**北京邮电大学**

毕业设计需求分析

题 目： HBASE与关系数据库同步工具的设计与实现

学 院： 计算机学院

专 业： 计算机科学与技术

班 级： 2012211308

学 号： 2012211320

姓 名： 胡官涛

指导教师： 王红熳

填表日期： 2016 年 3 月 20 日

需求规格说明书

[1．引言 3](#_Toc446197902)

[1.1编写目的 3](#_Toc446197903)

[1.2项目背景 3](#_Toc446197904)

[1.3参考资料 3](#_Toc446197905)

[2．任务概述 4](#_Toc446197906)

[2.1目标 4](#_Toc446197907)

[2.2运行环境 4](#_Toc446197908)

[2.3条件与限制 4](#_Toc446197909)

[3．功能需求 5](#_Toc446197910)

[3.1功能定位 5](#_Toc446197911)

[3.2功能描述 5](#_Toc446197912)

[4．性能需求 5](#_Toc446197913)

[4.1数据量级 5](#_Toc446197914)

[4.2时间特性 6](#_Toc446197915)

[4.3适应性 6](#_Toc446197916)

[5．运行需求 6](#_Toc446197917)

[5.1人机交互体验 6](#_Toc446197918)

[5.2软件接口 6](#_Toc446197919)

[5.3故障处理 6](#_Toc446197920)

[6．其它需求 7](#_Toc446197921)

# 1．引言

## 1.1编写目的

为明确软件需求、安排项目规划与进度、组织软件开发与测试，撰写本文档。

本文档供项目经理、设计人员、开发人员参考。

## 1.2项目背景

因特网的发展已经彻底改变了人们的生活，每时每刻都有大量数据被传送到互联网上。伴随着数据规模越来越大，越来越多的数据存取功能开始水平扩展，这样就可以删除和增加存储服务，从而适应多变的业务需求。而传统的关系数据库虽然在SQL查询方面表现良好，但是他们更侧重在一台服务器上，海量数据的存储成为瓶颈，单台服务器无法加载海量数据。而HBASE不同于传统的关系数据库，它是基于列存储的非关系数据库，在海量数据的存储和查询方面有很大优势。但是HBASE不支持SQL查询，而且对于繁复的业务流程难以满足需求。而传统的关系数据库恰好能和HBASE互补，如果能使用工具将二者随意转换，那么刚好在弥补彼此的缺点之后，在相应的需要的场景使用不同的数据模式，便能在大数据时代得心应手。

本次论文研究HBASE与关系数据库同步工具的设计与实现，旨在提供一个方便快捷的工具，从而使企业在使用的时候可以避免传统关系数据库和HBASE的不足点，充分利用二者的优点，适应大数据时代背景和多变的业务需求，从而更好的提供互联网服务。

## 1.3参考资料

[1]：周海汉 Apache Software Foundation 《HBASE官方文档中文版》http://abloz.com/hbase/book.html

[2]：明日科技 《JAVA从入门到精通（第三版）》 清华大学出版社，

[3]：[美] Lars George，译者:代志远/刘佳/蒋杰 《Hbase权威指南》 东南大学出版社

[4]：Baron Schwartz译者: 王小东/李军/康建勋 《高性能mysql》 电子工业出版社

[5]：肖丁 修佳鹏 《软件工程模型与方法》 北京邮电大学出版社

[6]：杨寒冰 同济大学毕业设计论文 2013 《关系型数据库与HBASE间的数据迁移介绍》

[7]：Abraham Silberschatz等 《数据库系统概念》 机械工业出版社

[8]：作者不详 百度文库 《HBase基本shell命令介绍》

[9]：HBASE视频教程（极客学院） http://www.jikexueyuan.com/course/hbase/

[10]：java课程视频（极客学院） http://search.jikexueyuan.com/course/?q=java

[11]：intelij idea 官方网站 http://www.jetbrains.com/idea/

# 2．任务概述

## 2.1目标

本软件旨在设计一个在Linux平台上的从关系数据库到HBASE数据库导出、同步软件，完成关系数据库到HBASE的数据同步和导出

项目名称：关系数据库到HBASE数据同步、导出工具的设计与实现

## 2.2运行环境

硬件要求：

CPU：不低于528MHz

内存：不低于2GB

软件要求

操作系统：ubuntu 12.04

开发环境：Java

运行软件：intelij idea

## 2.3条件与限制

该软件实现关系数据库到HBASE的数据同步，支持按时间、按表同步，并对过程进行监控

处理对象：HBASE数据 / 关系数据库数据（MySQL、sqlserver、oracle）

输入数据：MySQL / sqlserver / oracle数据表

输出数据： HBASE数据表，操作日志，控制台提示

# 3．功能需求

## 3.1功能定位

该工具用于数据迁移和迭代，在高效处理数据方面也有很好作用

系统中的位置为：



## 3.2功能描述

2.2.1关系数据库同步或者导入到HBASE

2.2.2导出：需要手动进行导出，并可以自定义设置

2.2.3 同步：在线同步，实时，可选择要同步的表

2.2.4 按表、按时间导出、同步，提供过程监控

# 4．性能需求

## 4.1数据量级

HBASE中的数据量级为TB级别，mysql中数据量级无要求

## 4.2时间特性

4.2.1对操作的响应时间(开始处理时间)

平均0.5s、最长1.5s。

4.2.2处理时间

即刻处理，总时延在30min以下

4.2.4其他

超时会警告并询问是否继续

## 4.3适应性

4.3.1操作方式

Linux shell 操作，有相应操作帮助

4.3.2运行环境

Java Hadoop

4.3.3与其他软件的接口以及开发计划发生变化时

具有适应能力，也具有一定的通用性，在数据结构、程序不变的情况下，能应用于不同的系统。

4.3.4其他补充

界软件由Java开发，具有良好的适应性，可以在Windows、Linux、osX等系统下运行

# 5．运行需求

## 5.1人机交互体验

指令简洁，操作简单，由于多数会应用于服务器，故没有图形化界面，但是命令行的操作方法对数据维护人员和开发者来说更加熟悉和快捷，具有良好的用户体验。

## 5.2软件接口

无

## 5.3故障处理

正常使用时不应出错，若运行时遇到不可恢复的系统错误，也必须保证数据库完好无损，并报告相应错误。在出现故障后应从整个系统数据库重新载入数据进行操作。

# 6．其它需求

1)系统的功能实现情况: 用户可在本系统下实现各种本文档中提及的的功能

2)系统的容错性: 用户输错数据都有提示信息，具有较好的容错性能。