BIG DATA

Aparajita, Diksha
I.T.-1, Dronacharya College Of Engg.

***Abstract*- The topic for our research is 'Big Data". 'Big Data' is a buzzword , or a catch-phrase , used to describe a massive volume of both structured and unstructured data that is so large that it becomes difficult to process it with traditional database and software techniques. Actually, it is a collection of data sets so large and complex that it sharing, transfer, analysis and visualization .It has increased the demand of information management specialists in the information technology and software sectors. Big Data requires exceptional technologies to efficiently process large quantities of data within tolerable elapsed time.**

I. INTRODUCTION

Big Data is an all-encompassing term for any collection of data sets so large and complex that it becomes difficult to process using traditional data processing applications.

The challenges include analysis, capture, curation, search, sharing, storage, transfer, visualization, and privacy violations. The trend to larger data sets is due to the additional information derivable from analysis of a single large set of related data, as compared to separate smaller sets with the same total amount of data, allowing correlations to be found to "spot business trends, prevent diseases, combat crime and so on.

II. PRIVATE SECTOR

eBay.com uses two data warehouses at 7.5 pet bytes and 40PB as well as a 40PB Hadoop cluster for search, consumer recommendations, and merchandising.

- Amazon.com handles millions of back-end operations every day, as well as queries from more than half a million third-party sellers. The core technology that keeps Amazon running is Linux-based and as of 2005 they had the world's three largest Linux databases, with capacities of 7.8 TB, 18.5 TB, and 24.7 TB.
- Walmart handles more than 1 million customer transactions every hour, which are imported into databases estimated to contain more than 2.5 pet bytes (2560 terabytes) of data – the equivalent of 167 times the information contained in all the books in the US Library of Congress.
- Facebook handles 50 billion photos from its user base.
- FICO Falcon Credit Card Fraud Detection System protects 2.1 billion active accounts world-wide.
- The volume of business data worldwide, across all companies, doubles every 1.2 years, according to estimates.
- Windermere Real Estate uses anonymous GPS signals from nearly 100 million drivers to help new home buyers determine their typical drive times to and from work throughout various times of the day.

III. QUALITY OF BIG DATA

Increasingly, the quest for quality data will centre on surrounding metadata like that found in data dictionaries and repositories, rather than the data itself. The goal of any big data initiative is to impact business outcomes positively. After determining the vision and strategy, IT leaders need to develop their technology approach. Understanding the differences in available technologies and how they interact is crucial to maximizing big data investment.

IV. WHY BIG DATA MATTERS?

- Determine root causes of failures, issues and defects in near-real time, potentially saving billions of dollars annually.
- Optimize routes for many thousands of package delivery vehicles while they are on the road.
- Analyze millions of SKUs to determine prices that maximize profit and clear inventory.
- Generate retail coupons at the point of sale based on the customer's current and past purchases.
- Send tailored recommendations to mobile devices while customers are in the right area to take advantage of offers.
- Recalculate entire risk portfolios in minutes.
- Quickly identify customers who matter the most.
- Use click stream analysis and data mining to detect fraudulent behavior.

V. BIG DATA IN ACTION

Perspective: UPS

UPS is no stranger to big data, having begun to capture and track a variety of package movements and transactions as early as the 1980s.The company now tracks data on 16.3 million packages per use data and analytics to optimize the efficiency of its 2,000 aircraft flights per day. day for 8.8 million customers, with an average of 39.5 million tracking requests from customers per day. The company stores more than 16 pet bytes of data.

Savings:

The project has already led to savings in 2011 of more than 8.4 million gallons of fuel by cutting 85 million miles off of daily routes. UPS estimates that saving only one daily mile per driver saves the company $30 million, so the overall dollar savings are substantial. The company is also attempting to

VI. BIG DATA IN BIG COMPANIES

Big data may be new for startups and for online firms, but many large firms view it as something they have been wrestling with for a while. Some managers appreciate the innovative nature of big data, but more find it "business as usual" or part of a continuing evolution toward more data. They have been adding new forms of data to their systems and models for many years, and don't see anything revolutionary about big data. Put another way, many were pursuing big data before it was big. For one multinational financial services institution, cost savings is not only a business goal, it's an executive mandate. The bank is historically known for its experimentation with new technologies, but after the financial crisis, it is focused on building its balance sheet and is a bit more conservative with new technologies. The current strategy is to execute well at lower cost, so the bank's big data plans need to fit into that strategy. The bank has several objectives for big data, but the primary one is to exploit "a vast increase in computing power on dollar-for-dollar basis.

VII. OBJECTIVES OF BIG DATA

1. Used for time reduction.
2. Used as Uninterrupted Power Service.
3. Developes new data-based offerings.
4. For internal business decision.

VIII. SUMMARY

It goes almost without saying that the skills, processes, and tools necessary to manage exploding amounts of non-standard data will become ever more scarce and important. For the most part, the companies we interviewed feel substantially less urgency than the startups we have encountered with regard to data science talent. For some, however, the talent shortage is beginning to bite. Many companies continue to rely on incumbent data warehouses for standard BI and analytics reporting, including regional sales reports, customer dashboards, or credit risk history. In this new environment, the data warehouse can continue with its standard workload, using data from legacy operational systems and storing historical data to provision traditional business intelligence and analytics results. But those operational systems can also populate the big data environment when they're needed for computation-rich processing or for raw data exploration. A company can steer the workload to the right platform based on what that platform was designed to do.