eBay is the world’s largest online marketplace, enabling the buying and selling of practically anything. Founded in 1995, eBay connects a diverse and passionate community of individual buyers and sellers, as well as small businesses. eBay currently serves over 112 million active users and 400+ million items for sale.

The Challenge
In today’s high-speed, always-on data economy, understanding how to best engage with your customers means understanding your data and how to transform it into actionable insights. eBay, the leading online retail and auction site, has proven that this is now the rule rather than the exception.

The key to eBay’s extraordinary success is its ability to turn enormous volumes of data into useful insights that customers can glean directly from the pages they frequent. To accommodate this explosive data growth, eBay’s data centers perform billions of reads and writes each day. This increasing demand to process data at blistering speeds meant eBay needed a solution that did not suffer from the bottlenecks, scalability limitations and transactional constraints associated with common relational database approaches. The company also needed to perform rapid analysis on a broad assortment of the structured and unstructured data it captured.

The Solution
The need to effectively manage massive volumes of dynamic data caused eBay to explore newer technologies designed for modern application requirements, specifically Apache Cassandra™ and DataStax Enterprise. Along with Cassandra and its high-velocity data capabilities, eBay was also drawn to the integrated real-time analytics bundled with DataStax Enterprise. The solution incorporates a scale out architecture that allows eBay to deploy multiple DataStax Enterprise clusters across several data centers using commodity hardware. The end result allows eBay to more cost-effectively process massive amounts of data at very high velocities, and achieve far more than they were able to with the higher-cost proprietary system they had been using. Currently, eBay is managing a sizable portion of its data center needs—250TBs+ of storage—in DataStax Enterprise clusters.

eBay leverages user activity data stored in Cassandra to power their recommendation engine.
"We have to be ready for disaster recovery all the time. It's great that Cassandra allows for active-active multiple data centers where we can read and write data anywhere."

Technical Architect
eBay

The Results
The ability to capture and interpret customer viewing and buying preferences gives eBay the unique ability to deliver an optimal user experience and drive purchases.

To achieve this level of insight, eBay relies on DataStax Enterprise to store user activity data and look up historical data quickly, so that they can model updates and deliver personalized recommendations in the fastest possible manner. In fact, DataStax allows eBay to handle huge amounts of write traffic more efficiently than any other RDBMS or NoSQL solution.

Currently, eBay handles over 6 billion writes per day across multiple DataStax Enterprise clusters and over 5 billion reads per day as well.

An example of how DataStax powers eBay's personalized experience is with its social features. DataStax Enterprise captures the social data eBay displays on each of its product pages and stores all the information needed to provide counts for 'like', 'own', and 'want' data on eBay product pages. It also provides the same data for the eBay "Your Favorites" page that contains all the items a user likes, owns, or wants, with Cassandra serving up the entire "Your Favorites" page. This data is loaded and stored in one of eBay's geographically dispersed clusters.

Load balancing and application availability are critical when it comes to managing high volumes of multi-structured data. DataStax Enterprise gives eBay architects the flexibility they need to design a system that enables any user request to go to any data center, with each data center having a single DataStax Enterprise cluster spanning those centers. This design feature helps balance the incoming user load and eliminates any possible threat to application downtime.

"We are able to analyze vast amounts of data in the same environment in real time, without needing to move it to a separate system for analysis," said eBay's technical architect. "This greatly accelerates the analysis process and eliminates the risk of losing data in transit."

eBay also uses DataStax Enterprise for many time-series use cases in which processing high-volume, real-time data is a foremost priority. These include mobile notification logging and tracking (every time eBay sends a notification to a mobile phone or device it is logged in Cassandra), fraud detection, SOA request/response payload logging, and RedLaser (another eBay sister company) server logs and analytics.

eBay is acutely aware of the need to keep their business up and open for business, and DataStax Enterprise plays a key part in that through its support of high availability clusters and no single point of failure.

About DataStax
DataStax is the fastest, most scalable distributed database technology, delivering Apache Cassandra to the world's most innovative enterprises. Datastax is built to be agile, always-on, and predictably scalable to any size.

With more than 500 customers in 45 countries, DataStax is the database technology and transactional backbone of choice for the worlds most innovative companies such as Netflix, Adobe, Intuit, and eBay. Based in Santa Clara, Calif., DataStax is backed by industry-leading investors including Lightspeed Venture Partners, Meritech Capital, and Crosslink Capital. For more information, visit DataStax.com or follow us @DataStax.

© 2014 Datastax, All Rights Reserved.