

集团大数据平台项目整体方案规划

大数据平台总路线：功能性-->稳定性-->智能化

功能化

1 平台增强功能：

1. HDFS容灾备份
2. HBase二级索引
3. HBase动态分割Region
4. YARN任务容器重用
5. YARN任务抢占式模型

1 多集群统一管理能力

1 Hadoop平台与云服务平台无缝整合

1 数据治理产品

1. 支持日、月、周、不定期、实时数据加载入库；
2. 支持结构化、非结构化数据处理；
3. 支持信息安全，对数据进行加密；
4. 支持结构化数据。如MPP，Oracle等关系型数据库的抽取加载；
5. 支持非结构化数据。如xml数据、json，html，xls等；
6. 支持大数据平台hadoop文件导出、导入；

1 ETL产品

1 自助分析产品

1 数据探索产品

1 爬虫引擎

1 图像识别工具

1 日志采集

稳定化

- 1 1. Yarn集群高可用
- 1 2. Hive的高可用
- 1 3. HBase的高可用
- 1 4. Spark的高可用
- 1 5. ETL高可用性
- 1 6. 应用高可用性

标准化

- 1 命名标准
- 1 发布标准
- 1 安全标准
 - 1. 系统环境安全
 - 2. 物理环境安全
 - 3. 网络安全
 - 4. 系统应用安全
 - 5. 数据安全
- 1 数据管控与平台管理能力

可视化

- 1 1. 平台运维可视化
- 1 2. 大数据基础能力可视化
- 1 3. 大数据开发可视化

智能化

- 1 自助分析
- 1 智能监控
- 1 数据探索

总结： 这是一些理论的东西，具体任务可以做成工作包，分解每一项工作。