

Oct 4, 2016, 08:00am EDT

# How eBay Uses Big Data and Machine Learning to Drive Business Value



**Maribel Lopez** Senior Contributor ⓘ

[Enterprise Tech](#)

*I help firms understand AI, mobile and cloud to improve their business*

⌚ This article is more than 4 years old.

## TWEET THIS



eBay is managing approximately 1 billion live listings and 164 million active buyers daily.



eBay receives 10 million new listings via mobile every week

Digital transformation, while not new, has changed tremendously with the advent of new technologies for big data analytics and machine learning. The key to most company's digital transformation efforts is to harness insights from various types of data at the right time. Fortunately, organizations now have access to a wide range of solutions to accomplish this goal.

How are leaders in the space approaching the problem today? I recently had a discussion with Seshu Adunuthula, Senior Director of Analytics Infrastructure at eBay, to discuss this matter. eBay was always a digital

1 of 4 free articles

**Introducing the Forbes.com subscription.** Unlimited content. Improved ad experien

Data is eBay's most important asset. eBay is managing approximately 1 billion live listings and 164 million active buyers daily. 🐦 Of these, eBay receives 10 million new listings via mobile every week 🐦 . Clearly, the company has large volumes of data, but the key to its future success will be how fast it can turn data into a personalized experience that drives sales.

## Designing and updating a technical strategy

The first challenge eBay wrestled with was finding a platform, aside from its traditional data warehouse, that was capable of storing an enormous amount of data that varied by type. Adunuthula stated that the type of data, the structure of the data and the required speed of analysis meant the company had to evolve from a traditional data warehouse structure to what it calls data lakes. For example, the company needs to keep roughly nine quarters of historical trends data to provide insights on items such as year over year growth. It also needs to analyze data in real-time to assist shoppers throughout the selling cycle.

The ability to support data at the scale of an internet company was a key consideration in the selection of technologies and partners. The company chose to work with Hortonwork's Hadoop product because it offered an open source platform that was highly scalable and the vendor was willing to work with eBay to design product enhancements. With a foundation of Hadoop and Hortonworks, the other two components of eBay's data platform strategy are what it calls streams and services.

A big technical challenge for eBay and every data-intensive business is to deploy a system that can rapidly analyze and act on data as it arrives into

the organization's systems (called streaming data). There are many rapidly evolving methods to support streaming data analysis. eBay is currently working with several tools including Apache Spark, Storm, Kafka, and Hortonworks HDF. The data services layer of its strategy provides functions that enable a company to access and query data. It allows the company's data analysts to search information tags that have been associated with the data (called metadata) and makes it consumable to as many people as possible with the right level of security and permissions (called data governance). It's also using an interactive query engine on Hadoop called Presto. The company has been at the forefront of using big data solutions and actively contributes its knowledge back to the open source community.

eBay's current big data strategy represents a few of the potential combinations and options that are available to companies seeking to process a large volume of data that aren't similar in format and combinations of data that may need to be analyzed in real-time or stored for analysis at a later date. Of course, the selection of big data solutions depends on what you are trying to accomplish as a business.

### **Using a big data and machine learning platform to deliver business value**

In the case of eBay, the company is using big data and machine learning solutions to address use cases such as personalization, merchandising and A/B testing for new features to improve the user's experience. For example, eBay models personalization on five quarters of structured (e.g. one billion listings, purchases, etc.) and unstructured (behavioral activity synopsis, word clouds, badges etc.) data. Merchandising improved by

using analytics and machine learning to help recommend similar items on key placements on site and mobile. Items, such as deal discovery, uses machine learning to find patterns in structured data. eBay's also creating predictive machine learning models for fraud detection, account take-over, and enabling buyer/seller risk prediction. Clearly, eBay has spent enormous time and resources attaining this level of expertise in data processing and business workflow enhancement. For eBay and many others, the journey is far from over. The company wants to continue to optimize streaming analytics and enhance data governance.

### **What should you do next?**

For those companies that are getting started, Adunuthula offered a few words of sage advice. The biggest challenge is data governance and preventing it from becoming the wild west. A business can't just dump everything into a system and worry about the governance later. If you're building a data strategy today, start with the governance.

Examples of this could include defining the process for allowing access to different people and how to enable PCI compliance in the data sets for retailers. The strategy should outline how to make data discoverable and how to evolve the process. He noted that there are new solutions, such as Atlas and Navigator, emerging today. However, the landscape continually changes. If you are starting the journey today, a business can put data governance in place before building massive datasets, data warehouses, and data lakes. It's easier to add data governance at the beginning of the process.

From discussions with my clients, I've learned there are several important steps in building a big data strategy that includes:

- **Defining a quick win and a longer term use case.** Building a tightly scoped use case is essential for acquiring funding and demonstrating immediate value from your data strategy efforts. For example, many companies define a use case that involves connecting and analyzing new data sources to understand buying behaviors. Selecting a narrow use case allows data analysts to test new technologies and deliver new insights to the business.
- **Evaluating what you need in a data partner.** eBay has a sophisticated engineering team and knows what it was trying to achieve. The company was looking for a partner to help deliver scale and assistance in improving open source solutions. A company might also need their partner to provide more training, consulting services and reference architectures based on industry.
- **Building the right ecosystem.** There isn't one data storage and analytics solution that will solve all of a company's use cases. In some areas, a company's existing data warehouse solutions work perfectly. In other cases, you'll need streaming analytics. Similarly, there isn't a single tool or vendor that will provide everything you need. Today's data analysis world requires an ecosystem of tools and partners. Look for partnerships between vendors that will ease integration challenges.
- **Looking for new use cases.** Instead of replicating what you have, a business should look for ways that new data can be acquired and analyzed to improve your business processes. Part of the benefit of these new data and analytics tools is discovering patterns, anomalies and new insights that didn't exist in your

legacy data analysis system. Business leaders should work with IT to look for ways that new data storage and analytics solutions can answer questions that weren't easy to answer in the past.

*Follow me on [Twitter](#) or [LinkedIn](#). Check out my [website](#).*



**Maribel Lopez**

Follow

I'm a technology industry analyst and strategic advisor at Lopez Research. I'm the author of "Right-Time Experiences: Driving Revenue with Mobile and Big Data" published...

**Read More**

Reprints & Permissions

ADVERTISEMENT

---