

Aster Data *n*Cluster 4.5

Deliver Ultra-Fast Analytics on Big Data

The continuing explosion in data sources and volumes strains and exceeds the scalability of traditional data management and analytic systems. At the same time, a new generation of analytics has emerged with requirements that are difficult or impossible for traditional architectures to meet: analysis of large data volumes, ultra-fast results, and deep data exploration through ad-hoc and interactive analysis. The demand for advanced analytic applications including fraud detection, customer behavior analysis, forecasting, scenario modeling and deep click-stream analysis requires a solution that overcomes the scalability, performance and complexity limitations of traditional approaches.

The Aster Data Solution

Aster Data Aster Data *n*Cluster delivers the first data-analytics server, a massively parallel (MPP) database with an integrated analytics engine to enable advanced analytics for deeper business insight. *n*Cluster's unique Applications-Within™ architecture runs analytic application logic inside the database, leveraging *n*Cluster's massively-parallel architecture and patent-pending SQL-MapReduce to fully parallelize processing for deep and ultra-fast analysis of massive data sets. *n*Cluster also delivers easy management of data and analytic applications, continuous availability, linear scalability and the cost advantages of commodity hardware. Combined with Aster Data's visual development environment and suite of optimized analytic modules—Aster Data Developer Express and Aster Data Analytic Foundation, *n*Cluster makes deep and rich analysis of big data that was previously difficult or impossible not only possible but also fast and easy.

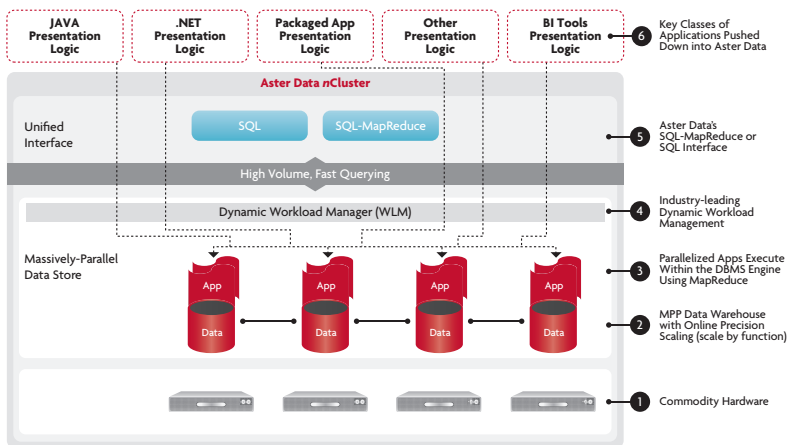


Figure 1: Aster Data's massively parallel data-analytics server enables applications to be embedded with data for ultra-fast analysis

Breakthrough Performance and Scalability

Aster Data *n*Cluster is designed to deliver leading performance and scalability for data and analytic processing. The result: 10x-1000x better performance that scales to terabytes and petabytes of data.

- **“Always-Parallel” Pervasive Parallelism.** *n*Cluster's massively parallel processing (MPP) architecture and in-database MapReduce enable end-to-end parallelism of data and analytic processing. *n*Cluster executes loads, queries, exports, backups, recoveries, installs and upgrades in parallel to take full advantage of all resources, optimizing performance for all data warehouse and analytic operations.
- **Unlimited Scalability.** *n*Cluster's unique Online Precision Scaling™ provides linear scalability across loads, queries, and backups independently or in unison to meet requirements. Granular splitting and load balancing of virtual partitions ensures maximum parallelism for massive “no limits” scalability.
- **“Always-On” Fault Tolerance.** *n*Cluster leads the industry in massive-scale fault tolerance with replication, automatic failover, NIC bonding, failure heuristics, and clustered backup to prevent unplanned downtime due to hardware or software failures.
- **In-Database MapReduce.** *n*Cluster deeply embeds an implementation of the MapReduce framework inside *n*Cluster, bringing the analytic power of MapReduce to developers of advanced analytics so that

Overview

- First visual integrated development environment for SQL and MapReduce analytic applications—Aster Data Developer Express.
- Suite of powerful building blocks for simplifying and accelerating development of analytic applications—Aster Data Analytic Foundation.
- Rich monitoring and easy management of data and analytic processing with the new Aster Data Management Console.
- Support for solid-state storage devices (SSD) to accelerate storage performance.
- Next-generation loading architecture for leadership throughput to minimize time needed for loading data.

Highlights

- Massively parallel processing (MPP) architecture for breakthrough performance and scalability
- Application processing services that allow applications to be fully parallelized inside the database
- Unlimited scalability to thousands of server cores and petabytes of active data
- Application embedding to enable ultra-fast, deep analytics on massive data scales
- Patent-pending SQL-MapReduce framework for powerful analytics
- Dynamic mixed workload management for thousands of concurrent users
- Powerful management tools for efficient administration
- Fault-tolerant design and online management to minimize downtime
- Up-to-the-second loading of fresh data
- Runs on certified commodity servers

“Aster Data's approach for interactive, big data applications is highly unique and allows us to store and process data in ways that was unimaginable in the past.”

Michael Brown, EVP of Software Engineering
comScore



they can easily deliver ultra-fast results on large data sets.

- **Dynamic Mixed Workload Management.** *nCluster* provides the first dynamic workload management capability for an MPP system that runs on commodity hardware. Automated policy controls ensure predictable performance and guaranteed service levels for diverse workloads, from interactive queries to large-scale reporting to batch data mining. Granular rule- and policy-based prioritization and dynamic resource allocation balance processing and compute resources for in-progress transactions, allowing administrators to adapt to changing priorities in real time.

Next-Generation Analytic Applications

Aster Data *nCluster* makes it easy to create advanced analytic applications that deliver ultra-fast analysis that scales to terabytes and petabytes of data. *nCluster* allows analytic applications to be embedded within the database for maximum performance, combines the analytic power of MapReduce with the familiarity of SQL, delivers the first visual development environment for SQL and MapReduce applications, and provides a suite of powerful analytic functions for accelerating development.

- **SQL-MapReduce.** Aster Data's patent-pending SQL-MapReduce (SQL-MR) framework makes it easy to leverage the power of MapReduce within the familiarity of SQL. SQL-MR allows developers to write powerful and highly expressive functions in languages such as Java, C#, Python, C++, and R and push them into the database where they can be called using standard SQL. SQL-MR functions are simple to write and are seamlessly integrated within SQL statements.
- **Embedded Applications.** *nCluster*'s Applications-Within approach makes it simple to embed analytic applications within the *nCluster* database to eliminate the overhead of moving large data sets from the database to analytic applications. Enterprises can easily push their existing applications into *nCluster* to see significant performance benefits.
- **Aster Data Developer Express.** Developer Express makes it easy to develop, validate, and deploy advanced analytic applications with the first visual development environment for SQL and MapReduce applications. It enables visual development through integration with the Eclipse integrated development environment, includes wizards to automate integration of applications with *nCluster*, provides a desktop testing environment for easy testing, and enables one-click push down of applications into *nCluster*.
- **Aster Data Analytic Foundation.** This suite of powerful, reusable analytic functions accelerates the development of rich analytics with ready-to-use modules optimized to take advantage of the power of Aster Data's SQL-MapReduce framework for fast analytics.

Efficient Management of Data and Applications

nCluster delivers rich visibility and control of not only the data but also the analytic applications inside *nCluster* to enable efficient management of data and applications, minimizing administrative work even as the system scales to 10s and 100s of servers. Rich visibility into data and application processing combined with capabilities for easy administration such as dynamic workload management and online maintenance make it easy to monitor and manage data and applications.

- **Powerful Console for Monitoring and Managing Data and Applications.** The new Aster Data Management Console makes it easy to configure, manage and monitor data, applications, and infrastructure. An intuitive graphical interface enables easy monitoring with summary dashboards, graphical views of query and process execution, and easy drill-down. It also makes administration easy with single-click scaling and point-and-click access to workload management policies.
- **"Always-On" Online Maintenance.** *nCluster* enables simultaneous load and export during queries, online backup and recovery, online restoration, and online scaling to avoid scheduled downtime.

About Aster Data

Aster Data is a proven leader in big data management and big data analysis for data-driven applications. Aster Data's *nCluster* is the first MPP data warehouse architecture that allows applications to be fully embedded within the database engine to enable ultra-fast, deep analysis of massive data sets. Aster Data's unique "applications-within™" approach allows application logic to exist and execute with the data itself. Termed a "Data-Analytics Server", Aster Data's solution effectively utilizes Aster Data's patent-pending SQL-MapReduce together with parallelized data processing and applications to address the big data challenge. Companies using Aster Data include Coremetrics, MySpace, comScore, Akamai, Full Tilt Poker, and ShareThis. Aster Data is headquartered in San Carlos, California and is backed by Sequoia Capital, JAFCO Ventures, IVP, and Cambrian Ventures, as well as industry visionaries including David Cheriton, Ron Conway, and Rajeev Motwani.

Copyright © 2010 Aster Data Systems. All rights reserved. Aster Data and the Aster Data logo are registered trademarks in the United States and/or other jurisdictions. All other marks and names mentioned herein may be trademarks of their respective companies. Information is subject to change without notice.

Technical Specifications

- **Server Hardware:** certified x86-based commodity server hardware
- **Operating Systems:** certified Linux-based operating systems
- **Drivers and APIs:** SQL, OLE DB, ADO.NET, ODBC, JDBC, Psycopg (Python)
- **SQL Standards:** ANSI SQL-92 compliant with SQL-99 and SQL-03 extensions
- **Compatible with leading business intelligence tools**
- **Hadoop connector** for rapid and parallel data transfer between *nCluster* and Hadoop
- **Data Integration:** leading tools including Informatica, GoldenGate, Microsoft SQL Server Integration Services (SSIS), Pentaho, Talend
- **Security:** Quest Authentication Services

"We're very excited to take advantage of the new Aster Data Management Console which provides deep visibility into the performance of the system and provides a deeper level of situational awareness and historical perspective for our entire operations and engineering staff."

Don Watters, Principal Architect, Data Services MySpace



For more information visit <http://www.asterdata.com> or call 1.888.ASTER.DATA