[分布式大数据管理系统 1](#_Toc21111)

[华东师范大学 1](#_Toc12616)

[第一章 Greenplum简介与安装 7](#_Toc2555)

[1.1 Greenplum发展历程 7](#_Toc26467)

[1.2 Greenplum与PostgreSQL 7](#_Toc983)

[1.2.1 PostgreSQL 7](#_Toc3812)

[1.2.2 Greenplum 8](#_Toc4402)

[1.3 MPP介绍 9](#_Toc4572)

[1.3.1术语： 9](#_Toc25424)

[1.3.2 Greenplum 和 PostgreSQL 的区别 13](#_Toc30221)

[1.3.3 MPP概述 13](#_Toc25654)

[第二章 Greenplum使用指导 36](#_Toc13201)

[2.1 Greenplum 编译安装和调试 36](#_Toc4401)

[2.1.1. 从源代码编译 Greenplum 37](#_Toc2472)

[2.1.2. 初始化 Greenplum 集群 41](#_Toc20202)

[2.2 集群初始化问题调试 42](#_Toc2853)

[2.3 Greenplum SQL执行流程概要 45](#_Toc10913)

[2.3.1 查询计划 46](#_Toc10000)

[2.3.2 查询执行 47](#_Toc23245)

[2.4. 调试 Greenplum MPP 数据库 49](#_Toc22381)

[2.4.1 调试 Master 节点Backend进程 49](#_Toc8499)

[2.4.2 调试 Segment 节点Backend进程 （QE) 52](#_Toc12122)

[2.4.3 使用 IDE 调试 53](#_Toc812)

[2.5 基础语法 56](#_Toc2771)

[2.5.1 基本操作 56](#_Toc12475)

[2.5.2 基本数据类型 59](#_Toc30174)

[2.5.3 常用函数 63](#_Toc31126)

[2.6 实验要求 66](#_Toc19960)

[第三章 Greenplum体系结构 67](#_Toc30706)

[3.1 分布式数据库架构 67](#_Toc3729)

[3.2 分布式数据库可靠性 70](#_Toc5534)

[3.2.1 备份机制 70](#_Toc29681)

[3.2.2 分布式事务 72](#_Toc16889)

[3.3 实验导引 79](#_Toc4747)

[3.4 实验要求 81](#_Toc20420)

[第四章 存储机制 83](#_Toc11900)

[4.1 数据组织方式 86](#_Toc9629)

[4.1.1 Greenplum 86](#_Toc17171)

[4.1.2 BigTable 88](#_Toc27666)

[4.1.3 MongoDB 90](#_Toc3868)

[4.1.4 Neo4j 91](#_Toc9560)

[4.2 行列存储方案 93](#_Toc6217)

[4.2.1 行列存储在不同场景下的应用 93](#_Toc17272)

[4.2.2 在Greenplum中使用存储 95](#_Toc1811)

[4.3 appendonly表和压缩技术 96](#_Toc14649)

[4.3.1 关于appendonly表和压缩表的基本介绍 96](#_Toc1683)

[4.3.2 与AOT有关的系统表 97](#_Toc8979)

[4.3.3 常用的压缩算法 100](#_Toc28698)

[4.3.4 ZLIB的压缩过程 104](#_Toc20709)

[4.3.5 压缩表的创建 107](#_Toc2194)

[4.4 实验：向gp中添加新的压缩算法 110](#_Toc32336)

[4.4.1 Greenplum创建压缩表及数据的压缩过程分析 110](#_Toc1525)

[4.4.2 snappy压缩算法 112](#_Toc14044)

[4.4.4修改源码（以snappy为示例） 113](#_Toc1365)

[4.4.5压缩效率比较 116](#_Toc3155)

[4.5 存储机制实验导引 118](#_Toc7754)

[4.5.1 存储机制实验实例 118](#_Toc14589)

[4.5.2存储机制实验要求](#_Toc26316)

[4.6参考文献及进一步阅读材料 121](#_Toc26316)

[第五章 索引 123](#_Toc21658)

[5.1 索引基本原理](#_Toc17880_WPSOffice_Level1) [2](#_Toc17880_WPSOffice_Level1)

[5.2 索引相关技术](#_Toc23270_WPSOffice_Level1) [2](#_Toc23270_WPSOffice_Level1)

[5.2.1 B-Tree 索引](#_Toc17880_WPSOffice_Level2) [2](#_Toc17880_WPSOffice_Level2)

[5.2.2 Hash 索引](#_Toc23270_WPSOffice_Level2) [2](#_Toc23270_WPSOffice_Level2)

[5.2.3 Bitmap 索引](#_Toc5837_WPSOffice_Level2) [2](#_Toc5837_WPSOffice_Level2)

[5.2.4 Gist 索引](#_Toc9487_WPSOffice_Level2) [2](#_Toc9487_WPSOffice_Level2)

[5.2.5 最新研究成果](#_Toc23253_WPSOffice_Level2) [2](#_Toc23253_WPSOffice_Level2)

[5.3 索引实验部分](#_Toc5837_WPSOffice_Level1) [2](#_Toc5837_WPSOffice_Level1)

[5.3.1 索引分析实例](#_Toc25746_WPSOffice_Level2) [2](#_Toc25746_WPSOffice_Level2)

[5.3.2 实验要求](#_Toc8225_WPSOffice_Level2) [2](#_Toc8225_WPSOffice_Level2)

[5.4 本章小结](#_Toc9487_WPSOffice_Level1) [2](#_Toc9487_WPSOffice_Level1)

[5.5](#_Toc23253_WPSOffice_Level1) [2](#_Toc23253_WPSOffice_Level1)

[第六章 查询执行 148](#_Toc26075)

[6.1 Greenplum中的查询计划生成 149](#_Toc13094)

[6.1.1 查询计划优化实例 150](#_Toc22756)

[6.1.2 查询计划的内部机制 159](#_Toc25406)

[6.1.3 分布式查询计划生成 163](#_Toc9248)

[6.1.4 ORCA查询优化器 169](#_Toc2455)

[6.2 Greenplum中的分布式查询执行 178](#_Toc18702)

[6.2.1 查询执行器相关概念 178](#_Toc17031)

[6.2.2 并行执行流程 182](#_Toc22576)

[6.2.3 执行模式 185](#_Toc21482)

[6.2.4 数据广播与重分布 186](#_Toc8655)

[6.2.5 Greenplum分析函数执行 188](#_Toc20307)

[6.3 Greenplum 源码修改实例 190](#_Toc12762)

[6.3.1 select-from子句优化 190](#_Toc30460)

[6.3.2 多表连接——贪心算法 191](#_Toc6047)

[6.4查询执行实验导引 193](#_Toc780)

[6.4.1查询执行分析实例 193](#_Toc16500)

[6.4.2查询执行实验要求 198](#_Toc4645)

[6.5查考文献 198](#_Toc9354)

[第七章 数据库管理 199](#_Toc10079)

[7.1用户和角色 199](#_Toc16260)

[7.1.1 角色、组、用户介绍 199](#_Toc26904)

[7.1.2 三者的区别 201](#_Toc23171)

[7.1.3 在Greenplum中使用新用户 203](#_Toc12096)

[7.1.4权限与回收 205](#_Toc7104)

[7.2 锁与MVCC 206](#_Toc9482)

[7.2.1 事务隔离 206](#_Toc12282)

[7.2.2 锁机制 209](#_Toc8494)

[7.2.3 MVCC 212](#_Toc17465)

[7.3 两阶段提交协议 214](#_Toc12567)

[7.4数据目录 218](#_Toc31409)

[7.4.1 数据库 219](#_Toc30392)

[7.4.2 表 220](#_Toc7123)

[7.4.3 表空间及其创建实例 221](#_Toc13793)

[7.5 数据库管理实验导引 224](#_Toc3275)

[7.5.1 数据库管理实例 224](#_Toc5440)

[7.5.2数据库管理实验要求 229](#_Toc7928)

[第八章 典型应用 230](#_Toc30842)

[8.1 海量轨迹数据管理系统 230](#_Toc10278)

[8.1.1轨迹数据分类 230](#_Toc17407)

[8.1.2 轨迹数据应用领域 232](#_Toc4493)

[8.1.3轨迹数据挖掘特点 234](#_Toc8324)

[8.1.4交通流量监控系统实例 235](#_Toc15338)

[8.2 面向POI的地理信息展示系统 245](#_Toc17225)

[8.2.1 PostGIS简介 247](#_Toc18725)

[8.2.2 Greenplum中使用postGIS 249](#_Toc10646)

[8.2.3 Greenplum中使用QuantumGIS 253](#_Toc8290)

[8.2.3 使用QuantumGIS可视化大众点评数据 254](#_Toc26898)

[8.3面向大规模流数据处理的农业大数据应用 256](#_Toc26485)

[8.4利用外部表读取xml格式数据 256](#_Toc994)

[8.4.1应用示例 257](#_Toc17675)

[8.4.2 实现方法 258](#_Toc26290)

[8.5 基于天气和交通事故的路径规划 261](#_Toc6409)

[8.5.1 Greenplum与MySQL分析对比 261](#_Toc10261)

[8.5.2 实验过程 262](#_Toc1259)

[8.6 实验：greenplum中加密算法的改进与实现 267](#_Toc7141)

[8.6.1 Greenplum加密过程分析 267](#_Toc11317)

[8.6.2 Greenplum加密算法的改进 268](#_Toc8831)

[8.7 分布式数据库负载生成器 272](#_Toc5412)

[8.7.1 GreenPlum Command Center 273](#_Toc6934)

[8.7.2负载生成器设计架构 275](#_Toc5197)

[8.6.3 实验过程 278](#_Toc22900)

[附录一 4.5实验代码 285](#_Toc16841)

[附录二 6.4.1实验代码 291](#_Toc17117)

[附录三 8.4实验代码 296](#_Toc11544)