日志分析系统

# 第一章项目概述

## 1.1 项目架构

使用kafka，spark，hbase开发日志分析系统。

项目技术架构

Kafka作为日志事件的消息系统，具有分布式，可分区，可冗余的消息服务功能。

Spark使用spark stream功能，实时分析消息系统中的数据，完成计算分析工作。

Hbase做为后端存储，存储spark计算结构，供其他系统进行调用

计算机生成了可选文字: }LogFile日LogFile…｝LogFile}.．口
Kafka
叮牛、
、＿竺巴夕
Spark
HBase
口口口、l一
一引口口L。nr一
贬白口―卜O吧一
砒臼映曰口口．。泪〉一
万口『｝口口一口匕r一
阵二西口』．『ee一

## 1.2 软件部署

### 1.2.1 hadoop 版本

Hadoop相关软件如zookeeper、hadoop、hbase，使用的是cloudera的 cdh 5.2.0 版本。

### 1.2.2 Kafka

Kafka 版本： 2.9.2-0.8.1.1

1. **安装部署**
   1. 部署kafka

**tar -xzf kafka\_2.9.2-0.8.1.1.tgz**

* 1. 编辑kafka 配置文件

**config/server-1.properties:**

**broker.id=0**

**port=9093**

**log.dir=/tmp/kafka-logs**

**config/server-2.properties:**

**broker.id=1**

**port=9093**

**log.dir=/tmp/kafka-logs**

**config/server-3.properties:**

**broker.id=2**

**port=9093**

**log.dir=/tmp/kafka-logs**

* 1. 启动kafka

**bin/kafka-server-start.sh config/server-1.properties &**

**bin/kafka-server-start.sh config/server-2.properties &**

**bin/kafka-server-start.sh config/server-3.properties &**

1. **创建kafka topic**

**> bin/kafka-topics.sh --create --zookeeper 10.10.102.191:2181, 10.10.102.192:2181, 10.10.102.193:2181 --replication-factor 3 --partitions 1 --topic recsys**

1. **查看是否创建成功**

**> bin/kafka-topics.sh --list --zookeeper localhost:2181**

**> bin/kafka-topics.sh --describe --zookeeper localhost:2181 --topic my-replicated-topic**

Topic:my-replicated-topic PartitionCount:1 ReplicationFactor:3 Configs:

Topic: my-replicated-topic Partition: 0 Leader: 1 Replicas: 1,2,0 Isr: 1,2,0

1. **kafka启动测试**

**> bin/kafka-console-producer.sh --broker-list localhost:9092 --topic test**

This is a message

This is another message

> **bin/kafka-console-consumer.sh --zookeeper localhost:2181 --topic test --from-beginning**

This is a message

This is another message

1. **注意事项**

在开发程序的时候，producer客户端必须要配置上broker的host映射信息，即使你的程序中使用的都是ip地址。

# 第二章项目开发

## 2.1 文件读取入Kafka

程序依赖Jar包

客户端依赖 libs目录：

Logback包：[**logback-classic-1.1.2.jar**](http://data.ztgame.com/redmine/attachments/download/243/logback-classic-1.1.2.jar)，**[logback-core-1.1.2.jar](http://data.ztgame.com/redmine/attachments/download/244/logback-core-1.1.2.jar" \o ")**

Kafka包（在kafka安装包lib目录中）

客户端配置文件 conf目录：

Logback：logback.xml

## 2.2 Spark\_Streaming 处理数据

## 2.3 HBase 保存数据

创建hbase表

create ‘recsys\_logs’,’f’

服务器端部署

jar xvf recsys-1.0.jar