aster data
more data. big insights.

Customer Success Story

myspace.com

MySpace.com Scales Analytics for All of Its Friends

Challenges with Massive Data Volumes

Since its founding in 2004, MySpace.com has grown into one of the internet's most popular websites with over 120 million active users every month. MySpace was looking for an analytic solution to help it better understand traffic on its site and optimize the experience of its social network members. MySpace is one of the most trafficked websites in the world with 10-11 billion events per day. Collecting and analyzing this amount of data posed serious challenges. MySpace needed a system which could scale easily as its traffic grew, load terabytes (TB) of data quickly 24x7 without user lock-out, and perform both simple and rich queries fast for intra-day analysis and response—all in a cost-effective manner. MySpace needed to analyze complete datasets—not samples or summaries—because sampling would completely miss infrequently occurring but highly profitable patterns.

MySpace's Approach to Large-Scale Data Warehousing

MySpace evaluated many options but only the Aster Data *n*Cluster massively parallel data-application server could provide the combination of scalability, speed, and analytic power to meet their needs. With Aster Data, MySpace has built a cost-effective MPP data warehouse that measures use of MySpace.com worldwide to optimize product features, marketing efforts, and site usage. Unlike monolithic architectures that are expensive to deploy and maintain, Aster provides an always-on, always-parallel MPP architecture with the first-ever SQL-MapReduce programming framework to embed and parallelize application logic with data for ultra-fast analysis and data processing.

Scalability Handles 100's of Terabytes of Data with Ease

The traffic generated by MySpace's 120 million+ active users yields 100s of TBs of data for analysis. With *n*Cluster, MySpace is able to collect 100% of its Web traffic data for analysis without the need for sampling data. The *n*Cluster MPP data warehouse at MySpace is used to analyze hundreds of TBs of data. Additional servers can be added quickly through one-click scaling as traffic on MySpace.com continues to grow. *n*Cluster also provides "always on" availability—reducing both planned and unplanned downtime through online fault-tolerance and recovery and live system administration. If a server at MySpace fails, *n*Cluster automatically recovers and restores the data replicas, ensuring queries do not fail and no data is lost. It is simple and transparent to remove the failed server to repair, replace, or upgrade as needed.

MySpace Background

- Social network with 120M active users
- Users share photos, videos, and interests
- 10-11 billion events per day

Challenges

- Collect 100% of the traffic generated from MySpace.com
- Load two to three terabytes of new data per day
- Analyze 100s of terabytes to improve user experience on MySpace

"Our key business performance metrics are all powered out of the Aster system... If all that went away, I think it's kind of like going back to an age where there was no light."

Hala Al-Adwan, VP of Data
MySpace

Maximum System Performance and Speed

MySpace needed to understand traffic on its site intra-day, requiring the daily loading of terabytes of data into Aster Data *n*Cluster. More importantly, MySpace needed queries to return results quickly. Aster Data uniquely met both of these challenges. The *n*Cluster MPP architecture is based on independently-scalable tiers, each of which adds a degree of freedom to MySpace. The *n*Cluster Worker tier can be scaled to increase query performance and volume. The *n*Cluster Loader tier can scale independently to increase load throughput. This enables high-volume loading and exporting. Even as MySpace loads data, 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 not at all.

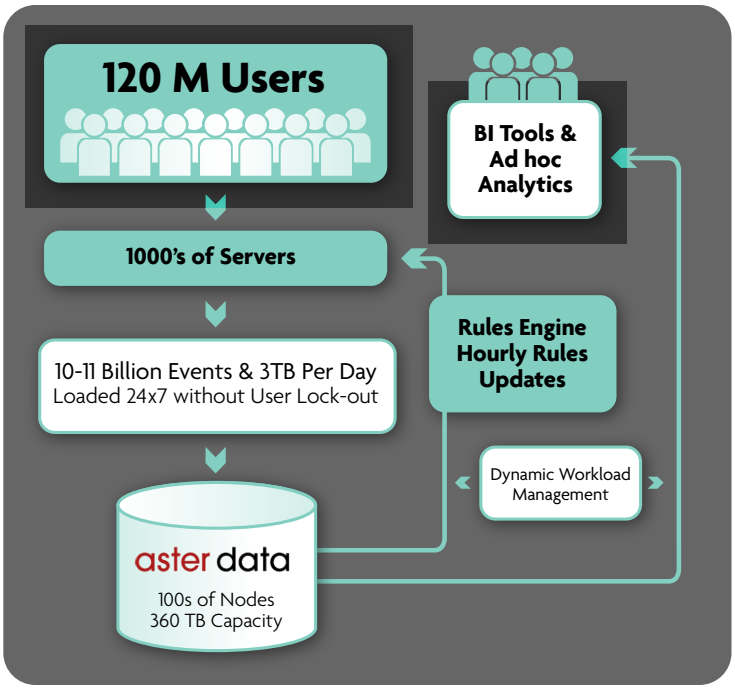

Figure 1: The MySpace MPP data warehouse manages 100s of terabytes of active data.

Analytic Power

MySpace was looking for a solution to take them beyond simple query and reporting for the deep insights that drive significant business results. With Aster Data's SQL-MapReduce, 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.

"Integrating Aster Data and including them from the very beginning in the MySpace Music project ... is what allowed that to be the most successful data warehouse implementation we've had to date... We should definitely use it as a blueprint for any future implementations we do."

Christa Stelzmuller, Chief Data Architect
MySpace

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, or call 1.888.Aster.Data.

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