# 代码svn

1. bigdata-manager

描述：大数据任务平台WEB工程，负责流程创建和管理、定时调度任务的创建和管理、计算结果状态和数据质量监控的展现，配置信息写入mysql数据库。

Svn地址：https://192.168.0.126/svn/tech\_web/bigdata/bigdata-manager

开发工具：IDEA

开发语言：java、jsp

2. bigdata-scheduling

描述：大数据任务平台定时调度工程，读取mysql中bigdata-manager配置的定时任务，定时拉起spark任务，处理入库或统计计算，监控spark任务状态，保存任务运行结果到mysql中。

Svn地址：https://192.168.0.126/svn/tech\_web/bigdata/bigdata- scheduling

开发工具：IDEA

开发语言：java

3. bigdata-processing

描述：大数据任务平台spark处理工程，读取WEB工程的流程配置，解析为spark程序，执行具体的处理逻辑，此工程为老版本（CDH5）集群的工程，能运行在新集群（CDH6）下，只包含统计相关的功能，需要使用sbt assembly打成完整的包。

Svn地址：https://192.168.0.126/svn/tech\_web/bigdata/bigdata-processing

开发工具：IDEA

开发语言：java、scala

4. gbsp-processing

描述：大数据任务平台spark处理工程，读取WEB工程的流程配置，解析为spark程序，执行具体的处理逻辑，基于新版本（CDH6）集群开发，包含统计相关功能、批量数据入库hive、实时数据入hive、数据入kudu功能，需要使用sbt assembly打成完整的包。

Svn地址：https://192.168.0.126/svn/tech\_web/bigdata/gbsp-processing

开发工具：IDEA

开发语言：java、scala

5. flume-plugin

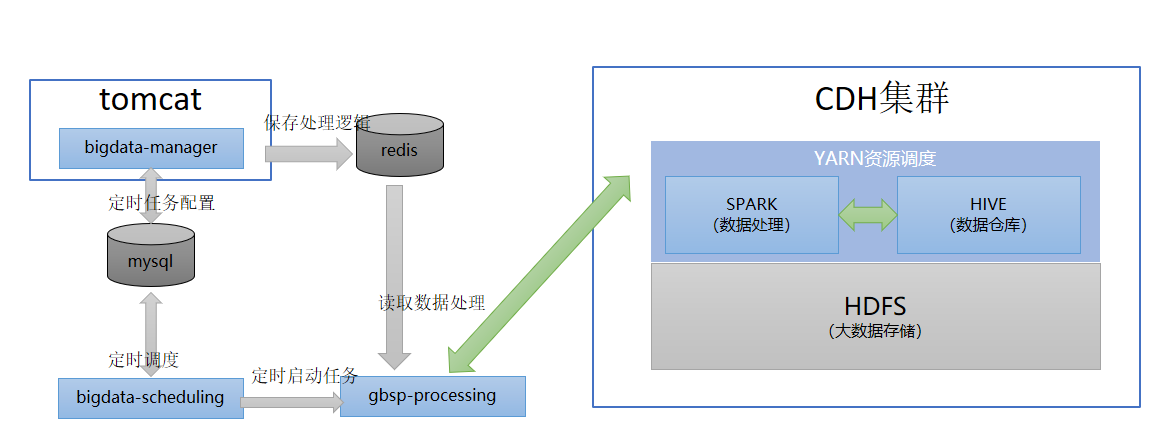
描述：flume采集数据插件，扩展flume的功能，用于实时数据采集，主要功能包括：支持通过jdbc采集mysql数据；从数据记录中提取字段到flume的header中；对数据记录中的数据进行转换、过滤处理等。

Svn地址：https://192.168.0.126/svn/tech\_web/bigdata/flume-plugin

开发工具：IDEA

开发语言：java

# 部署说明



1. bigdata-manager，web工程，部署在tomcat里面，使用腾讯云mysql。

统计相关的所有库的连接脚本：

114.132.169.71 服务器：/data/bigdata/script

2. bigdata-scheduling，普通java工程，主要是定时调度相关，里面包含了通过excel转义自动生成trino导入配置sql和hive建表语句的代码的功能。

3. gbsp-processing，通过spark处理数据的工程，注意打完整的包（sbt assembly），要把CDH集群没有的依赖jar打进去，通过sbt配置文件配置是否需要打到jar包里面。

provided 参数为不打到jar里面，一般为cdh集群已经包含的jar，不加这个参数就会打入。

