

How big data is used in retail industry: etsy

[Technology](#), [Computer](#)



Big Data can be defined as the inability of traditional data architectures to efficiently handle the new datasets. Characteristics of Big Data that force new architectures are:

1. Volume: Size of dataset
2. Variety: Data from multiple sources
3. Velocity: Rate of flow
4. Variability: Change in other characteristics

We all are aware of the 4 v's and they are considered as the core characteristics of Big Data. NIST perfectly explains these characteristics and further puts emphasis on needing a higher system architecture for higher performance. The system scaling is done using techniques called vertical and horizontal scaling. While researching definition of big data I realized most of the websites talk about the V's of big data but none talk about how this is a full field but NIST explains us in detail about scaling and what it does. One thing to note is that this article was written in September of 2015 and since then a lot of things have changed and like any technology, big data is evolving and its definition and characteristics are evolving too. The four v's has evolved and are now there are seven v's which include [5] -

5. Veracity: Refers to the completeness and accuracy of the data
6. Value: How much value does the data has
7. Visualization: The most important part where the processed data is presented so that readers can understand it.

<https://assignbuster.com/how-big-data-is-used-in-retail-industry-etsy/>

Big data has become a common name and it is being used in e-commerce for numerous ways. Big data is being used in e-commerce to give personalized buying experience to the customers. Using a customer's browsing and buying habits to provide them with personalized recommendations can result in increased sales. We all aware of Amazon who provides their customers with " Customers who bought this item also bought" section which resulted in 30 percent increase in sales. Big data along with click-stream data can be used to monitor prices of products in real time and adjust the prices accordingly. Amazon uses different tools to monitor and adjust pricing of their own products accordingly. By this, they are making sure the customer gets the best price and no other competitor beats them with a lower price. Big data can also be used to send personalized offers which can be in form of emails or even pop-ups while they are trying to abandoned the cart.

Etsy is an online e-commerce website which is a platform for selling handmade and vintage items. Most of the items they sell are handmade and made by individuals like you and me. Etsy was created back in 2005 in Brooklyn apartment by Rob Kalin, Chris Maguire and Haim Schoppik in their Brooklyn apartment [Citation]. Within two years Etsy had nearly 450, 000 registered users and generated \$26 million in annual sales. After that Etsy went through many changes in their structure and two of the creators left the company and then the hired Chad Dickerson, senior director of product at Yahoo to lead the company. Dickerson was hired as the chief technology officer and he took the company in an upward direction and was given the position of CEO later on. In 2013 Dickerson proposed tweaking its Terms of

Service and allowed selling of manufactured goods which were not taken positively by other sellers but in the end, it helped the company and boosted their sales and revenue. According to VentureBeat in 2013 Etsy sales grew from \$895 million to \$1.34 billion. And in 2014 the numbers went up 43 percent to a total of \$1.93 billion. In 2015 Etsy had more than 1.5 million active sellers and debuted its IPO.

Chart is taken from - Marketingcharts.com

In 2014 Etsy was number eight on top ten shopping websites and this shows us how big and how much market they share and had about 22 million buyers [citation].

According to Chris Bohn a Senior Data Engineer at Etsy and according to him they want to use big data to understand more about their customers which include both sellers and buyers. They would like to provide a rich and smooth experience and their end goal is buyers find their products easier and sellers be able to reach the right buyers. According to Bohn, they want to use Big Data, "To know how people are different in their shopping habits across the geography of the world."

Let's first start with what type of data architecture Etsy used to use and what changes they made for been able to change with the time and to accommodate big data analysis. Initially, Etsy used to use monolithic Postgres database which consisted of listings, users, sellers, buyers, conversations and forums. As the company grew and their user database grew so they had to shard horizontally. The front end was driven by PHP.

<https://assignbuster.com/how-big-data-is-used-in-retail-industry-etsy/>

Ross Snyder, a senior software engineer said: " The site's uptime was not that great and regular maintenance windows and site deploys often dissolved into outages." This all lead to Etsy creating a middleware which would help with scaling the website performance and at the same time the middleware would decrease the number of SQL calls. Etsy named this middleware Sprouter which they planned on making open-source and using it for a long time but after using it for while they decided to abandon it as it required DBAs to write stored procedures for nearly every piece of site functionality" and created a bureaucracy developers had to go through to get functionality made. It was never open-sourced and was rest to death. Then they moved from Postgres to sharded MySQL databases. According to Synder the reason they used MySQL at that time was " Flickr is using it on an enormous scale. It scales horizontally, basically, to near infinity, and there's no single point of failure-it's all master to master replication."

During this process, Etsy decided to do some analytics and copied the data from SQL back to a Postgres server which they called a BI server but what they did not realize is that they went back to the original thing they wanted to go away from and it was all back to zero. They also realized that Postgres is not the best option for performing analytic queries and it was really hard to get a huge amount of data into the database. Here Etsy unknowingly faced the Volume characteristic of the big data. They again started their hunt for finding an appropriate replacement and came across HP Enterprise Vertica. One of the first reason they selected HPE Vertica is because it derives from Postgres and it has a Berkeley license which can enable Etsy to take it private and make changes to the code accordingly and do not have to

republish it to the community. Using hp Vertica boosted the efficiency of their queries by running them 50x-1000x faster. With Vertica, there is unlimited scalability and this is a great feature for Etsy as they were a growing company. Vertica also stores 10x to 30x more data per server and also has compression. Etsy at start faced a problem with outages mentioned before and Vertica can guarantee maximum uptime and eliminate failures. It is also an open architecture with support for Hadoop, R and other range of BI tools. Etsy used a data replication tool to copy the data over to Vertica used Vertica's open architecture feature to build their internal tools for doing analytics on the data. The big data problem here was to prepare their database to be able to do some analytic work and use the data they have been collecting. With Vertica, Etsy was able to quickly and efficiently analyze 30 terabytes of data. Bohn says that the greatest benefits are accessibility and speed and that use of the tool has spread to all departments. The fact that queries that previously took many days to run, now run in minutes, provides a dramatic example of the level of increased productivity gained company-wide. Apart from all the amazon functionality, Vertica was able to save Etsy \$80, 000 a month by switching from amazon cloud. By using Vertica Etsy did not have to hire any new people as Vertica uses a lot of similarity with Postgres and their developers already had experience with it.

Velocity -It is the measure of how fast the data coming in and for Etsy, it been top ten websites for shopping it was generating a massive number of clicks which needed to be stored. Etsy wanted to use this data to figure out where their customers are clicking and at what point they are leaving the website. Etsy took it to next step and used this clickstream data and joined it

with their data to find details about the customer and what their buying history was in the past. This is how Etsy can get some value out of the clickstream data as it is the just path of click what a consumer goes through. The second type of data that Etsy has is the transactional data which includes order values, the category of sales, purchase frequency, the amount paid and shipping preferences.

Variety -Etsy had clickstream data, data about sellers, buyers, forums, messages and lot of different types and before they used to use Postgres which is not ideal for handling the variety of data. But with Vertica, they now can store any variety of data.

Volume -When we talk about volume we are talking about an insanely large amount of data and in our case, Etsy already had 30 TB of data which needed to be transferred and stored.

By entering the world of big data, the employees were able to do much more than before with very little time. The first result of this whole change was all employees started using it as getting results was much better and way faster than a traditional database. It was not that they were getting different results but the time from entering the query and getting the output was shortened. By adding Vertica Etsy was then able to get information in real time right away. This feature was used by them when they introduced postage on their website where sellers can use postage service provided by Etsy to ship their products. The engineers wanted to keep an eye on this feature and know how it is performing in real time which was made possible using Vertica. All the departments were able to use this functionality and the

people from finance department said, “ Wow, I can run these financial reports that used to take days in literally seconds.”. Within very short time Etsy had 200 Vertica accounts and had a total of 750 employees which shows us how much popular this new change was.

One of the surprises which Etsy faced according to Chris Bohn is that when they installed Vertica they thought only their analysts would use but it was so easy and so popular that they had to buy more licenses for their users. Vertica was being used for many other ways such as for their internal dashboards, running financial reports and for testing too.

The result of this project resulted in total revenue of \$119 million for the first half of 2015, up 44% on the same period in 2014. The number of active buyers grew to 21.7 million and the number of active users grew to 1.5 million. Etsy is the go-to place for unique products and gifts. None of this would have been possible without their keen embracing of Big Data and analytics.

This project was successful as not only it led to increase in revenue but it also led to change in company’s culture. Generally, the change in culture leads to change in technology but at Etsy, the technology changed the way people did their job. According to Bohn “ This is technology that has driven the culture. It’s really changed the way people do their job at Etsy. It really has been impactful.”

After this project, Etsy realized that they were spending too much on AWS and they can save that money by buying their own servers. Bohn said “ Wait

a minute. This is crazy. We could actually buy our own servers. This is commodity hardware that this can run on, and we can run this in our own data center. We will get the data in faster because there are bigger pipes.” That’s what Etsy did by creating Estydoop which has 200 plus nodes and they ended up saving a lot of money and it would not have been possible if they did not do the big data project. Another thing which Etsy realized is that the market was changing and now smartphones were becoming common and been used for e-commerce. Etsy was able to use big data to figure out what each customer on their smartphone doing on their website and used that data to find crossover points and to change things accordingly. By this way, Etsy was moving along with the developing technology and not been left behind.

Citation

[1] “ 5 Benefits of Big Data for E-Commerce Companies and Shoppers.” SmartData Collective. Accessed March 25, 2017. <http://www.smartdatacollective.com/seanmallonbizdaquk/410001/5-benefits-big-data-e-commerce-companies-and-shoppers>.

[2] “ A brief history of Etsy, from 2005 Brooklyn launch to 2015 IPO.” VentureBeat. March 05, 2015. Accessed March 25, 2017. <http://venturebeat.com/2015/03/05/a-brief-history-of-etsy-from-2005-brooklyn-launch-to-2015-ipo/>.

[3] Akter, Shahriar, and Samuel Fosso Wamba. “ Big data analytics in E-commerce: a systematic review and agenda for future research.”

<https://assignbuster.com/how-big-data-is-used-in-retail-industry-etsy/>

SpringerLink. March 16, 2016. Accessed March 25, 2017. <https://link.springer.com/article/10.1007/s12525-016-0219-0>.

[4] April 8, 2014 – by MarketingCharts staff. “ Top 10 Shopping and Classifieds Websites #8211; March 2014.” MarketingCharts. April 08, 2014. Accessed March 25, 2017. <http://www.marketingcharts.com/updates/top-10-shopping-and-classifieds-websites-march-2014-41856/>.

[5] “ Big Data’s Role in Etsy’s Product Development.” InfoQ. Accessed March 25, 2017. <https://www.infoq.com/interviews/big-data-etsy-product-development>.

[6] “ How Etsy Uses Big Data for eCommerce to Put Buyers and Sellers in the Best Light.” BriefingsDirect Transcripts. Accessed March 25, 2017. <http://www.briefingsdirecttranscriptsblogs.com/2016/04/how-etsy-uses-big-data-for-ecommerce-to.html>.

[7] Sean Gallagher – Oct 3, 2011 1: 59 pm UTC. “ When “ clever” goes wrong: how Etsy overcame poor architectural choices.” Ars Technica. October 03, 2011.

[8] Accessed March 25, 2017. <https://arstechnica.com/business/2011/10/when-clever-goes-wrong-how-etsy-overcame-poor-architectural-choices/>.

[9] “ The biggest surprise Etsy encountered when applying HP Vertica to search | #HPBigData2014.” SiliconANGLE. August 13, 2014. Accessed March

<https://assignbuster.com/how-big-data-is-used-in-retail-industry-etsy/>

25, 2017. [10] <http://siliconangle.com/blog/2014/08/13/the-biggest-surprise-etsy-encountered-when-applying-hp-vertica-to-search-hpbigdata2014/>.

[11] "The biggest surprise Etsy encountered when applying HP Vertica to search | #HPBigData2014." SiliconANGLE. August 13, 2014. Accessed March 25, 2017. <http://siliconangle.com/blog/2014/08/13/the-biggest-surprise-etsy-encountered-when-applying-hp-vertica-to-search-hpbigdata2014/>.

[12] "Vertica Advanced Analytics Offerings." Vertica Big Data SQL Analytics Platform Free Software & Trials | Hewlett Packard Enterprise. Accessed March 25, 2017. <http://www8.hp.com/us/en/software-solutions/advanced-sql-big-data-analytics/try-now.html>.

[13] "What does Etsy's architecture look like today? - High Scalability -." High Scalability. Accessed March 25, 2017. <http://highscalability.com/blog/2016/3/23/what-does-etsys-architecture-look-like-today.html>.