Assignment: Critical Review of Big Data Trends and Technologies

Course Name:
Big Data Analytics Winter 2025
Date: 08/12/2025 In class Discussion + upload to blackboard

1. **What is the aim of the paper?**

   The paper studies how Big Data technologies have evolved from early batch-processing systems to modern cloud-native, real-time, and AI-driven architectures. It not only pinpoints significant technological developments but also tackles the current issues related to scalability, governance, and privacy, and finally, directs future research by means of systematic literature review and bibliometric analysis of recent studies.

2. **According to the paper, what are the major trends in Big Data technologies?**

   The paper highlights recent trends in Big Data technologies. One of the primary directions is the transition from older, batch-oriented processing systems like Hadoop to the faster and more dynamic platforms, particularly in-memory engines like Apache Spark. Research indicates that Spark's capability of performing data processing in memory makes large-scale analytics more effective than the traditional MapReduce method.

   Moreover, the paper underlines the trend of the growing prominence of real-time and streaming data processing indicated by the development of technologies such as Apache Kafka and Apache Flink, which are key for applications that require immediate insights rather than waiting for batch jobs. Sectors like healthcare, finance, IoT, and smart cities heavily rely on real-time analytics.

   Additionally, the paper observes that there is a strong trend in the direction of the uptake of cloud-based Big Data platforms that was characterized by the fact that organizations are moving their storage, analytics, and processing tasks to the cloud to enjoy the benefits of scalability, less infrastructure management, and high availability. Distributed analytics, which is one of the advantages provided by cloud computing, is also a requirement for the growing data volume and velocity being handled.

   Another trend is the inclusion of Artificial Intelligence and Machine Learning into the Big Data workflows. According to the paper, modern Big Data systems are increasingly turning to ML and data mining techniques for such tasks as prediction, classification, clustering, anomaly detection, and decision support. The merger of these technologies facilitates the transformation of large volumes of data into profitable insights for various sectors.

The ascending trend of IoT and sensor-based applications is another important factor. With the increasing number of connected devices producing nonstop data flows, Big Data technologies must support the incoming data at a very high rate and in various forms.

Finally, the paper mentions the increasing interest in data governance, privacy, and ethics. The growth of Big Data has made data protection, responsible data use, and trust key topics in both research and industry along with the global struggle for data privacy and ethics.

3. **According to the paper, what are the main Big Data challenges?**

The report indicates that an array of main obstacles still exist which adversely affect the progress and application of Big Data technologies.

- Privacy and security - The collection and processing of personal or sensitive information is very much part and parcel of Big Data; hence the organizations must be very careful to make sure that data can't be accessed, stolen, or misused in any way. The paper highlights that inadequate security measures, feeble access control, and absence of proper precautions can put large datasets at high risk.
- Ethical issues - With the impact of Big Data systems on decision-making across various sectors like healthcare, finance, education and smart cities, the emergence of issues like fairness, transparency and even bias in data or algorithms becomes crucial. The paper states that organizations must not only ensure the ethical usage of data but also build and sustain trust with users.
- Scalability - The continuous increase in data volume, velocity and variety is the driving force behind the demand for Big Data systems to deal with huge datasets, very quick ingestion, and diverse formats.
- Data quality - Many datasets are incomplete, inconsistent, noisy, or unstructured. Poor data quality reduces the accuracy of analytics and machine learning models, requiring more advanced preprocessing and validation methods.
- Infrastructure complexity – Today, Big Data setups involve many tools and platforms for storage, streaming, and analytics, making the infrastructure complex.

Overall, the article highlights that the factors of privacy, ethics, scalability, and data quality along with complexity pose the most critical issues that still determine the directions of big data research and practices.

4. **What methods, frameworks, or algorithms are used in Big Data analytics today?**

The article dives into the integration of processing frameworks, storage systems, and analytical algorithms as the core pillar for Big Data analytics. It stresses the importance of the Hadoop ecosystem which brought about distributed storage through HDFS and batch processing using MapReduce. Although Hadoop remains relevant, Apache Spark—

with in-memory computation, faster processing, and broader analytical capabilities—is increasingly used.

The paper reviews the use of several tools in the context of real-time data processing, including Apache Kafka which is widely used for the ingestion and streaming of data, and Apache Flink which is known for low-latency, event-driven analytics. These tools allow companies to process data continuously instead of waiting for batch jobs to finish.

Data management is mentioned and illustrated through the concepts of Data Lakes and Data Warehouses. Data Lakes accommodate huge volumes of raw and semi-structured data, while Data Warehouses deal with structured, cleaned, and query-ready information. Most of the Big Data ecosystems make use of both storage types based on the data analytics needed.

The machine learning and data mining techniques are added as an emphasis of the paper which includes clustering algorithms such as K-Means and DBSCAN, classification procedures like logistic regression and Random Forests, and dimensionality reduction techniques such as Principal Component Analysis. Deep learning is used for complex analytics on unstructured or very large datasets.

The paper also stresses the role of statistical analysis, predictive modelling, and data mining applications in various sectors, including healthcare, finance, and smart city technologies. In general, Big Data analytics releases the hidden valuable insights from immense data volumes through the combination of distributed processing frameworks, scalable storage, and advanced analytical algorithms.