

4.6: Big Data

Definition: Big Data

Big data refers to large, complex datasets typically involving : a high **volume**, **velocity**, and **variety** of data that require cost-effective, innovative forms of information processing for enhanced insight and decision making.

Big Data refers to capturing large complex data sets that conventional database tools do not have the processing power to analyze. Storing and analyzing that much data is beyond the power of traditional database management tools. Understanding the best tools and techniques to manage and analyze these large data sets is a problem that governments and businesses alike are trying to solve. Big data comes from many different areas such as text, images, audio, and videos. Businesses use this data and refer to it as predictive analytics or user behavior analytics. Companies such as Walmart and Amazon are now collecting big data, to see what searches their customers are looking at. Think of the number of customers and products these two powerhouses have and the amount of data generated.

✓ Use Case Big Data 4.6.1

Amazon wants to collect and analyze massive amounts of data on customers, from different sources and formats to derive user analytics to help improve the customers' experience. The data includes customers' purchases, browsing history, reviews, other activity on the site. The data sets can be unstructured such as customer reviews, social media conversations, or chat logs. They can also be semi-structured such as purchase history, web server logs, or product catalogs.

What are the insights gained by Amazon to increase sales, marketing outreach, or efficiency?

Solution

- Analyzing purchase history and browsing patterns to provide personalized product recommendations. The insights will allow Amazon to take actions to convert to higher sales.
- Applying machine learning techniques across customer data to segment users based on behaviors. This enables targeted marketing outreach.
- Tracking reviews and other metrics to identify trends and issues with products. This feedback leads to improvements.
- Processing log data to detect fraud in real-time before orders are fulfilled. This reduces risk.
- Indexing text data like product descriptions and customer service transcripts to efficiently handle text searches. This will improve customers' experience in searching and selecting products.

This page titled [4.6: Big Data](#) is shared under a [CC BY 4.0](#) license and was authored, remixed, and/or curated by [Ly-Huong T. Pham and Tejal Desai-Naik](#) ([Evergreen Valley College](#)) .

- **4.6: Big Data** by Ly-Huong T. Pham, Tejal Desai-Naik, Laurie Hammond, & Wael Abdeljabbar is licensed [CC BY 3.0](#).