



第九届中国系统架构师大会
SYSTEM ARCHITECT CONFERENCE CHINA 2017

京东亿级流量海量数据搜索架构

尹德位 @ 京东搜索

yindewei@jd.com

2017/10/21

SACC
2017

北京·新云南皇冠假日酒店

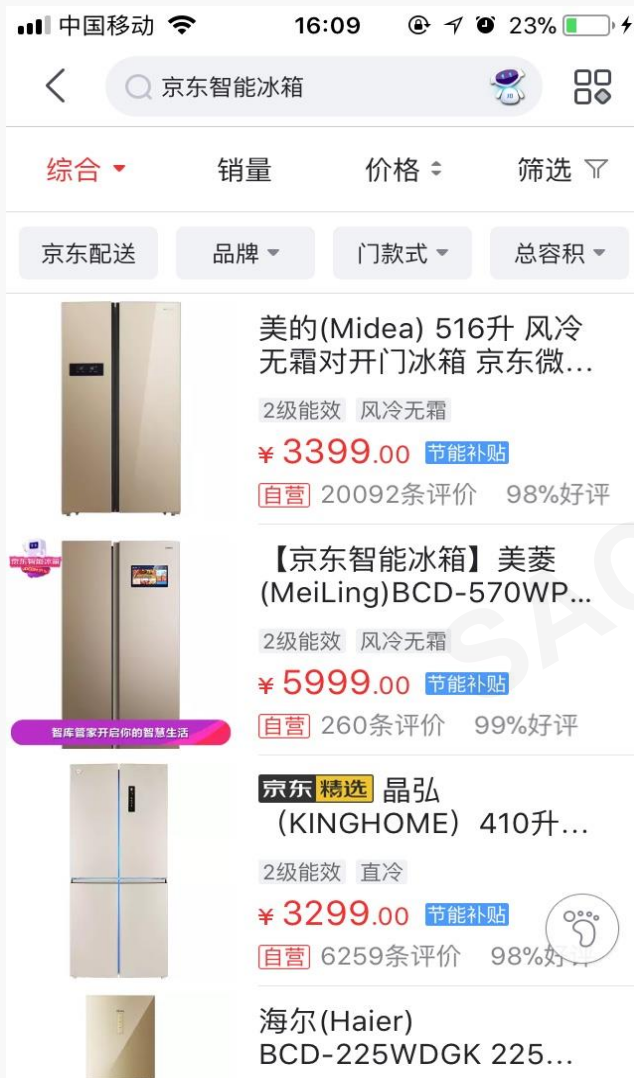


- 搜索系统简介
- 搜索架构演进之路
- 搜索系统的发展与挑战

SACC2017

	电商搜索	门户搜索
索引量	普通	大
召回率	高	普通
精准率	高	普通
排序	多维度 分流/提权	文本相关性 竞价排名等
实时性	高 (部分信息秒级)	普通
个性化	高 (千人千面)	普通
多样性	丰富 (排序/导航)	普通

移动搜索页 — 智能终端入口



京东开普勒系统 — 京X计划



PC搜索页 — 重要入口之一

搜索我的购物车 6

3c产品 | 京东3c | 3c数码 | 手机 | 小米 | 小米盒子 | 3c数码产品 | 诺基亚 | 3c认证 | 3c频道 | 小米3c | 华为 |

全部商品分类 | 服装城 | 美妆馆 | 超市 | 生鲜 | 全球购 | 闪购 | 拍卖 | 金融

全部结果 > "3c"

品牌:										更多 +多选
前置摄像头像素:	1600万及以上	800万-1599万	500万-799万	200万-499万	120万-199万	120万以下	无	+多选		
后置摄像头像素:	后置双摄像头	1200万-1999万	500万-1199万	500万以下	无	+多选				
颜色:		+多选								
高级选项:	屏幕尺寸	网络	热点	系统	机身内存	运行内存	CPU核数	电池容量	其他分类	

商品精选 广告

综合 销量 评论数 新品 价格 ¥ - ¥

配送至 京东配送 货到付款 仅显示有货 可配送全球

共76万+件商品 1/100 在结果中搜索 确定

京东搜索系统规模概况

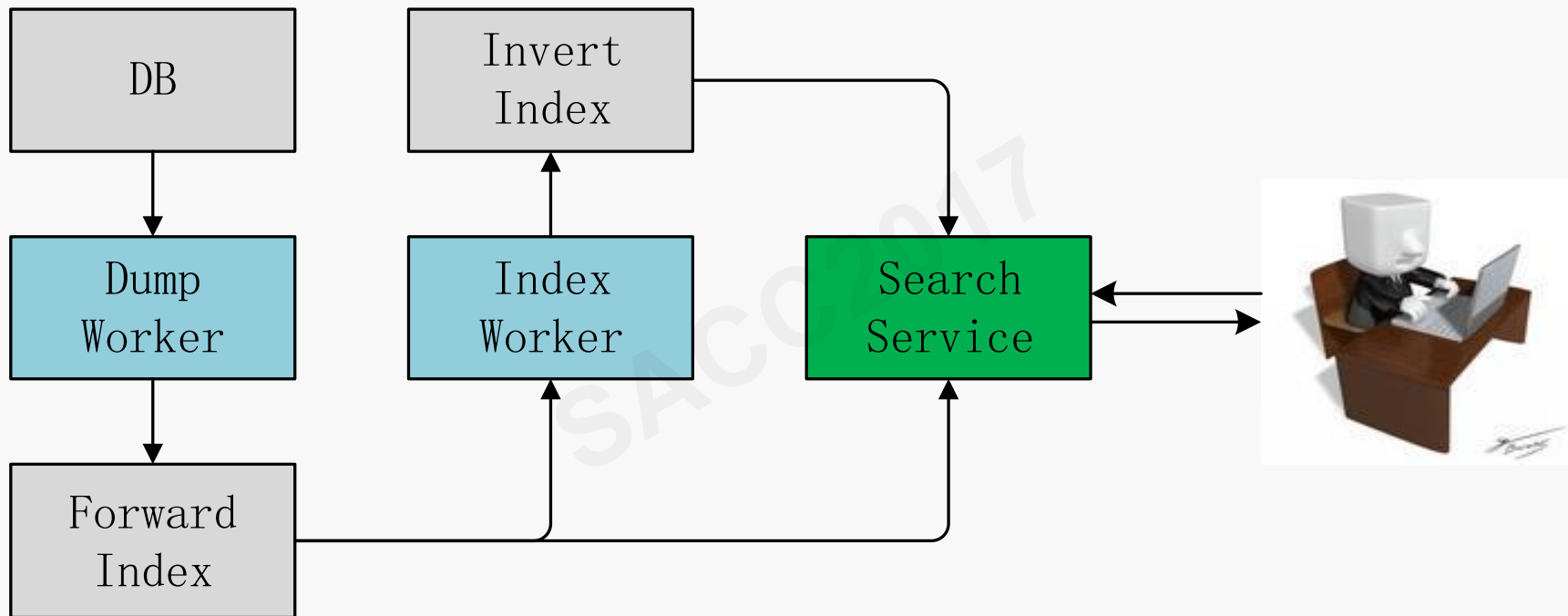
流量入口	APP/PC/WX/QQ 头条/百度/网易
索引量	十亿级
日更新	亿级
用户数	亿级
日PV	十亿级

- 搜索系统简介
- 搜索架构演进之路
- 搜索系统的发展与挑战

SACC2017

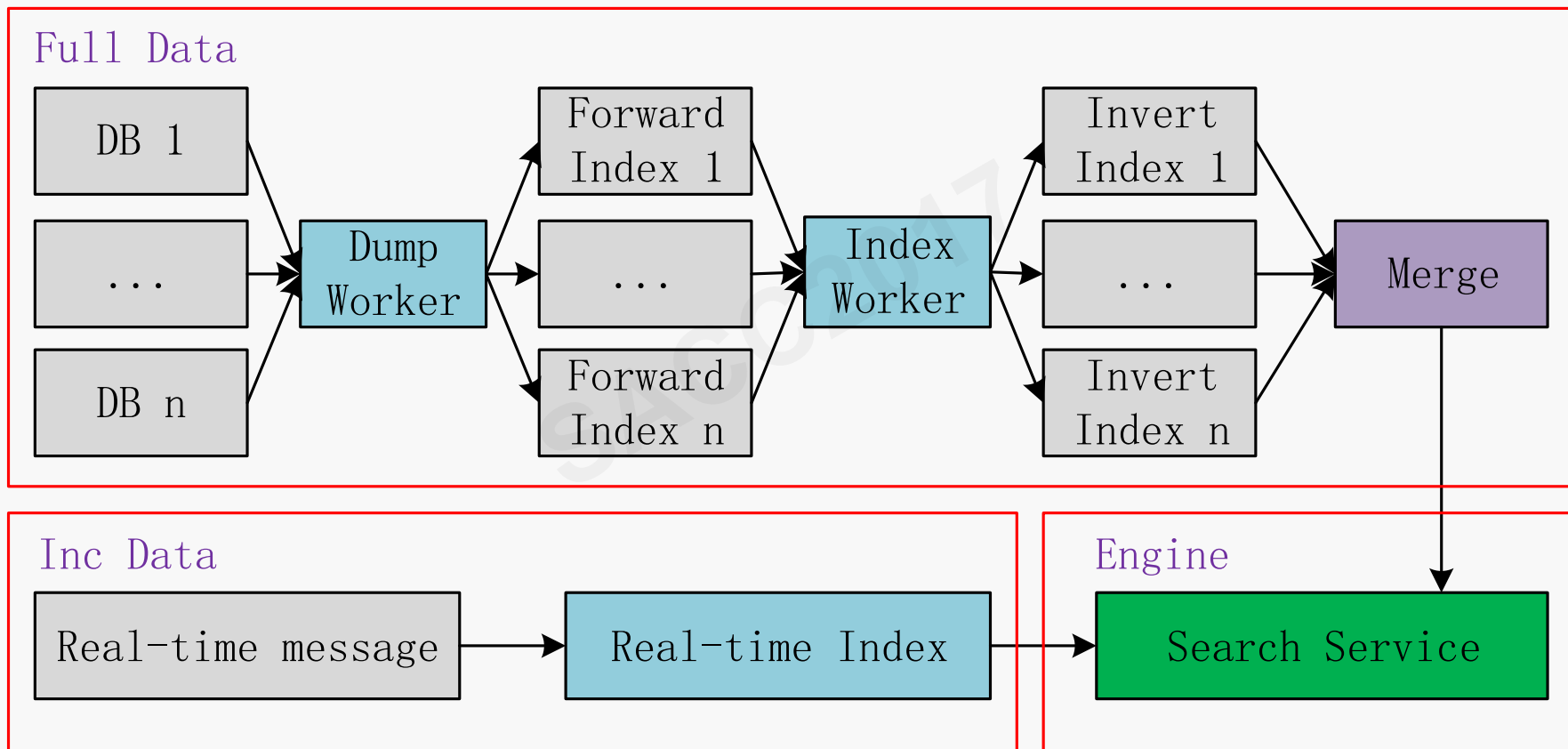
搜索架构V1 — 2011年

系统结构：简单
数据量级：百万
更新方式：全量



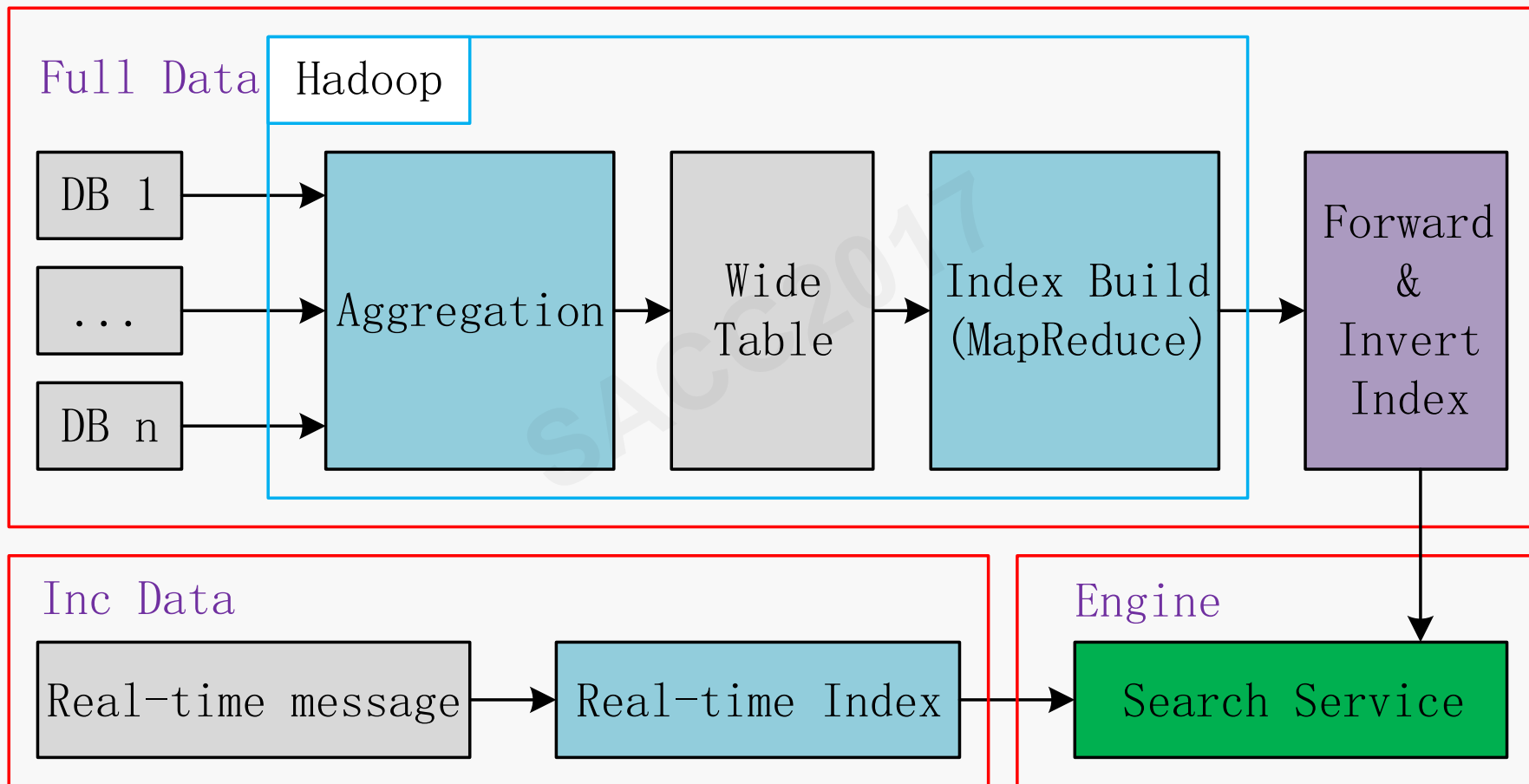
搜索架构V2 — 2013年

系统结构：分库分表+实时索引
数据量级：千万
更新方式：全量+增量

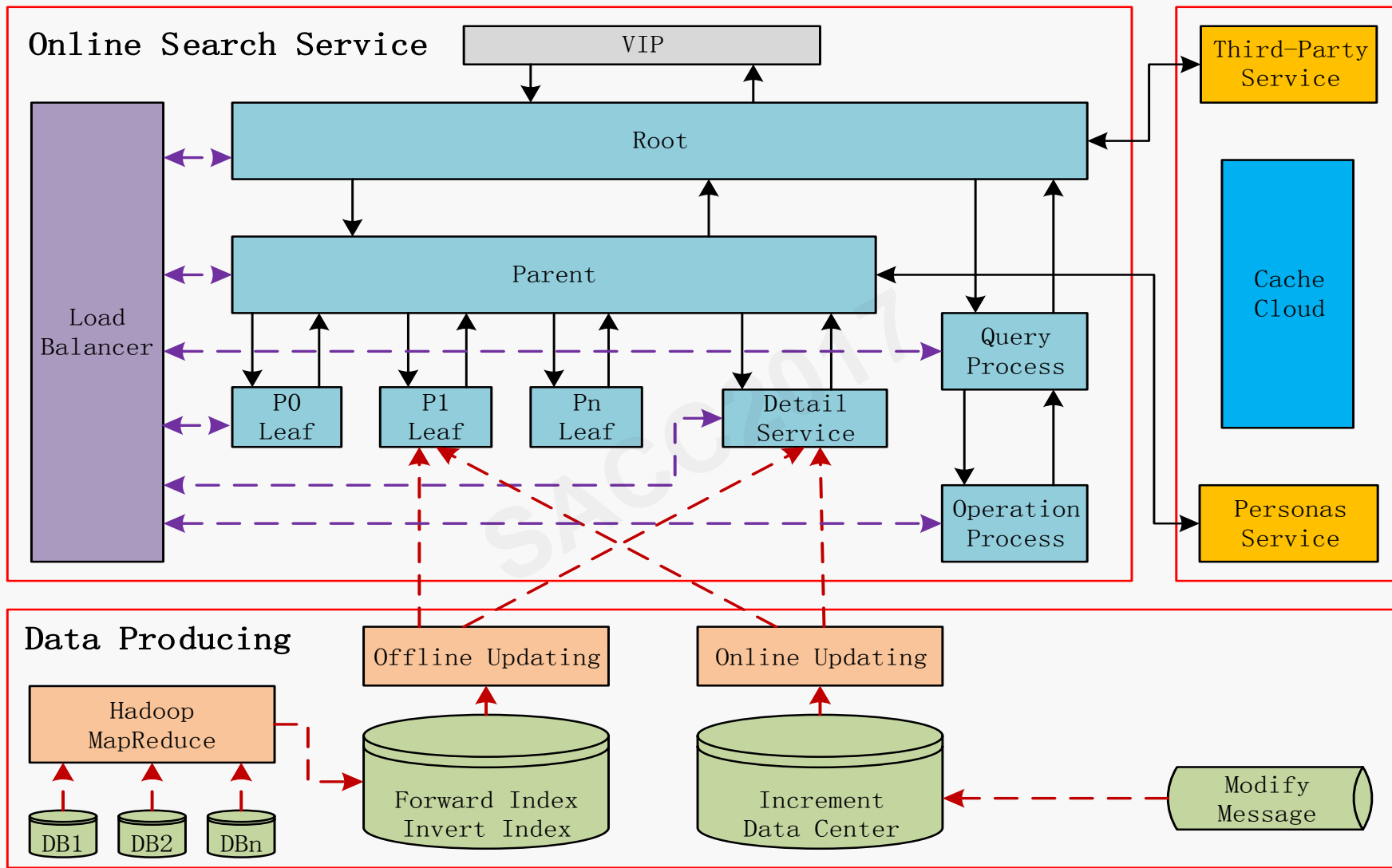


搜索架构V3 — 2014年

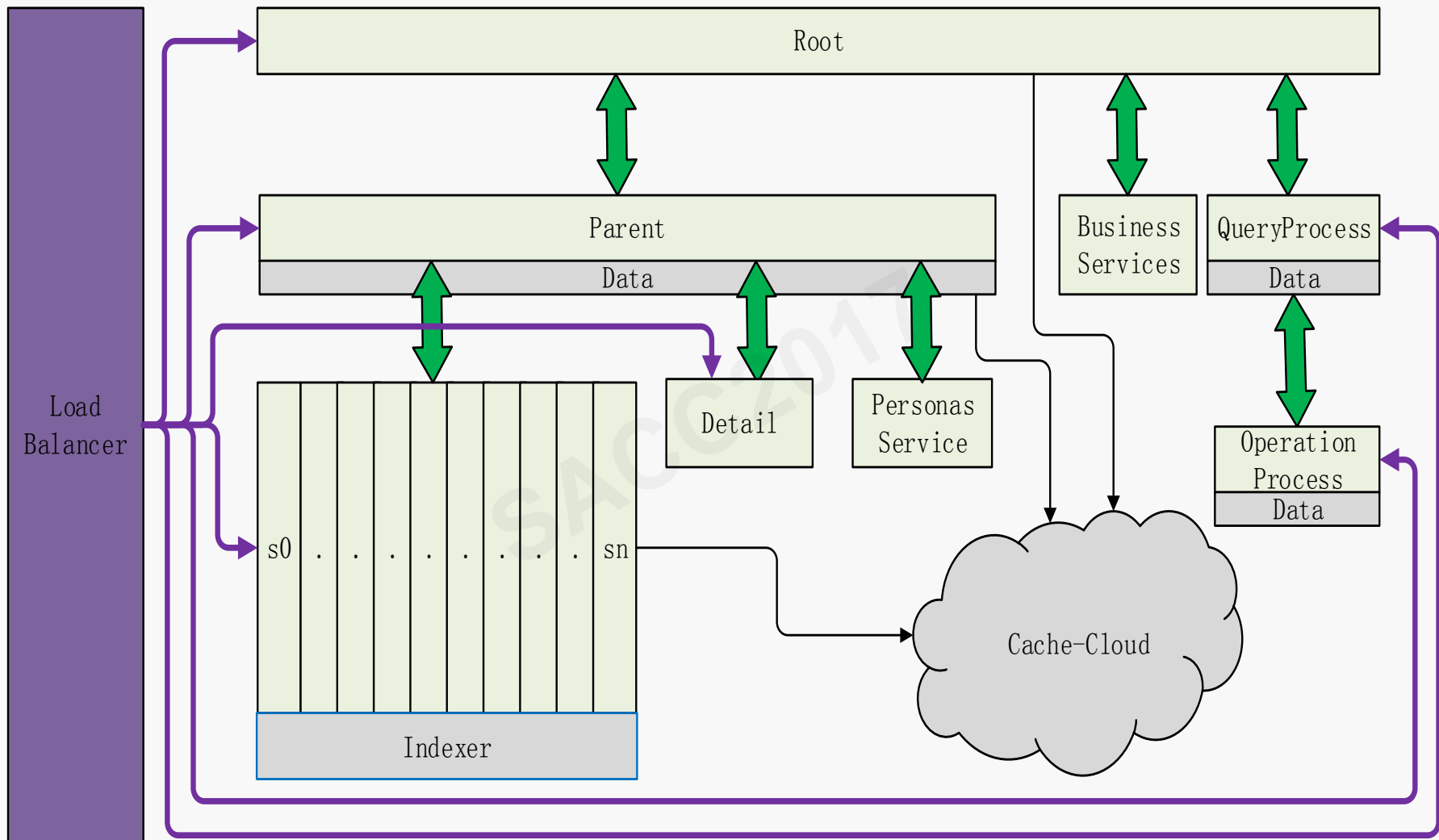
系统结构：离线hadoop处理
数据量级：亿级
更新方式：全量+增量



平行搜索(分布式)引擎 – 2016年



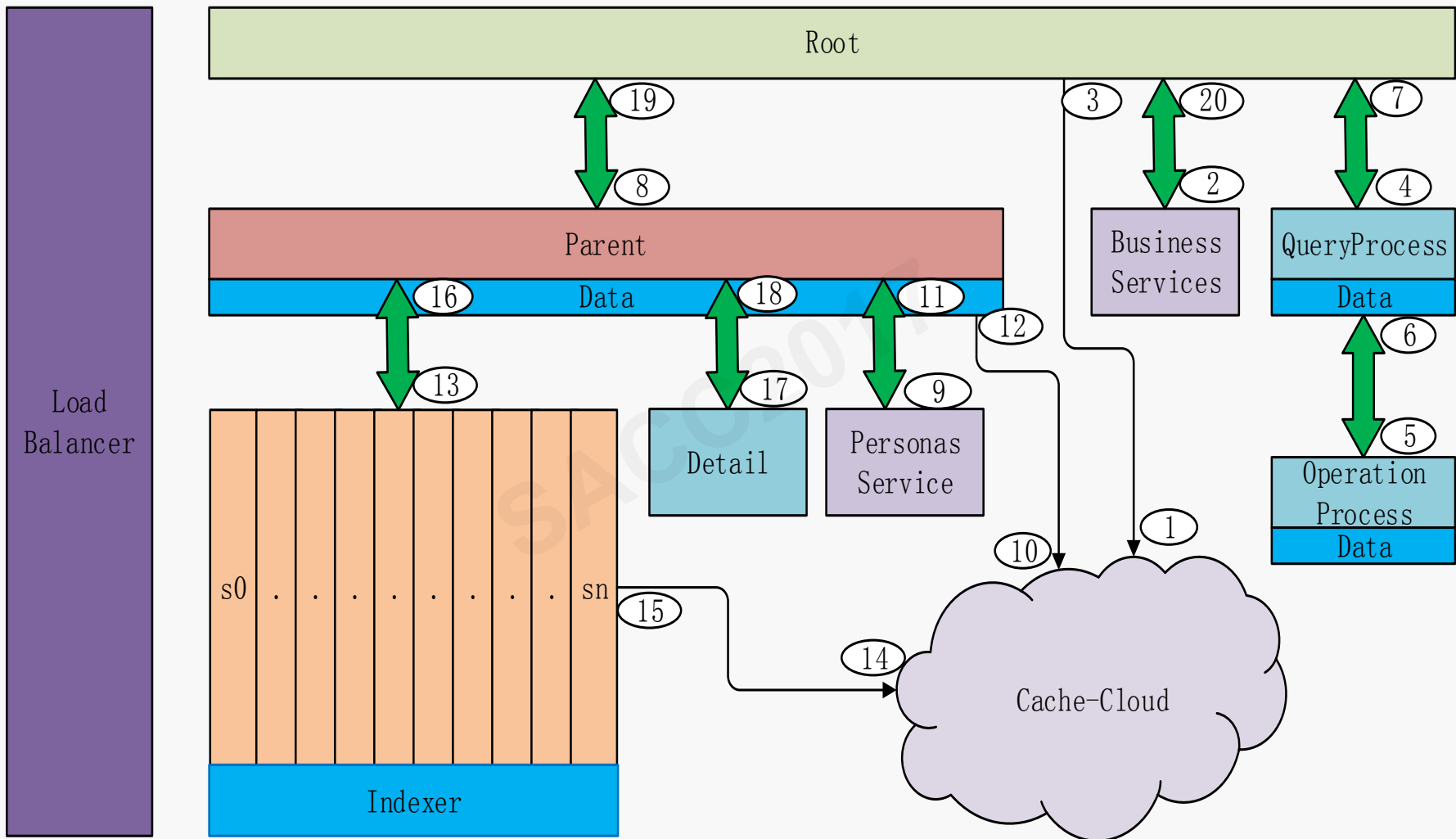
平行搜索(分布式)引擎 - 负载均衡管理



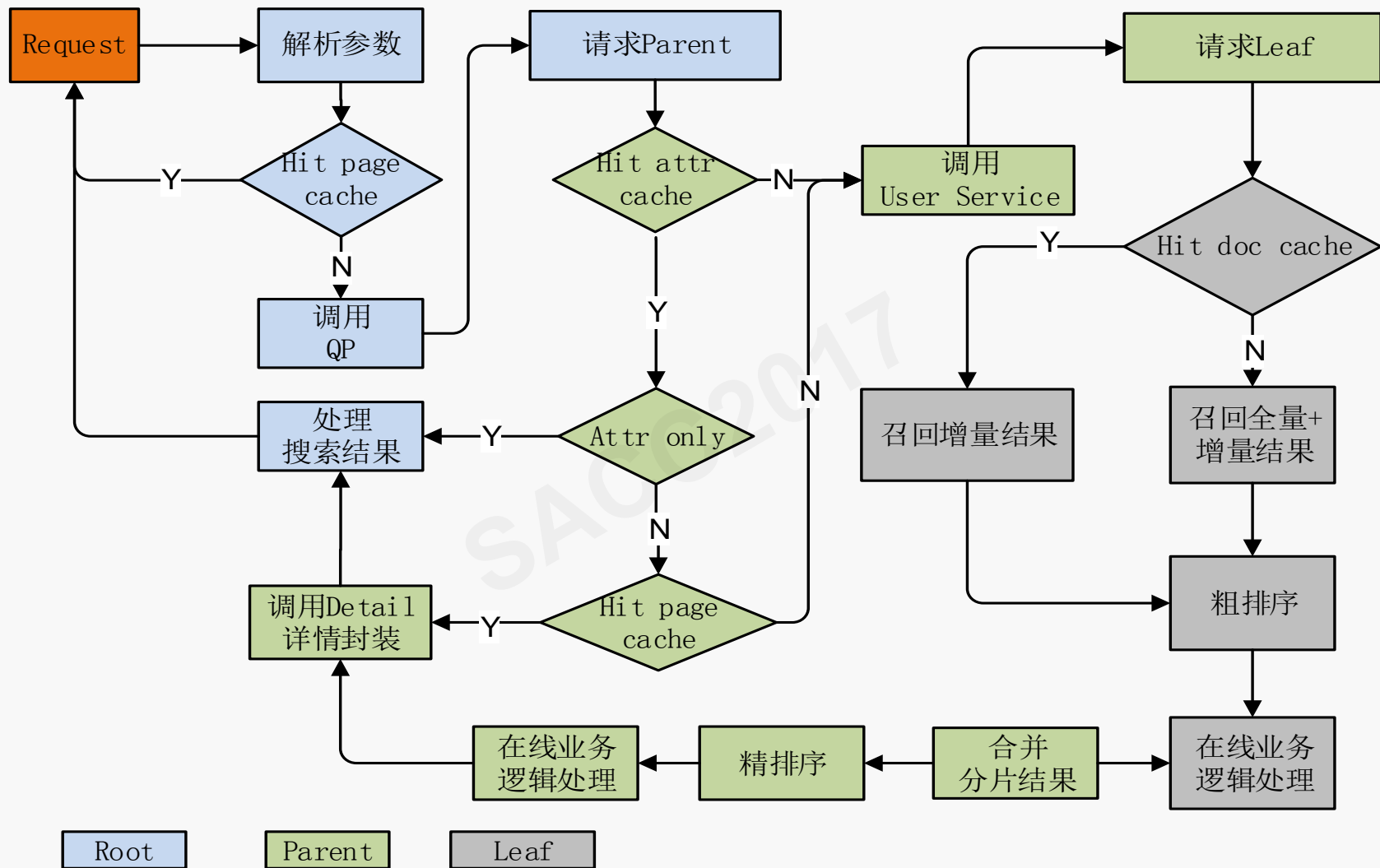
平行搜索(分布式)引擎 – 负载均衡管理

1. 轮询
2. 负载窗口机制
3. 哈希
4. 动态伸缩
5. 集群灾备

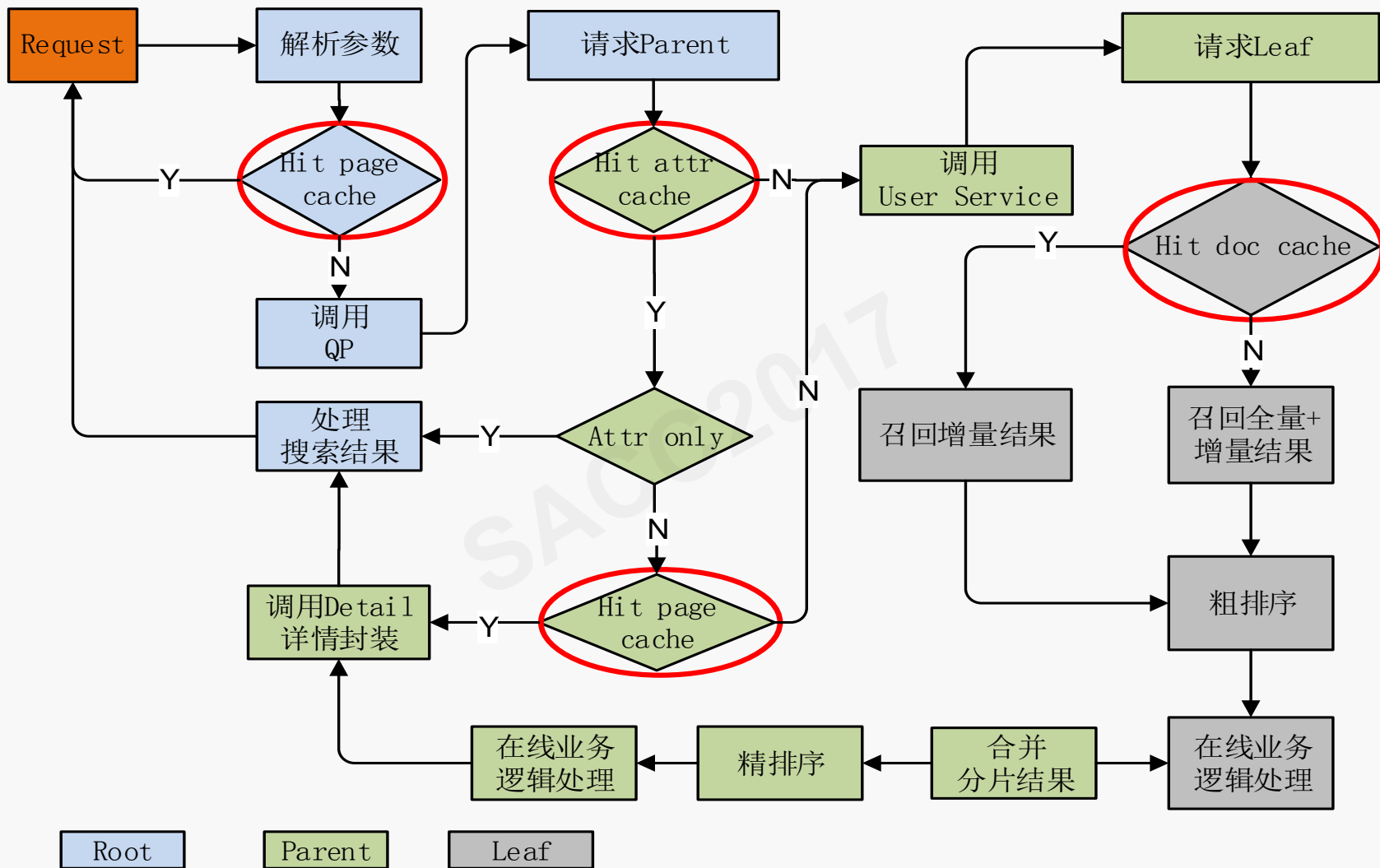
平行搜索(分布式)引擎 — 搜索流程示意



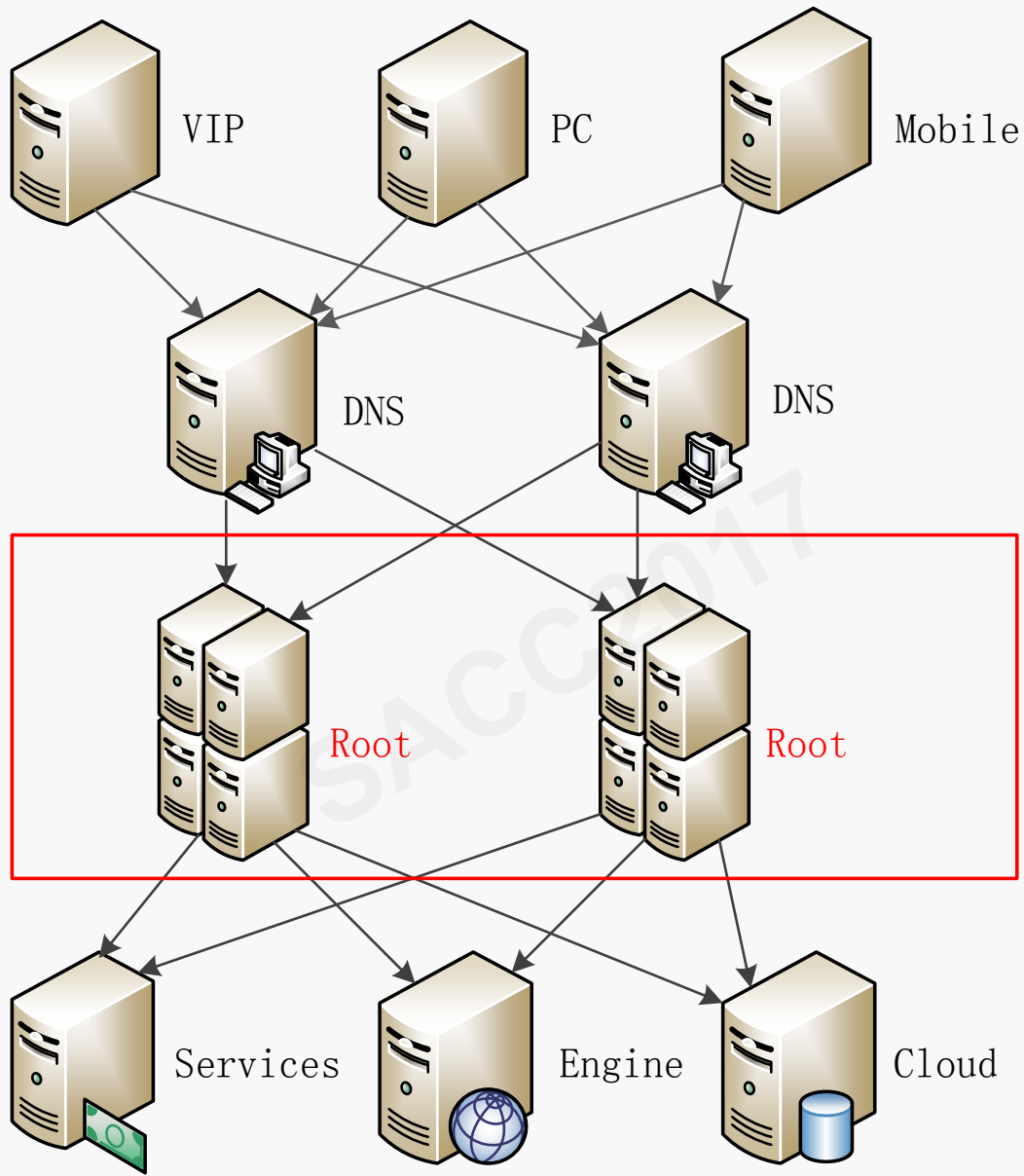
request flow



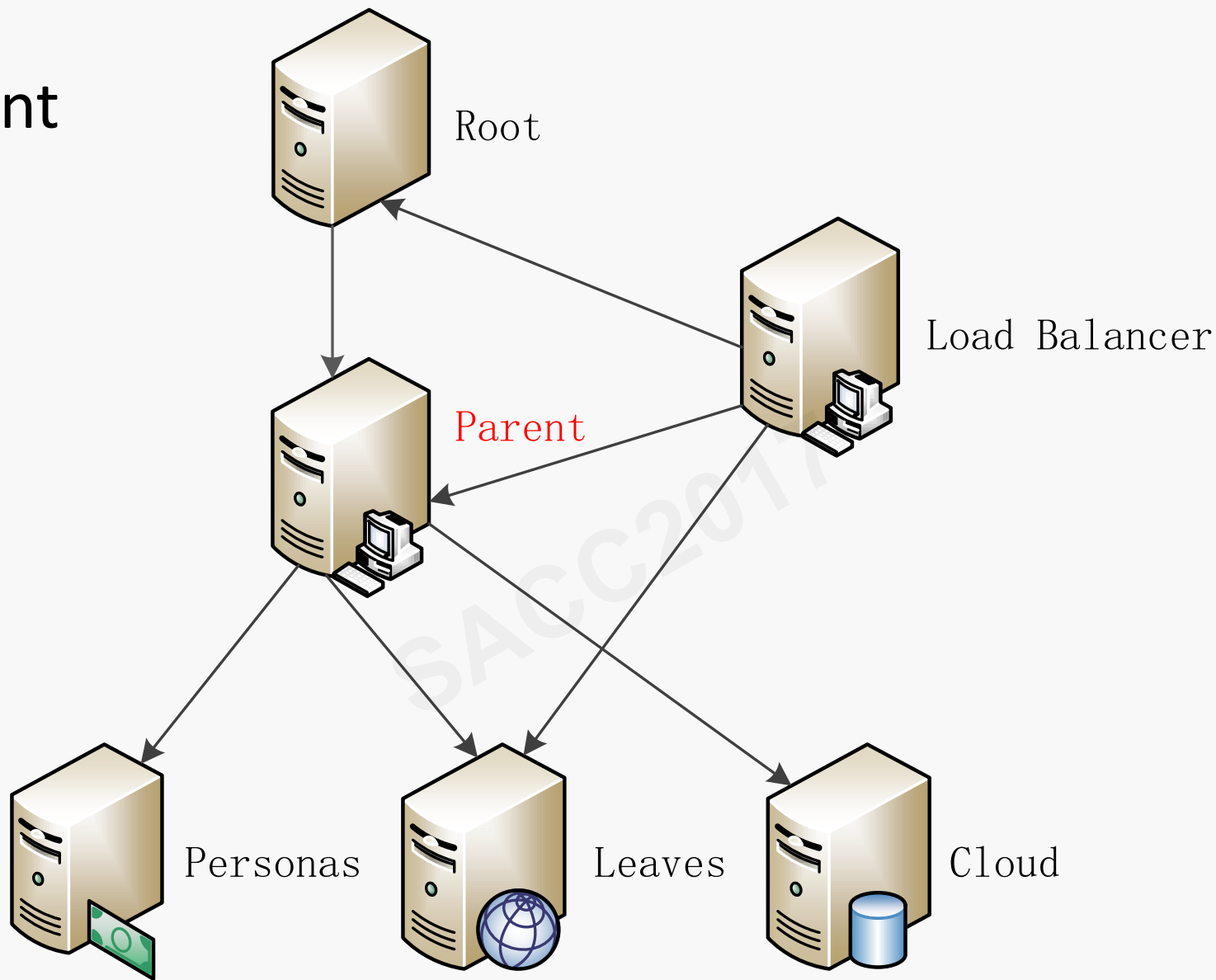
trilevel cache



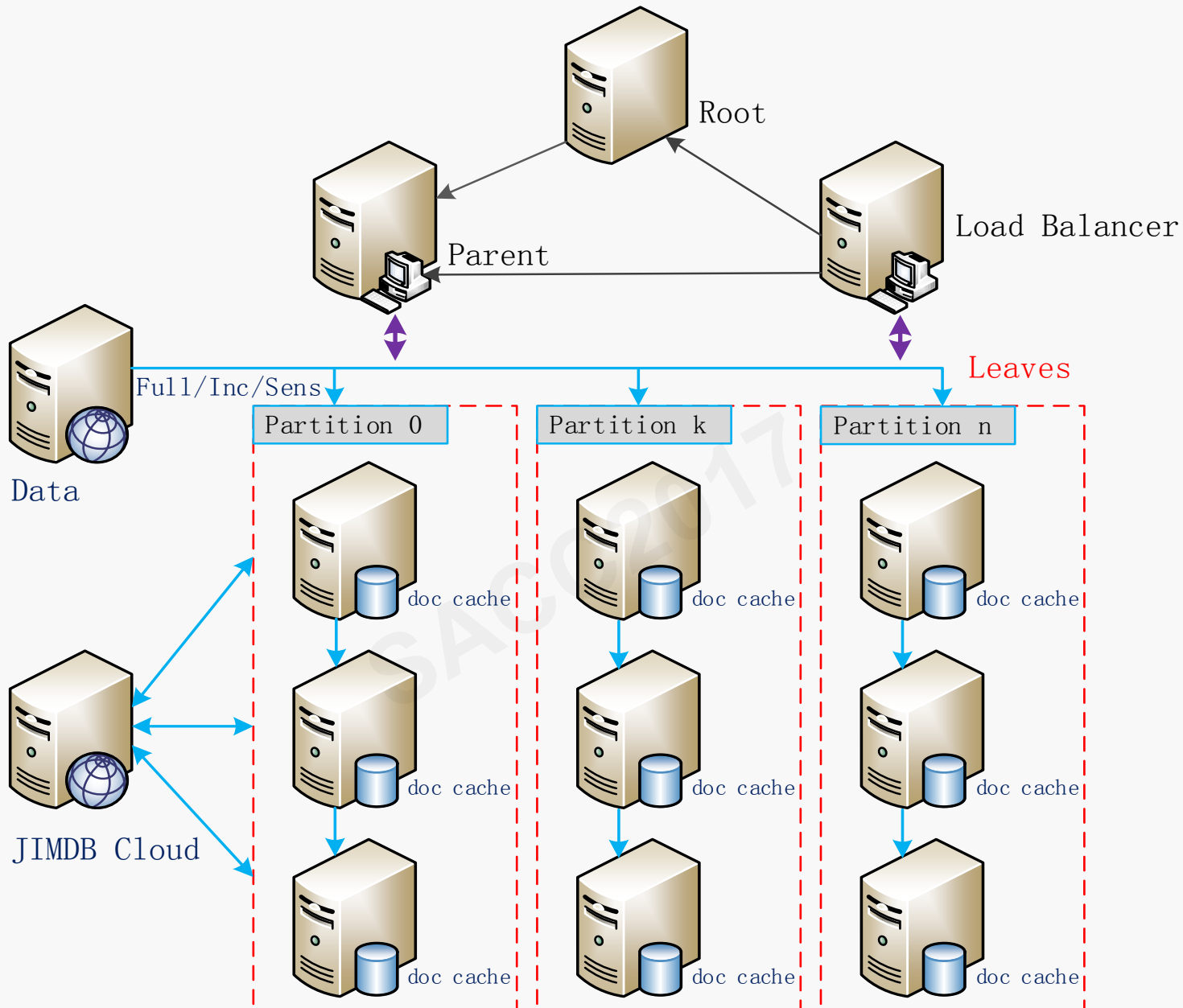
Root



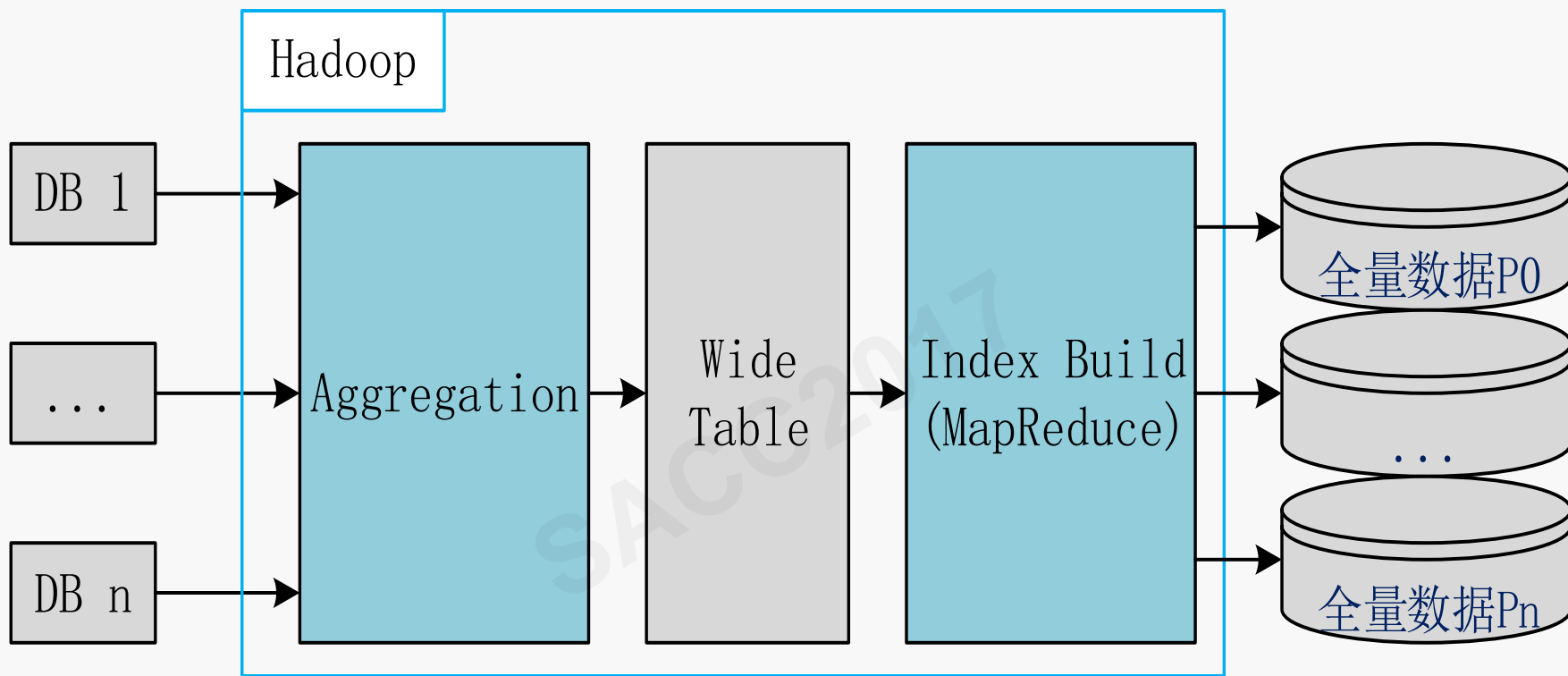
Parent



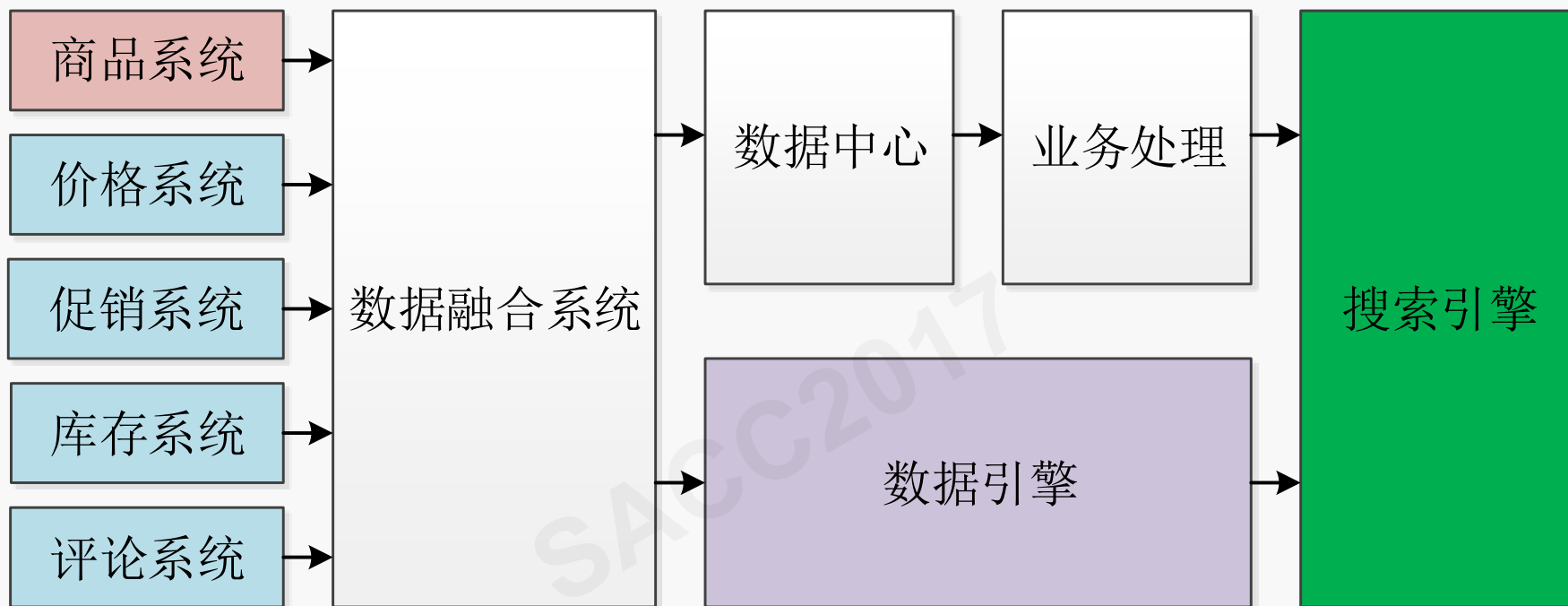
Leaf



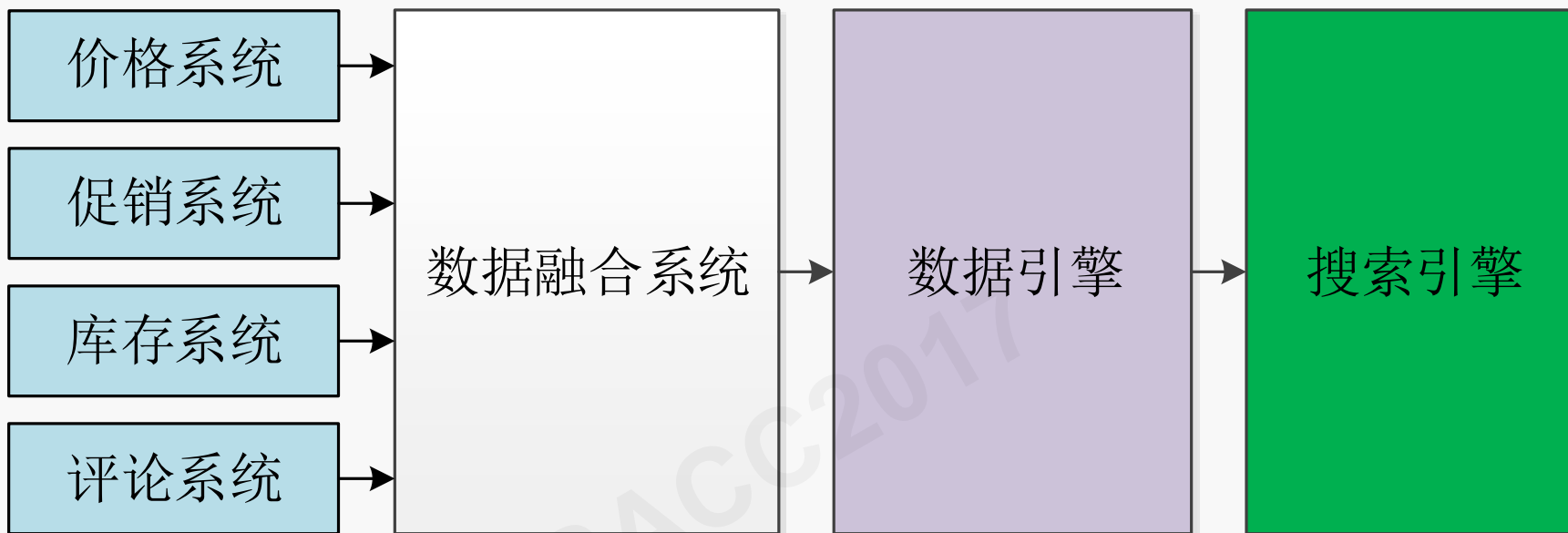
索引数据生产 — 全量



索引数据生产 — 增量



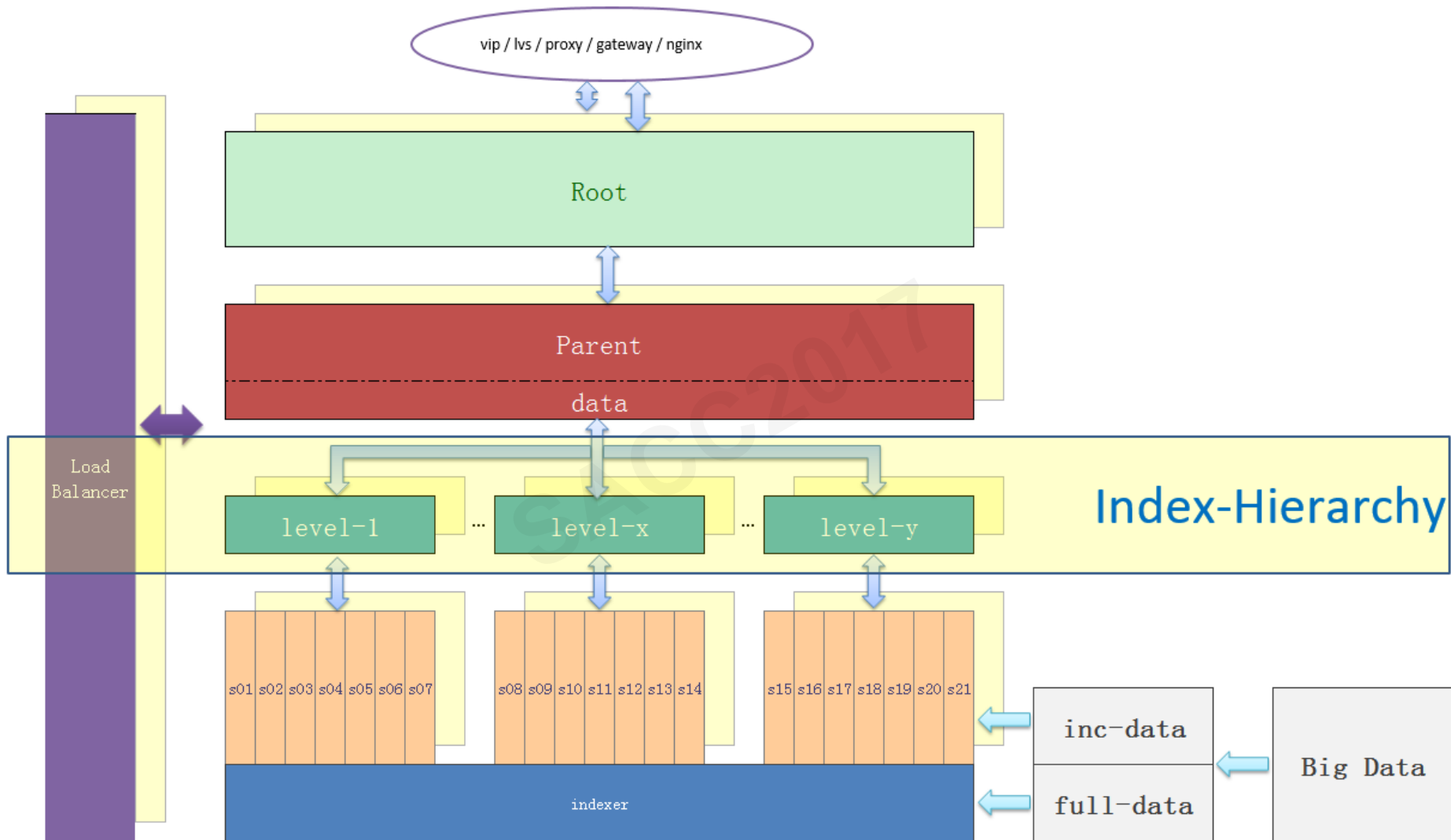
索引数据生产 — 敏感



- 搜索系统简介
- 搜索架构演进之路
- 搜索系统的发展与挑战

SACC2017

挑战一：数据量持续增长



挑战二：搜索多维化

1. 语音视频多维立体搜索
2. 智能交互
3. 高度个性化

SACC2017

THANKS



尹德位@京东搜索

yindewei@jd.com