

MySpace.com



## MySpace Scales Analytics for All of Their Friends

“With Aster, MySpace has the analytic horsepower and necessary scalability to meet our needs. The platform easily scales with our traffic growth and helps us better understand our users and improve their experience.”

- Jim Benedetto,  
SVP of Technology,  
MySpace

### Challenges with massive data volumes

Since its founding in 2003, MySpace.com has grown to one of the Internet’s most popular Websites with over 110 million active users every month. MySpace was looking for an analytic solution to help them better understand traffic on their site and optimize the experience of their social network members.

MySpace is one of the most trafficked Websites in the world with more than 1.2 billion pageviews a day (comScore, Dec., 2007). Collecting and analyzing this amount of data posed serious challenges. MySpace needed a system which could scale easily as their traffic grew, load terabytes (TB) of data fast, and perform both simple and complex queries quickly for intra-day analysis and response – all in a cost-effective manner. MySpace needed to analyze complete datasets – not just samples or summaries. Sampling would completely miss infrequently occurring but highly profitable patterns.



### Approach

MySpace evaluated many options but only Aster could provide the combination of scalability, speed, and analytic power to meet their needs. With Aster, MySpace has built a cost-effective data analytics platform that measures usage of MySpace.com worldwide to optimize product features, marketing efforts, and site usage.

Unlike a monolithic architecture that is very expensive to deploy and maintain, Aster is a cluster of analytic processing nodes based on commodity servers. Aster provides multiple tiers of scalability, automated management features including intelligent provisioning and failover, and specialized analytic extensions that result in fast queries that are simply not possible with alternative solutions.

### Scalability

The traffic generated by MySpace’s 110 million+ active users translates into 100s of TBs of data for analysis. With Aster, MySpace is able to collect 100% of their Web traffic data for analysis without the need for sampling data. The Aster system at MySpace is a 100-node cluster capable of analyzing hundreds of TBs of data. Additional nodes can be added quickly through one-click scaling as the traffic on MySpace.com continues to grow exponentially. With Aster, scaling the system is easy – going from raw hardware to a balanced system in one click.

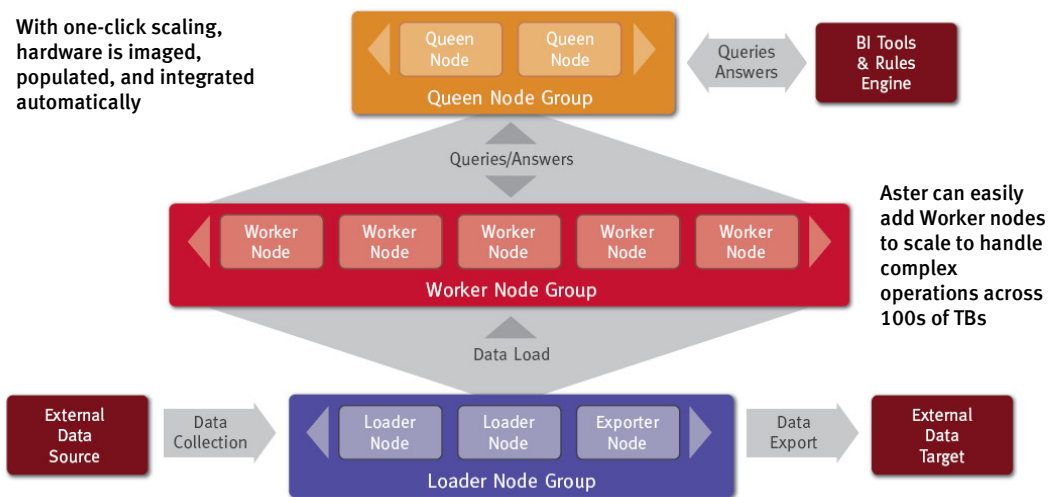
In addition to one-click scaling, Aster provides zero-click, zero-downtime failover – automatically detecting issues and isolating the hardware within seconds. The internally-maintained level of data replication is restored in minutes, ensuring queries do not fail and no data is lost. It is simple and transparent to remove the commodity node to repair, replace, or upgrade as needed.



## Speed

MySpace needed to understand traffic on their site intra-day, requiring terabytes of data to be loaded in their data warehouse every day. More importantly, they needed queries to return results quickly. Aster uniquely met both of these challenges.

Aster's architecture is structured in independently-scalable tiers, each of which adds a degree of freedom to the customer. The Aster Worker tier, where data is stored on locally attached disks, can be scaled to increase query performance and volume. The Aster Loader tier can scale independently to increase load throughput. This enables massively parallel processing for extraction and loading. Once the data is loaded, user queries are intelligently routed to each node to process only relevant data. This enables query load-balancing to eliminate hot-spots and increase performance, returning results in seconds or minutes versus hours or "did not finish".



Independent scaling of Loader nodes enables control over data loading speeds. With MySpace, data from thousands of Web servers are collected and loaded for analysis within minutes

## Analytic Power

MySpace was looking for a solution that can take them beyond simple query and reporting to provide the deep insights that can drive significant business results. With Aster, MySpace can use powerful analytic extensions to simplify and speed analysis. One example of this expressive power is the ability to quickly identify and reinforce the "golden path" for marketing campaigns as well as understand and streamline the complaint-cycles that impact customer loyalty. This analysis is conducted in minutes or seconds compared with queries that do not finish when using alternative analytic solutions.

### 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.

