

Matt Hyatt

Jan 24

Assignment 1 - COMP 358

Big Data Platform: Amazon retail

Business Case Evaluation

One motive for Amazon's use of Big Data Analytics is to improve the user experience. Some potential benefits include better product search recommendations, better choice of delivery options, and better feedback for Amazon sellers. User experience in general has an ongoing scope, however analysis decisions/outcomes need to be instantaneously available in order to be applicable to the user. Parameters can include user search history, user ratings, price information, Amazon membership information, browser cookies, order history, timestamps, time spent on a certain page/screen frame, delivery locations, timestamps, and product return history.

Identification of Data

Sources of data will most likely come directly from the user, indirectly from the user's browser, Amazon delivery vehicles, and from third part sellers.

Data Filtering

Data will be cleansed and preprocessed for analysis. Missing fields can be substituted for the mean or most frequent value for that field. Catagorical data can be normalized with one hot encoding, and textual data like product names and user search history can be tokenized. Scalar values can be normalized with a max-abs scaler or fit on a normal distribution.

Data Extraction

Amazon can utilize data lakes and the flexible nature of noSQL databases to ensure that data will be available to retrieve in batches and analyze; regardless of the type or shape of the data.

Data Aggregation

Data from different sources or tables are aggregated In order to aggregate data from different sources, Amazon can utilize the Hadoop Ecosystem and HDFS management. They will likely use a cloud-based distribution of Hadoop in order to use their own in-house AWS analytics frameworks.

Data Analysis and Visualization

Amazon will likely use Apache R and Mahout software to analyze data in batches and distribute some of the results to third party sellers through Apache Sqoop. Data can be visualized in heat-maps in order to show variable correlations or in histograms to show the frequency of occurrences with respect to other variables like time, location or product.

Analysis Outcome

After analyzing the data Amazon will be more equipped to suggest more relevant products to their users and provide better feedback to their suppliers / third party sellers. They will be able to optimize the routes of their delivery services, enhance the user experience of the Amazon website, and more intelligently price their products to take advantage of opportunities in the economy.