

本科毕业论文（设计）

文献综述

|  |  |
| --- | --- |
| **学 生 姓 名** | **吴光宇** |
| **学号** | **2014051016** |
| **专业** | **计算机科学与技术** |
| **年级班级** | **2014级计算机（应用）1班** |
| **指导教师** | **张欢（讲师）** |
| **所在学院** | **计算机学院** |
| **提交日期** | **2018年4月18日** |

2018 年 4 月

成都信息工程大学 计算机学院

基于Hadoop自适应光学分析系统的设计与开发

# 基本概念的概述

## 自适应光学

自适应光学是在1953年由Horace W.Babcock提出，主要构想是用闭环矫正波前误差来补偿天文视宁度[1]。造成波前误差的主要原因是大气湍流，而波前误差会极大的影响最后的成像质量，被动光学以及早期的主动光学无法很好的解决这个问题，而自适应光学的出现为解决这个问题提供了可能。

自适应光学系统主要由波前传感器，波前控制器以及波前矫正器组成。其中波前传感器主要负责捕捉光因大气湍流和其他原因造成的波前误差，波前控制器通过分析、计算波前传感器传输的数据，并将其转化为波前矫正器的电压参数，而波前矫正器负责控制控制矫正单元，使望远镜镜面发生形变，以矫正相位，从而获得成像质量较好的图形。

## Hadoop

Apache Hadoop软件库是一个框架，允许使用简单的编程模型跨分布式计算机集群处理大型数据集[2]。

Hadoop是google分布式系统的开源版本，hadoop有三个核心：

HDFS文件系统，该文件系统直接参照了Google的GFS文件系统，在设计上有诸多相似之处。HDFS运行在廉价的商业软件上，因此HDFS文件系统是高度容错的[3]。HDFS通过存放多份数据来避免因为硬件错误造成的数据丢失，通过相应的算法来保证数据的一致性。

第二个核心是mapreduce计算框架，hadoop的开发人员在HDFS文件系统的基础上，实现了google的mapreduce计算框架，使hadoop真正成为了一个可以实际运用的分布式计算框架，mapreduce的本质在于并行运算，mapreduce将耗时的计算任务进行划分，并将该任务传送到计算数据所在的节点，然后多节点并行运算，然而这种并行几乎不需要付出代价，除了需要满足该编程模型的一些要求之外，这使得没有任何并行和分布式系统经验的程序员可以轻松利用大型分布式系统的资源[4]。

最后一个核心是hadoop的资源调度框架yarn，可以保证hadoop集群的负载均衡。

# 国内外研究现状

现在基于分布式框架的程序已经有了较为广泛的应用，以google为代表的大量公司已不同程度的将自己的基础服务向分布式进行迁移，比如google搜索引擎的数据存储在GFS分布式文件系统中。

在自适应光学系统短时间运行后，将累计大量的数据，这些数据主要由波前传感器捕获的斜率数据和波前控制器计算获得电压数据组成。而通常的做法是将这些数据存储在单机的关系数据库中，当数据量变得庞大，便需要对数据库进行横向扩展，而关系型数据库的横向扩展又极为困难。

解决这类问题的通常做法是将数据由关系型数据库向非关系型数据库迁移，虽然现在有很多其他系统与分布式结合的例子，但是还没有自适应光学分析系统与分布式相结合的例子。

# 自适应光学分析系统改进方式及技术路线

## 现有的自适应光学分析系统

现有的自适应光学分析系统通过数据的导入，对实验数据进行存储，存储的方式是传统的关系型数据库，后续的数据查询，数据编辑等都是基于传统数据库，同时可以基于这些存储的数据进行计算分析，获取数据的相关信息和特征。

上述系统存在一个问题，不能支持实时数据的导入，或者实时数据的导入收到单台物理机的限制，无法支持大量数据的并行查询，计算。数据的存储量受到单台物理机的限制，而通过横向扩展突破该现在的代价将及其高昂。

## 改进方案

考虑到大量数据的存储及实时查询需求，可以将数据存储方式由关系型数据库替换为非关系型的分布式数据库，并实现以下一些核心要点：

实时数据的并行化插入集群。

分布式数据库上的数据查询。

多方式的数据展示。

相关特征计算的并行化。

以此可以作为后续脱离传统光学分析方法而采用大数据分析方法的基础。

## 技术路线简介

### Spring

支撑spring的两大核心原理，一是控制反转(依赖注入)，在传统的java程序中，系统模块之间必然存在这各种各样的依赖关系，过强的耦合会使软件结构复杂、维护困难[5]。而spring的控制反转(依赖注入)则可以解决这个问题。另一个是面向切面编程，面向切面编程主要实现的目的是针对业务处理过程中的切面进行提取，它所面对的是处理过程中的某个步骤或阶段，以获得逻辑过程中各部分之间低耦合性的隔离效果[6]。这些切面，也叫关注点，一个关注点，就是一个特定的目的，或者说一个系统需要实现的目标[7]，而这个关注点可能横跨几个模块。

在spring中实现控制反转(依赖注入)的方式是jvm级别维护一个Ioc容器，该Ioc容器通过配置文件来生成、管理、销毁类实例，而在程序中需要获取类实例都在Ioc容器中进行获取，即组件之间的依赖关系由容器在运行期间决定，形象的说，就是就是容器在运行期间将对象实例注入到组件中，而组件只需要关心自己的业务逻辑，而不需要关心使用的资源来自何方。

而面向切面编程主要解决代码重复和关注点分散的问题，面向切面编程比较常见的例子就是系统的日志接口，几乎每个类都需要使用日志接口，如果将日志功能单独抽象为一个模块，然后再每个类中都对其进行实例化，这样会造成所有模块都和日志模块耦合，而如果将日志功能的代码嵌入到每个需要使用的地方则会造成该段代码重复率极高，而对于原本的类来说，日志功能的实现却不是该模块应该关注的内容，为了解决这些问题，spring提供了面向切面编程的功能，通过声明的方式将功能模块应用到需要他们的模块中去，进而改变他们的行为。

Spring是基于配置来实例化对象、应用对象、销毁对象，随着项目规模的不断扩大，spring需要的配置文件越来越繁杂，以至于维护配置信息变得极其困难，spring boot的出现就是为了解决这个问题，spring boot默认习惯优于配置，即spring默认配置好了符合大多数项目的配置文件，而使用者无需关心，直接使用即可，而调整默认配置也变得极其简单。

### Spring Security

Spring security是为基于spring的应用程序提供声明式安全保护的安全性框架[8]。

认证和权限控制是Spring Security要处理的两个主要领域[9]。认证需要处理这样一些问题：

1.用户具有哪些权限

2.访问对应资源需要什么权限

3.用户是否具有访问该资源的权限

授权需要解决这些问题：

1.该系统权限种类

2.如何授予用户权限

Spring security支持广泛的认证模型，其中大部分既支持标准机构发布的模型，也支持第三方发布的模型。

Spring security的权限认证主要集中在以下三个方面：

1.WEB请求的认证

2.方法执行的认证

3.独立域资源的授权

### Hbase

Hbase是一个开源的非关系型分布式数据库（NoSQL），它参考了google的bigTable[14]建模，实现的语言为java[10]。

Hbase适合于存储大数据，支持单条记录的快速查询，在任意指定位置单条或批量添加，删除数据[11]。Hbase将数据作为字节块存储在分布式文件系统中，如果列值为空，将不会被存储，这将大大节省存储空间。而行键按LSM结构进行存储，进而支持单条记录的快速查询。

分布式的非SQL数据库提供了人们十分需要的可伸缩特性[15]。这使得横向扩展变得容易，从而能够存储更多的数据。

# 程序中涉及的相关特征值的计算方式

## 常见统计数据

## Zernike

通过大气湍流的光线会产生波前畸变，严重影响成像质量，为了计算波前矫正器的电压，需要获取到连续的波前形状，而恢复波前形状的算法叫波前重构算法，现在常见的波前重构算法有模式法、区域法、直接斜率法[12]。

由于 Zernike 多项式的各项与光学像差有相应的对应关系 ,用 Zernike 多项式对镜面面形数据进行处理的方法已经广泛应用于工程项目[13]。

# 总结

本文对比分析了基于传统关系型数据和基于分布式关系数据库Hbase的自适应光学分析系统各自的优势和劣势，总结了当前自适应光学分析系统的研究现状。由于关系型数据库本身必须严格按照范式进行设计，从而导致表之间复杂的关联关系，当数据量达到一定的数量级，维护这种关联关系将变得极为困难，且准确获取相关信息在这种情况下将变得极为耗时，这将影响到系统的所有功能，而将关系型数据库置换为非关系型数据库后，可以有效解决上述提到的问题，而且在将整个系统构建在hadoop环境下，为后续进行数据的分析提供了可能。

参考文献

1. 林旭东,薛陈,刘欣悦等.自适应光学波前矫正器技术发展现状[J].中国光学,2012年,第5卷,第4期.Marvin M. Theimer, Keith A. Lantz, and David R. Cheriton Preemptable Remote Execution Facilitiesfor the V-System 2002
2. Apach Hadoop:what is apache hadoop[EB/OL]. <http://hadoop.apache.org/>
3. BORTHAKUR D. HDFS architecture guide[EB/OL]. (2013T04T04) [2014 T09 T04]. http://hadoop.apache.org/docs/r1.2.1/hdfs\_design.html.Performance of PVM with the MOSIX Preemptive Process Migration Scheme
4. Dean J, Ghemawat S. MapReduce: Simplified data process-ing on large clusters/ / Proceedings of the 6th Symposium onOperating System Design and Implementation (OSDI 04) .San Francisco, California, USA, 2004: 137-150Christopher Clark, Keir Fraser, Steven Hand, Jacob Gorm Hansen†, Eric Jul†, Christian Limpach, Ian Pratt, Andrew Warfield Live Migration of Virtual Machines NSDI'05 Proceedings of the 2nd conference on Symposium on Networked Systems Design & Implementation - Volume 2 2005
5. 薄奇,许林英.Spring框架中IoC的实现[J].微处理机,2008年,第1期
6. 百度百科：面向切面编程[EB/OL]. https://baike.baidu.com/item/面向切面编程
7. 张国平,万仲保,刘高原.Spring Aop框架在J2EE中的应用[J].微计算机信息,2007,第23卷，第12-3期
8. Craig Walls. Manning.Spring.in.Action[M].4th.Edition.北京市丰台区成寿寺路11号：人民邮电出版社。
9. 蒋丛萃,史卓.Spring Security在电子商务中的设计应用[J].电脑知识与技术,2016，第5x期
10. 维基百科：Apache Hbase[EB/OL]. <https://zh.wikipedia.org/wiki/Apache_HBase>
11. 张叶,许国艳,花青.基于HBase的矢量空间数据存储与访问优化[J].计算机应用,2015,35(11):3102-3105
12. Becherer R J,Horwitz B A. Adaptive optics systems and technology[C]// International Society for Optical Engineering.Meeting sponsored by SPIE. Proceedings of SPIE-The International Soc.for Opt.Eng, San Diego,CA:[s.n.], 1982,365:1-150.
13. 单宝忠,王淑岩,牛憨笨,刘颂豪. Zernike 多项式拟合方法及应用[J].光学 精密工程,2002,第3期
14. F. Chang, J. Dean, S. Et al, Bigtable: A Distributed Storage System for Structured Data, OSDI 2006.
15. Chao-Hsien Lee, Yu-Lin Zheng, ‘ Automatic SQL-to-NoSQL Schema Transformation over the MySQL and HBase Databases’, 2015International Conference on Consumer Electronics