

# Greenplum 介绍

Greenplum 是全球领先的开源大数据平台，是能够提供包含实时处理、弹性扩容、混合负载、云原生和集成数据分析等强大功能的大数据引擎。

著名分析机构 Gartner 2019 年报告中，在经典数据分析领域 Greenplum 全球排名第三，实时分析领域全球排名并列第四。Greenplum 是两个领域中排名前十的产品中的唯一一款开源产品。

Greenplum 基于 MPP (大规模并行处理) 架构构建，具有良好的弹性和线性扩展能力，并内置并行存储、并行通讯、并行计算和优化技术。同时，Greenplum 还兼容 SQL 标准，具备强大、高效、安全的 PB 级结构化、半结构化和非结构化数据存储、处理和实时分析能力，可部署于企业裸机、容器、私有云和公有云中。值得一提的是，作为 OLAP 型的大数据平台，Greenplum 同时还能够支持涵盖 OLTP 型业务的混合负载，从而帮助客户真正打通业务-数据-洞见-业务的闭环。

目前，Greenplum 已经为国内外各行各业客户所广泛使用，支撑着全球各大行业的核心生产系统，其涉及领域涵盖金融、保险、证券、通信、航空、物流、零售、媒体、政府、医疗、制造、能源等。其中，国际客户包括摩根斯坦利、摩根大通、美国国家税务局、美联储、三星、戴尔、福特、爱立信等，国内客户包括深交所、建设银行、民生银行、光大银行、浦发银行、航旅纵横、中国移动、华为等。自 2015 年开源以来，更是吸引了包括阿里云、百度云、中移动、旷世、去哪儿网、易观、腾云科技、饿了么、金风科技在内大量开源用户。

## Greenplum 大数据平台的优势

- **一次打包到处运行的平台**：部署灵活，不受限于硬件环境和平台，无论裸机、私有云、公有云均可部署。硬件环境的普适性，提供了极大的灵活性，解放了硬件平台的制约和绑定，从而允许客户灵活选择最适合的方案，降低未来的迁移代价，而开发、运维人员无需要学习新的数据库处理技术，人力成本也能够大大降低。

- **处理和分析各种数据源的数据的平台**：支持各种数据源，包括 Kafka、Hadoop、HIVE、HBase、S3、Gemfire、各种数据库和文件等，不需要移动数据，避免了数据加载的复杂性，和其带来的数据不一致的问题。
- **支持各种数据格式的平台**：不管是结构化、半结构化 ( XML、JSON、KV ) 还是非结构化，譬如文本数据、GIS 数据、图数据等。
- **具有强大内核的平台**：Greenplum 具有强大的内核技术，包括数据水平分布、并行查询执行、专业优化器、线性扩展能力、多态存储、资源管理、高可用、高速数据加载等。
- **具备强大灵活性和可扩展性的平台**：支持扩展 ( Extension )、自定义类型和函数、PXF 和外部表技术。可以使用多种语言实现用户自定义函数和聚集，包括 PL/Python、PL/R、PL/Java、PL/Perl、PL/PGSQL 和 C 等。
- **支持标准的平台**：支持 SQL、JDBC 和 ODBC 等行业标准。经过半个多世纪的发展，SQL 成为了数据平台的万向头，向上可以连接各种 BI 工具、可视化工具和数据分析工具，向下可以连接各种 ETL 工具、各种数据源和各种格式的数据等。
- **集成数据分析平台**：支持商业智能 ( BI )、文本、GIS、图、图像等。流式支持也在开发中。通过 Pivotal 开源的 Apache 顶级项目 MADlib，Greenplum 可以在数据内部运行 50 多种数据分析和机器学习算法。MADlib 提供 SQL 接口进行数据分析，大大降低了数据分析的门槛；MADlib 内建于数据库内，使用 MPP 的优势，提高了分析的效率；MADlib 可以在全量数据而不是抽样数据上进行分析，提高了精度。
- **开放源代码且持续大力投入的平台**：2017 年 Pivotal 在 github 的开源贡献列表中全球排名第四左右。采用开源方案，不担心后门问题，不担心被锁定。开源还可以构建更好的生态。
- **采用敏捷软件开发方法开发的平台**：Greenplum 采用敏捷方法开发，实现了快速迭代、持续发布和质量内建。2017 年 Greenplum 发布了 10 个版本，以前发布一个版本需要 1 个月左右，现在只需要十几个小时。
- **具备企业级稳定性的平台**：Greenplum 经过十多年发展，有大量活跃客户，大量数百节点集群为全球 2000 强企业生产系统提供服务，稳定性非常高。
- **具备成熟生态系统的平台**：Greenplum 生态非常完善，有大量的合作伙伴。

## 发展历程

Greenplum 公司成立于 2003 年，2008 年发布 Greenplum 数据库产品。2010 年 Greenplum 被高端存储领域巨头 EMC 收购，同年 Greenplum 借助 EMC 中国卓越研发集团进入中国，并创建 Greenplum 中国研发团队。2014 年 Greenplum 从 EMC 独立出来创立 Pivotal，并获得通用电气、福特汽车和微软的投资。2018 年 4 月 Pivotal 在纽约证券交易所上市，目前市值 60 多亿美元。

## 开源社区和生态

Greenplum 基于世界上最先进的开源数据库 PostgreSQL。

2015 年 10 月 Pivotal 正式把投资超过 10 年的 Greenplum 开源，采用 Apache 2 协议，成为世界上第一款成熟的开源 MPP 数据库。开源之后，Greenplum 社区发展活跃，短短 2 年左右，在全球有上百来自美国、中国、俄罗斯、日本、英国、德国、芬兰、瑞士等国家的贡献者，获得社会各界的广泛关注。其中 Greenplum 中文社区尤为活跃，目前约有半数的贡献来自中国开发者，社区贡献者包括阿里云、中移动等大公司，也有诸多中小公司和数据库爱好者。

开源之后，Greenplum 把敏捷软件开发方法学引入到分布式数据库的开发中，通过使用站立会议、回顾会议、结对编程、持续集成、测试驱动、单周迭代等敏捷方法建立了高效的快速反馈系统，大大提高了产品的质量和客户的满意度。Greenplum 5.0 是开源之后发布的第一个稳定版本，大约保持 1 个半月一个版本的发布速度。Greenplum 主干分支 ( master ) 开发非常活跃，众多社区期待的特性稳健推进中，包括内核升级、新课题，处于业界领先地位。

Greenplum 社区邀请 Pivotal、阿里云、百度等社区人士共同定期举办各种技术研讨会、技术沙龙、101 培训等活动，获得良好反馈。国内多所知名大学开设了 Greenplum 相关数据库课程，并在学术领域取得了突出成就。

Greenplum 生态完善。得益于对 SQL 标准的出色支持，Greenplum 可以与上下游的各种产品实现集成，包括 PgAdmin、Informatica、Cognos、SAS、Talend、Qlik、Tableau、Anaconda、Microstrategy、Boundless、Zattset、Datometry 等，涵盖 ETL、商业智能、高级分析、可视化、集成分析、GIS 数据处理、迁移、安全和管理等各个领域。

更多信息请访问 [greenplum.cn](http://greenplum.cn)。