

# EMC GREENPLUM HD ENTERPRISE EDITION

## 为企业进一步完善 Hadoop

### 重点

EMC Greenplum HD Enterprise Edition 为您提供：

- 一个完整的企业级解决方案
- 与 Apache Hadoop 百分之百的兼容性
- 消除了 HDFS 的常见问题
- 通过网络文件系统 (NFS) 直接访问
- 通过 JobTracker 和 NameNode 增强功能实现的高可用性
- 比基于 Apache 的版本快二至五倍的性能
- 一个用于多站点部署的高级管理工具
- 可靠的数据保护功能，包括快照和群集间镜像
- 一个包含丰富的企业级商情工具的网络
- 方便地与 EMC Greenplum MPP 数据库集成

最近的技术趋势和 Internet 的发展，催生了数量庞大而复杂的数据。结果是，那些希望获得竞争优势的企业必须找到有效的方法，分析从新来源获得的非结构化和机器生成的数据。通过采用 Hadoop 来存储和分析数量巨大的电话呼叫数据、Twitter 信息、电子邮件、图片和视频，企业将可以获得新的真知灼见，以推动业务发展和利润增长。

Apache Software Foundation 和开源社区最近的发展将 Hadoop 提升到一个新位置 — 企业现在将其看作是一个对现有基础架构进行补充和完善的关键数据处理平台。现在，Hadoop 是任何大数据分析计划中不可或缺的一部分。

EMC® Greenplum® HD 产品系统让您能够充分利用大数据分析的优势，而又可以避免从头开始构建项目涉及的开销和复杂性。Greenplum HD 有 Community 和 Enterprise 版以及 Greenplum HD Data Computing Appliance 模块三种形式，可提供一个完整的平台。

经验丰富的 Hadoop 用户往往发现，易用性、性能和可管理性方面存在的问题会大大延缓他们的内部开发流程，同时会长时间占用宝贵的资源。为解决 Hadoop 的这些核心问题，EMC Greenplum 将提供一种新的版本，最终满足该平台高级企业级用户的需求。EMC Greenplum HD Enterprise Edition 提供一个完整的平台，包括安装、培训、全球支持，以及超越 Apache Hadoop 版本简单组包的附加价值。有了 Greenplum HD，Hadoop 变得更快、更可靠且更易于使用了。

## GREENPLUM HD ENTERPRISE EDITION 组件

### 应用程序可移植性方面的创新

HD Enterprise EDITION 这一实现与 Apache Hadoop 堆栈有百分之百的接口兼容性。通过维护 Hadoop 接口兼容性，HD Enterprise Edition 提供了无缝的应用程序可移植性，同时还提供大型企业需要的高级功能：

### MAPR 文件系统

MapR 文件系统已替换 HDFS 以提高 MapReduce 性能并提供多维可扩展性。无锁存储服务支持任何文件上的多个并发读写操作。

### MAPREDUCE

MapReduce 是一个 Hadoop 框架，可用于方便地编写应用程序，从而以可靠且容错的方式并行处理大量非结构化和结构化数据。该框架具备硬件故障适应能力，可相对于用户应用程序透明地处理这些故障。

## HIVE

Hive 是 Hadoop 的一个数据仓库系统，允许轻松地执行数据摘要、即兴查询和分析 Hadoop 兼容文件系统中存储的大型数据集。SQL 风格的界面为用户提供基于行的存储功能，此功能与压缩功能一起，提高了存储数据的压缩比率。

## PIG

Pig 是一种编程语言，用于借助 Hadoop MapReduce 平台处理大型半结构化数据集。它可以为开发人员提供一种替代 Java 的编程语言，从而让他们更轻松地编写 MapReduce 作业。

## HBASE

HBase 是一个有版本控制、以列为导向的分布式存储平台，可为用户应用程序提供对大数据的随机实时读 / 写访问。

## ZOOKEEPER

Zookeeper 是一个高可用性的系统，用于协调分布式进程。分布式应用程序使用 ZooKeeper 来存储和传递对关键配置信息的更新。

## MAPR HEATMAP

MapR Heatmap 可提供深入见解、完整访问和强大的工具，以帮助您轻松了解您的群集。图形化和程序化界面可扩展以适应最大的群集。

## GREENPLUM HD ENTERPRISE EDITION 特性

### 为企业提供一个功能强大的数据协同处理解决方案

Greenplum Database 和 Greenplum HD 是两项互为补充的技术，它们联合起来，为分析结构化、半结构化和非结构化数据提供一个功能强大的解决方案。结合使用二者，分析师可利用 Greenplum SQL 和高级分析函数的强大功能，对 Hadoop 和 / 或 Greenplum Database 中的数据进行复杂、高速、交互式的分析。使用外部表可实现快速、轻松的并行数据共享，从而简化数据访问和移动。综合解决方案提供了业内唯一一个完整的大数据分析平台。

### 更易于使用

为让 Hadoop 有效地用于更大的用户组和工作负载，它必须在较大规模下仍能够轻松地使用、调配、操作和管理。EMC Greenplum 已投资开发了一些关键的技术突破，让客户只需要很少的员工即可更轻松地将数据移入 / 移出群集、调配群集资源，甚至管理非常大的 Hadoop 群集。

- Greenplum HD EE 允许通过 NFS 协议实现实时读 / 写数据流。有了 NFS 访问权限，任何远程客户端都可以直接装载群集。应用程序服务器可将它们的日志文件和其他数据直接写入到群集中，而不必先将其写入直连或网络连接存储中。
- 您可以使用图形化文件浏览器访问并操作群集数据。您可以直接浏览文件，通过鼠标点击自动打开关联的应用程序，或者将文件和目录拖放到 / 拖放出群集。
- 管理界面允许用户以图形方式深入了解按群集拓扑组织的节点运行状况、服务状态和资源利用率，从而提供对群集资源和活动的全面了解。只需扫一眼，便可看到整个群集中数千个节点的运行状况。

## 更加可靠

对于生产数据，企业对可靠性、可用性和存储功能有极高的期望。通过重新设计 Hadoop 以提供高可用性，HD Enterprise Edition 可防止因单点故障而带来的常见 Hadoop 难题。

- 分布式高可用性内置于整个系统中。Greenplum HD EE 体系结构提供 JobTracker 高可用性和分布式 NameNode 高可用性，以防止作业丢失、令人扫兴的重启操作和麻烦的故障切换事件。
- 远程镜像提供了在远程站点保留群集数据同步副本的能力，这样在发生灾难时业务运营不会中断。
- 快照可针对应用程序错误或误删除为数据提供保护，同时还使您能够恢复到一个特定时间点。恢复操作非常轻松：只需将目录或文件从一个快照目录复制到当前目录即可。

## 更快

在 HD Enterprise Edition 的整个开发期间，该产品的体系结构都以提高 Hadoop 的性能为目标。从包含单个节点到数千个节点，任何规模的群集都将体验到显著超过基于 Apache 的版本的 I/O 和性能优势。

- Greenplum HD EE 提高了 MapReduce 性能，并提供多维可扩展性。数据将直接写入到数据块设备中，而不用通过 HDFS 或 Linux 文件系统层。
- 文件系统是以 C/C++ 语言编写的。除了效率和性能更高以外，还避免了 HDFS 存在的与 Java 相关的常见问题，如垃圾回收问题。
- 自动化和透明的客户端压缩可降低开销和空间占用。

## 安装和培训服务

EMC 针对 Greenplum HD 提供一个包括安装和培训服务的实施包，以确保系统按预期正常运行。无论您的环境是简单还是复杂，专家级规划、设计和实施服务都可帮助您快速实现大数据分析所带来的价值。

通过专业化的内容和课程设置，Greenplum Hadoop 培训计划可为您企业中所有阶层的人员提供帮助。不管您是想要了解 Hadoop 价值并使用它的高级管理人员，还是具有 Hadoop 开发经验的 MapReduce 编程人员，EMC Greenplum 都可提供一个适合于您的培训计划。

## EMC GREENPLUM HD 合作伙伴体系

业界主流的技术公司都已与 EMC 建立了合作伙伴关系，以形成一个庞大而且活力四射的合作伙伴体系，以基于 EMC 的 Hadoop 产品提供更多商情、数据传输和其他技术能力。这一广泛的支持是 EMC 为 Hadoop 带来价值的见证。那些一向信任 EMC 的技术公司和企业，现在还可以信任 EMC 的开源数据分析工具。

## 企业级商情和分析建模工具支持

Greenplum HD Enterprise Edition 合作伙伴体系包括以下领域的行业开拓者：提取、转换和加载 (ETL)；商情 (BI)；以及分析工具包。这些合作伙伴为您成功进行大数据分析提供关键组件。

## 利用 EMC 全球服务让 EMC GREENPLUM 发挥最大效益

EMC 针对 EMC Greenplum Data Computing Appliance 系统硬件和软件提供全套服务，以确保系统在您的环境中按预期正常运行，同时最大限度地降低业务和预算风险。无论简单还是复杂，专家级的规划、设计和实施服务都可以帮助您让硬件和软件在您的环境中快速实现其价值。完成实施后，EMC 的数据迁移服务可以帮助您规划、设计并安全地将重要数据跨任意远的距离迁移到新的系统。EMC 还将帮助您将新系统集成到您的信息体系结构以及商情和分析应用程序中（如 SAS、MicroStrategy、Business Objects 和 Tableau），并在完成集成后帮助您管理新环境。

通过利用 EMC 全面的数据仓储和商情部署最佳做法，并在我们行之有效的方法体系指导下，我们经过全面培训的专业服务人员和项目管理团队将帮助您更快地看到业务成效，而又不会让您感到资源紧张。

## EMC GREENPLUM 的数据计算产品

EMC 的数据计算产品部正在通过推出突破性的产品来推动数据仓储和分析的发展，这些产品有 EMC Greenplum HD、EMC Greenplum Data Computing Appliance、EMC Greenplum Database，还有 EMC Greenplum Chorus — 业内第一个企业数据云平台。该部门的产品实现了开放式系统、云计算、虚拟化和社会协作的强大威力，让全球各地的企业能够更深入地了解他们的数据，实现前所未有的更大价值。

### 联系我们

如欲详细了解 EMC Greenplum 产品、服务和解决方案如何帮助您实现大数据带来的商机，请访问我们的网站：  
[www.greenplum.com](http://www.greenplum.com)。

EMC<sup>2</sup>、EMC、Greenplum 和 EMC 徽标是 EMC Corporation 在美国和其他国家 / 地区的注册商标或商标。此处使用的所有其他商标均为其各自所有者的资产。© 版权所有 2011 EMC Corporation。保留所有权利。中国印刷。10/11 产品介绍 H8892

易安信电脑系统（中国）有限公司  
中国北京朝阳区霄云路 38 号 现代汽车大厦 15 层  
邮编：100027 电话：(8610) 8438 6000  
传真：(8610) 8453 8174  
售前服务热线：400 650 6006  
网站：[www.EMC2.com.cn](http://www.EMC2.com.cn)

**EMC<sup>2</sup>**