BIG DATA MANAGEMENT IN IOT SYSTEMS

Name:- Avinash Kumar Chaurasia
Matriculation Number: 6896835
Paderborn | 17.09.2021

Fraunhofer IEM
HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

AGENDA

1. Motivation
2. Introduction
3. Problem Analysis
4. Approaches for Big Data Management
5. My Contribution
6. Conclusion

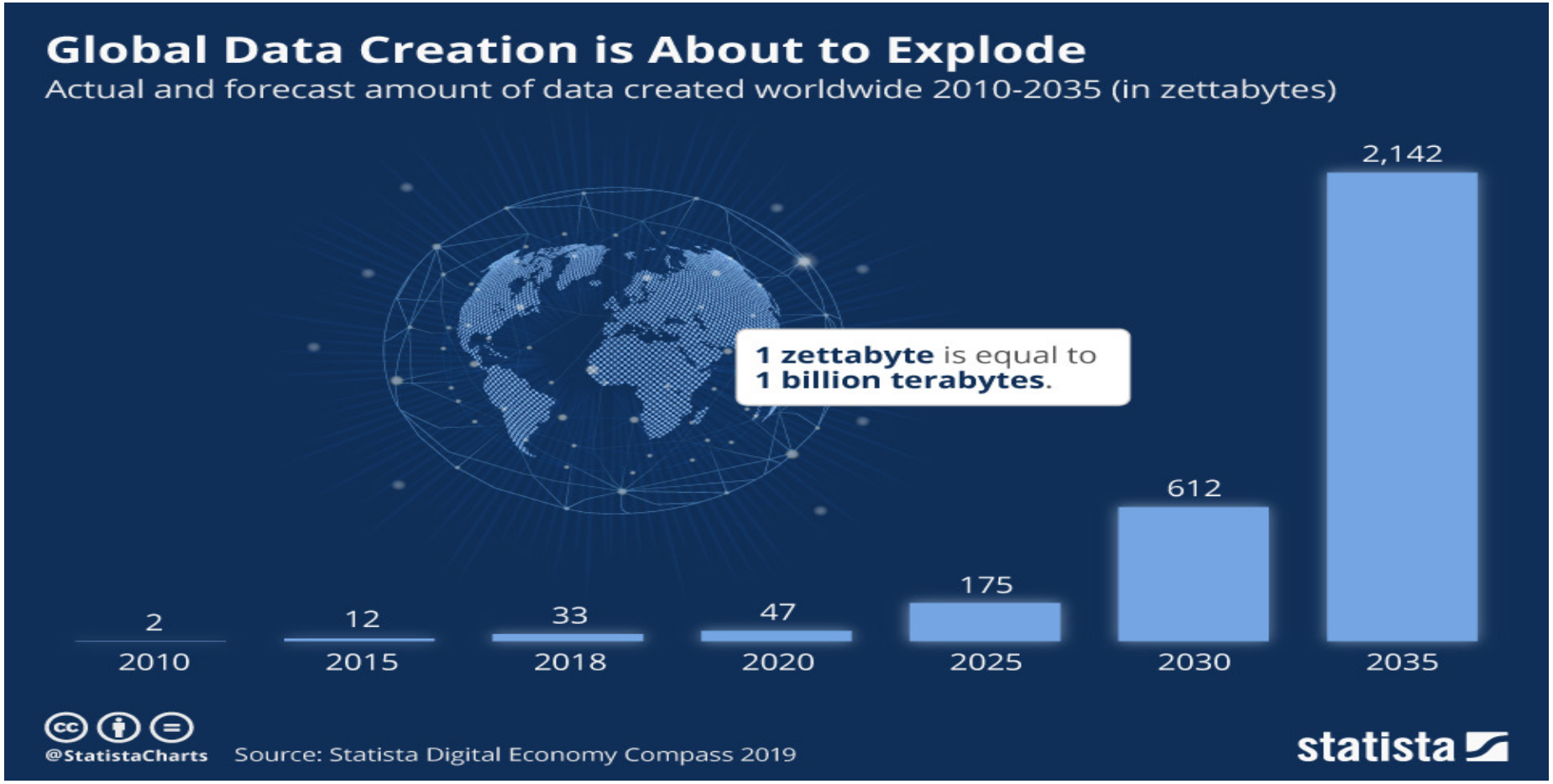
MOTIVATION

1

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Fraunhofer
IEM

Why data management is vital in Green Programming



Source: https://www.digitalinformationworld.com/2019/04/chart-global-data-creation-forecasts.html

INTRODUCTION

2

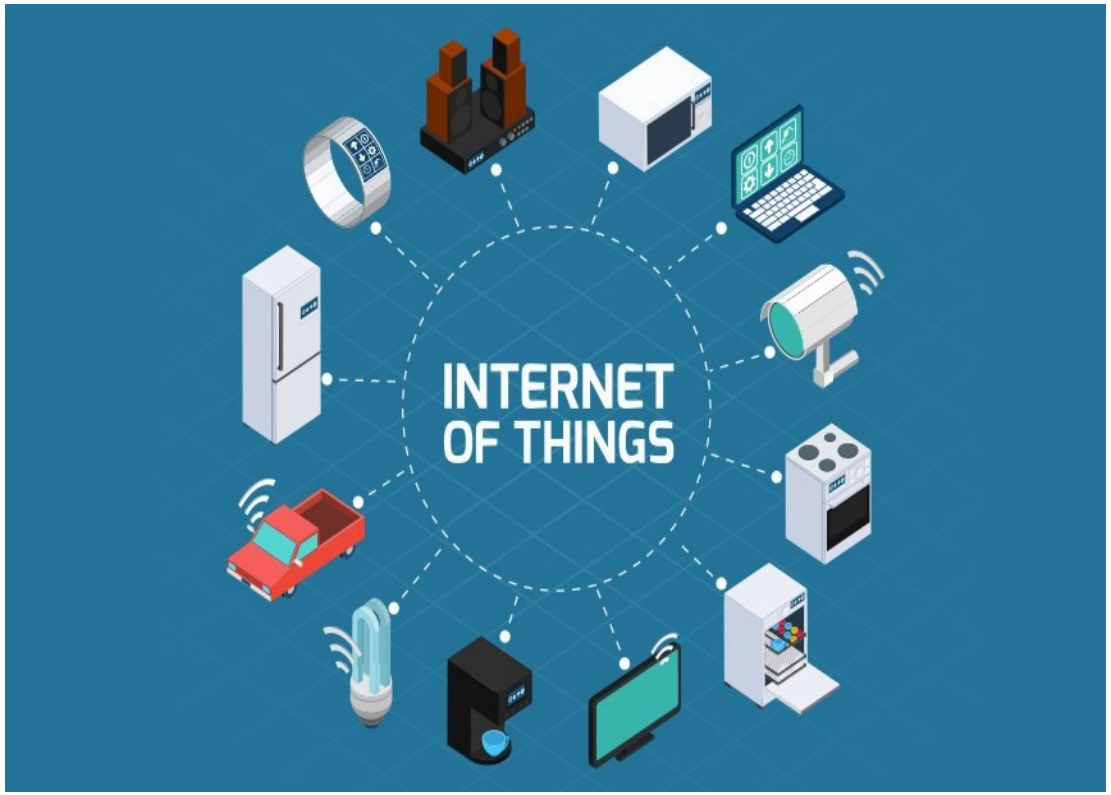
HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Fraunhofer
IEM

Introduction

Internet of Things(IoT)

- The term Internet of things (IoT) came into existence when Kevin Ashton used the term as the title of a presentation in 1999.
- An evolutionary concept that aims to connect billions of physical items around the world.
- Embedded systems, such as processors, sensors, and communication hardware, are used to collect, send, and act on data collected from their environments by web-enabled smart devices.
- Some of the applications of IoT are smart Buildings, Smart homes, Smart cities or Smart Grid



Source: https://medium.com/agileinsider/internet-of-things-iot-101-7588a388ef70

Big Data Problem

IoT impacts in the form of Big Data

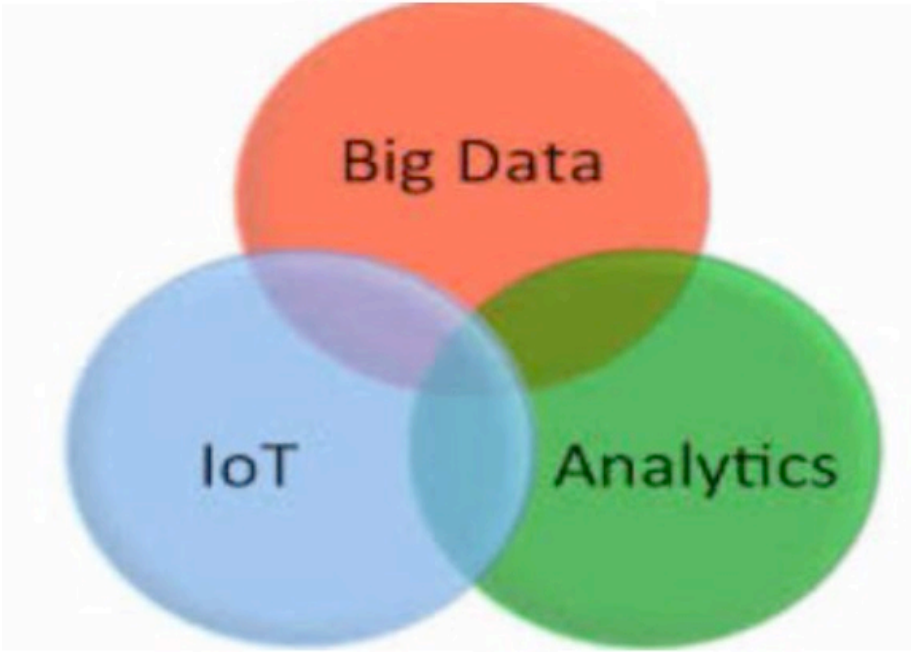
- An enormous amount of data is getting accumulated every day by the internet from various sources.
- Data from multifarious networked embedded devices turns into a Big Data Problem.
- Traditional data processing software or databases cannot handle Big Data.

Source: https://www.linkedin.com/pulse/importance-big-data-analytics-evolution-iot-hamed-sadegh-pour-mba

Relation between IoT, Big Data and Analytics

Analyzing IoT data for Big Data Management

- Massive data generated by IoT needs to be explored and investigated in real-time in order to not compromise the quality of data.
- Provides predictive analysis to foresee future problems and provide solutions.
- Proper analysis can save lot of data by finding the structure of data and helps conduct analyses and improve decision making.



Source: B Sobhan Babu, T Ramanjaneyulu, I Lakshmi Narayana, and K Srikanth. Data management in iot using big data technologies and tools.

PROBLEM ANALYSIS

3

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

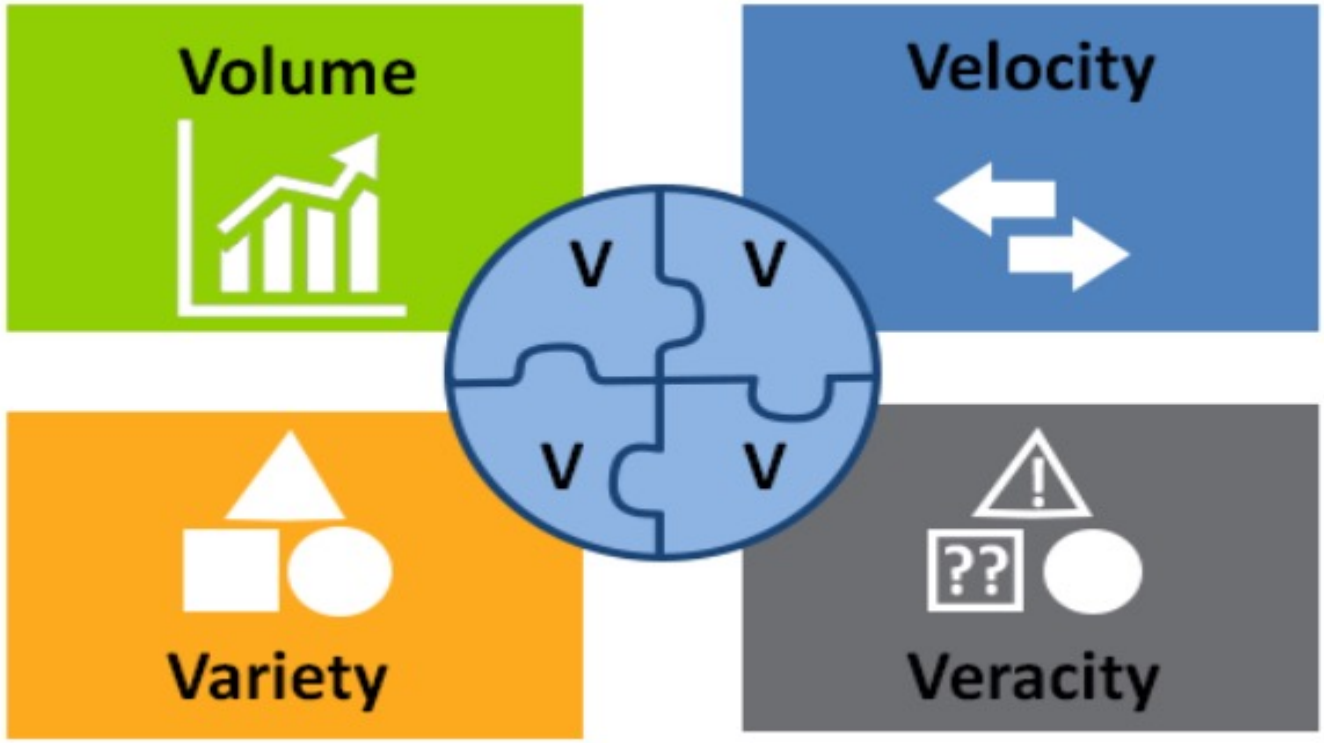
Fraunhofer
IEM

4V's of Big Data

- The four V's model, which is well-known in the world of big data, contains four dimensions. The four V's have the following characteristics.

1. Volume
The Size of data

2. Variety
Different types of data



3. Velocity
Rate of information processing

4. Veracity
Correctness of data

Source: https://www.zarantech.com/blog/the-4-vs-of-big-data/?utm_source=blog&utm_medium=blog-header-search-box&utm_term=

Challenges in 4V's model of Big Data

Volume

- Difficult for the organizations to collect entire volumes of data.
- Overwhelming volumes of data require big storages and hence huge financial burden on organizations.

Variety

- The presence of structured, unstructured, and semi-structured data with different formats in one place is a problem.
- Difficult to integrate data.

Velocity

- Data must be managed forthwith to provide a way for the new data to breeze in.
- Operational and process delays for data driven companies due to slow processing of data.
- Risk of using outdated information, which is inconsistent with the organizational immediate needs and requirements.

Veracity

- Distinguishing between reliable and unreliable data is critical.

APPROACHES FOR BIG DATA MANAGEMENT

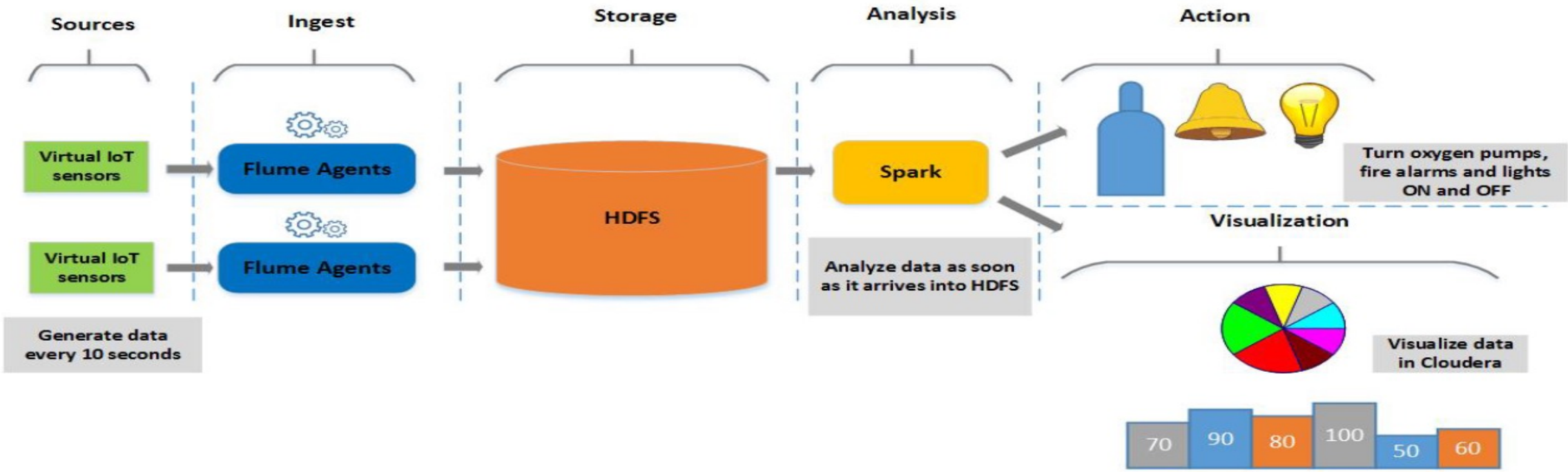
4

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Fraunhofer
IEM

Approach 1

IoT Big Data Analytics for Smart Buildings Systems

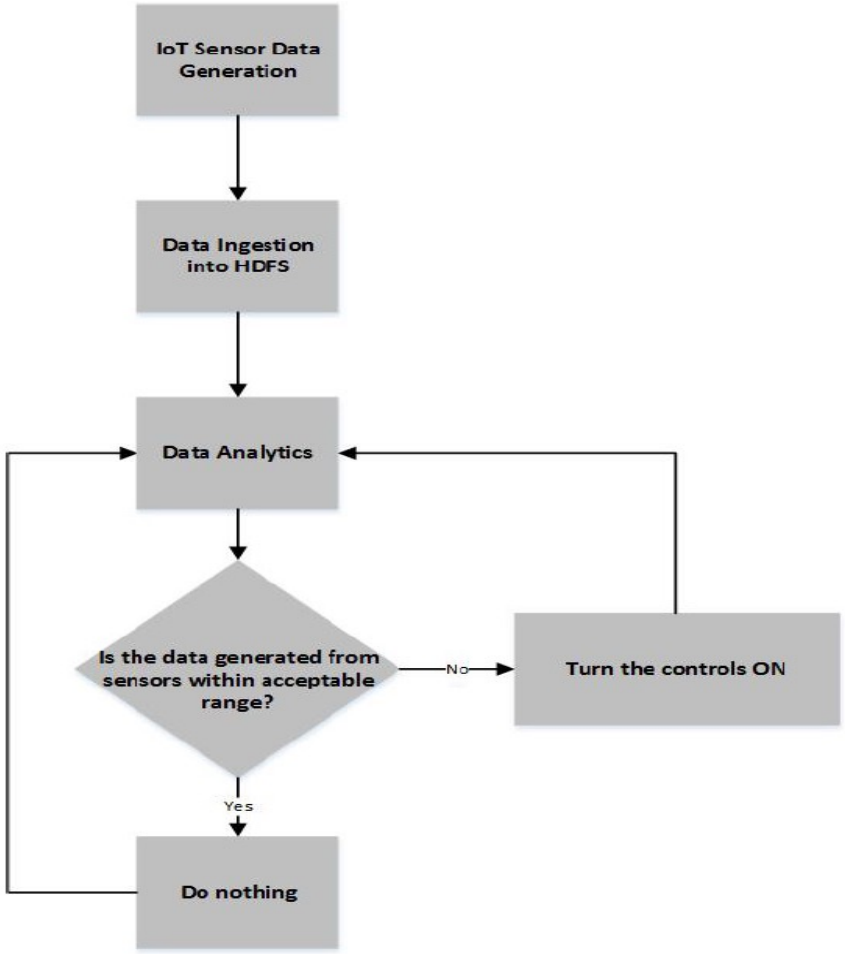


Source: Muhammad Rizwan Bashir and Asif Qumer Gill. Towards an iot big data analytics framework: Smart buildings systems. In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pages 1325–1332, 2016.

- Authors Bashir and Gill proposed an effective framework called IoT Big Data Analytics (IDBA) to implement Big Data Analytics for smart devices using technologies like Hadoop Distributed File system(HDFS) for data storage and PySpark for Analysis.

Workflow of IDBA

- Sensors data is generated using python code.
- Generated data is stored on HDFS using Apache Flume.
- Based on the analysis of data, actions like switching ON the controls of smoke, heat and luminosity if data is below threshold level.



Source: Muhammad Rizwan Bashir and Asif Qumer Gill. Towards an iot big data analytics framework: Smart buildings systems. In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pages 1325–1332, 2016.

Approach 2

Deep Learning for Edge-Cloud Data Analytics in IoT

- Ghosh proposes a deep learning-based approach for data reduction on the edge with machine learning on the Cloud.

- **Deep Learning**
  - A subclass of Machine Learning techniques that allows computer models with several processing layers to learn multiple degrees of abstraction for data representations.
  - Autoencoders are used in Neural Networks to reduce the size of data.

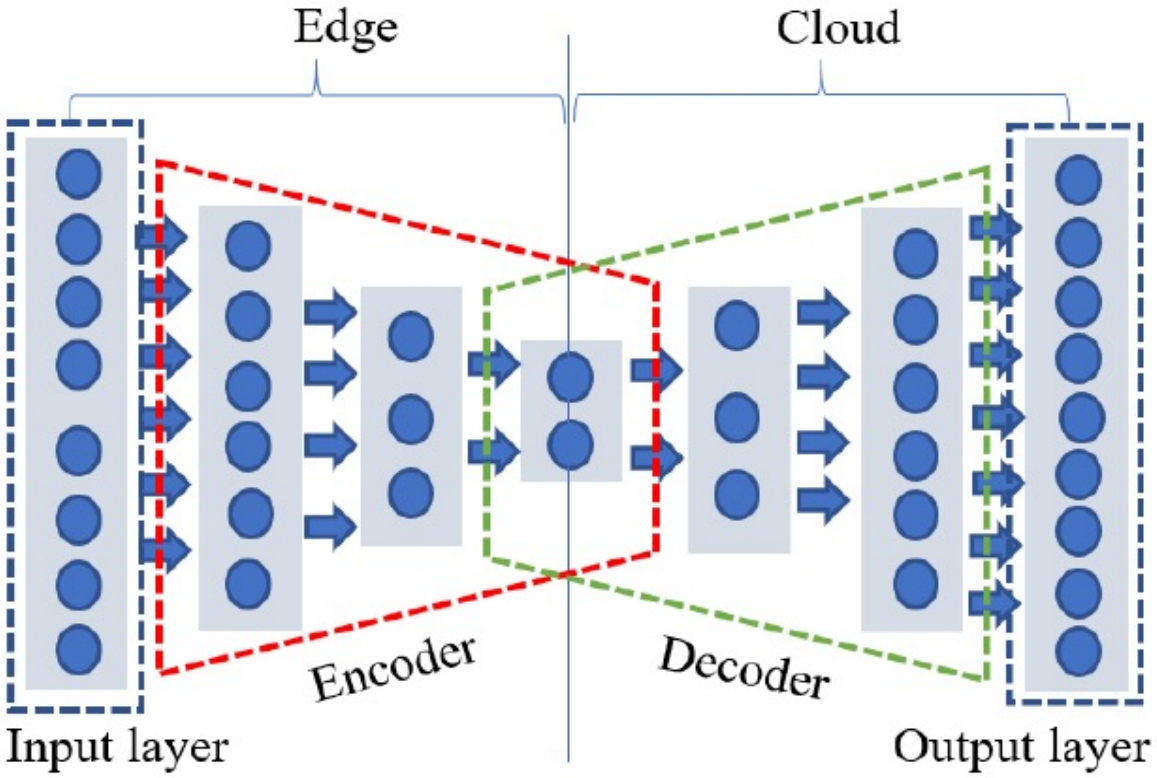
- **Edge Computing**
  - Edge computing is a distributed computing paradigm in which computation and data storage are brought closer to the data sources. This should increase response times while also conserving bandwidth.

- **Dimensionality Reduction**
  - It refers to the methods that reduce the number of features or attributes in a data set and this method is important when we are working on large data sets. For example Principle Component Analysis(PCA).

Approach 2

Workflow of Deep Learning for Edge-Cloud Analytics

- Data collected from various smart phones.
- High dimensional data is put on the edge where autoencoders reduce the data into smaller number of dimensions.
- The reduced data is then put on the cloud for Machine Learning or data analytics tasks.
- Processing reduced data for ML tasks is much faster.

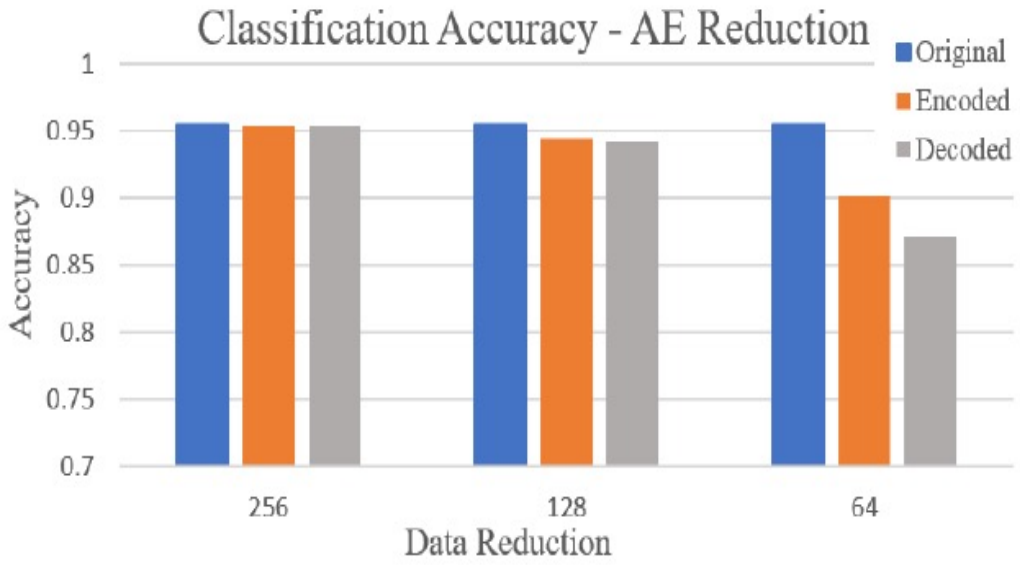


Source: Ananda M. Ghosh and Katarina Grolinger. Deep learning: Edge-cloud data analytics for iot. In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pages 1–7, 2019.

Approach 2

Classification Accuracy between Original and reduced data

- The encoder reduces the features from 561-265-128 to 256-128-64 features.
- Data is reduced from 11.2MB for original data to 5.754MB for 256 features, 2.877MB for 128, and 1.4386MB for 64 features.
- 50% reduction in data did not have a significant influence on classification accuracy.
- 77% reduction only resulted in a 1% change.



Source: Ananda M. Ghosh and Katarina Grolinger. Deep learning: Edge-cloud data analytics for iot. In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pages 1–7, 2019.

MY CONTRIBUTION

5

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Fraunhofer
IEM

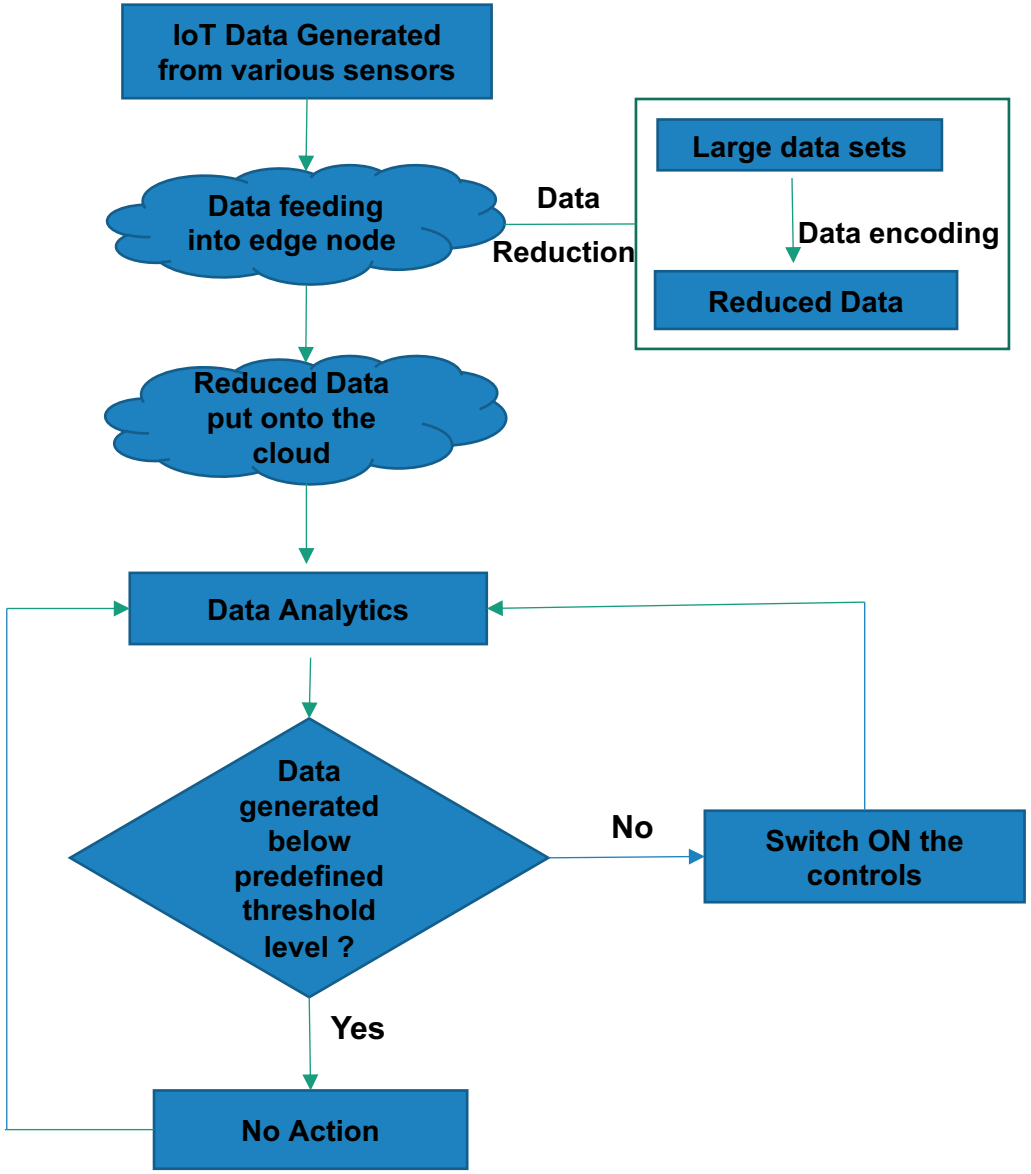
Drawbacks of IDBA framework

- Latency issue in HDFS because MapReduce takes the set of data and converts it into another set of data, where an individual element is broken down into a key-value pair.
- MapReduce works by breaking the processing into phases: Map and Reduce. So, MapReduce requires a lot of time to perform these tasks, thus increasing latency. Hence, reduces processing speed.
- Processing Overhead which makes the read/write operations very expensive.

My Proposal

- Merging of two approaches in order to mitigate the drawbacks of IDBA approach.
- The generated data should be placed on the edge cloud instead of HDFS.
- Data is reduced on edge cloud using autoencoders.
- The reduced data is then put on cloud for further data analytics tasks.
- Since the data will be reduced on edge itself, the data analytics process performed on the cloud will be much faster and robust.

Wokflow of Adapted approach



CONCLUSION

6

HEINZ NIXDORF INSTITUT
UNIVERSITÄT PADERBORN

Fraunhofer
IEM

Conclusion

- The proliferation of IoT devices has resulted in massive volumes of data, which is anticipated to continue.
- IoT has emerged as one of the biggest sources of Big Data, which is unavailing if not managed and analyzed properly.
- Challenges in the four dimensions of Big Data were addressed.
- Different tools like Hadoop Distributed File System(HDFS) and Deep Learning can be used for the effective data management of IoT data.
- A proposal was suggested to mitigate some of the drawbacks in HDFS.

Acknowlegments

1. B Sobhan Babu, T Ramanjaneyulu, I Lakshmi Narayana, and K Srikanth. Data management in iot using big data technologies and tools.

2. Shivanjali Khare and Michael Totaro. Big data in iot. Pages 1–7, 07 2019.

3. Owais Khalid and Suntharalingam Senthilananthan. A review ofdata analytics techniques for effective management of big datausing iot. In2020 5th International Conference on InnovativeTechnologies in Intelligent Systems and Industrial Applications(CITISIA), pages 1–10, 2020.

4. Muhammad Rizwan Bashir and Asif Qumer Gill. Towards an iot big data analytics framework: Smart buildings systems. In 2016 IEEE 18th International Conference on High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pages 1325–1332, 2016.

5. Ananda M. Ghosh and Katarina Grolinger. Deep learning: Edge-cloud data analytics for iot. In 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), pages 1–7, 2019.

Thank you