

Aster Data® nCluster®: TCO Advantage

Today's data-driven systems and applications face exponential data growth, but the inability of traditional database systems to economically scale and deliver high performance has made the management and analysis of such data volumes extremely costly or, in some cases, unaffordable.

The biggest contributors to the total cost of ownership (TCO) for data management systems are:

- Cost of hardware (infrastructure including storage, network, servers, etc.)
- Cost of management (hardware and software)
- Cost of scaling (effort required to scale out)
- Cost of downtime (business disruption due to downtime)

Aster Data nCluster delivers the first analytic platform, a massively parallel (MPP) row and column database with an integrated analytics engine. It 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 an "analytic platform," Aster Data's solution effectively uses Aster Data's patent-pending SQL-MapReduce® together with parallelized data processing and applications to address the big data challenge.

nCluster is architected to not only enable high-performance data-driven applications, but also to fundamentally change the economics of big data management and big data processing. Since it is designed to analyze terabytes to petabytes of data, the nCluster architecture includes optimizations that make it economically viable at these massive data sizes. Using such optimizations, nCluster minimizes both initial acquisition costs and ongoing operational costs, including cost of maintenance, scale-out, downtime, etc.

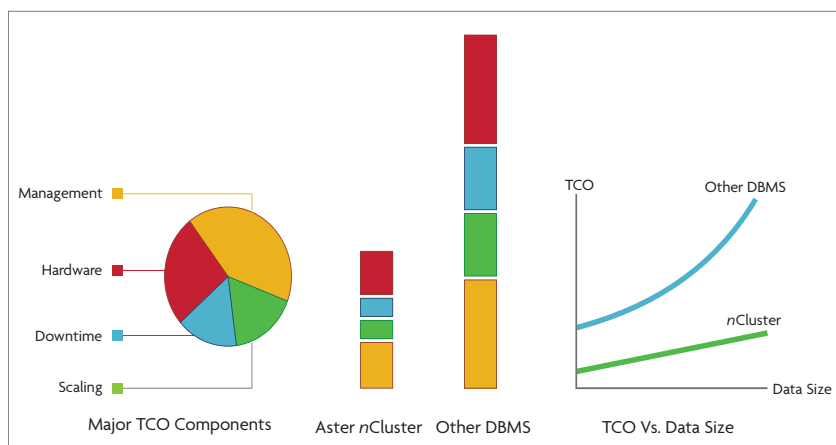


Figure 1: Aster nCluster TCO advantage.

Quick Overview

The Aster Data nCluster massively parallel analytic platform is built on an architecture that makes it economically viable to analyze terabytes to petabytes of data that today's data warehouses and data-driven applications need to manage. nCluster leverages its "Always Parallel" architecture to provide high performance and scalability using inexpensive commodity hardware as building blocks, and "Always On" capabilities to reduce cost of downtime, scaling, and management.

Highlights

- Off-the-shelf commodity servers and networking hardware reduce hardware costs
- Coexistence of heterogeneous hardware protects investments over years
- Discrete-purpose servers enable choice of cost-effective configuration for each task
- Self-managing, self-healing capabilities along with a single system view reduce management costs
- Single-click online scaling reduces cost of scaling
- Major causes of planned and unplanned downtime are eliminated, reducing business disruption costs

"Aster Data allows us to have superior scalability on commodity hardware—giving us a tremendous cost and performance advantage. Aster Data was the only company we could find that gave us this capability."

Shawn Farshchi, Chief Operating Officer
Coremetrics



Cost of Hardware

Traditional database systems were architected 10 to 30 years ago and were optimized for hardware technology available at that time. Over the years, there has been a tremendous increase in the power of components such as processors, memory, network switches, and disk controllers. At the same time, the fundamental architecture of traditional database software has remained unchanged, failing to fully leverage the advances in hardware technology. The scale-up approach of traditional databases involves costly investment in hardware but does not provide the desired scalability and performance. Similarly, several MPP database vendors use purpose-built hardware that is costly to acquire and even costlier to maintain.

Modern commodity hardware is extremely powerful with configurations such as two quad-core processors, 16 to 32 GB RAM, and multi-terabyte storage available in a single, inexpensive server. The *nCluster* MPP data warehouse architecture is designed to leverage such commodity hardware. It uses patent-pending technology that enables a large number of commodity hardware building blocks to work as a single, unified system that delivers extremely high performance and scalability while requiring significantly smaller hardware investments.

nCluster servers are inexpensive commodity servers that can be procured from the vendor of an organization's choice; even existing servers can be repurposed. Similarly, *nCluster* uses standard Gigabit Ethernet switches for inter-server communication. Storage costs are one of the biggest contributors to data center costs and *nCluster* provides a significant advantage over other systems here as well. Data is stored on direct attached storage of the commodity servers. *nCluster* does not require costly network-based storage (SAN/NAS) infrastructure. *nCluster* Smart Compression™ further reduces the cost of storage. *nCluster* can commingle heterogeneous hardware in a single cluster to protect hardware investments over time. Organizations can leverage the power of Moore's law and other hardware advances by adding new-generation hardware without throwing away existing hardware. This is different from many other MPP systems that require "forklift upgrades" to leverage new-generation hardware or to scale a system beyond the capacity of existing hardware.

Also, *nCluster* can dedicate special-purpose servers for different tasks, allowing organizations to pick the most cost-efficient configuration for each task. For example, Loader servers require very little persistent storage—their cost can be reduced by choosing a configuration with minimal storage. Similarly, Backup servers can have a storage-heavy configuration to provide low cost per terabyte. Further, *nCluster*'s software modularization can help organizations run different functions on the same server. For example, if load requirements are not high, the loader software can coexist on the Queen servers.

Similarly, *nCluster* provides a hybrid row and columnar architecture, which lets you store data in the format that best matches your needs and reduces the hardware footprint required. Use a columnar database for data that will be accessed for canned business intelligence reports. Use a row-based database for massive datasets accessed for exploratory or advanced analytics.

Cost of Management

At current data volumes, traditional database systems are extremely expensive to manage because they aren't fundamentally designed to handle such volumes. Large data volumes require bigger hardware, more complex deployment architectures, and continuous tuning of a large number of system parameters, leading to high management costs.

nCluster significantly reduces the cost of managing and administering the MPP database software and hardware. The *nCluster* architecture is built on the fundamental

"In these challenging economic times, many customers are looking for a way to do more in analytics and data warehousing without making a major investment in new hardware."

Aster Data is offering that opportunity: not only can you run on your existing commodity infrastructure, you can even mix server models and generations to come up with enough capacity.

*Richard Winter, President
WinterCorp*



principles of self-management and self-healing. Leveraging its Recovery-Oriented Computing architecture, it transparently handles software and hardware errors, minimizing administrative burden. Similarly, its “Always Parallel” architecture provides high performance and scalability out-of-the-box, obviating the need to continuously monitor and optimize the large number of system parameters that impact performance of traditional databases.

Further, *nCluster* provides a “single system view” for managing the whole cluster, which makes the task of administering the cluster very easy. The management effort is independent of the cluster size, eliminating the effects of increasing data volumes on the cost of management, so managing a 100-server cluster is as easy as managing a three-server cluster.

Similarly, installation and upgrade of software across the cluster is done by the Queen server—administrators do not need to directly install any software on any other server. Due to the simplified architecture, *nCluster* also eliminates the need to separately administer hardware sub-systems like storage, significantly reducing management costs.

Cost of Scaling

With continuous increase in the size of data, organizations are often required to undertake costly capacity expansion projects for MPP data warehouses. The notion of TCO over the life of a data warehouse should not be limited to the steady-state costs at a given size. It should also include the costs incurred in the process of scaling the system. If a data warehousing environment needs to scale, then the total cost of scaling (TCS) can be defined as:

$$\begin{aligned} \text{TCS} = & \text{Incremental cost of new hardware and software} \\ & + \\ & \text{Cost of human effort required to scale} \\ & + \\ & \text{Cost of downtime required to scale} \\ & + \\ & \text{Cost of data migration} \\ & + \\ & \text{Cost of software upgrades} \end{aligned}$$

nCluster is designed to scale easily and keep the total cost of scaling at a minimum. With its Online Precision Scaling feature, *nCluster*’s capacity can be increased by adding more commodity servers, so organizations can make small incremental investments and control costs. This is in contrast to many other systems that require investments in large steps to increase capacity. Expansion of *nCluster* does not require throwing away old hardware—creating value where traditional scaling projects would recognize sunk costs.

Further, *nCluster* Online Precision Scaling provides a unique online capacity expansion capability—bare-metal servers can be plugged into the network and *nCluster* automatically installs the software, while the system continues to be available for business. Thus, the effort required is minimal and the scale-out does not require the lengthy downtime associated with capacity expansion projects of other MPP database systems. Data migration during scale-out is automatically managed by *nCluster*—data is redistributed across the cluster in the background without manual intervention.

Software installation and upgrades are also done in a massively parallel way in *nCluster*. The administrator only needs to install the software on the Queen, which automatically installs software on all other servers in parallel. *nCluster* Upgrade Manager automates most tasks involved in software upgrades and runs the upgrade process in parallel on each server in the cluster, significantly reducing the cost and downtime associated with

Technical Specifications

- **Hardware** – All x86-based commodity servers with direct-attached storage, including Dell, HP, IBM, and Sun
- **Networking** – All GbE and 10 GbE commodity network switches, including Cisco and Juniper Networks
- **Operating Systems** – Leading Linux-based systems including Gentoo, RedHat (RHEL), CentOS, Novell SUSE, and Ubuntu
- **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
- **BI Tools** – SAS, MicroStrategy, SAP Business Objects, IBM Cognos, Microsoft SQL Server Reporting Services (SSRS) and Analysis Services (SSAS), Pentaho, and JasperSoft
- **Hadoop Data Connector** – Two-way connector for rapid, parallel data transfer between Aster Data and Hadoop
- **Data Integration** – Leading tools including Informatica, GoldenGate, Microsoft SQL Server Integration Services (SSIS), Pentaho, Talend

upgrades. The parallel execution of upgrades makes the time required for the upgrade process independent of the cluster size.

Cost of Planned and Unplanned Downtime

As analytics-intensive applications play a critical role in business operations, the cost of system downtime can be very high. *nCluster* has market-leading “Always On” capabilities that minimize both planned and unplanned downtime, significantly reducing the cost of associated business impact.

The most common causes of planned downtime have been either completely eliminated or significantly reduced in *nCluster* through features such as:

- *nCluster* Online Precision Scaling for single-click incorporation of new servers
- Online concurrent loads and exports eliminate the need for a dedicated loading window
- *nCluster* Online Backup eliminates the need for a dedicated backup window
- *nCluster* Upgrade Manager for automated software distribution minimizes downtime during software upgrades

Issues that cause unplanned downtime in traditional database systems have been addressed in *nCluster* with the following capabilities:

- Automatic software-based data replication within the cluster for high availability
- Completion of queries even in case of hardware failure through automatic fail-over in a completely user-transparent manner
- Online restoration of replicas lost due to server failure or data corruption, with no system down-time
- *nCluster* Online Resync for fast online resynchronization between original server and replica in case of transient server failures, with no system downtime
- Online restore from table-level backups
- Smaller time-to-recovery due to massively parallelized restores and network-efficient backup compression

About Aster Data

Aster Data is a market leader in data management and advanced analytics for diverse and big data, enabling the powerful combination of cost-effective storage and ultra-fast analysis of relational and non-relational data. Aster Data *nCluster* is an analytic platform that incorporates a massively parallel processing (MPP) hybrid row and column database with an integrated analytics engine, allowing application logic to execute with data to deliver breakthrough performance and scalability. Aster Data's solution utilizes Aster Data's patent-pending SQL-MapReduce to parallelize processing of data and applications and deliver rich analytic insights at scale. Companies including Barnes & Noble, Intuit, LinkedIn, Akamai, Full Tilt Poker, and MySpace use Aster Data to deliver applications such as deep clickstream analysis, recommendation and personalization analytics, real-time fraud detection, and churn analysis.