EMC DATA COMPUTING APPLIANCE

Integrated Platform Driving the Future of Big Data Analytics

ADDRESSING THE NEW CHALLENGES OF A DATA-DRIVEN WORLD

Exploding data volumes, new data types, and ever-growing competitive challenges have led to radical changes in analytical technologies and a new approach to exploiting data.

Decades-old legacy architectures for data management have reached scale limitations that make them unfit for processing big data. The fast-growing data assets, broad diversity in data type and structure, and the need for complex analytics to unlock value from these data assets have overwhelmed traditional architectures.

The EMC® Data Computing Appliance (DCA) is an integrated analytics platform that accelerates analysis of Big Data assets within a single integrated appliance by using the Pivotal Greenplum™ database for analytics-optimized SQL on structured data. Delivery as a preconfigured appliance assures rapid deployment, simplified administration, and industry-leading TCO.

Modular Design for Scalability and Flexibility

Designed as a modular platform, DCAs can be scaled at any time by adding new modules. Adding modules provides linear scalability of storage and compute capacity for database capabilities. The structured data processing of SQL in the Pivotal Greenplum database delivers maximum flexibility and scalability for organizations that require fast analysis of diverse data sets.

Performance Architecture

The DCA employs a massively parallel processing (MPP) architecture for fast SQL processing, plus the fastest data loading rates in the industry—without the complexity and constraints of proprietary hardware. DCAs are purpose-built for analytics, and provide scalable computation, storage, and interconnect, delivered as a pre-configured appliance.

DCA FEATURES

Extreme Analytical Performance

At the heart of DCA is Pivotal Greenplum, a shared-nothing, massively parallel SQL relational open source data warehouse optimized for advanced SQL and predictive analytics on big data. The core principle of DCA is to move analytical processing dramatically closer to the data, running analytics in parallel atop an MPP architecture

ESSENTIALS

- Purpose-built, high-performance big data analytics appliance

- Includes Pivotal Greenplum for advanced SQL and predictive analytics on big data

- Modular design offers unequaled flexibility and scalability

- MPP architecture accelerates execution through transparent parallelism

- Preconfigured for faster deployment

while flowing data efficiently between resources as needed. The result is industry-leading performance for big data analysis at an affordable TCO.

Industry-Leading Flexibility and Scalability

EMC DCAs are configured to match users' needs, delivered ready to run and deployable within a few hours. Once deployed, DCA capacity can be scaled linearly by adding modules. DCAs can be configured from ¼ rack to 11 racks.

Coherent Appliance-Wide Administration

DCAs are easily administered -- regardless of configuration -- as all modules are managed and monitored through Pivotal® Greenplum Command Center, an appliance-wide administration tool. Integration with SNMP network management systems helps DCAs fit easily into most data center management frameworks.

PIVOTAL GREENPLUM FEATURES

Analytics-Optimized SQL Processing

Database modules in the DCA run Pivotal Greenplum, an SQL database completely optimized for analytical processing. By installing available algorithms that have been redesigned to run in parallel, the Greenplum database provides 10- to 100-times faster execution of statistical and analytical algorithms than traditional SQL databases.

Industry-Leading Parallel Data Loading

Unlike competing appliances, the DCA with the Greenplum database ingests data in parallel, achieving two- to five-times faster rates for loading. These rates scale linearly with system size, making DCA the industry leader for data ingests.

Linear Scalability of Pivotal Greenplum Using DCA Modules

The Greenplum database is designed for the Massively Parallel Processing (MPP) environment provided by the DCA. MPP allows users to linearly scale Greenplum database capacity, load rate, and performance by adding modules. The easy addition of modules limits service interruptions, allowing users to schedule data redistribution to occur during maintenance intervals.

Enterprise Availability

Data storage in the DCA is protected at three levels. Data is stored in RAID disk arrays that continue operating after a drive failure. Hot spare disks are provided and, in the event of a drive failure, are swapped in automatically by RAID controllers to avoid service interruptions. All database data are mirrored, with mirror copies of all data residing independently in the DCA. RAID, hot spares, and mirrors support resynchronization processes for automated self-healing recovery of storage failures.

Query processing is also protected. All segment servers are redundant, protected by automatic failover to assure that a server failure does not result in a system outage. Redundant master servers with automatic failover assure that host nodes also present no single points of failure.

Fully redundant 10GB Ethernet switching and connections link segment processors and master servers. This redundant interconnect provides automatic failover eliminating single points of failure between nodes in the DCA.

Reliable Backup and Disaster Recovery

Database modules in the DCA can be backed up using EMC Data Domain® for backup and recovery, as well as for remote disaster recovery using replicated data. EMC Data Domain's deduplication technology enables databases in the DCA to achieve backup rates as fast as 14TB per hour. Once backed up, Data Domain wide-area replication can remotely replicate Greenplum database data to remote sites for maintenance of warm standby systems to be used if a disaster affects the primary DCA.

System Performance and Capacity

Performance and capacity of DCA with Pivotal Greenplum modules (see Table 1):

Table 1.   GREENPLUM DATABASE CAPACITY, SCAN RATES, AND LOAD RATES

ATTRIBUTE	1 RACK	11 RACKS
Number of modules	4	44
Useable capacity (uncompressed)	184TB	22024B
Useable capacity (4x compression)	736TB	8096TB
Scan rate	48GB/sec	528GB/ sec
Data load rate	17TB/ hour	50TB/ hour

DCA Pivotal Greenplum Module Specifications

- 8U (¼ rack)

- Greenplum database

- 96 CPU cores

- 1024GB total memory

- 1.8TB 10K RPM SAS drive

- 96 total storage drives

- 46TB useable capacity (uncompressed)

- 184TB useable capacity (user data, 4x compression)

Physical Specifications

The DCA requires modest power and cooling (see Table 2), simplifying the task of provisioning DCAs into the data center.

Table 2.   DCA ENVIRONMENT PHYSICAL SPECIFICATIONS

	DCA QUARTER RACK	DCA HALF RACK	DCA FULL RACK
PHYSICAL DIMENSIONS	Height: 75 in/190 cm Width: 24 in/61 cm Depth: 42 in/107 cm		
WEIGHT	831lbs/ 377 kgs	1039 lbs/ 471 kgs	1455 lbs/ 660 kgs
MAX POWER DRAW VA	4030	6530	11490
COOLING (BTU/HR)	12781	21025	37512

LEARN MORE

Contact your EMC sales representative or authorized reseller to learn more about how EMC DCA can benefit your organization. You can also learn more about this solution at http://pivotal.io/big-data/emc-dca

CONTACT US

To learn more about how EMC products, services, and solutions can help solve your business and IT challenges, contact your local representative or authorized reseller, visit www.emc.com, or explore and compare products in the EMC Store.

EMC²