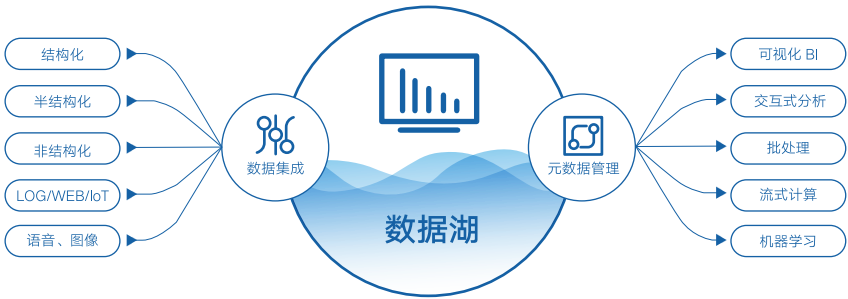


# 百度MapReduce

企业级大数据基础平台BMR

# 百度智能云“天算”大数据平台

基于百度先进的大数据技术，结合数十年在大数据方面的应用实践及 To B 服务的经验积累，沉淀出端到端、开源开放、高性价比的百度智能云“天算”大数据平台，致力于让企业的大数据应用更高效、更低门槛。百度智能云“天算”大数据平台包括：覆盖数据集成、数据治理、数据管理、数据开发、交互式分析、数据服务的 EDAP 数据湖管理与分析平台，兼容开源 Hadoop 生态的大数据基础平台 BMR，国内首款分析型数据库 Apache 孵化项目 Palo(Doris)，日志分析 BLS、实时分析 BSC、检索分析 BES 等场景化数据分析产品，以及兼具敏捷 BI 及可视化大屏能力的 Sugar。覆盖数据的全生命周期管理过程，帮助用户管理和分析数据，挖掘数据价值，驱动业务增长。



## 📄 开源开放

- ▶ Doris 进入 Apache 孵化器
- ▶ BMR 完全兼容开源组件

## ¥ 高性价比

- ▶ BOS 四级存储，成本低
- ▶ BMR 弹性伸缩，高性价比

## 📄 端到端

- ▶ 集成、治理、开发、分析、应用全场景
- ▶ 批、流、BI、ML 一站式

百度智能云 BMR 为“天算”大数据平台数据计算、分析引擎，致力为客户提供安全可靠、组件全面、简单易用的大数据计算分析服务，帮助企业低门槛洞察数据价值，快速实现业务数据驱动。

# 关于BMR

百度 MapReduce BMR 企业级托管大数据平台具有组件完备、高性价比、开源开放等特性，可以帮助用户高效构建企业级大数据基础平台。

BMR 包括丰富的 Hadoop 生态组件，如 Hadoop、Spark、Hive、Presto、Flink、Hbase、Druid、Pig、Phoenix 等。用户可通过界面灵活选择所需组件和配置，一键创建所需集群，并且通过可视化界面、API 方式灵活管理集群。此外，先进的计算存储分离技术、弹性伸缩技术在确保高可靠的同时，真正帮助用户做到了用时高效获取资源、闲时释放资源，帮助用户以最低成本获得最高计算性能。



# 产品优势



## 完备的大数据生态组件

BMR 完全兼容开源 Hadoop 生态，集成并优化了完备的大数据组件，包括 Spark、Hive、Hbase、Flink 等等，用户可以根据业务分析场景，灵活选择、组合大数据组件，高效实现批量数据处理、实时流式计算、一键搭建企业级实时数仓等，并为用户提供了丰富好用的数据处理、开发工具。

## 一键可视化创建、高效统一运维

### 一键创建部署

- ▶ 根据模板一键创建、开箱即用

### 按需灵活配置

- ▶ 选组件、选版本、选套餐

### 全方位可视运维

- ◆ 平台运行状态
- ◆ 作业执行分析
- ◆ 资源使用监控
- ◆ 多租户配置管理
- ◆ 监控报警配置

用户可以通过可视化的点选操作按需选择组件、版本、套餐等快速创建大数据平台，无需考虑复杂的环境搭建过程。提供统一的可视化监控运维工具，帮助用户高效管理各类大数据处理分析组件和平台，并且开放运维 API 接口。



## 智能弹性伸缩，提供高性价比的计算服务

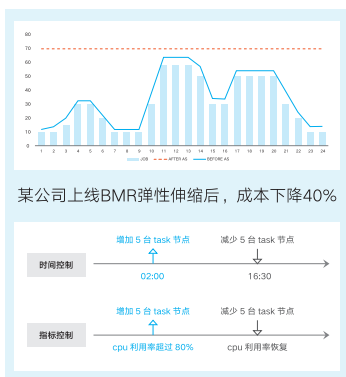
通过智能弹性伸缩，可以根据作业负载情况或业务周期时间，自动增减计算资源，不会为闲置的计算资源付费，降低用户成本。

### 基于场景按时间自由定制伸缩规则

- ▶ 指定时间点
- ▶ 按天、周、月等周期配置

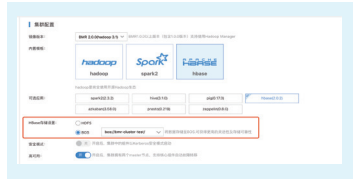
### 自定义规则，弹性伸缩

- ▶ CPU 指标
- ▶ 作业延迟指标等
- ▶ 自定义指标



## 先进的计算存储分离架构，集群可用性更高 存储更可靠、扩展更灵活

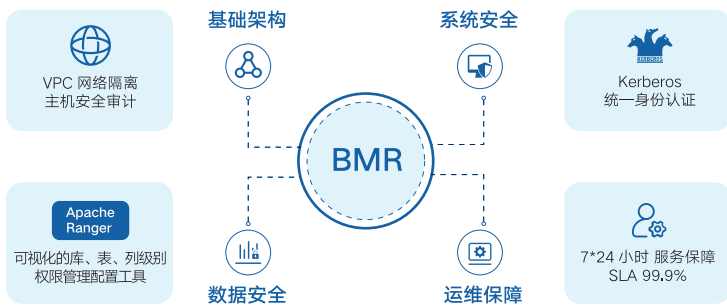
通过先进的计算存储分离架构，集群不可抗力异常时可快速恢复，可用性更高；数据存储更可靠，存储弹性扩展，更方便。



- ▶ 集群节点灵活选配 CDS 云盘，多副本保证数据可靠不丢。此外，集群节点不可抗力异常可快速拉起恢复，不用额外考虑数据恢复问题。
- ▶ 多组件数据存储支持挂载 BOS 对象存储，支持存储灵活弹性扩展，按需归档等。

## 多层次、全方位安全机制

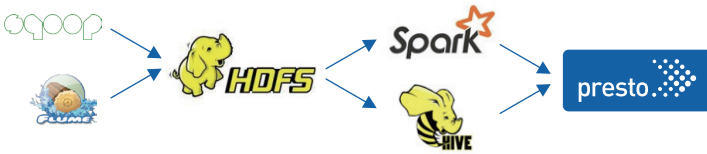
BMR 提供多层次的安全机制保障集群平台和数据的安全。基础架构上，用户基于 BMR 创建的大数据基础平台运行在独立的 VPC 网络环境下，从网络上进行隔离，并支持用户主机安全审计。提供 Kerberos 和 Ranger 组件，为用户提供基于用户身份、作业提交和资源使用，以及数据访问的认证和鉴权，细粒度保障平台和数据的安全。



# 案例与应用场景

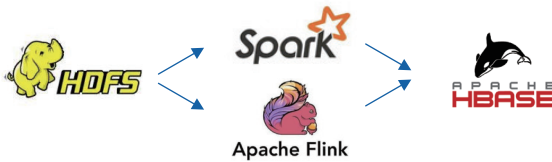
## 数仓建设

通过 Sqoop 同步业务数据库或 Flume 采集日志至 HDFS 中，利用 Spark 进行处理后导入 Hive，通过 Presto 进行查询，快速搭建数据仓库。



## 实时推荐系统

通过 Flink 进行实时数据的近似计算，通过 Spark 定时进行全量数据精确计算，通过 Lambda 架构，满足推荐系统的实时 + 精确的需求。



## 在线分析

通过 Kafka 传输实时数据至 Druid 中，进行实时读写，快速构建实时 OLAP 分析系统。

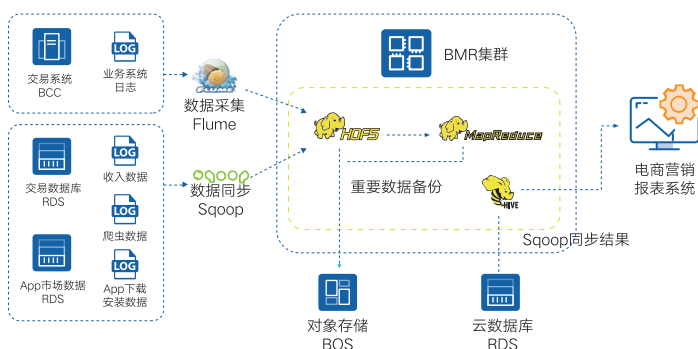


## 电商行业应用场景 - 电商运营数仓系统

电商行业需要基于营销数据、市场数据、用户行为数据等进行综合分析，制定有效运营策略。某电商网站基于 BMR，结合百度智能云大数据解决方案，一周完成云上运营数仓系统建设，使得运营策略更加有针对性、运营转化更高。

### 解决的问题：

- ▶ 运营数仓系统搭建从数据采集、到存储、处理、分析需要依赖许多组件，技术门槛高、运维成本大，完备搭建周期长，BMR 丰富的大数据组件，可支持完成高效搭建。
- ▶ 基于 BMR 及百度智能云大数据平台的云化弹性伸缩能力，计算周报表及月报表时快速拉伸资源、计算完成后快速释放资源，上线 1 年省去 IT 成本达百万。



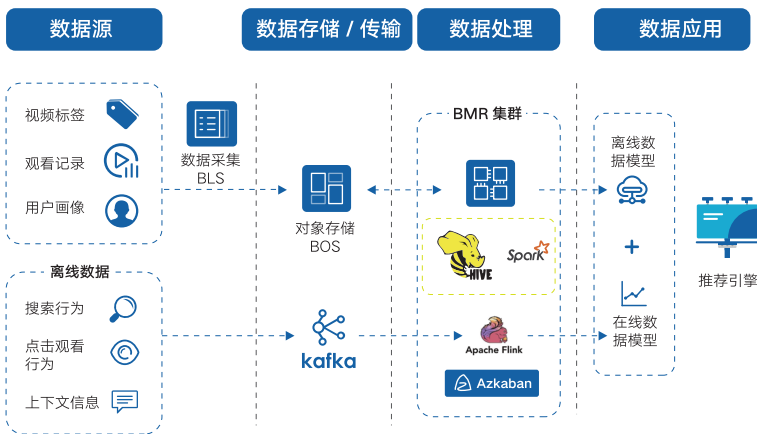
## 视频行业应用场景 - 视频推荐系统搭建

某互联网视频公司基于 BMR，利用百度智能云的企业级大数据解决方案，快速搭建推荐引擎，推荐准确率提升 20%。通过分析挖掘丰富的用户数据，可以为用户提供更加精准的视频推荐，达到推荐效果的“千人千面”。

### 解决的问题：

- ▶ 基于 BMR 及百度智能云大数据生态产品，快速搭建稳定、可靠的大数据处理平台，将基础平台搭建周期从月级降低到天级，并且省去了运维成本。
- ▶ 通过 BMR 的多级安全防护机制，保障平台及数据的安全性。



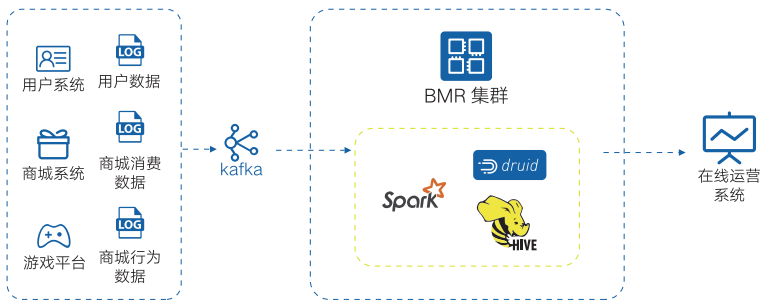


## 游戏行业应用场景 - 用户行为实时分析

在网络游戏行业，需要分析用户的行为数据，及时调整游戏机制和运营策略，保持游戏对用户的吸引力。某互联网游戏公司利用 BMR，结合百度智能云大数据解决方案，高效搭建在线运营系统，实时监控用户行为数据和运营数据情况，及时更新运营策略，驱动游戏机制创新，ARPU 值提升了约 60%。

### 解决的问题：

- ▶ 通过统一的消息中间件Kafka实现了海量、多种用户行为分析数据源的实时接入。
- ▶ 通过BMR中的Druid组件构建了秒级查询的在线运营系统，对运营效果极速反应。





百度智能云官方微信公众号

[cloud.baidu.com](http://cloud.baidu.com)

4008-777-818