THE ROLE OF BIG DATA IN MECHANICAL ENGINEERING DESIGN

[1]Kolluru Venkata Nagendra, [2]Spurgeon Pynam , [3]Bharath Potlapati
[1]Assistant Professor, [2]UG-Student, [3]UG-Student
[1]Computer Science & Engineering,
[1]Geethanjali Institute of Science and Technology, SPSR Nellore,INDIA

Abstract: A large segment of Mechanical Engineering focuses on domains such as product design and development, manufacturing and energy, which are likely to benefit from Big Data. Big Data has deeply reached in different areas of Mechanical Engineering and the most powerful use of big data can be felt in the processing and manufacturing. The role of Big Data in Mechanical Engineering can be felt by the fact that the best quality data are generated by the sensors that are further used in various fields of Mechanical Engineering. In this paper explains the Big data applications and benefits in the design and describes the Hadoop Distributed file system advantages. It deals with the problems in big data and types of daemons in Hadoop.

Index Terms - **Big Data, Design, Machine Learning, Hadoop, Map Reduce, Name node, Data node**

I. INTRODUCTION

Big Data is a term associated with complex and large datasets. A relational database cannot handle big data, and that's why special tools and methods are used to perform operations on a vast collection of data. Big data enables companies to understand their business better and helps them derive meaningful information from the unstructured and raw data collected on a regular basis. Big data also allows the companies to take better business decisions backed by data. The five V's of Big data is as follows:

- **Volume –** Volume represents the volume i.e. amount of data that is growing at a high rate i.e. data volume in Peta bytes
- **Velocity –** Velocity is the rate at which data grows. Social media contributes a major role in the velocity of growing data.
- **Variety –** Variety refers to the different data types i.e. various data formats like text, audios, videos, etc.
- **Veracity –** Veracity refers to the uncertainty of available data. Veracity arises due to the high volume of data that brings incompleteness and inconsistency.
- **Value –**Value refers to turning data into value. By turning accessed big data into values, businesses may generate revenue.

Mechanical Engineers are prepared for the mathematical and scientific principles to design, development and operational evaluation of physical systems. Finally, it deals with the flow of material, such as air and water, and heat transfer. Practical applications, theory, computer simulated courses and evaluations are all methods used to distribute skills. Now mechanical engineers highly rely on Big Data to complete their task efficiently and thus increase the productivity. The below figure 1 shows the five V's of Big Data.

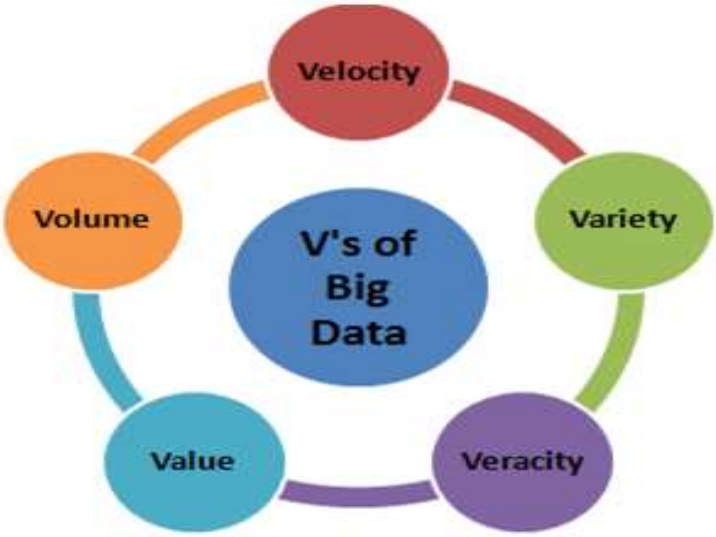

Figure:1 Five V's of Big Data

II. DESIGN AND MANUFACTURING ENGINEERING DATA

Mechanical activities and frameworks frequently create a constant stream of sensor information, occasion information and relevant information through sensors, smart machines and instrumentation. In a factory, data sources possibly include Computer-Aided Design (CAD) models, Computer-Aided Manufacturing (CAM) models, Computer-Aided Engineering (CAE) models, sensors, instruments, Internet transactions and simulations. The information is regularly extensive, quick moving and complex. The information is frequently in substantial assortment, including records, test information, item disappointment information, CAD/CAM/CAE information, unstructured CAD illustrations and details and item and process execution information, and so forth.

III. BENEFITS AND IMPACTS OF BIG DATA IN MECHANICAL ENGINEERING DESIGN

Design and Mining Big Data can offer advantages in outline, for example, better discovery of imperfections in outline and plan change, sparing plan time and costs, quick reaction to advertise, creating imaginative items and coordinating clients' needs and picking up client fulfillment through extricating critical client necessities from the client created information to refresh also, refine existing design manufacturing Engineering Data. The manufacturing sector generates a great deal of text and numerical data in product development processes. Big Data offers the following benefits in manufacturing in Table (1).

Table: 1 BENEFITS IN MANUFACTURING

S.No	Benifits	Working Example
1	Improvements in supply planning	• Unlock significant value and unearth valuable insights by performing Big Data analytics and making information transparent • Better forecast products, production and manufacturing output • Better forecast sales volumes through semantic based Big Data analytics • Improve relationship with suppliers and conduct better contract negotiations according to collected supplier performance data • Improve decision-making and minimizes risks in supply
2	Improved product manufacturing processes	• Provide an infrastructure for transparency in manufacturing • Analyze sensor data from production lines, creating self-regulating processes that cut waste, avoid costly (and sometimes dangerous) human interventions and ultimately lift output • Better monitor and control manufacturing processes by tracking every detail about every part and procedure, better information visibility and BigData analytics for data in motion • Perform predictive manufacturing and optimize manufacturing processes • Better simulate and test new manufacturing processes

3	Driven efficiency across the extended enterprise	• Increase the efficiency of the manufacturing processes • Increase energy efficiency • Enable effective and consistent collaboration through integrating datasets from multiple systems • and divisions • Facilitate innovative design for manufacturing and integration of CAD/CAE/CAM; reduce unnecessary iterations in product development cycles; and finally reduce production and development costs
4	Improved service	• Determine what manufacturing parameters most influence customer satisfaction • Develop new products and make products better match customers' needs through sentiment analysis and recommendation systems for Big Data • Enable mass-customization in manufacturing • Better correlate manufacturing and business performance information together • Reduce warranty costs through warranty analysis based on Big Data analytics • Better perform remote intelligent services
5	Defect tracking and product quality	• Perform predictive diagnostics for product failure • Monitor product data quality • Early detect quality problems • Better detect product defects • Provide real-time alerts based on analyzing • manufacturing data • Reduces defects during manufacturing processes • Boost quality

IV.APPLICATIONS OF BIG DATA IN MECHANICAL ENGINEERING DESIGN

Some manufacturing firms, such as General Electric, view Big Data from sensors in manufactured products (e.g., locomotives, jet engines and gas turbines in GE's case) as key to effective and efficient servicing strategies. In the same mode, automobile manufacturers such as General Motors created self driving cars based on the analysis of Big Data from sensors and machine vision technologies (Davenport, 2013). Big Data has become the key asset for the whole production and manufacturing cycle, as well as the provision of services in the automotive and mobility space. Big Data is actually at heart of how the extracted sensor data and location data are combined to provide services.

Cloud-Based Design Manufacturing (CBDM) is a service-oriented networked product development model. Based on this model, service consumers can configure, select and utilize customized product realization resources and services. CBDM uses the Internet of Things (IoT) (e.g., RFID), smart sensors and wireless devices (e.g., smart phone) to collect real-time design and manufacturing-related data. IoT allows engineers to have access to data such as equipment condition, machine utilization and the percentage of defective products from any location. Engineers can use Big Data analytics for forecasting, automation and proactive maintenance.

A computational method that is based on time series analysis was proposed to assess engineering design processes using a CAD tool. Educational data mining and learning analytics were studied to assess student performance in learning and designing in a project-based setting. Machine learning has been used in Big Data. Massive Parallel-Processing (MPP), distributed file systems and cloud computing, etc. are supporting technologies of Big Data (Zaslavsky *et al*., 2012). Besides general cloud infrastructure services (storage, compute, infrastructure/Virtual Machine (VM) management), the following are some of the applications:

- Big Data in General Electric, General Motors and the Automotive Industry
- Big Data in Semiconductor Manufacturing and Integrated Circuits
- Big Data at Work for a Missile Plant
- Big Data in Cloud-based Design and Manufacturing
- Big Data and Additive Manufacturing
- Technology Integration Based on Big Data for More Value
- Production Process Monitoring, Maintenance, Quality Assurance and Logistics for manufacturers
- Big Data in CAD/CAE/CAM and CAD Educational Assessment
- Medical Device Design and Manufacturing

V.INTRODUCTION TO HADOOP

Hadoop is an open source framework for writing and running distributed applications that are capable of batch processing large sets of data. Hadoop framework is mainly known for Map Reduce and its distributed file system. The Map Reduce algorithm consists of two basic operations: map and reduce. It is a distributed data processing model that runs on a large cluster of machines. Hadoop includes three parts: Hadoop Distributed File System (HDFS), Hadoop Map Reduce and Hadoop Common. Hadoop is changing the perception of handling Big Data especially the unstructured data. Apache Hadoop enables surplus data to be streamlined for any distributed processing system across clusters of computers using simple programming models. It truly is made to scale up from single servers to a large number of machines, each and every offering local computation, and storage space. Instead of depending on hardware to provide high-availability, the library itself is built to detect and handle breakdowns at the application layer, so providing an extremely available service along with a cluster of computers, as both versions might be vulnerable to failures.

TYPES OF DAEMONS IN HADOOP

Hadoop is comprised of FIVE separated daemons

- NameNode
- Secondary NameNode
- DataNode
- Job Tracker
- Task Tracker

a).Namenode: NameNode in HDFS Architecture is also known as Master node. HDFS Namenode stores meta-data i.e. number of data blocks, replicas and other details. This meta-data is available in memory in the master for faster retrieval of data. NameNode maintains and manages the slave nodes, and assigns tasks to them. It should deploy on reliable hardware as it is the centerpiece of HDFS.

b).Secondary Namenode: In HDFS, when NameNode starts, first it reads HDFS state from an image file, FsImage. After that, it applies edits from the edits log file. NameNode then writes new HDFS state to the FsImage. Then it starts normal operation with an empty edits file. At the time of start-up, NameNode merges FsImage and edits files, so the edit log file could get very large over time. A side effect of a larger edits file is that next restart of Namenode takes longer.

Secondary Namenode solves this issue. Secondary NameNode downloads the FsImage and EditLogs from the NameNode. And then merges EditLogs with the FsImage (FileSystem Image). It keeps edits log size within a limit. It stores the modified FsImage into persistent storage. And we can use it in the case of NameNode failure.

c).Datanode: DataNode in HDFS Architecture is also known as Slave. In Hadoop HDFS Architecture, DataNode stores actual data in HDFS. It performs read and write operation as per the request of the client. DataNodes can deploy on commodity hardware.

d).Job Tracker: The JobTracker is the service within Hadoop that farms out MapReduce tasks to specific nodes in the cluster, ideally the nodes that have the data, or at least are in the same rack.

- Client applications submit jobs to the Job tracker.
- The JobTracker talks to the NameNode to determine the location of the data
- The JobTracker locates TaskTracker nodes with available slots at or near the data
- The JobTracker submits the work to the chosen TaskTracker nodes.

e)Task Tracker: A TaskTracker is a node in the cluster that accepts tasks - Map, Reduce and Shuffle operations - from a JobTracker. Every TaskTracker is configured with a set of slots, these indicate the number of tasks that it can accept. When the JobTracker tries to find somewhere to schedule a task within the MapReduce operations, it first looks for an empty slot on the same server that hosts the DataNode containing the data, and if not, it looks for an empty slot on a machine in the same rack. The TaskTracker spawns a separate JVM processes to do the actual work; this is to ensure that process failure does not take down the task tracker. The TaskTracker monitors these spawned processes, capturing the output and exit codes. When the process finishes, successfully or not, the tracker notifies the JobTracker. The TaskTrackers also send out heartbeat messages to the JobTracker, usually every few minutes, to reassure the JobTracker that it is still alive. These message also inform the JobTracker of the number of available slots, so the JobTracker can stay up to date with where in the cluster work can be delegated.

VI.HADOOP DISTRIBUTED FILE SERVER

The Hadoop Distributed File System (HDFS) is based on the Google File System (GFS) and provides a distributed file system that is designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. It is highly fault-tolerant and is designed to be deployed on low-cost hardware. It provides high throughput access to application data and is suitable for applications having large datasets.

HDFS holds very large amount of data and provides easier access. To store such huge data, the files are stored across multiple machines. These files are stored in redundant fashion to rescue the system from possible data losses in case of failure. HDFS also makes applications available to parallel processing. The below figure: 2 shows the Architecture of HDFS.

Features of HDFS

- It is suitable for the distributed storage and processing.
- Hadoop provides a command interface to interact with HDFS.

- The built-in servers of namenode and datanode help users to easily check the status of cluster.
- Streaming access to file system data.

The step by step procedure:

- Input data is broken into blocks
- Which are of size 128 Mb and then blocks are moved to different nodes.
- Once all the blocks of the data are stored on data-nodes, the user can process the data.
- Resource Manager then schedules the program on individual nodes.
- Once all the nodes process the data, the output is written back to HDFS.

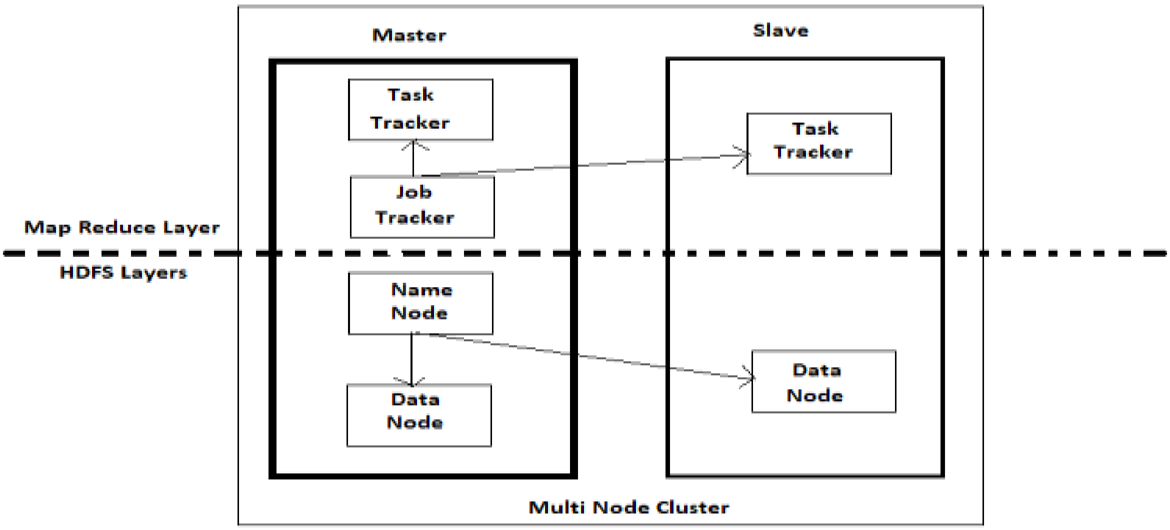

Figure: 2 The architecture of HDFS

VII.PROBLEMS REGARDING BIGDATA

Big Data is massive, poorly structured, heterogeneous data including and beyond the petabyte. This data is indecipherable to human scale. Hadoop is the most popular tool to manage the Big Data. There are others as well like spark, Apache strom, Apache Samoa etc. Tasks and challenges related with the Big Data are:

- Recognition of Big Data
- Discover new and innovative ways to find the specific data that can help you.
- Modeling and simulation of intelligent ways to model the problems that can be solved by Big Data.
- Effective and efficient ways to contextualize the data so that it gets relevant to specific individuals and groups
- Efficient ways to Analyze and visualize the results that come up due to Big Data.
- Storage, streaming and processing of Big Data to extract insights from it.

VIII.CONCLUSION

Big Data is large in volume, velocity, variety, value, variability and veracity. Big Data is used to combine different types of data in the field of Mechanical Engineering. In this study, we conclude that Big data plays a vital role in Mechanical Engineering and Design. The applications of Mechanical Engineering and Design are very much useful in Nowadays. The methods and technologies like Hadoop and Hadoop Distributed File system adds additional advantage to Big Data. Big data not only useful for Computer Science, it can be very much useful for Mechanical Engineering and Design also.

REFERENCES

[1].Agrawal, D., P. Bernstein and E. Bertino, 2011.Challenges and opportunities with Big Data. PurdueUniversity Purdue e-Pubs.

[2].Bellini, P., M. di Claudio, P. Nesi and N. Rauch,2013.Tassonomy and Review of Big Data

Solutions Navigation. In: Big Data Computing,Akerkar, R. (Ed.), CRC Press, ISBN-10: 1466578386, pp: 57-101.

[3].Bizer, C., P. Boncz, M.L. Brodie and O. Erling, 2012. The meaningful use of Big Data: four perspectives-- four challenges. SIGMOD Rec., 40: 56-60.

[4].Bottles, K., E. Begoli and B. Worley, 2014. Understanding the pros and cons of Big Data

analytics. Physic. Executive, 40: 6-12.

[5].Brown, B., M. Chui and J. Manyika, 2011. Are you ready for the era of 'Big Data'? McKinsey Global Institute.

[6].Camilli, R. and A. Duisberg, 2013. Big Data and the automotive industry. Report, Bird and Bird.

[7].Chardonnens, T., 2013. Big Data analytics on high velocity streams. MSc Thesis, University of Fribourg.

[8].Cho, Y. and S. Hwang, 2015. A study on LBS technique makes for CRETA LBS platform service. Int. J. Signal Process. Syst., 3: 82-87.

[9]. Davenport, T., 2013. The future of the manufacturing workforce, report three: Contemporary manufacturing technologies and the talent to use them. Technical Report, Manpower Inc.

[10]. Dayal, U., M. Akatsu, C. Gupta, R. Vennelakanti and M.Lenardi, 2014. Expanding global Big Data solutions with innovative analytics. Hitachi Rev., 63: 333-339.

[11]. Cottle, M., Kanwal, S., Kohn, M., Strome, T., and Treister, N.: Transforming health care through big data. strategies for leveraging big data in the health care industry new york: Institute for health technology transformation, 2013.

[12]. Cruz-Neira, C., Sandin, D.J., and DeFanti, T.A., Surroundscreen projection-based virtual reality: The design and implementation of the cave. In: Proceedings of the 20th annual conference on computer graphics and interactive techniques, SIGGRAPH '93, pp. 135-142. New York: ACM, 1993. doi:10.1145/166117.166134.

[13].Cruz-Neira, C., Sandin, D.J., DeFanti, T.A., Kenyon, R.V., and Hart, J.C., The cave: Audio visual experience automatic virtual environment. Commun. ACM 35(6):64–72, 1992. doi:10.1145/129888.129892.

[14].Cunha,J.,Silva,C.,andAntunes,M.,Healthtwitterbigbatamanagement with hadoop framework. Procedia Comput. Sci 64:425–431, 2015.

[15].Dang, S., Coon, J.P., and Simmons, D.E., Combined bulk and pertone relays Election in super dense wireless net works. In: Communication workshop (ICCW), 2015 IEEE international conference on, pp. 2200–2205. IEEE, 2015.

[16]. Davis, J.C., and Hunt, T., Hyperbaric oxygen therapy. J. Intensive Care Med 4(2):55–57, 1989.

[17].Hansik Ryu, Issues and Prospect of Big data Business, Issue & Trend, 2012.

[18] Society of Digital Policy and Management, Big data Analystics for Business, Gwangmungak, 2014.

[19] McKinsey Global Institute, Big data: the next frontier for innovation, competition and productivity. McKinsey & Company, 2011.

[20] Min-Jeong Song, Business Future Loadmap Made by Big data, Hansmedia, Korea, 2012.

[21]. Gartner,'Hyper Cycle for Analytic Applications', 2011.

[22]. Lidong Wang and Cheryl Ann Alexender, Big Data in Design and Manufacturing Engineering, science publications.2015

BIOGRAPHIES

K.Venkata Nagendra, working as Assistant Professor in the Department of Computer Science Engineering at Geetanjali Institute of Science and Technology, Nellore, Andhrapradesh, India. He has 8 years of experience in the field of teaching. He is a research scholar in Vikrama Simhapuri University. He did his M.Tech in ANU, Guntur. His areas of interests are Data warehousing and Data Mining and Cloud Computing.

Spurgeon Pynam, is an UG Student in the Department of Mechanical Engineering in Geethanjali Institute of Science and Technology,SPSR Nellore,India. He had more intrest on computer science. He is doing the research on integration of Mechanical Engineering and Design with Computer Science. His areas of interests are Aeronautics, Machine Learning and Big Data.

Bharath Potlapati, is an UG Student in the Department of Mechanical Engineering in Geethanjali Institute of Science and Technology,SPSR Nellore,India. He had more intrest on computer science. He is doing the research on integration of Mechanical Engineering and Design with Computer Science. His areas of interests are Aeronautics, Machine Learning and Big Data.