



Greenplum Datasheet

Greenplum Database 3.3

Greenplum Database is a software solution built to support the next generation of data warehousing and large-scale analytics processing. Supporting SQL and MapReduce parallel processing, Greenplum Database offers industry-leading performance at a low cost for companies managing Terabytes to Petabytes of data.

Greenplum Database utilizes a shared-nothing MPP (massively parallel processing) architecture optimized for business intelligence and analytical processing. It allows a cluster of servers to operate as a single database supercomputer -- automatically partitioning data and parallelizing queries -- to achieve performance 10s or 100s of times of faster than traditional databases.

Greenplum Database 3.3 provides:

Extreme scale on commodity hardware

- From low TBs to the largest multi-Petabyte data warehouses -- scale is no longer a barrier
- Flexible hardware vendor choices including HP, Dell, Sun, EMC and IBM

Elastic expansion and sandboxing

- Add servers while online for more storage capacity and performance (- no costly appliance upgrades)
- Supports agile data warehousing practices -- easily create private sandboxes for business analysts

Massively parallel analytic processing

- Unified parallel engine supports SQL and MapReduce processing across 100s or 1000s of CPU cores
- Comprehensive SQL support (SQL-92, SQL-99, SQL-2003 OLAP extensions)

Greenplum Database moves processing power as close as possible to the data, so processing always occurs in parallel, delivering unmatched query and load performance. Greenplum Database is simply the industry's fastest and most affordable high-end data warehousing solution. With Greenplum Database, users have the power to answer complex questions, running analyses that used to take days with traditional solutions, in literally just seconds.

SYSTEM HARDWARE COMPATIBILITY

Operating System

Greenplum Database 3.3 supported for production on SUSE Linux Enterprise Server 10.2 (64-bit), Red Hat Enterprise Linux 5.x (64-bit), CentOS Linux 5.x (64-bit), Sun Solaris 10U5+ (64-bit). Server hardware includes HP, Dell, Sun and IBM.

Greenplum Database 3.3 supported for non-production (development and evaluation) on Mac OSX 10.5, Red Hat Enterprise Linux 5.2 or higher (32-bit), CentOS Linux 5.2 or higher (32-bit).

Processor

Optimized for x86 64-bit. (Supports x86 32-bit for non-production usage)

Memory

Min: 1GB per CPU Core
Recommended: 4GB per CPU Core

Storage

Direct Attached [SATA / SAS]

Storage Area Network

Network

Gigabit Ethernet or
10 Gigabit Ethernet

Greenplum Database Features

Shared-nothing massively parallel processing architecture

Responsible for the distribution of data and parallel execution of Shared-nothing MPP (massively parallel processing) architecture Greenplum Database's architecture provides automatic parallelization of data and queries – all data is automatically partitioned across all nodes of the system, and queries are planned and executed using all nodes working together in a highly coordinated fashion.

Multi-level fault tolerance

Greenplum Database utilizes multiple levels of fault-tolerance and redundancy that allow it to automatically continue operation in the face of hardware or software failures.

Online system expansion

Add servers to increase storage capacity, processing performance and loading performance. The database can remain online and fully available while the expansion process takes place in the background. Performance and capacity increases linearly as servers are added.

Workload management

Provides administrative control over system resources and their allocation to queries. Allows users to be assigned to resource queues that manages the inflow of work to the database. Also allows priority adjustment of running queries.

Petabyte-scale loading

High-performance loading utilizing MPP Scatter/Gather Streaming technology. Loading speeds scale with each additional node to greater than 4TB/hour.

Anywhere data access

Allows queries to be executed from the database against external data sources, returning data in parallel, regardless of their location, format, or storage medium.

In-database compression

Utilizes industry-leading compression technology to increase performance and dramatically reduce the space required to store data. Customers can expect to see a 3-10x disk space reduction with a corresponding increase in effective I/O performance.

Multi-level partitioning

Allows flexible partitioning of tables based on date, range or value. Partitioning is specified using DDL and allows an arbitrary number of levels. The query optimizer will automatically prune unneeded partitions from the query plan.

Centralized administration

Provides cluster-wide management tools and utilities that allow administrators to manage the database as if it was a single system.

SQL standard

Comprehensive SQL-92 and SQL-99 support with SQL 2003 OLAP extensions. All queries are parallelized and executed across the entire system.

MapReduce support

MapReduce has been proven as a technique for high-scale data analysis by Internet leaders such as Google and Yahoo. Greenplum natively runs MapReduce programs within its parallel engine.

Programmable parallel analytics

Offers a new level of parallel analysis capabilities for mathematicians and statisticians, with support for R, linear algebra and machine learning primitives.

Greenplum Performance Monitor

View the performance of your Greenplum Database system including system metrics and query details. The dashboard view allows you to monitor the system utilization during query runs. Drill down into a query's detail and plan to understand its performance.

pgAdmin3 for GPDB

pgAdmin III is the most popular and feature rich Open Source administration and development platform for PostgreSQL. Greenplum Database 3.3 ships with an enhanced version of pgAdmin III that has been extended to work with Greenplum Database and provides full support for Greenplum-specific capabilities.

Support For Indexes

Greenplum supports a range of index types including B-Tree and Bitmap.



Industry standard interfaces

Supports standard database interfaces (SQL, ODBC, JDBC, DBI) and is interoperable with market-leading business intelligence (BI) and extract/transform/load (ETL) tools.ns.

About Greenplum

Greenplum is the pioneer of Enterprise Data Cloud™ solutions that bring the power of self-service to large-scale data warehousing and analytics, providing customers with flexible access to all their data for business intelligence and analytics. Greenplum offers industry-leading performance at a low cost for companies managing terabytes to petabytes of data. Data-driven businesses around the world, including NASDAQ OMX, NYSE Euronext, Reliance Communications, Skype and Fox Interactive Media/MySpace, have adopted the Greenplum Database to support their mission-critical business functions. For more information visit greenplum.com.

Contact

Greenplum

1900 S. Norfolk Drive,
Suite 224

San Mateo, CA 94403

United States

+1 650 286 8012 (ph)

+1 650 286 8010 (fax)