

同样提供了 Ranger、Kerberos 等权限管控组件。当然，除了这些开源组件以外，最近几年出现了很多成熟的商业化大数据计算产品，无论是在计算性能、运行稳定、技术支撑等方面都优于开源社区产品，比如阿里巴巴提供的 MaxCompute（原 ODPS），可以轻松支持 PB 甚至 EB 级数据量计算及存储。

6.3.4 搭建运维管理平台

运维管理平台提供了界面化操作集群的能力。它帮助开发、运维人员简单快速地搭建和管理集群，同时也对集群健康状态进行监控，并在异常出现时及时作出告警。

当前云平台服务商及一些第三方 Hadoop 发行版都提供了运维管理平台。如：阿里云的 EMR、Hortonworks 的 HDP、Cloudera 的 CDH。

6.3.5 搭建研发管理平台

数据研发管理平台作为平台工具，从整个数据开发流程上辅助开发人员进行数据中台的建设。其应包含以下能力：数据集成、数据模型、数据开发 / 算法开发、数据质量、数据服务、数据资产、任务运维等。

数据集成作为数据中台数据采集层，在数据时效性上提供（准）实时、离线采集能力。在数据源格式上应支持结构化、半结构化和非结构化数据的采集能力。其中，应支持主流关系型数据库 CDC 日志、SDK 上报数据、文件流数据的（准）实时数据采