

# EMC GREENPLUM DATA COMPUTING APPLIANCE

## Data Warehousing der nächsten Generation

### WESENTLICHE PUNKTE

- Speziell entwickelte, hochgradig skalierbare Data Warehousing-Hardware- und -Software, die Greenplum Database, Greenplum HD und Drittanbieteranwendungen sowie Computing-, Speicher- und Netzwerkfunktionalität in einem einzigen, einfach zu implementierenden System der Enterprise-Klasse vereint
- Optimierte für die schnelle Ausführung von Abfragen, unerreichte Datenladegeschwindigkeiten und lineare Skalierbarkeit
- Erweiterte modulare „All-in-One“-Analyselösung für das Management von strukturierten und unstrukturierten Daten, ETL- und BI-Prozessen
- Eine einzige Plattform für Data Warehousing, Data Marts, Text-Mining und statistische Berechnungen
- Besserer Einblick in und höherer Wert aus Daten dank erweiterter Analysefunktionen und einheitlichem Datenzugriff
- Hohe Verfügbarkeit, Speicher und Disaster Recovery der Enterprise-Klasse mit bewährten EMC Lösungen

### HERAUSFORDERUNGEN EINER DATENORIENTIERTEN WELT MEISTERN

Angesichts steigender IT-Kosten, explodierender Datenmengen und immer neuer Herausforderungen im Wettbewerb ist es Zeit für neue, effektivere Datenanalysesysteme. Diese Entwicklungen haben grundlegende Veränderungen in der Datenbanktechnologie und einen neuen Ansatz zur Datennutzung hervorgebracht.

In vielen Unternehmen werden noch jahrzehntealte Datenmanagement- und Analysesysteme genutzt, die für die heutigen enormen Datenmengen völlig ungeeignet sind. Die EMC® Greenplum® Data Computing Appliance (DCA) vereint eine leistungsstarke MPP-Architektur (Massively Parallel Processing) mit der branchenweit höchsten Datenlade- geschwindigkeit und dem besten Preis-Leistungs-Verhältnis – ohne die Komplexität und Einschränkungen von proprietärer Hardware. Es handelt sich um eine speziell entwickelte, hochgradig skalierbare parallele Datenanalyse-Appliance, die Datenbank-, Computing-, Speicher- und Netzwerkfunktionen in einem einzigen, einfach zu implementierenden System der Enterprise-Klasse vereint.

Die Greenplum DCA ist eine einheitliche Appliance für Big Data-Analysen – eine modulare Lösung für strukturierte und unstrukturierte Daten sowie Anwendungen von Greenplum-Partnern, beispielsweise Business Intelligence (BI) oder Extract, Transform, Load (ETL). Unternehmen können mit einem Primär-Rack beginnen, das ein Greenplum Database-Modul („Standard“ oder „High Capacity“) enthält, und die Appliance je nach Bedarf an Verarbeitungskapazität in 1/4-Rack-Schritten auf insgesamt bis zu sechs Racks erweitern. Dabei können das Greenplum Database Standard Module, Greenplum Database High Capacity Module, Greenplum HD Module und das Greenplum Data Integration Accelerator Module in beliebiger Reihenfolge und Anzahl kombiniert werden. Alle Module sind über eine leistungsfähige High-Speed-Verbindung mit geringer Latenz aneinander angebunden.

Die Greenplum DCA ermöglicht die schnelle und einfache Implementierung von Big Data-Analysen. Mit einer integrierten Appliance, die sich durch optimierte Performance, einfache Bereitstellung, erweiterte Systemüberwachung, einfacheres Management und geringeren Platzbedarf auszeichnet, erzielen Sie schneller Ergebnisse. Durch die Greenplum DCA-Module wird die Kapazitäts- und Performance-Erweiterung der Systemkomponenten Greenplum Database (Analysedatenbank) und Greenplum HD (Hadoop) erheblich vereinfacht. Diese Datenmanagement-Appliance bietet maximale Flexibilität und Skalierbarkeit für Unternehmen, die sich mit Datenmengen im Terabyte- und Petabyte-Bereich auseinandersetzen müssen.

## FUNKTIONEN DER DATA COMPUTING APPLIANCE

### EXTREME UND ZUVERLÄSSIGE PERFORMANCE MIT ELASTISCHER SKALIERBARKEIT

Das Herz der Greenplum Data Computing Appliance (DCA) ist die Greenplum Database mit einer speziell für Business Intelligence und Analyseverarbeitung entwickelten Shared-Nothing-MPP-Architektur. Das Kernprinzip der Greenplum Database besteht in der Annäherung der Verarbeitung an die Daten und Benutzer. Auf diese Weise können alle Abfragen vollständig parallel verarbeitet, alle Speicherverbindungen gleichzeitig genutzt und die Daten gemäß der Abfrageanforderungen effizient auf die Ressourcen verteilt werden. So können zahlreiche komplexe Verarbeitungsprozesse so nahe wie möglich an den Daten erfolgen, was die Verarbeitungseffizienz maximiert und für eine unübertroffene Aussagekraft sorgt.

### SCATTER/GATHER STREAMING ZUM LADEN VON DATEN

Bei der Greenplum DCA wird der Datenfluss an alle Nodes der Appliance mit der EMC Greenplum-MPP-Technologie Scatter/Gather Streaming™ (SG Streaming) gemanagt. Bei der durchgängig parallelen Übertragungsmethode fließen die Daten von einem oder mehreren Quellsystemen ohne aufeinanderfolgende Engpässe zu allen Nodes der Datenbank. So erreicht die Greenplum DCA Ladegeschwindigkeiten von mehr als 10 Terabyte pro Stunde/Rack und lädt Daten zwei- bis fünfmal schneller als andere Appliance-Lösungen.

### KONFIGURATION, MASTER-SERVER UND SEGMENTSERVER

Zur Steigerung der Performance bei Analyseabfragen kann der Greenplum DCA-Cluster durch Hinzufügen von bis zu sechs Schränken mit automatischer Datenverteilung erweitert werden. Jedes Primär-Rack enthält zwei Master-Server und vier Segmentserver. In einer Multi-Rack-Konfiguration enthalten die Erweiterungs-Racks keine Master-Server. Als Teil der Greenplum Database sind die Master-Server zuständig für die Authentifizierung, Abfrageoptimierung, den Lastausgleich zwischen den Segmentservern und das Management des Fehlertoleranzmechanismus der Daten.

### HOCHVERFÜGBARKEIT DER ENTERPRISE-KLASSE

Durch ihre Multilevel-Fehlertoleranz mit automatischer Fehlerkorrektur, die automatisches Failover und vollständige Onlinesynchronisierung sowie mehrere Ebenen der Redundanz und Integritätsprüfung umfasst, erfüllt die Greenplum DCA die Zuverlässigkeitsanforderungen der geschäftskritischsten Projekte. Die Datenverfügbarkeit stützt sich auf Hardware-RAID-Schutz auf Festplattenebene und die Spiegelung von Daten zwischen den verschiedenen Segmentservern. Durch diese Systemzuverlässigkeit werden Datenverluste bei Ausfall einer Festplatte oder eines Servers verhindert.

### SCHNELLE BEREITSTELLUNG UND ZUVERLÄSSIGE PERFORMANCE

Die Greenplum DCA ist eine speziell entwickelte Daten-Appliance auf Basis von Open-Source-Systemen. Sie vereint Datenbank-, Server- und Speicherfunktionen in einem einzigen, einfach zu implementierenden System, das innerhalb weniger Tage (nicht etwa Wochen oder Monate) bereitgestellt werden kann. Das System ist in Schritten von jeweils einem Modul auf eine Multi-Rack-Konfiguration erweiterbar. Der Aufbau und das Pre-Tuning der Appliance sorgen für eine zuverlässige Performance. Durch die Straffung der Data Warehouse- und Analyseinfrastruktur wird der administrative Overhead verringert.

### ZUVERLÄSSIGES BACKUP UND DISASTER RECOVERY

Für eine stabile und zuverlässige Data Protection in der Greenplum DCA-Analyseumgebung kommen EMC Data Domain® und EMC Symmetrix® zum Einsatz. Dank der Deduplizierungs- und Backup-Technologie von EMC Data Domain bietet die Greenplum DCA eine schnelle, zuverlässige Daten-Recovery mit Backup-Durchsatzgeschwindigkeiten von bis zu 14 TB pro Stunde. Darüber hinaus kann die Data Domain-Replikationstechnologie für große Entfernungen (WAN) auch zur Remote-Replikation von Greenplum-Datenbanken verwendet werden. Bei der SAN-Spiegelungslösung der Greenplum DCA ermöglichen EMC Symmetrix VMAX™, EMC TimeFinder®/Snap und Symmetrix Remote Data Facility (SRDF®) die erweiterte Speicher- und Datenreplikation zwischen zwei Standorten im synchronen Modus.



Die EMC Greenplum DCA bietet eine optimale Kombination aus Flexibilität, günstigem Preis und Performance und ist die richtige Lösung für Unternehmen, die funktionale, intelligente Anwendungen zur Analyse von Big Data bereitstellen und dabei keine Zeit verlieren möchten.

#### GREENPLUM DB STANDARD MODULE

- Branchenweit bestes Preis-Leistungs-Verhältnis
- Unterstützt lineare Skalierbarkeit

#### GREENPLUM DB HIGH CAPACITY MODULE

- Bietet Kapazität für mehrere Petabyte Daten ohne ein Mehr an Platzbedarf, Stromverbrauch oder Kosten
- Beste Data Warehouse Appliance bezüglich der Kosten pro Einheit

#### GREENPLUM HD MODULE

- Branchenweit erste hochleistungsfähige Hadoop Appliance für das Co-Processing von Daten
- Ermöglicht das Co-Processing strukturierter und unstrukturierter Daten

#### GREENPLUM DIA MODULE

- Hosting und rasche Integration von Partneranalyseanwendungen in die Greenplum Data Computing Appliance

#### PROAKTIVE EMC ONE-SUPPORTSTRUKTUR

Der EMC Customer Service bietet Ressourcen und Services für die schnelle, proaktive Klärung lösungsbezogener Probleme und Fragen. So wird eine hohe Business Continuity und Datenverfügbarkeit ermöglicht. Der weltweit verfügbare Wartungs- und Support-Service von EMC ist rund um die Uhr erreichbar und kann auf vielfältige Weise kontaktiert werden: über Online-Support-Tools wie Live-Chat und online gemanagte Service-Requests, Live-Support am Telefon und Vor-Ort-Support durch den branchenführenden globalen Außendienst.

Darüber hinaus ist die Data Computing Appliance mit der Funktion „Secure Remote Support“ (Dial-Home) ausgestattet, die präventiven Remote-Support rund um die Uhr ermöglicht, indem das EMC Support Center automatisch über kritische Hardware- und Softwarefehler benachrichtigt wird. Das EMC Support Center führt dann eine Remote-Fehlerdiagnose durch, um Systemausfälle zu vermeiden bzw. so kurz wie möglich zu halten. Zur schnellstmöglichen Behebung von Hardwarefehlern werden automatisch Kundentechniker geschickt.

#### DIE MODULE DER DATA COMPUTING APPLIANCE

Zur Greenplum Data Computing Appliance (DCA) gehören folgende Module:

- Greenplum Database Standard Module: Ein speziell entwickeltes, hochgradig skalierbares Appliance-Modul für die Datenanalyse, das Datenbank-, Computing-, Speicher- und Netzwerkfunktionalität in einem einfach zu implementierenden System der Enterprise-Klasse vereint. Dieses Modul ist Branchenführer hinsichtlich Preis und Performance.
- Greenplum Database High Capacity Module: Dieses Modul wurde entwickelt, um mehrere Petabyte an Daten zu bewerkstelligen, ohne dass dabei zusätzlicher Platz belegt wird oder der Stromverbrauch bzw. die Kosten steigen. Für Unternehmen, die eine ausführliche Analyse extrem großer Datenmengen benötigen oder auf der Suche nach einer Langzeitarchivierung sind, erweist sich dieses Modell als das günstigste bezüglich der Kosten pro Einheit im Bereich Data Warehouse.
- Greenplum HD Module: Das branchenweit erste hochleistungsfähige Hadoop Appliance-Modul für das Co-Processing von Daten. Durch die Kombination von Hadoop mit der Greenplum Database ermöglicht die DCA eine Mitverarbeitung (Co-Processing) strukturierter und unstrukturierter Daten in einer nahtlosen Lösung.
- Greenplum Data Integration Accelerator (DIA) Module: Dieses Modul ermöglicht die rasche Integration von Partneranalyseanwendungen in die Greenplum Data Computing Appliance. Es wird u. a. eingesetzt, um die Herausforderungen beim Laden von Daten in einem parallelen, skalierbaren Modell zu bewerkstelligen, um Batch-Ladezeiten zu verkürzen oder um Micro-Batch-Loading zu implementieren.

In dieser Tabelle sind die technischen Daten der vier Greenplum DCA-Module zusammengefasst:

Modultyp	Greenplum DB Standard Module	Greenplum DB High Capacity Module	Greenplum HD Module	Greenplum DIA Module
Software	Greenplum Database	Greenplum Database	Greenplum HD CE	Zertifizierte Partnersoftware
Segmentserver	2 Sockel/12 Cores			
Arbeitsspeicher gesamt	192 GB			
Speichertyp	600 GB	2 TB	2 TB	2 TB
Speicherlaufwerke gesamt	48			
Verfügbare Kapazität (unkomprimiert)	9 TB	31 TB	28 TB	70 TB
Verfügbare Kapazität (komprimiert)	36 TB	124 TB	112 TB	–

Beispielkonfigurationen eines Greenplum DCA-Clusters mit den Modulen Greenplum Database Standard und Greenplum Database High Capacity:

Modultyp	GP DB Standard Module		GP DB High Capacity Module	
Anzahl Module	4	24	4	24
Anzahl Racks	1	6	1	6
Verfügbare Kapazität (unkomprimiert)	36 TB	216 TB	124 TB	744 TB
Verfügbare Kapazität (komprimiert)	144 TB	864 TB	496 TB	2.976 TB
Scan-Rate	24 GB/s	144 GB/s	14 GB/s	84 GB/s
Datenladerate	10 TB/h	60 TB/h	10 TB/h	60 TB/h

## MAXIMIERUNG DES NUTZENS DER EMC GREENPLUM DCA MIT EMC GLOBAL SERVICES

Mit einem umfassenden Angebot an Support und Services für die EMC Greenplum-Hardware- und -Softwarekomponenten sorgt EMC Global Services dafür, dass Ihr System wie gewünscht in Ihrer Umgebung arbeitet und gleichzeitig die Risiken für Ihr Unternehmen und Ihr Budget minimiert werden. Die von Experten durchgeführten Planungs-, Design- und Implementierungsservices ermöglichen Ihnen eine schnelle Wertschöpfung aus den Hardware- und Softwarekomponenten in Ihrer Umgebung, unabhängig davon, wie komplex diese ist. Nach der Implementierung unterstützt der Datenmigrationsservice von EMC Sie bei der Planung, dem Design und der Durchführung einer sicheren Migration kritischer Daten über beliebige Entfernungen hinweg auf Ihr neues System. EMC hilft außerdem bei der Integration Ihres neuen Systems in Ihre Informationsarchitektur und Ihre BI- und Analyseanwendungen (z. B. SAS, MicroStrategy, Business Objects oder Tableau) sowie beim Management der neuen Umgebung.

Unsere hochqualifizierten Servicemitarbeiter und Projektmanagementteams greifen auf einen umfangreichen Bestand an EMC Best Practices für die Bereitstellung von Data Warehousing- und Business Intelligence-Lösungen sowie bewährte Methoden zurück, sodass Sie schneller von Ihren Investitionen profitieren, ohne dabei Ihre Ressourcen zu belasten.

## EMC GREENPLUM DATA COMPUTING PRODUCTS DIVISION

Die EMC Data Computing Products Division prägt die Zukunft von Data Warehousing und Datenanalyse mit bahnbrechenden Produkten wie EMC Greenplum HD, EMC Greenplum Data Computing Appliance, EMC Greenplum Database und EMC Greenplum Chorus, der branchenweit ersten Cloud-Plattform für Unternehmensdaten. Diese Produkte bieten die Vorzüge von Open-Source-Systemen, Cloud Computing, Virtualisierung und Social Collaboration und versetzen international operierende Unternehmen in die Lage, mehr Nutzen und Erkenntnisse als je zuvor aus ihren Daten zu ziehen.

### KONTAKT

Informationen über die Produkte, Services und Lösungen von EMC und deren Vorteile für Ihr Unternehmen erhalten Sie bei Ihrem EMC Vertriebsmitarbeiter, autorisierten Händler oder auf unserer Website [www.emc2.de](http://www.emc2.de).

EMC<sup>2</sup>, EMC, Data Domain, EMC Greenplum, EMC Greenplum MPP Scatter/Gather Streaming, SRDF, Symmetrix, TimeFinder, VMAX und das EMC Logo sind eingetragene Marken oder Marken der EMC Corporation in den USA und anderen Ländern. Alle anderen in diesem Dokument erwähnten Marken sind das Eigentum ihrer jeweiligen Inhaber. © Copyright 2011 EMC Corporation. Alle Rechte vorbehalten. Veröffentlicht in den USA. 9/11 Lösungsüberblick H7419.5