

# Aster Data *n*Cluster: Manageability

As data sizes grow, data warehouses managing large volumes of data become unwieldy and unmanageable. In an attempt to meet the current scalability requirements, traditional database systems have required complex deployment architectures and ongoing tuning of a large number of system parameters. While such deployment architectures and tuning have been unable to meet the scalability requirements of analytically-intensive, big data applications and data warehouses, the task of managing such systems has become extremely complex, leading to high maintenance costs and slow performance. The complexity of managing such systems also has a direct bearing on system availability—complex systems lead to higher chances of human errors and system failures that reduce availability.

Aster Data *n*Cluster delivers the first data-analytics server, a massively parallel (MPP) 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 *n*Cluster provides “Always On” and “Always Parallel” capabilities for big data management and processing. At the same time, it provides powerful manageability features that make managing large-scale data warehouses extremely easy and error-free.

The fundamental manageability principles of *n*Cluster are:

- **Size-Independent Manageability** – Management effort is not dependent on the number of servers in the cluster, i.e. the effort required for administration does not increase with cluster size.
- **Self Management** – The system is self-managing and minimizes the need for administrator intervention. *n*Cluster leverages its Recovery-Oriented Computing architecture for self-healing capabilities.
- **Single System View** – The system always presents a single system view to the administrator. Even as the system scales to several hundred servers, the administrator

## Quick Overview

Aster Data *n*Cluster provides manageability features that make big data management and big data processing extremely easy and error-free. With its self-management and “single system view” capabilities, managing *n*Cluster is as easy at the scale of hundreds of servers as it is at a few servers.

## Highlights

- Single-click online scalability
- Automated storage management
- Heterogeneous hardware management
- Bare-metal installs and single-click upgrades
- System-wide statistics
- Dynamic mixed workload management
- Table compression and multi-level partitioning
- Online backup and recovery
- Intuitive management tools with powerful administrative features
- Monitoring and alert framework

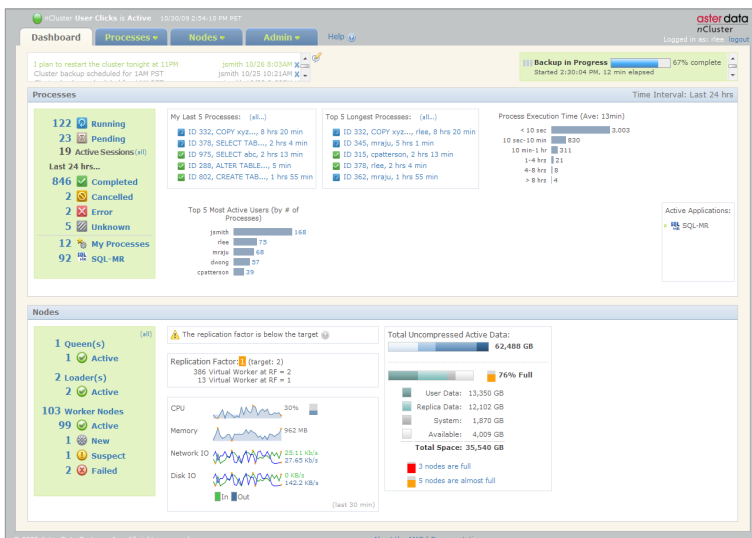


Figure 1: Aster Data Management Console

sees the system as a single entity. All the internal details of the cluster are transparently managed by *nCluster* software.

## Hardware Management

### Single-Click Online Scalability

*nCluster* is designed from the ground up to be an elastic cluster. Its Online Precision Scaling™ feature enables adding and removing servers from the cluster with a single click in Aster Management Console (AMC), eliminating large migration effort and cost required to expand the capacity of traditional systems.

### Automated Storage Management

As *nCluster* uses direct attached storage, storage management is automatically performed by the *nCluster* software. The system automatically load balances data, formats drives, propagates exceptions such as I/O errors back to the management console, and generates alerts on disk failures. When compared to the large network-based storage systems required by traditional database systems, the effort of managing storage in *nCluster* is minimal—even at the multi-terabyte or petabyte scale.

### Heterogeneous Hardware

*nCluster* is designed to protect existing hardware investments and can commingle different generations of hardware in the same cluster. Manageability of the system is not impacted by the existence of different hardware in the same cluster.

### Network Management

*nCluster* Network Aggregation makes the task of network configuration and failure handling extremely easy. The *nCluster* administrator only needs to assign a subnet to the cluster and configure the VLAN on the switch. Other aspects of network management are completely managed by the system. These include:

- Assignment of IP addresses
- Auto-discovery of all network interfaces
- Aggregation of links for maximum performance
- Failover of network interfaces during boot-up and steady-state

### SQL-Based Hardware Statistics Repository

*nCluster* provides a single-system interface to pull out hardware statistics (CPU, disk, network, memory, processes, etc.) from the entire cluster. The SQL model helps DBAs overlay hardware statistics with system statistics to gain valuable insights for performance management and diagnosis.

## Software Management

### Bare-Metal Install

When adding a new machine to the cluster, zero preparation is required for the machine—the administrator can simply input the MAC address of the first network interface and power on the machine. The system automatically installs the software, formats the drives, configures the network, and rebalances the existing data and workload. This is in contrast to traditional systems that require significant manual effort for installing and configuring the software on every server.

### Single-Click Upgrade

*nCluster* Upgrade Manager makes the process of upgrades extremely simple and fast. The administrator only needs to upload the upgrade software to the Queen server, which automatically distributes the software to other servers. Upgrades are run in

parallel on all servers, making the effort and time required for upgrading the software on a 100-server cluster as minimal as a three-server cluster.

## Workload Management

Aster Data provides the first-ever dynamic workload management (WLM) capability for an MPP system that runs on commodity hardware. Automated policy controls ensure predictable performance and guaranteed service levels for the most diverse workloads from interactive queries to large-scale reporting to batch data mining. *nCluster* also features breakthrough capabilities in low-latency interactive queries and trickle-feed loading to meet the most demanding real-time analytic requirements. *nCluster* Dynamic Workload Management provides configurable policies and mechanisms to manage workloads efficiently:

- **Dynamic Workload Management** – A unique dynamic workload manager effortlessly balances processing and compute resources to in-progress transactions, allowing administrators to seamlessly adapt to changing priorities in real time. Includes granular rule- and policy-based prioritization and dynamic resource allocation and re-allocation.
- **Quality of Service Policies** – Different users can be assigned to different priority classes and assigned different fair-shares within those classes. Thus prioritization and fair-share scheduling can be used to provide QoS to *nCluster* users.
- **Bulk Load Management** – ETL load is automatically distributed among loader servers. Aster Bulk Feeders contact a single cluster entity that routes the session to an appropriate loader based on the current system load, providing very high throughput that is required for loading large volumes of data into frontline data warehouses.
- **Workload Repository** – *nCluster* provides an active SQL-based workload repository to help DBAs gain insights into the current activities within the cluster, and analyze historical statistics.
- **Multiple Databases** – Administrators can host multiple databases in a cluster as separate autonomous domains to help isolate workloads.

## Data Management

Big data applications require terabytes to petabytes of data and hence require efficient and scalable data management capabilities. Traditional database systems impose complex management requirements for tasks such as undo management, file management (data files, control files, log files), space allocation, etc. Such a management model is simply infeasible at the large scales of today's analytically-intensive big data applications.

*nCluster* eliminates the complexity of such data management and processing tasks through its self-management principle—all such routine but cumbersome and error-prone tasks are auto-managed by *nCluster* software. At the same time, *nCluster* provides powerful data management options in areas that require context-specific administrative decisions. These include:

- **Table Compression** – *nCluster* provides table compression capabilities for reducing the storage footprint. Administrators can choose from HIGH, MEDIUM, and LOW levels of compression, depending on their storage and performance needs.
- **Multi-Level Partitioning** – Partitioning is inherently built into the system and enables classic divide-and-conquer techniques of data management. Administrators can choose as many levels of data partitions as desired by creating child tables that automatically inherit parent table properties. Partitioning provides ease of data management and also gives performance gains through “partition pruning”—a process that completely leaves out unnecessary table partitions during query execution.
- **Configurable Replication Factor** – *nCluster* does not overly rely on RAID1+0 configuration of underlying storage—it implements replication at a software level. Administrators

can choose the desired replication factor and the system automatically maintains an appropriate number of replicas across the cluster. This provides data protection not only against disk failures, but also against failures of all other components of servers.

## Backup Management

*nCluster* Online Backup provides enterprise-class data protection capabilities, while keeping the task of managing large volume backups extremely simple. It implements backup and recovery features using principles of massively parallel computing—dedicated Backup servers form Aster Data *nCluster* Backup Cluster and run with software that provides a single system view for backup management. Like *nCluster*, Backup Cluster also provides self-management of storage and failures, minimizing administrative burden. New capacity can be easily added to Backup Cluster by adding new servers.

## Management Tools

*nCluster* provides a complete range of system management tools built on the principles of usability engineering. The administrative interfaces are designed to be simple and intuitive, yet provide a rich set of features. Since many of the routine but complex tasks are auto-managed by *nCluster*, administrators can focus on high-value decisions and improve productivity. This significantly reduces the effort required for learning as well as ongoing maintenance of the system.

### Aster Data Management Console (AMC)

- Graphical system administration interface
- Single-click cluster scaling and load balancing
- Graphical insight into system metrics, such as performance and storage
- Access to system and server logs, as well as *nCluster* configuration options

### Aster Data *nCluster* Admin

- Graphical database administration interface
- Access to a complete set of administrative functionality
- Works on Windows, Mac, and Linux

### Aster Data *nCluster* Terminal (ACT)

- Command line interactive terminal
- Access to administrative functionality

### Aster Data *nCluster* Backup Terminal

- Addition of new Backup servers for capacity and performance management
- Management of backup and restore jobs along with history
- Backup data management

### Nagios® Plug-ins

- Full suite of plug-ins for monitoring connectivity, availability, and performance of *nCluster* using open source Nagios® monitoring platform

## 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. For more information please visit [www.asterdata.com](http://www.asterdata.com), or call 1.888.Aster.Data.