

Big Data Technologies in E-Commerce

¹Vaibhav Devidas Chaudhari, ²Dipesh Dattaram Khatate, ³Prof. Nikhita Mangaonkar

^{1,2}Student, ³Assistant Professor
Sardar Patel Institute of Technology, Mumbai, India

Abstract— Hadoop is one of the best technology to manage data in huge quantity and it uses multiple different sub technologies into it like HDFS and Map-Reducing. Map Reducing uses key value pair and manages data and HDFS is Hadoop Based File System which is helpful to manage and organize data while saving it. It is also one of the helpful technology to be used by the E-commerce website now-a-days to save time and manage data in effective manner.

I. INTRODUCTION

The Existence of Big Data came in consideration when large amounts of relational data was difficult to handle and it was impossible to manage data from unstructured format.

Unstructured and large amount of data like videos, audios, social updates, photos etc which are diverse in their variety can be termed as Big Data.

Hadoop, Map Reduce, Apache Hive, No SQL and HPC are the technologies used by big data application to deal with the massive data.

The space managed by these technologies handle massive amounts of data in PB, YB, ZB and TB.

E-commerce is a growing sector generating huge amounts of data which needs to be managed properly for the purpose of CRM's.

To solve this problem Big Data came into picture and was a boon to E-commerce sector.

The mapReduce and HDFS of Hadoop found to be being helpful towards e-commerce in order to manage data in effective manner and to find out the results quickly.

II. RESEARCH METHODOLOGY AND ALGORITHM

Big data allows E-commerce companies to:

1. COMPETE ON VALUE RATHER THAN PRICE.

Amazon is the e-commerce standard when it comes to smart, effective pricing. It can easily access its competitors' pricing data and respond quickly with its own deals — changing some items' prices up to 10 times a day. The industry-wide shift to dynamic pricing means that companies will no longer be competing on price alone. They will now need to establish a reputation for offering their customers the best value and the best experience.

2. PERSONALIZE EVERY INTERACTION.

Amazon is also the leader when it comes to using big data to recommend products its customers may like. Amazon's sales increased nearly 30% when it was first implemented. This is a simple and incredibly effective way to keep customers on a retail site and keep them buying.

My company sells custom menswear online, and we're building a database of men's body measurements and style

preferences. In the future, when a customer provides his height and weight, we'll be able to generate a standard measurement that's a perfect fit without ever having to measure him. The helps to remove the extra work for the customer and making it easy to buy

3. PREDICT TRENDS.

Recording our customers' measurements and preferences doesn't just help us tailor the shopping experience to each individual. We can also analyze the most popular fabrics and cuts so we know what to keep in stock. This allows us to negotiate favorable rates with suppliers and tailors.

Today, companies are using social media and their own resources to track trends on what and how consumers are buying. Record label EMI, for instance, pulls data from third-party sites and apps to understand listening and buying patterns to predict its product demand and target its advertising.

4. REDUCE SHOPPING CART ABANDONMENT.

Companies can also use cross-device tracking to reduce shopping cart abandonment rates. EBay research found that the average consumer uses as many as three or five devices or platforms during the course of her buying journey. Mapping this journey with data allows retailers to help customers transition from one device to the next and complete their purchases.

5. IMPROVE THE CUSTOMER EXPERIENCE.

Consumers might have reservations about their favorite retailers knowing intimate details about their lives, but they're going to love the results in practice. Sharing all those personal tidbits is helping companies like CNA identify fraud and prevent customers from having their identities compromised. Retailers can use information from live transactions and other sources (such as social feeds and geo data from apps) to prevent credit card fraud in real time.

While it might seem eerie for Target to know about major life events before your family does, the applications of big data are, by and large, major wins for both retailers and their customers. Competing on price alone will soon be a thing of the past as companies rely on data to create more seamless, personalized shopping experiences.

MapReduce:

Google introduced Map Reduce in order to process and store huge data on product hardware. Large-scale data records in clusters are processed using Map Reduce processing model. The map() function and reduce() function are the two functions contained in the Map Reduce programming model. Users can simulate their own processing logics having well defined map() and reduce() functions.

The master node takes the input in the Map function which performs the task, divide into smaller sub modules and distribute into slave nodes. The hierarchical tree structure is created when the slave node divides the sub modules again. The result is then passed to the master Node after the slave node processes the base problem. The system (Map Reduce) arranges together all intermediate pairs based on the keys (intermediate) and refer them to reduce() function for producing the final output. The master node collects the results from all the sub problems and combines them together to form the output. Here Reduce function is termed as the master node.

Map(in_key, in_value)---

>list(out_key, intermediate_value) Reduce(out_key, list(intermediate_value))---

>list(out_value)

The parameters of map () and reduce () function is as follows:

map (k1, v1) ! list (k2, v2) and reduce (k2, list(v2)) ! list (v2)

A Map Reduce is a framework that represents a master and slave architecture where one master node handles a number of slave nodes. The input data set is divided into even-size data blocks for equal load distribution. The result is generated when each data block is assigned to one slave node and is processed by a map function. The master node is interrupted by slave node when it is idle. The slave node is assigned to new tasks by the scheduler. The scheduler accepts data resources and locality into consideration when it disseminates data blocks.

The advantage of map reduce:

As Map reduce cluster of machines and computations handle thousands of nodes and fault-tolerance large variety of problems are easily expressible.

The disadvantage of map reduce:

Processing real time, not always very easy to implement, batch processing, shuffling of data.

Map Reduce Components:

1. Data Node: stores blocks of HDFS—default replication level for each block: 3.
2. Name Node: Doesn't deal with files directly, manages HDFS meta data.
3. Job Tracker: Allocates and monitors job execution on slaves, schedules—Task Trackers.
4. Task Tracker: Is useful run Map Reduce operations.

MapReduces Job Tend to be very small as far as code is consider IdentityReducer is very common "Utility" jobs can be composed Represent a data flow, more than a procedure Sort which Inputs set of files, one value per line.

Mapper key is file name, line number Mapper value is one of the part of the line Sort Algorithm which Takes advantage of reducer properties:

(key, value) pairs are processed in order by key; reducers are themselves ordered

Mapper: Identity function for value

(k, v)

(v, _)

Reducer: Identity function (k', _) -> (k', "")

There are three main steps MapReduce Algorithm uses:

Shuffle Function

Reduce Function

Map Function

Map Function

It is the first step in MapReduce Algorithm. It takes input functions (say Data Sets) and divides them into smaller sub-functions. Then it performs required computation on each sub-task in parallel.

Following two sub-steps are performed by this step:

Mapping

Splitting

Mapping step takes those smaller Sub Data Sets and perform required action on each Sub Data Set.

Map Function output is the set of key and value pairs as <Key, Value>.

Splitting step takes input Data Set from Source and divide into smaller Sub Data Sets.

Shuffle Function

This is the second step in MapReduce Algorithm. This Function is also know as "Combine Function".

It performs the following two sub-steps:

Sorting

Merging

The outputs coming from "Map Function" are considered and then these two sub-steps are performed on each and every key-value pair.

Merging combines all key-value pairs which have same keys (grouping key-value pairs by comparing "Key"). This step returns <Key, List<Value>>.

Sorting takes input from Merging step and then sort all key-value pairs by using Keys. This step also returns <Key, List<Value>> output but with sorted key-value pairs.

Reduce Function

This is the last step in Map-Reduce Algorithm.

It performs only one action and that is Reduce step.

It takes list <Key, List<Value>> as list of sorted pairs from Shuffle Function to perform reduce operation.

reduce-function

MapReduce Final Step Outcome:

reduce-function-output

Final step that is output which looks like first step outcome. However

first step <Key, Value> pairs are different than final step <Key, Value> pairs. Final step <Key, Value> pairs are sorted and computed pairs.

Result Analysis

1. Personalization

Collection of more data from customers through various touch points such as visitor browsing patterns, loyalty programs and past purchase behavior, companies process this information to carry out customer segmentation and thus push out promotions and personalized content.

2. Dynamic pricing/offers

A comparatively perfect profile can be generated for each and every customer after collecting and processing data. This profile will thus offer details as to what price would re-engage the customer and influence him to make another purchase. Note that this has proven the most effective way for customer retention strategies. This would also help accelerate the process of concluding whether a \$10 off or 20% discount would work the finest on any particular customer.

3. Customer service

Because of poor customer service 68% of online visitors leave, so improving this aspect is critical in increasing sales. Big Data helps in coordinating various communication channels, including emails, phone calls, or live chat features. Considering the issues in which customers are concerned with, or classifying time periods where customer support is most needed can also help companies best to allocate resources more efficiently. Customer conflict can thus be solved more effectively within a short period of time.

4. Supply chain management

Any E-Commerce company knows that for careful management and transparent visibility of its supply chain process are important. Big Data allows for retailers to decide patterns which can be useful to forecast any potential hiccups and interruption to the process, and thus quickly act on protection measures. For example, if any changes in shipping or warehousing updates can be taken as real-time and communicated to the retailer immediately.

5. Predictive Analytics:

Big Data allows you to extract a more in-depth knowledge of the different channels in your business, including inventory, and sales. The added flexibility is given by Knowing how to forecast. Deciding the next steps to be taken of your business operations.

III. RESULT

As all E-commerce web sites deals with the huge amount of data so it becomes necessary for them to manage it in effective manner in order to get result in quick response time and it also

becomes important to manage and store only that data which is necessary to store in this scenario. Map Reducing helps a lot to the e commerce and also manages huge amounts data in effective manner.

Also HDFS(Hadoop Based File System) helps to manage and organize data to store it in effective manner so it becomes a helpful tool to the

E-commerce websites as far as huge data is concern, they can rely on the big data technologies like map reducing and HDFS to store the data in effective manner and also give the output in quick response time so time saving can be done and also vendor can focus on his main business rather than managing data and other things.

IV. CONCLUSION

Big data will influence e-commerce strategies in sales and marketing, the deep study of relevant data will become more important.

Big data offers solutions for a modifications of online stores, a real-time portfolio optimization and individualized customer support

The MapReduce model is successfully adopted and used for many different purposes in E-Commerce . the model is easy to use without experience of distributed and parallel systems, since it hides the details of parallelization, locality optimization, fault tolerance and load balancing. It also a large variety of problems are easily expressible as MapReduce computations.

So, through this study we can say that each E-commerce website must use Big Data technologies like Map Reducing and HDFS which can help them in managing data in effective manner so they can primarily focus on business rather than doing secondary tasks like managing data which can ultimately provide effective functioning of business with good throughput.

REFERENCES

- [1] http://static.usenix.org/events/osdi04/tech/full_papers/dean/dean_html/index.html
- [2] <http://www.bigdatalandscape.com/news/why-big-data-is-a-must-in-ecommerce>.
- [3] https://www.researchgate.net/publication/264129339_Big_Data_Analytics_and_its_Application_in_E-Commerce.
- [4] Aditya B. Patel, Manashvi Birla, Ushma Nair, "Addressing Big Data Problem Using Hadoop and Map Reduce".