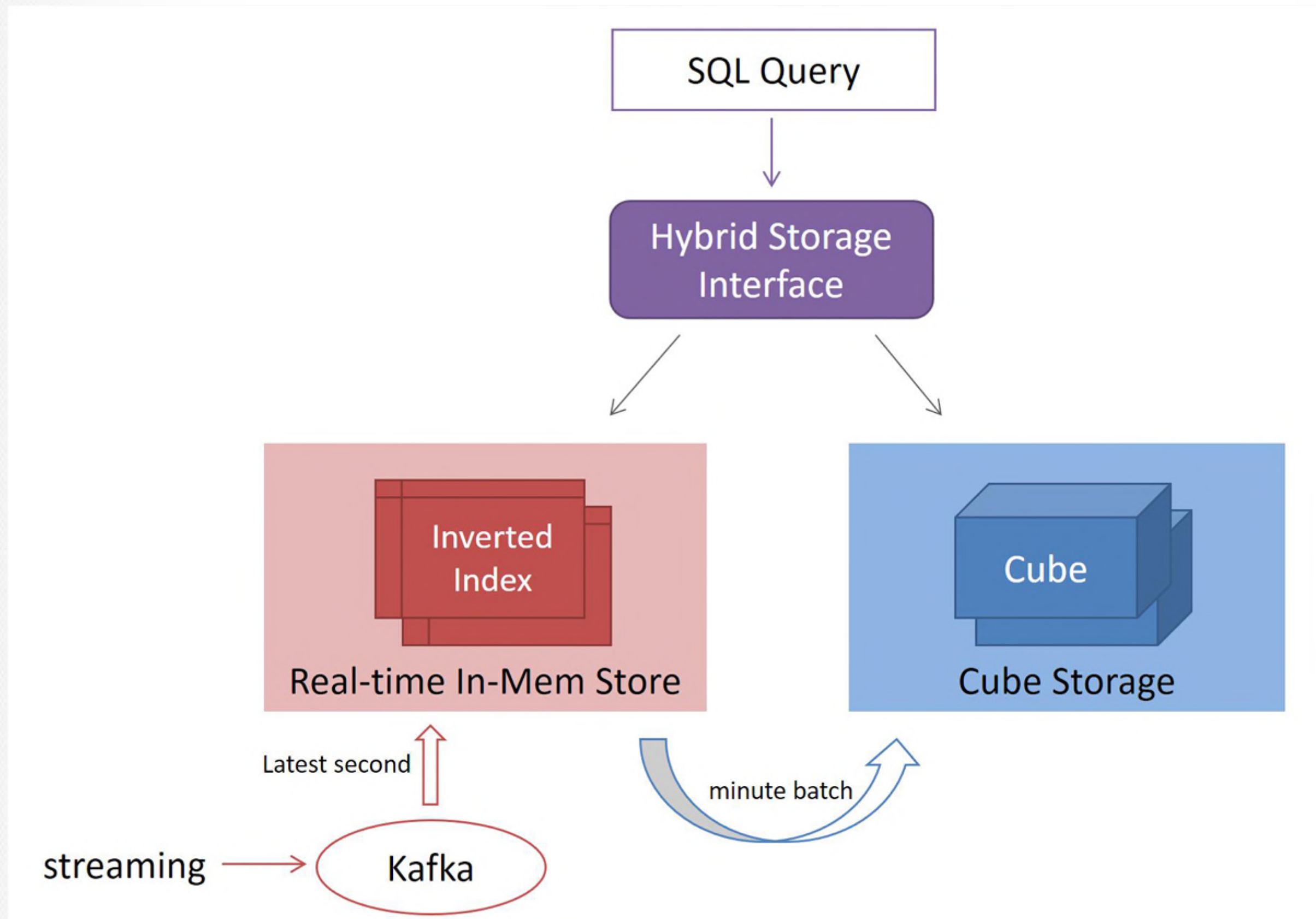
Cube ↕	Progress ↕	Last Modified Time ▼	Duration ↕	Actions
embedded_cube	73.68%	2016-10-10 21:52:28 PST	1.65 mins	Action ▾ <input type="button" value="Refresh"/>
embedded_cube	4%	2016-10-10 21:52:23 PST	0.23 mins	Action ▾ <input type="button" value="Refresh"/>
twitter_tag_cube2	100%	2016-10-10 21:44:19 PST	5.37 mins	Action ▾ <input type="button" value="Refresh"/>
embedded_cube	100%	2016-10-10 21:44:14 PST	3.27 mins	Action ▾ <input type="button" value="Refresh"/>
twitter_tag_cube2	100%	2016-10-10 21:38:15 PST	7.28 mins	Action ▾ <input type="button" value="Refresh"/>
embedded_cube	100%	2016-10-10 21:37:38 PST	2.83 mins	Action ▾ <input type="button" value="Refresh"/>
embedded_cube	100%	2016-10-10 21:34:26 PST	3.48 mins	Action ▾ <input type="button" value="Refresh"/>
embedded_cube	100%	2016-10-10 21:24:14 PST	3.27 mins	Action ▾ <input type="button" value="Refresh"/>



Twitter测试数据

- 测试环境: AWS 8台vm的Hadoop集群, Kafka 3个broker
- 测试数据: Twitter sample数据流, 每秒1万+条消息
- 测试Cube: 9个维度, 3个度量
- 构建时间: 每10分钟触发, 每次构建耗时约4分钟;
每2分钟触发, 耗时约3分钟

还差多少？



最新功能 / 未来工作

v1.5.4

- Window / Grouping 函数
- JMX Metrics 监控

v1.6

- Streaming Cubing

不久的将来

- 支持Snowflake雪花模型
- TPC-H性能测试
- 补完实时节点
- 优化MR引擎，尝试持续构建引擎



Thanks

See you on our next meeting