

bilibili

Bilibili数据平台实践

——薛赵明



个人介绍



07年本科毕业

混迹过阿里，网易等

16.9加入B站

目录

● 部门介绍

● 平台实践

● 总结展望



部门介绍

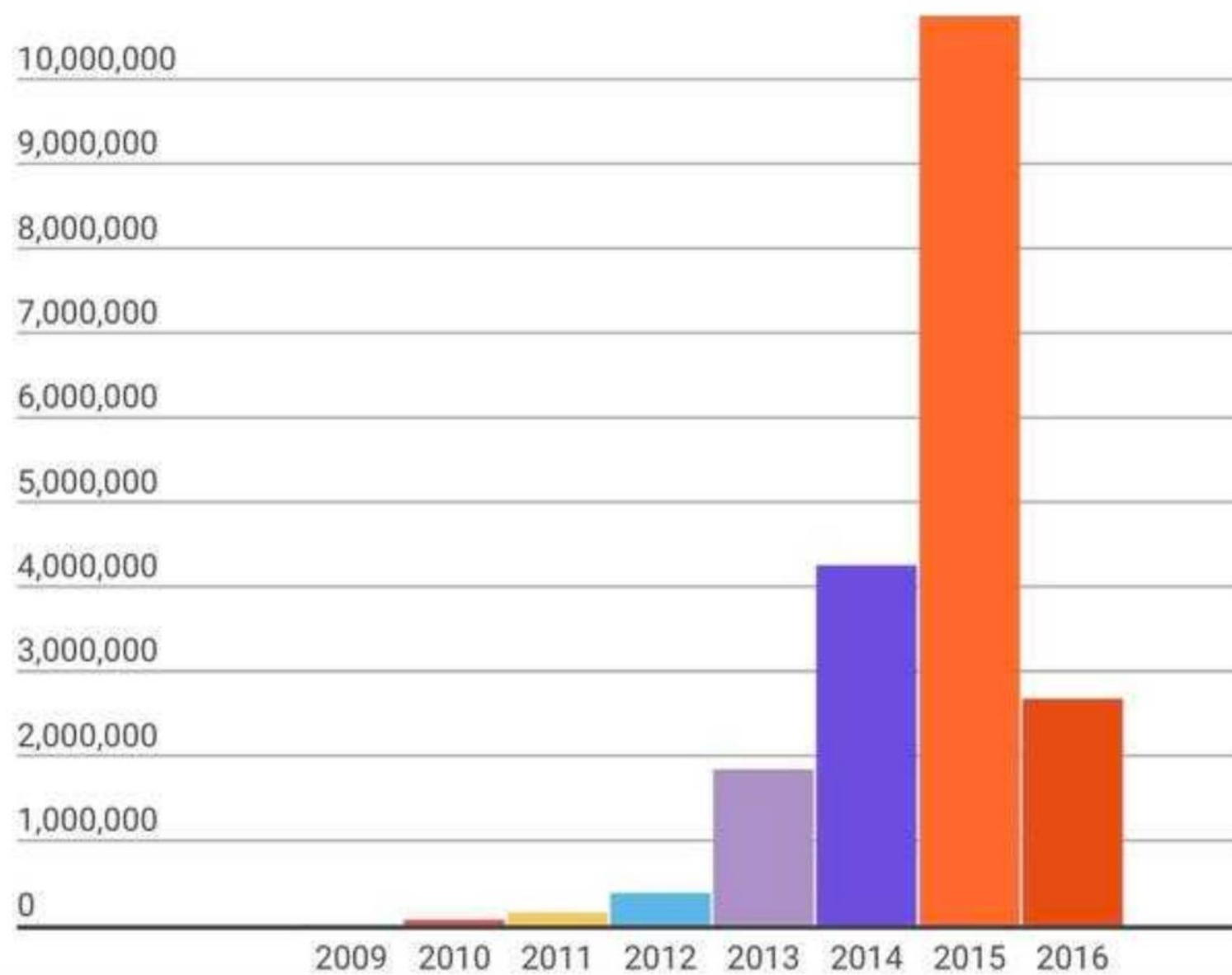


平台开发

数仓建设

BI分析

用户注册(非官方数据)
截止时间: 2016-02-18



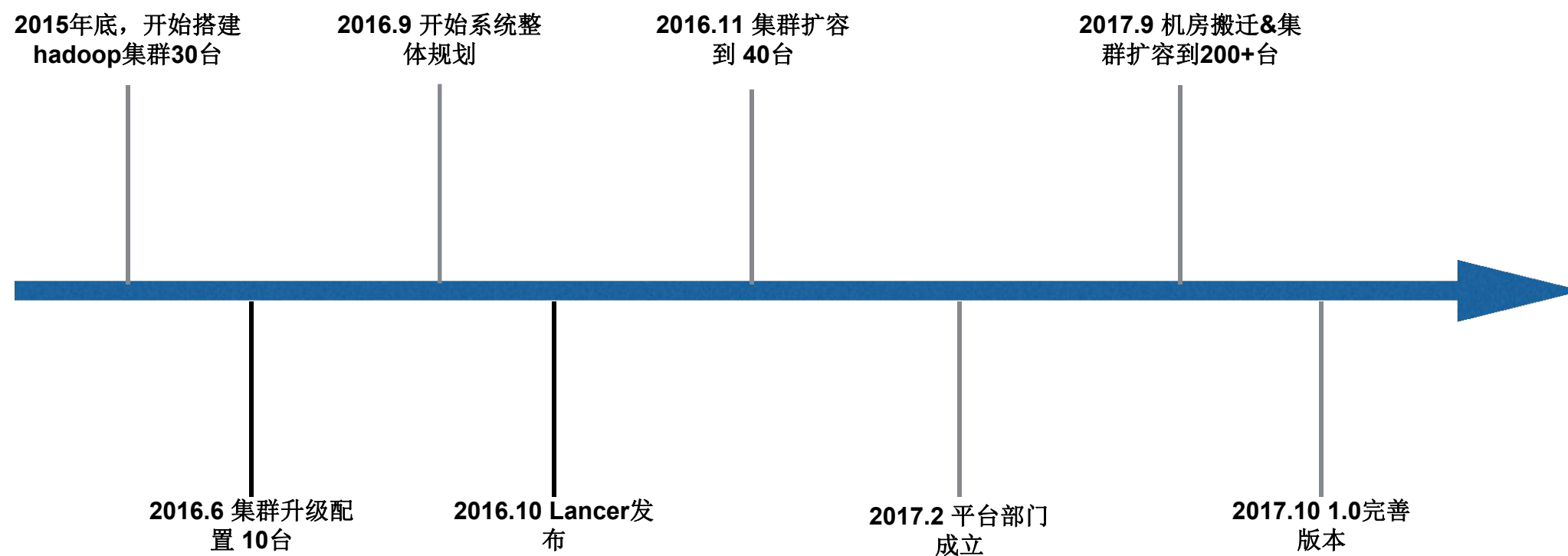
核心数据

DAU:3000万

MAU:1.5亿



历史发展



目录

● 部门介绍

● 平台实践

● 未来展望



数据平台解决什么问题？



高层：全局的统计报表


产品&运营：及时的数据统计分析

.....

数仓：灵活的系统调度

数据开发：分布式的计算能力

数据平台怎么解决问题？

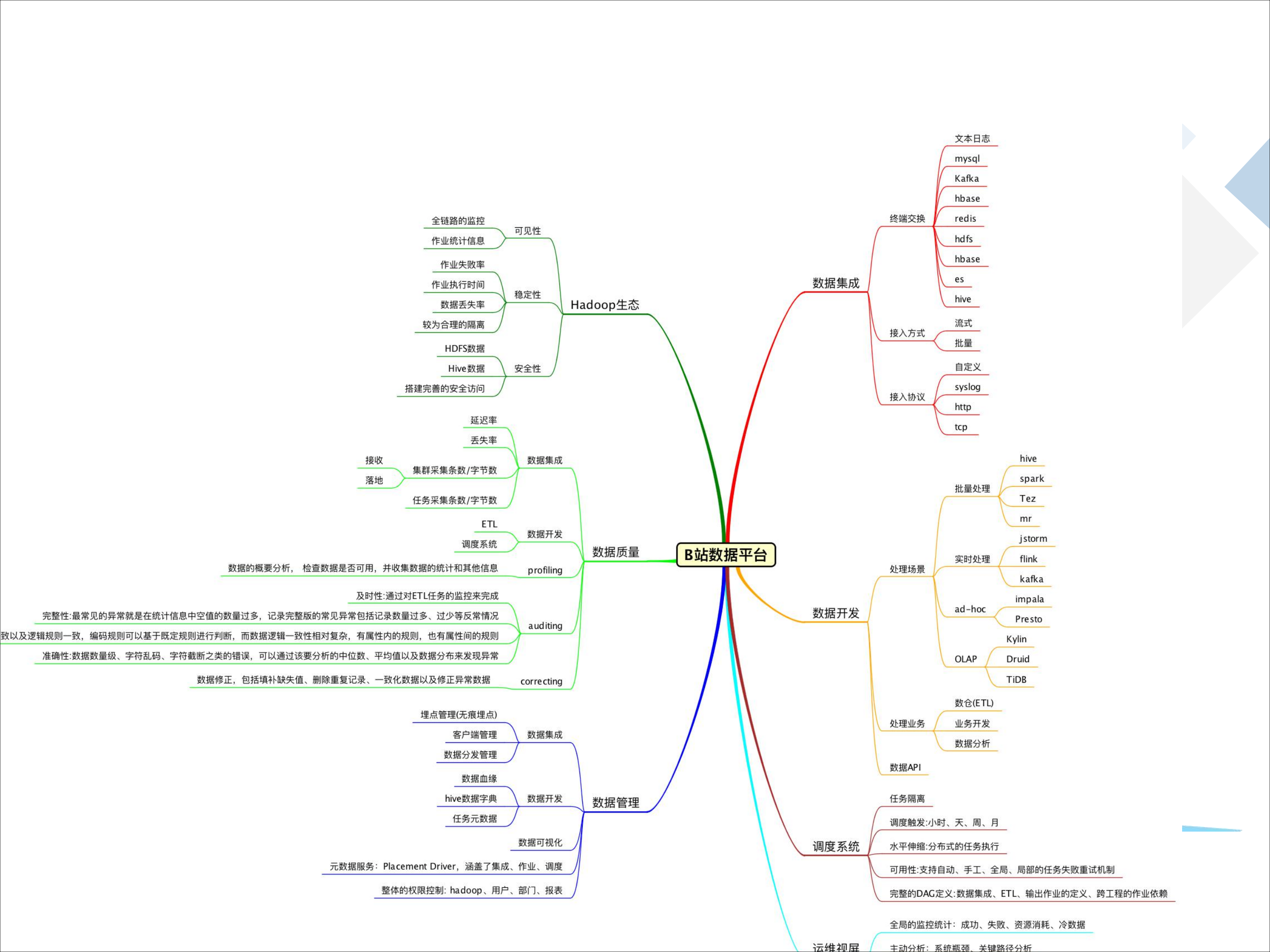


存：数据大集中，存入HDFS/HBase

通：统一规范，打通各个业务单位

用：推动数据开放和交换

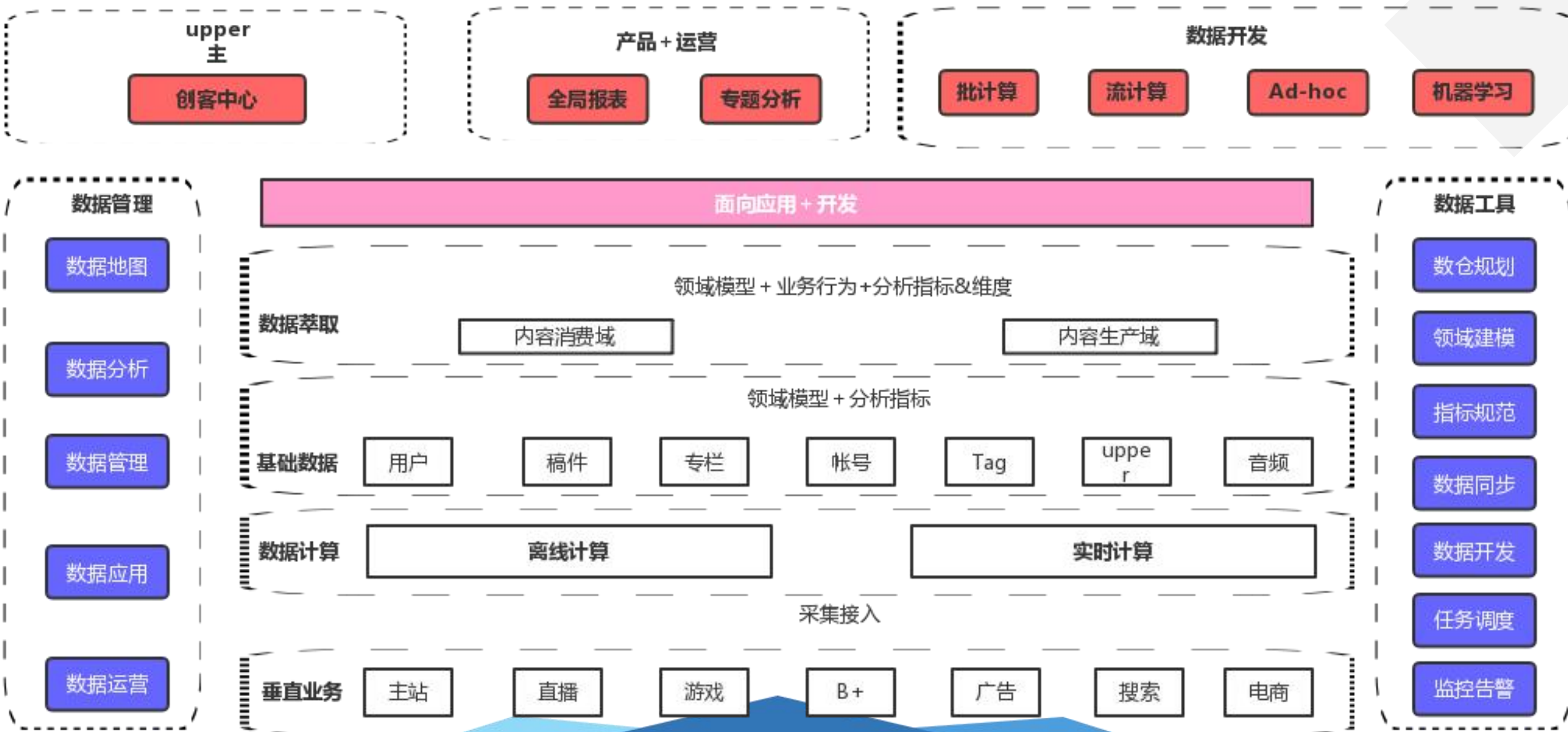




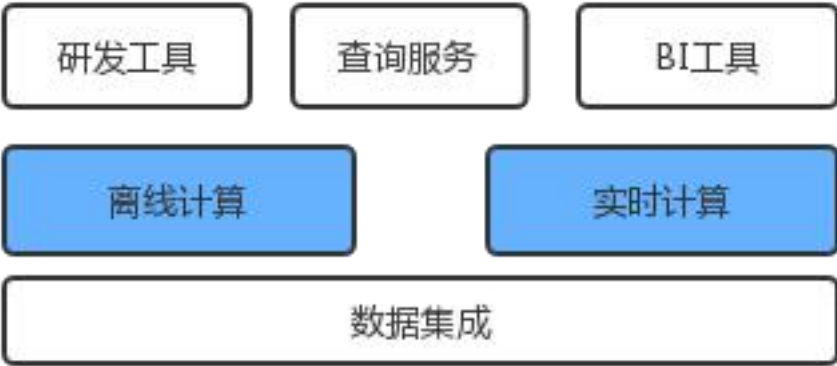
数据中台



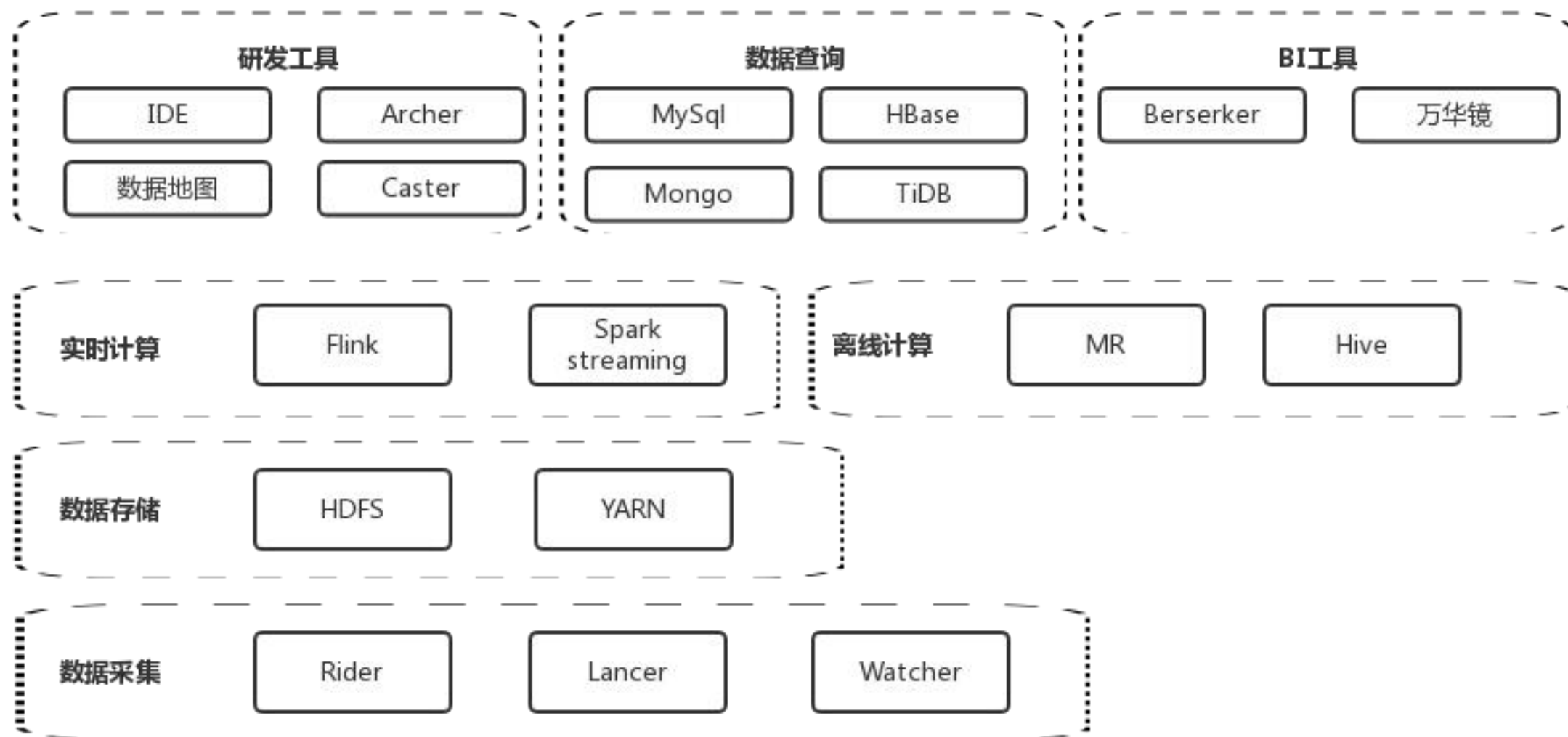
数据中台



数据中台



数据中台



Solomon:实时数仓

Prism:明细数据查询

Berserker:一站式的数据开发, BI平台

Avenger:数据盘, 用于外部数据交换

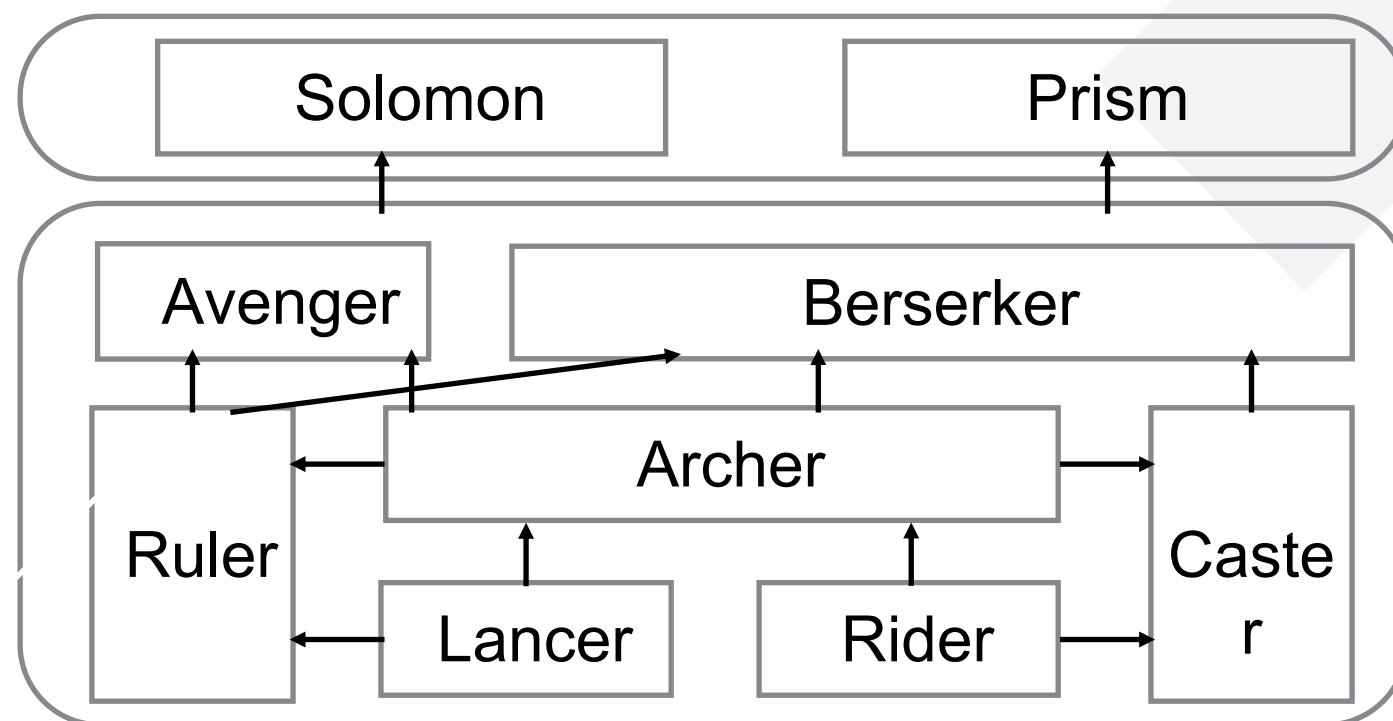
Archer:调度平台

Ruler: 元数据服务

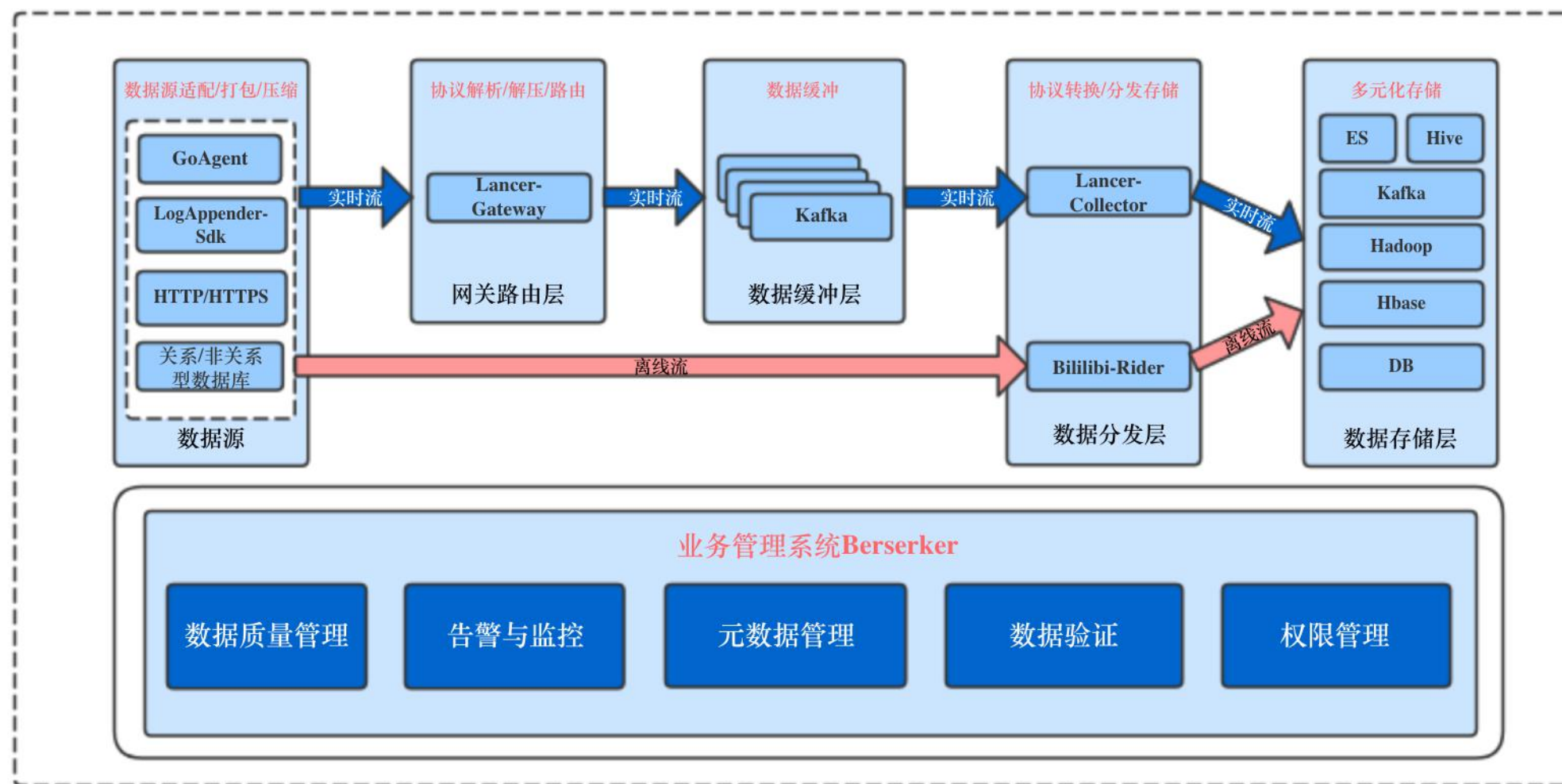
Lancer:流式流数据采集

Rider:批量数据采集

Caster:运维监控



Lancer



智能BI

数据平台

万华镜

数据BI

数据集成

数据开发

数据监控

权限管理

不持

消息

帮助

数据BI

报表展示区

报表管理

平台报表分类

报表权限申请

数据模型管理

首页 / 数据BI / 报表展示区

输入报表名称

搜索

查看收藏

周边

周边-主站账号统计

周边-主站账号视频统计

商城数据

1

支付系统

充值统计

充值订单统计

支付渠道统计

支付统计

1

文章区数据

投稿-成功投稿

投稿-投稿位置来源

投稿-创作中心位置点击

投稿-编辑器功能

专栏访问

1 2 3

PGC

PGC-web页面访问统计

PGC-web页面事件统计

PGC-移动页面事件统计

PGC-移动页面访问统计

1

商务中心

app-商务中心地域数据

H5-商务中心渠道数据

开放平台

订单TOP10

流量漏斗

搜索数据

黑马词排行榜TOP100

搜索基础数据

账号

三无会员

关注粉丝数UP主分布

数据集集成

输入部门名称

筛选数据源

生效状态	任务名称	部门	创建时间	创建人
生效	数据上报	live	2017-12-01 12:12:17	kuang
生效	归属信息	live	2017-11-30 14:25:35	caimen
生效	归属信息	live	2017-11-30 14:21:10	caimen
生效	数据上报	search	2017-11-29 14:28:44	zhangz
生效	数据上报	search	2017-11-29 14:23:19	zhangz
生效	数据上报	bigdata	2017-11-29 12:02:14	chenji
生效	数据上报	video	2017-11-28 19:23:52	xuxi

数据开发

bilibili 数据平台

SQL查询

定时项目

执行进程

执行历史

数据血缘关系

Hive表权限申请

HDFS文件下载

万华镜

数据BI

数据集成

数据开发

数据监控

权限管理

首页 / 数据开发 / SQL查询

新建SQL查询

输入查询内容、报表名称

查询时间	创建人	查询内容
2017-12-04 22:00:03	maojian	select mid as USER_INTERNAL_ID, fid as USER_CONCERNED en attribute in (2,6) and status=0 then co when (attribute in (2 atas=1) or attribute=100 then
2017-12-04 22:00:03	maojian	select USER_ACCOUNT,MEDIA_ID,MEDIA_NAME,concat('ht bili.com/video/av/MEDIA_ID/') as MEDIA_URL,MEDIA_TAG ME,IP from (select USER_ACCOUNT,MEDIA_ID,MEDIA_NAME,IP from (select
2017-12-04 22:00:02	maojian	sel (select USER_ACCOUNT,USER_TYPE,LOGIN_I _TYPE,TERMINAL_ID,TERMINAL_OS_VERSION,ACTION_M TYPE from (select case when mid=0 then mid select head CCO

运维监控

数据监控

[首页](#) / [数据监控](#) / [监控总览](#)

监控总览

选择队列

🕒 正在执行作业数

i 被关闭作业数

部门资源监控

2017-11-04 - 2017-12-04

队列资源监控

☆ 总作业数

✔ 成功作业数

❗ 问题作业数

任务实时监控

节点状态监控

冷数据监控

问题作业监控



目录

● 部门介绍

● 平台实践

● 总结展望



总结



尽早的机房规划(机柜, 网络)

架构搭建以大厂(BATH)为标杆

系统优化以商业化公司(cloudera, hortonworks, mapR)为标

杆



总结&展望

产

- 1.数据决策(高管&决策者 数据辅助决策)
- 2.应用分析，专题分析
- 3.数据监控

技术面：全面的Spark化，docker on yarn

安全：全链路安全的打通

THANKS

哔哩哔哩 - (° - °)つ口 乾杯~

