Smartbi安装组件介绍

本文档介绍Smartbi及其各个组件的功能。

名称	说明
Smartbi	Smartbi主体应用。
知识库	存储Smartbi系统相关的数据信息。必不可少的一个重要组件。
Smartbi-MPP	高速缓存库，分布式的数据缓存库，分析人员在遇到大数据量查询场景下，可将数据抽取到Smartbi-Mpp进行加速，使用亿级数据查询。
Smartbi-UnionServer	跨库联合查询，分布式SQL查询引擎，处在Smartbi的SQL引擎层，为不同的数据源提供统一的SQL解析、跨库查询能力。
Smartbi-Proxy	负载均衡服务器。用于smartbi集群代理转发。
导出引擎	主要用于自助仪表盘等场景导出图片或者pdf等。
数据挖掘-实验引擎	负责接收smartbi 发送实验执行请求。 通过解析实验定义，生成spark 计算任务或python计算任务，分别发送给spark集群或python集群。 本身并不承担计算任务，只负责计算任务的调度跟分发。
数据挖掘-服务引擎	提供模型预测服务给第三放系统调用。
Python执行节点	python 执行环境，主要承担实验引擎发送过来的python计算任务。 可以横向扩张，实验引擎会根据负载均衡策略把python计算任务分发到不同节点。
Spark	分布式计算平台，主要承担实验引擎发送过来的spark计算任务，其中Worker实例可以横向扩展。
Hadoop	分布式系统基础平台，主要存储计算任务的中间结果数据
Zookeeper	用于实现数据挖掘实验引擎和数据挖掘服务引擎的高可用。