



Aster *n*Cluster

Quick overview:

Aster *n*Cluster makes it easy to derive deep insights from large data sets. The shared nothing *n*Cluster architecture transforms off-the-shelf, commodity hardware into a powerful, self-managing, and scalable analytic database capable of handling the most demanding data warehouse environments.

Key Benefits:

- Gain deep insights on massive data sets
- Scale incrementally from 100s of GBs to 100s of TBs and beyond with low-cost, commodity nodes
- Increase data capacity, query performance, or loading bandwidth independently
- Automate database administration and maintenance tasks
- Ensure high availability and resiliency with pervasive self-management

Deep Insights on Massive Data

Data warehouses today have difficulty analyzing the exploding amount of data, much less meeting new business analytics demands. You are either forced to discard potentially high-value data or simplify your analysis – until now. Aster *n*Cluster is a massively parallel processing (MPP) database which helps you better understand customer behavior, monetize new opportunities, and control risk by using all of your data. Aster *n*Cluster transforms off-the-shelf commodity hardware into a high-performance analytic database capable of handling the most demanding data warehousing environments by providing:

- Powerful analytics
- Incremental scalability
- Easy manageability

Powerful Analytics

Aster *n*Cluster allows you to get deep insights from your data quickly – answers you can’t get with traditional approaches that are overwhelmed with complex queries. Behind Aster’s analytic power are patent-pending algorithms designed to efficiently perform complex multi-terabyte joins and aggregations at breakthrough speed for processing large-scale queries, reporting, and analysis. Aster *n*Cluster provides fast, scalable loading that ensures your data and results remain fresh and up-to-date.

Incremental Scalability

Traditional data warehouses require you to plan out analytic capacity requirements months or years in advance, as well as make large capital purchases (i.e., “super-size”) up-front. Aster *n*Cluster lets you start with minimal investment (right size) and scale incrementally when needed (right time). Aster *n*Cluster is built on a unique, multi-tiered *n*Cluster architecture which consists of three separate classes of nodes: Queens, Workers, and Loaders. The three-tier design encapsulates a clean separation of roles for analytic processing. Each tier can be independently and incrementally scaled in response to the workload characteristics – adding more capacity (Workers), loading bandwidth (Loaders), or concurrency (Queens) on an as-needed basis. Your data warehouse can now scale on-demand to adapt to your changing business conditions.

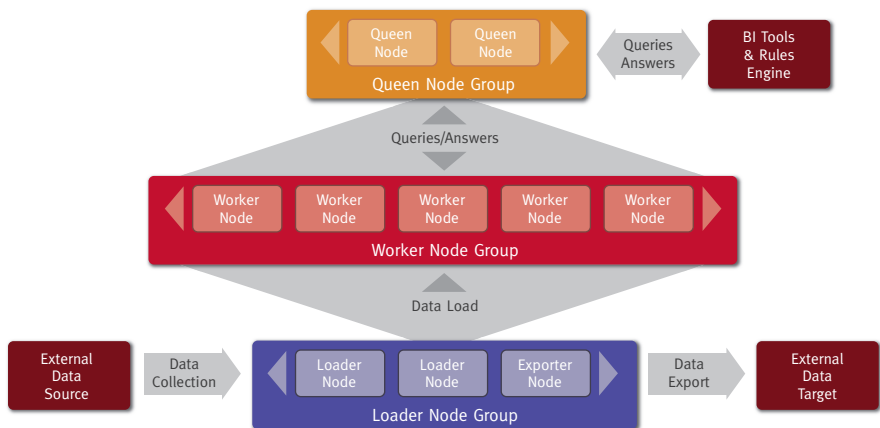


Figure 1: Aster *n*Cluster provides a multi-tiered architecture for independent, linear, and incremental scalability

Technical Specifications

- Linux-based operating system
- Massively parallel processing (MPP)
- Supported server hardware: all x86-based industry-standard servers
- Supported network hardware: 1 GbE or 10GbE industry-standard switches
- Supported APIs: SQL, JDBC, ODBC
- Industry-standard ANSI SQL-92 with SQL-99 extensions
- Interoperable with leading BI and ETL tools
- Fully portable from Teradata, Oracle, Microsoft SQL Server, MySQL, Sybase, IBM DB2, and other relational databases

Easy Manageability

Data warehouses are expensive to operate due to support costs for administration, tuning, and maintenance. The iterative loop of installation and tuning continues relentlessly as data grows and hardware generations change. Aster nCluster provides a unifying cluster fabric to ensure high availability, adaptation, and resiliency with minimal effort via pervasive self management. By providing a single system image to administrators, Aster nCluster automates heterogeneous cluster administration while still providing complete system visibility and control. Your data warehouse can now adapt to its hardware infrastructure transparently.

Key Features

POD (Performance Optimized Dimensional Partitioning)TM – patent-pending algorithms that intelligently place data to minimize network traffic across nodes and enhance query performance.

POD TransportTM – optimize query performance by combining compression and aggressive buffering for network-efficient shuffling of data segments at query run-time.

Network-optimized query planning – algorithms combine global and local optimizers for maximum parallel processing performance and network efficiency.

Precision scaling – incrementally increase capacity or performance across multiple functional tiers (query, loading, export, etc). Scale independently or in unison to meet business requirements via commodity hardware.

Heterogeneous commodity scale-out – extend multiple generations of commodity hardware to scale easily over time – no proprietary appliance hardware lock-in or rip-and-replace upgrades.

Efficient parallel loading – parallel loading scales linearly with additional nodes to meet any required SLA. Loading is concurrent

with query processing – queries can continue uninterrupted during loading due to workload isolation.

Single-system manageability – unifying cluster fabric provides single-system interface to administrators. Automates routine cluster administration (OS and DBMS installation, configuration, and load balancing) while providing visibility and control.

One-click scaling – go from raw hardware to a fully-balanced system within minutes – compared to days with other data warehouse alternatives. Provisioning, cluster rebalancing, and data replication are handled automatically.

Self-healing high availability and recovery – highly resilient system continues query processing, even in the event of failures. Predictive heuristics and healing algorithms provide fast, automated recovery from both hardware and transient software failures.

Online replica restoration – in the event of node failure, system recovery initiates replica restoration in the background with no system down-time. Change-tracking restoration algorithms offer ultra-fast recovery times.

About Aster Data Systems

Aster Data Systems is a proven innovator in massively parallel processing (MPP) databases for data warehousing and analytics – bringing deep insights on massive data analyzed on clusters of commodity hardware. Co-founded by three colleagues in the Stanford Computer Science Ph.D. program, Aster's nCluster database provides patent-pending innovations in scalability, manageability, availability, and analytics. Aster is headquartered in Redwood City, California and is backed by Sequoia Capital, Cambrian Ventures, and First-Round Capital.

