

Greenplum 企业应用实战

何勇 陈晓峰 著

Enterprise Application with Greenplum

- 阿里巴巴资深技术工程师撰写，完全展现阿里巴巴的Greenplum企业实战经验
- 系统介绍Greenplum的功能特性、使用方法、高级应用；详细讲解Greenplum的系统架构、运维管理、性能优化和各种技巧；包含大量企业级应用案例和实操指导



机械工业出版社
China Machine Press

Greenplum 企业应用实战

Enterprise Application with Greenplum

何勇 陈晓峰 著



机械工业出版社
China Machine Press



图书在版编目 (CIP) 数据

Greenplum 企业应用实战 / 何勇, 陈晓峰著. —北京: 机械工业出版社, 2014.10
(数据库技术丛书)

ISBN 978-7-111-48100-3

I. G… II. ①何… ②陈… III. 关系数据库系统 IV. TP311.138

中国版本图书馆 CIP 数据核字 (2014) 第 226297 号

Greenplum 企业应用实战

出版发行: 机械工业出版社 (北京市西城区百万庄大街 22 号 邮政编码: 100037)

责任编辑: 姜 影

责任校对: 董纪丽

印 刷: 北京市荣盛彩色印刷有限公司

版 次: 2014 年 10 月第 1 版第 1 次印刷

开 本: 186mm×240mm 1/16

印 张: 21.75

书 号: ISBN 978-7-111-48100-3

定 价: 69.00 元

凡购本书, 如有缺页、倒页、脱页, 由本社发行部调换

客服热线: (010) 88378991 88361066

投稿热线: (010) 88379604

购书热线: (010) 68326294 88379649 68995259

读者信箱: hzjsj@hzbook.com

版权所有·侵权必究

封底无防伪标均为盗版

本书法律顾问: 北京大成律师事务所 韩光 / 邹晓东

为什么写作本书

阿里巴巴是国内最早使用 Greenplum 作为数据仓库计算中心的公司。从 2009 年到 2012 年 Greenplum 都是阿里巴巴 B2B 最重要的数据计算中心，它替换掉了之前的 Oracle RAC，有非常多的优点。

- ❑ Greenplum 的性能在数据量为 TB 级别时表现非常优秀，单机性能相比 Hadoop 要快好几倍。
- ❑ Greenplum 是基于 PostgreSQL 的一个完善的数据库，在功能和语法上都要比 Hadoop 上的 SQL 引擎 Hive 好用很多，对于普通用户来说更加容易上手。
- ❑ Greenplum 有着完善的工具，相比 Hive，整个体系都比较完善，不需要像 Hive 一样花太多的时间和精力进行改造，非常适合作为一些大型的数据仓库解决方案。
- ❑ Greenplum 能够方便地与 Hadoop 进行结合，可直接把数据写在 Hadoop 上，还可以直接在数据库上写 MapReduce 任务，并且配置简单。

从 2010 年毕业加入阿里巴巴 B2B 的数据仓库起，我就开始接触 Greenplum 数据库，并有幸维护了一年多的 Greenplum 数据库，积累了很多数据库的相关知识。Greenplum 在国内的应用相对比较少，尤其是网上资料相当匮乏。在使用 Greenplum 的过程中，阿里巴巴遇到了很多困难，也积累了很多宝贵经验。

由于学习资料的匮乏，我和何勇有了将阿里巴巴使用 Greenplum 的一些经验技巧汇聚成书的想法，这样既总结和沉淀了自身知识，同时也可以给国内使用 Greenplum 的同行们提供一点帮助。

本书组织结构

本书从实战角度出发，结合了大量实践案例（附有详细的代码），由浅入深介绍了

Greenplum。本书由 15 章组成，主要分为 3 篇。

上篇（第 1 ~ 3 章）——基础篇

基础篇目的是帮助读者快速了解 Greenplum，从实战的角度介绍一些入门必备的基础知识。从如何安装部署 Greenplum 开始，一步步引导读者搭建自己的 Greenplum 数据库，然后介绍基本的语法及相关操作。本篇最后通过分析两个具体的数据仓库 ETL 的案例，加强读者对 Greenplum 功能特性的了解，提高实践能力。

中篇（第 4 ~ 7 章）——进阶篇

进阶篇重点介绍 Greenplum 的数据字典、执行计划、架构以及一些高级特性。

数据字典是 Greenplum 对元数据信息的组织方式，执行计划是数据库执行 SQL 的灵魂，高级特性则是 Greenplum 的优势所在。本篇结合了大量案例对以上内容进行了深入分析。通过对这些内容的学习，可以深入理解数据库的工作原理，是进阶的必经之路，可以让读者游刃有余地应对各种 Greenplum 的日常操作。

下篇（第 8 ~ 15 章）——管理篇

管理篇主要介绍一些与数据库管理员相关的知识，包括线上部署、性能优化、权限控制、监控、容灾 / 扩容方案、常用脚本以及常见问题等。这些更偏向于后台管理，是 DBA 必修的课程。

本书面向的读者

阅读本书需要读者对关系型数据库有基本了解，最好也了解一些 Linux 的基本操作。本书面向的读者主要有：

- ❑ Greenplum 数据库管理员
- ❑ 使用 Greenplum 的 ETL 开发工程师
- ❑ 数据库爱好者、分布式系统爱好者
- ❑ 数据分析师、商业智能分析师

如何阅读本书

对于刚接触 Greenplum 的读者来说，建议从第 1 章读起，书中有很多结合实践的例子，帮助读者加深对 Greenplum 的理解。

如果对 Greenplum 已经比较熟悉，可以从第 4 章开始读起，应重点关注第 5 章讲解分布

式执行计划。了解执行计划是熟悉数据库工作原理的一种有效方法，分布式执行计划与普通执行计划还是有较大差别的。

如果只是想通过本书了解 Greenplum 数据库与普通关系型数据库的差异，可以先阅读第 7 章，了解 Greenplum 架构，然后再选取自己喜欢的章节进行阅读。

本书偏向于实战，理论方面的知识相对较少，感兴趣的读者可以结合其他介绍数据库的资料进行阅读，如“PostgreSQL8.2 官方文档”、《数据库系统概念》、《数据库系统实现》。

虽然 Greenplum 并不是开源的，但是 Greenplum 是基于 PostgreSQL 开发的，而 PostgreSQL 是开源的。还有，Greenplum 自带了很多维护脚本，大部分采用 Shell 和 Python 编写。因此，对于想深入理解 Greenplum 的读者来说，建议结合 PostgreSQL 源码及 Greenplum 安装目录下的一些脚本代码进行阅读。通过阅读这些源码，可以加深对 Greenplum 底层的了解。

本书大部分例子都是基于 Greenplum 4.1 版本进行编写的，在写作的过程，Greenplum 已经升级到 Greenplum 4.3，因此在书稿快完成时，笔者对全书进行了审核。由于 Greenplum 4.3 与 Greenplum 4.1 大部分的内容都是相同的，因此，笔者在书稿中必要的地方采用修改或备注的方式加入了一些 Greenplum 4.3 的新特性，尽量保证了本书的时效性。

勘误和支持

本书第 1、3、7、8、14 章由何勇撰写，其余章节由陈晓峰撰写，由于是第一次写作，在写作上肯定会有很多的不足，加上笔者能力有限，难免会有遗漏或错误。如果读者发现书中有遗漏或错误的地方，可以发邮件到 greenplum_book@163.com，同时欢迎大家通过邮件一起交流关于 Greenplum 的问题。

致谢

感谢姜迅、吴伟两位老大，我在阿里巴巴 B2B 数据平台工作时他们给予了我很多帮助，是他们给予了我充足的时间来学习和锻炼；感谢孔令西、孙伟光、张耀耀、江小辉、曾春秋等各位同事对我工作的支持和帮助；感谢一起维护 Greenplum 数据库的何勇、唐成、王海、任振中，他们带给我一段难忘的共同奋斗的经历。特别感谢女友李梅花在写书过程中的支持和体谅，她作为一个数据分析师，从用户的角度，在本书写作过程中给予了很多宝贵的建议。没有这些支持，我们不可能完成本书的写作。

最后，感谢杨福川、姜影在本书出版过程中提供的帮助，尤其是姜影，她在本书审阅过程中非常认真和仔细，指出了书稿中很多的问题。

陈晓峰

目 录 *Contents*

前言

上篇 基础篇

第1章 Greenplum 简介	2
1.1 Greenplum 的起源和发展历程	2
1.2 OLTP 与 OLAP	3
1.3 PostgreSQL 与 Greenplum 的关系	3
1.3.1 PostgreSQL	3
1.3.2 Greenplum	5
1.4 Greenplum 特性及应用场景	6
1.4.1 Greenplum 特性	6
1.4.2 Greenplum 应用场景	7
1.5 小结	8
第2章 Greenplum 快速入门	9
2.1 软件安装及数据库初始化	9
2.1.1 Greenplum 架构	9
2.1.2 环境搭建	11
2.1.3 Greenplum 安装	13
2.1.4 创建数据库	20
2.1.5 数据库启动与关闭	20
2.2 安装 Greenplum 的常见问题	22

2.2.1	/etc/hosts 配置错误	22
2.2.2	MASTER_DATA_DIRECTORY 设置错误	24
2.3	畅游 Greenplum	25
2.3.1	如何访问 Greenplum	25
2.3.2	数据库整体概况	27
2.3.3	基本语法介绍	28
2.3.4	常用数据类型	35
2.3.5	常用函数	37
2.3.6	分析函数	43
2.3.7	分区表	46
2.3.8	外部表	49
2.3.9	COPY 命令	51
2.4	小结	52
第 3 章	Greenplum 实战	53
3.1	历史拉链表	53
3.1.1	应用场景描述	53
3.1.2	原理及步骤	54
3.1.3	表结构	55
3.1.4	Demo 数据准备	57
3.1.5	数据加载	58
3.1.6	数据刷新	61
3.1.7	分区裁剪	64
3.1.8	数据导出	64
3.2	日志分析	65
3.2.1	应用场景描述	65
3.2.2	数据 Demo	65
3.2.3	日志分析实战	66
3.3	数据分布	68
3.3.1	数据分散情况查看	69
3.3.2	数据加载速度影响	69
3.3.3	数据查询速度影响	72

3.4 数据压缩	73
3.4.1 数据加载速度影响	73
3.4.2 数据查询速度影响	74
3.5 索引	75
3.6 小结	75

中篇 进阶篇

第4章 数据字典详解	78
4.1 oid 无处不在	78
4.2 数据库集群信息	80
4.2.1 Gp_configuration 和 gp_segment_configuration	80
4.2.2 Gp_id	82
4.2.3 Gp_configuration_history	84
4.2.4 pg_filespace_entry	84
4.2.5 集群配置信息表转化	84
4.3 常用数据字典	85
4.3.1 pg_class	85
4.3.2 pg_attribute	88
4.3.3 gp_distribution_policy	89
4.3.4 pg_statistic 和 pg_stats	90
4.4 分区表信息	90
4.4.1 如何实现分区表	91
4.4.2 pg_partition	91
4.4.3 pg_partition_rule	92
4.4.4 pg_partitions 视图及其优化	93
4.5 自定义类型以及类型转换	94
4.6 主、备节点同步的相关数据字典	95
4.7 数据字典应用示例	96
4.7.1 获取表的字段信息	96
4.7.2 获取表的分布键	96
4.7.3 获取一个视图的定义	97

4.7.4	查询 comment (备注信息)	98
4.7.5	获取数据库建表语句	99
4.7.6	查询表上的视图	103
4.7.7	查询表的数据文件创建时间	104
4.7.8	分区表总大小	106
4.7.9	如何分析数据字典变化	108
4.7.10	获取数据库锁信息	111
4.8	Gp_toolkit 介绍	112
4.9	小结	114
第 5 章	执行计划详解	115
5.1	执行计划入门	115
5.1.1	什么是执行计划	115
5.1.2	查看执行计划	116
5.2	分布式执行计划概述	116
5.2.1	架构	116
5.2.2	重分布与广播	117
5.2.3	Greenplum Master 的工作	119
5.3	Greenplum 执行计划中的术语	120
5.3.1	数据扫描方式	120
5.3.2	分布式执行	121
5.3.3	两种聚合方式	122
5.3.4	关联	123
5.3.5	SQL 消耗	126
5.3.6	其他术语	126
5.4	数据库统计信息收集	128
5.4.1	Analyze 分析	128
5.4.2	固定执行计划	129
5.5	控制执行计划的参数介绍	130
5.6	规划器开销的计算方法	131
5.7	各种执行计划原理分析	133
5.7.1	详解关联的广播与重分布	133

5.7.2	HashAggregate 与 GroupAggregate	137
5.7.3	Nestloop Join、Hash Join 与 Merge Join	141
5.7.4	分析函数：开窗函数和 grouping sets	142
5.8	案例	144
5.8.1	关联键强制类型转换，导致重分布	144
5.8.2	统计信息过期	145
5.8.3	执行计划出错	145
5.8.4	分布键选择不恰当	147
5.8.5	计算 distinct	148
5.8.6	union 与 union all	150
5.8.7	子查询 not in	152
5.8.8	聚合函数太多导致内存不足	154
5.9	小结	155
第 6 章 Greenplum 高级应用		156
6.1	Appendonly 表与压缩表	157
6.1.1	应用场景及语法介绍	157
6.1.2	压缩表的性能差异	157
6.1.3	Appendonly 表特性	158
6.1.4	相关数据字典	164
6.2	列存储	165
6.2.1	应用场景	165
6.2.2	数据文件存储特性	166
6.2.3	如何使用列存储	166
6.2.4	性能比较	166
6.3	外部表高级应用	168
6.3.1	外部表实现原理	168
6.3.2	可写外部表	171
6.3.3	HDFS 外部表	173
6.3.4	可执行外部表	177
6.4	自定义函数——各个编程接口	179
6.4.1	pl/pgsql	180

6.4.2 C 语言接口	182
6.4.3 plpython	185
6.5 Greenplum MapReduce	187
6.6 小结	193
第 7 章 Greenplum 架构介绍	195
7.1 并行和分布式计算	195
7.2 并行数据库	197
7.3 Greenplum 架构分析	198
7.4 冗余与故障切换	199
7.5 数据分布及负载均衡	200
7.6 跨库关联	202
7.7 分布式事务	203
7.8 其他大数据分析方案	205
7.9 小结	208

下篇 管 理 篇

第 8 章 Greenplum 线上环境部署	210
8.1 服务器硬件选型	210
8.1.1 CPU	211
8.1.2 内存	211
8.1.3 磁盘及硬盘接口	211
8.1.4 网络	213
8.2 服务器系统参数调整	213
8.2.1 Solaris 参数修改	214
8.2.2 Linux 参数修改	216
8.2.3 系统参数及性能验证	217
8.3 计算节点分配技巧	221
8.4 数据库参数介绍	221
8.5 数据库集群基准测试	225
8.6 小结	227

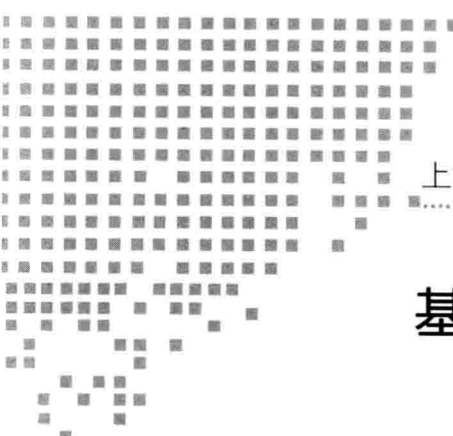
第 9 章 数据库管理	228
9.1 用户及权限管理	228
9.1.1 Greenplum 数据库逻辑结构	228
9.1.2 Grant 语法	229
9.2 登录权限控制	231
9.3 资源队列及并发控制	232
9.4 Greenplum 锁机制	236
9.5 数据目录结构	238
9.6 数据文件存储分布	240
9.7 表空间管理	241
9.8 小结	244
第 10 章 数据库监控及调优	245
10.1 Linux 监控工具介绍	245
10.1.1 监控磁盘	245
10.1.2 监控网络	246
10.1.3 监控 CPU	247
10.1.4 监控内存	247
10.2 安装 Performance Monitor	248
10.3 监控 Segment 是否正常	252
10.4 VACUUM 系统表	253
10.5 数据倾斜排查	255
10.6 查看子节点的 SQL 运行状态	258
10.7 自动加分区	261
10.8 自动赋权	266
10.9 清理过期数据	266
10.10 小结	267
第 11 章 解读 Greenplum 维护脚本	268
11.1 添加 Greenplum Contrib 模块	268
11.2 启动和关闭脚本 gpstart 和 gpstop	270
11.3 初始化系统脚本 gpinitssystem	272

11.4 集群操作脚本 gpssh 和 gpscp	274
11.5 数据库状态检查脚本 gpstate	275
11.6 数据库升级脚本 gpmmigrate	276
11.7 参数修改脚本 gpconfig	281
11.8 数据库一致性检查脚本 gpcheckcat	282
11.9 小结	284
第 12 章 备份及恢复策略	286
12.1 Greenplum 3.x	286
12.2 Greenplum 4.x	287
12.3 gp_dump 和 pg_dump	290
12.4 Greenplum Master 备份策略	294
12.4.1 增加 Standby Master	295
12.4.2 重新同步 Standby Master	296
12.4.3 启用 Standby Master	296
12.5 小结	297
第 13 章 数据库扩容	299
13.1 迁移计算节点	299
13.1.1 两种备份方案	300
13.1.2 数据迁移实战	301
13.2 增加计算节点	306
13.3 小结	311
第 14 章 基于 Greenplum 的海量数据实时分析服务平台	312
14.1 需求概述	312
14.2 典型方案	313
14.2.1 NoSQL	313
14.2.2 分布式数据库 / 集群	314
14.2.3 分表分库	315
14.2.4 方案优劣分析	316
14.3 基于 Greenplum 的混合架构	316

14.3.1 架构分析	317
14.3.2 实施要点	317
14.4 小结	318

第 15 章 使用 Greenplum 的常见报错及小技巧.....319

15.1 分析常见报错.....	319
15.1.1 找不到类型 705 对应的操作符	319
15.1.2 SQL 占用的资源超过了资源队列限制	321
15.1.3 自定义函数不能在 Segment 上执行	321
15.1.4 子查询没有加别名	322
15.1.5 字段名有歧义	322
15.1.6 字段重名	323
15.1.7 gpfdist 错误：无法读取文件	323
15.1.8 事务被中止	324
15.1.9 网络异常错误	324
15.1.10 无法删除表	324
15.1.11 内存不足	325
15.1.12 文件名在 pg_class 中已存在	325
15.1.13 不能对分布键执行 Update	325
15.1.14 网络错误	326
15.1.15 无法找到数据文件	326
15.2 常见问题及解决办法	326
15.3 常用的一些小技巧	329
15.3.1 显示 SQL 执行的时间	330
15.3.2 获取某个 schema 下所有的表或视图	330
15.3.3 查找分区最多的表	330
15.3.4 连接 Segment 节点	331
15.3.5 psql 默认密码登录	331
15.3.6 查看数据库启动时间	331
15.3.7 查看在 psql 中 \d 到底查询了哪些数据字典	331
15.4 小结	332



上篇 *Part 1*

基础篇

- 第1章 Greenplum 简介
 - 第2章 Greenplum 快速入门
 - 第3章 Greenplum 实战
-



本章先介绍 Greenplum 的产生背景、特性及应用场景、与 PostgreSQL 关系，以及发展历程。

1.1 Greenplum 的起源和发展历程

短短十多年，互联网在中国经历了从门户网站、搜索、即时通信、游戏娱乐、垂直细分……到电子商务、Web 2.0，再到社会化网络、移动互联网的一系列进化和变革。无论是互联网还是移动互联网，都是由海量的数据构成。对海量数据分析的需求开始突破传统边界，不再局限于电信、移动、金融、保险、制造等传统企业，涌现出大批将海量、庞杂的数据转化为知识，提供业务经营决策支持的企业。针对数据密集型计算中的海量数据处理这一问题，研究者开始考虑如何利用大规模集群系统所具有的可伸缩性和容错性的优势，实现高效的数据管理功能。比较典型的解决方案有 Teradata、Greenplum、Hadoop Hive、Oracle Exadata、IBM Netteza 等。

Greenplum 是一家总部位于美国加利福尼亚州，为全球大型企业用户提供新型企业级数据仓库（EDW）、企业级数据云（EDC）和商务智能（BI）提供解决方案和咨询服务的公司。选择 Greenplum 的产品的国际大客户有：纳斯达克、纽约证券交易所、Skype、FOX、T-Mobile 等，在中国，中信实业银行、东方航空公司、阿里巴巴、华泰保险、中国远洋（Cosco）、李宁公司等大型企业用户也选择了 Greenplum 的产品。Greenplum 的发展历程简要如下。

- ❑ 2003：Greenplum 由 Scott Yara 和 Luke Lonergan 成立。
- ❑ 2005：Greenplum 数据库第一个版本发布。