

EMC GREENPLUM HD

Delivering Enterprise-Ready Apache Hadoop

ESSENTIALS

EMC Greenplum HD offers you:

- A complete, enterprise-grade solution
- Up to two to five times faster performance than other packaged offerings
- Advanced management for multi-site deployments
- Robust data protection features, including snapshots and inter-cluster mirroring
- A rich network of enterprise business intelligence tools
- Simple integration with the EMC Greenplum MPP database

Recent computing and business trends have triggered an explosion in the amount of unstructured data companies generate each day. That's why big data analytics can give you a competitive advantage. By extracting the knowledge wrapped within unstructured machine-generated data, your enterprise can make better decisions that drive revenue and reduce costs.

Hadoop has rapidly emerged as the preferred solution for big data analytics across unstructured data. But the fast-changing Hadoop ecosystem can present challenges to any company that wants to standardize on core functionality and build repeatable processes. It is often up to you to determine which packages to use, provide version control, test for compatibility, and deploy into production. This can lead to many time-consuming failed attempts—and, ultimately, a dead end.

The EMC Greenplum HD product family enables you to take advantage of big data analytics without the overhead and complexity of a project built from scratch. Available in Community and Enterprise editions as well as Greenplum HD Data Computing Appliance, Greenplum HD provides a complete platform, including installation, training, global support, and value-add beyond simple packaging of the Apache Hadoop distribution. Greenplum HD makes Hadoop faster, more dependable, and easier to use.

GREENPLUM HD FEATURES

APPLICATION PORTABILITY

Greenplum HD provides a 100 percent compliant Hadoop interface that eliminates the need to change applications developed against the Apache distribution. Examples include:

MapReduce

MapReduce is a Hadoop framework for easily writing applications that process large amounts of unstructured and structured data in parallel in a reliable and fault-tolerant manner. The framework is resilient to hardware failures, handling them transparently from user applications.

Hive

Hive is a data warehouse system for Hadoop that facilitates easy data summarization, ad-hoc queries, and the analysis of large datasets stored in Hadoop-compatible file systems. This SQL-like interface gives users a columnar storage capability, which, along with compression, results in an improved compression ratio for storing data.

HBase

HBase is a distributed, versioned, column-oriented storage platform that delivers random real-time read/write access to Big Data for user applications.



THE GREENPLUM STORY

Data warehousing and analytics innovator Greenplum was acquired by EMC in July 2010, becoming the foundation of the new EMC Data Computing Division. Greenplum's software and appliance products deliver the Big Data analytics platform of the future, enabling actionable insight from vast and diverse datasets.

HIGH AVAILABILITY SUPPORT

Fault-Tolerant Name Node and Task Tracker

Greenplum HD fault-tolerance capabilities protect users from the single point of failure inherent in Apache Hadoop. Automatic recovery from failures improves on the platform's availability and lowers your total cost of ownership.

INTEROPERABILITY

Integration with the Greenplum Database

Greenplum Database supports Hadoop External tables, enabling Greenplum Database users to access data residing on Hadoop Distributed File System (HDFS) without materializing the data. Administrators can read and write files in parallel from Greenplum to HDFS, enabling fast, easy data sharing. You can perform cross-platform analysis using the power of Greenplum SQL and advanced analytic functions accessing data on HDFS. The combined solution delivers the industry's only complete Big Data analytics platform.

Enterprise Business Intelligence and Analytic Modeling Tool Support

The Greenplum HD partner ecosystem includes leading industry innovators in ETL, BI, and analytic toolkits. These partners provide critical components for your success in big data analytics.

INSTALLATION AND TRAINING SERVICES

EMC delivers an implementation package of installation and training services for Greenplum HD to ensure that the system will perform as expected. Expert planning, design, and implementation services help you quickly realize the value of Big Data analytics, no matter how simple or complex your environment may be.

Greenplum Hadoop training programs are designed to empower all levels of your organization with specialized offerings and curriculum. Whether you're a senior executive trying to understand the value and use of Hadoop, or a MapReduce programmer building on the Hadoop experience, EMC Greenplum has a program for you.

GREENPLUM HD EDITIONS

The EMC Greenplum HD product family offers enterprise-ready Apache Hadoop to:

- Deliver great reliability and manageability.
- Provide an easier deployment and use paradigm.
- Enable faster processing and real-time data interaction.

Greenplum HD is available in two fully supported software editions: Community and Enterprise.

HD SOFTWARE COMMUNITY EDITION

HD Community Edition is a 100 percent open source certified and supported version of the Apache Hadoop stack that includes HDFS, MapReduce, Zookeeper, Hive, and HBase. EMC Greenplum adds fault tolerance for the Name Node and Job Tracker, both single points of failure in Hadoop.

HD SOFTWARE ENTERPRISE EDITION

HD Enterprise Edition is a 100 percent interface compatible implementation of the Apache Hadoop stack. By maintaining Hadoop interface compatibility, Enterprise Edition provides seamless application portability while delivering advanced features required by larger organizations. These include:

- Data management features such as snapshots and wide area replication.
- Simple data loading and access using a native NFS interface.
- End-to-end manageability, including simple cluster deployment, automatic failure detection and notification, multi-site management, and rolling upgrades.

Best of all, these capabilities are delivered along with a two to five times faster performance improvement over the standard packaged versions of Apache Hadoop.

Compare the features of the available Hadoop Distributions:

Hadoop Features	Apache Hadoop	Greenplum HD Community Edition	Greenplum HD Enterprise Edition
Scalability	Yes	Yes	Yes
Compatibility	100%	100%	100%
Language	Java	Java	C
Commodity Hardware	Yes	Yes	Yes
Certified Stack	No	Yes	Yes
Installation Services	No	Yes	Yes
Performance	Standard	Standard	2-5X Faster
NFS Support	No	No	Yes
Fault Tolerance	No	Yes	Yes
Rolling Upgrades	No	No	Yes
Snapshot	No	No	Yes
Wide-Area Replication	No	No	Yes

EMC GREENPLUM HD PARTNER ECOSYSTEM

Several technology companies have partnered with EMC to create a vibrant and powerful ecosystem, offering additional business intelligence, data transfer, and other technology capabilities on top of EMC's Hadoop offerings. This breadth of support is a testament to the value EMC brings to Hadoop. Technology companies and enterprises can now extend the trust they have in EMC to the open source data analytics tool.

EMC Greenplum HD partners include Concurrent, CSC, Datameer, Informatica, Jaspersoft, Karmasphere, Microstrategy, Pentaho, SAS, SnapLogic, Talend and VMware.

MAXIMIZE EMC GREENPLUM BENEFITS WITH EMC GLOBAL SERVICES

EMC global services deliver a full range of support and services for EMC Greenplum hardware and software to ensure that your system performs as expected in your environment, while minimizing risk to your business and budget. Expert planning, design, and implementation services help you quickly realize the value of your hardware and software in your environment—no matter how simple or complex. After implementation, EMC's data migration services can help you plan, design, and safely migrate your critical data over any distance to your new system. EMC will also help you integrate your new system into your information architecture and Business Intelligence and analytics applications (ex: SAS, MicroStrategy, Business Objects, Tableau, etc.), and manage your new environment when it is complete.

Extensively trained professional services personnel and project management teams, leveraging EMC's extensive data warehousing/business intelligence deployment best practices and guided by our proven methodology, accelerate the business results you need without straining the resources you have.

CONTACT US

To learn more about how EMC Greenplum products, services and solutions can help you realize Big Data opportunities visit us at www.greenplum.com.

ABOUT EMC GREENPLUM'S DATA COMPUTING PRODUCTS DIVISION EMC's new Data Computing Products Division is driving the future of data warehousing and analytics with breakthrough products such as EMC Greenplum HD Family, EMC Greenplum Data Computing Appliance Family, EMC Greenplum Database, EMC Greenplum Database Community Edition, EMC Greenplum HD and EMC Greenplum Chorus—the industry's first Enterprise Data Cloud platform. The division's products embody the power of open systems, cloud computing, virtualization, and social collaboration—enabling global organizations to gain greater insight and value from their data than ever before.

EMC², EMC, and where information lives are registered trademarks or trademarks of EMC Corporation in the United States and other countries. All other trademarks used herein are the property of their respective owners. © Copyright 2011 EMC Corporation. All rights reserved. Published in the USA. 2011-0329