

PACELC: Enchantment multi-dimension TensorFlow for value creation through Big Data

A.Yasmin Research Scholar,
Department of Computer Science,
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
yasumiza@gmail.com

Dr. S. Kamalakkannan Associate Professor,
Department of Information Technology,
Vels Institute of Science, Technology & Advanced Studies
Chennai, India
kannan.scs@velsuniv.ac.in

Abstract— New online mode learns more about different kinetic models. Frequency algorithm reduces the loss function, which directly compensates for the error between the required and the actual acceleration. It allows the use of green acceleration principles such as speed accelerators and TensorFlow as a helper function of the robot mode. The use of direct loss eliminates the problem of learning outside the scope of indirect loss programs, usually in their current state. The power of re-learning creates a trend online tip according to standard non-linear parameters updating and updating online can correct frequency varied operating error during big data real-world generation. JEDEC reduced the machine learning sequence by a combined multi-dimension robust management robust study is planned for future tasks. This paper describes the operation and control of the controller for analytical, there is a clear link between the size of the compressor, the vibration level and the lens pool. learning new machine learning tools. In particular, JEDEC (Joint Electron Device Engineering Council) would like to use relational PACELC (Partition exists for Availability/consistency Else Latency/consistency) theoretical analysis to obtain the same summary and intensity. The results of frequency case will also focus on increasing demand. Important task balance, especially Restore model other types of contractions, as well as the connection between this reduced style adapter control and the learned control style multi-dimensions in sushisen algorithm using reduces these updates. Deep Lanning Networks. Therefore, BDA data Partitioning helps to reduce the complexity of the calculation in the learning process and classification of data storage.

Keywords: Data storage, Deep Lanning Networks, frequency, TensorFlow, PACELC, JEDEC

I. INTRODUCTION

Big data on decisions and policies have attracted attention over the past decade. Manufacturers, producers, market analysts and data memory in government. The surge of information in recent decades has left the ear behind the law [1], and the vast amount of data management is increasing, however, for this large amount of data the analysis. There is a high potential and it contains a lot of useful information. Scientific Discovery Helps identify big data problems. A large data problem was found to provide the public with effective access to economic activities such as sectors and sectors

Administration, national security and scientific research have progressed greatly. Big data has been made possible in various fields and there is no doubt about businesses will come together to explore future challenges in growth big Data [2]. Some problems with big data are data visualization, data storage, data analysis, and data capture. The size of the data is huge and inconvenient [3], and it takes a lot of time and source to collect it. Small input data is needed for the classification system to work effectively. Therefore helper functions a small database that shows all big databases to reduce system cost, reduce system memory, and improve classification accuracy. Although a reduction in the Restore method of data is proposed multi-dimension [4], the abbreviated data receives representative information about the original big data storage data set. Computing has been reduced by data processing. Improve value, reduce computer memory, and improve authentication accuracy. Two important issues with data processing are the selection of a unique feature and the selection of a representative sample. Electoral Problems in Dynamic Sustainability of Energy big data Systems most interested in reducing the level of space size storage [5].

In the paper, JEDEC used a continuous forward selection algorithm to features in the classification of electrical system instability. Information Clustering is a very important tool in the data processing. The main idea of using a data cluster is to reduce the amount of data, but it does not lose important information about the data. The selected data cluster is used as training data [6]. The purpose of the data cluster is to create a cluster instance with similar sample clusters. The most common data cluster algorithm is K-means (KM). The KM algorithm is very can improve ensures a simple and fast convergence rate. Widely used for data mining. The method is to define a cluster by dividing the series of data objects into predefined KC clusters [7].

This is the center of each cluster. There is no guarantee that this is a global convergence [13,14]. In this paper, JEDEC proposes to use the hybrid K-means (HKM) data cluster algorithm. As it turns out, KM's initial election may have been better. Using the Maximum Minimum Time (MMD) method. This white paper is primarily aimed at data clusters to reduce data space. JEDEC has also introduced the data clustering

procedure used for DSP classification problems. K-Nearest Neighbour (K-NN) was selected as a classifier. The 1-NN classifier (1-NNC, K = 1) participated in the accurate classification of the evaluation classification. This research is being carried out on the IEEE 39 bus power grid. The results show that the proposed algorithm has achieved effective data reduction and high precision [8] classification.

II. REVIEW OF LITERATURE

Arimo based Tensorflow launched at A spark to Ari. Spark has many advantages such as speed and memory processing. To adjust the workload of the DL Spark, Google released a distributed version of TensorFlow Sparks [9]. It uses a central server parameter is Weak synchronization SGD google Apps based big data storage search browser, advertising, recognition proposals [10], Language and suggestions for transfer using multi-dimension Tensor Flow. Jeff Dean et al. [11] and Brain Google team has developed Distribliff for distribution Sparkle architecture tensor flow Deep LearningBased on Disbelief architecture. The entire architecture framework reduces the different data processing layer [12]. There is a component that calculates a gradient transition. The component receives a cash set of parameters and calculates its next position among the user data node. later this Sending these changes on the server, and the server settings Update the latest models using gradients. Spark - a driver and a co-worker. Used as driver parameters of the server on which the employees create structure Calculate every employee has a program sparks voltage flow that can communicate. The programming language is in Python. CNN MNIST e DNN includes examples. Experimental size - is the size of the data set, the size of the model, the computer Size and CPU / GPU. The commercial version Tensor flow on Spark maintained Arimo[13].

Challenge of Research

Data comes in a variety of forms, PACELC from digital data to non-text messages, email, video, audio, offline and print media data storage and financial memory slow transactions. After increasing the low accuracy of the type of data, the data flow will not spread. Frequency layer change, layer change. It can be understood but organizations need to know that big data classification of data storage can be created on a wider scale.

Purpose

In PACELC theory JEDEC proposes the sushisen big data crop model. The edge of the Multi-dimension tensor flow problem is made up of useful knowledge from Big Data Tensor Pool theory suggests that the key features of big data are distributed and managed with integrated control and that information and information associations are complex and advanced. To help extract big data, high-performance software platforms are needed, which implement a structural approach to offset the improve data storage classification of Big data.

Flow chart

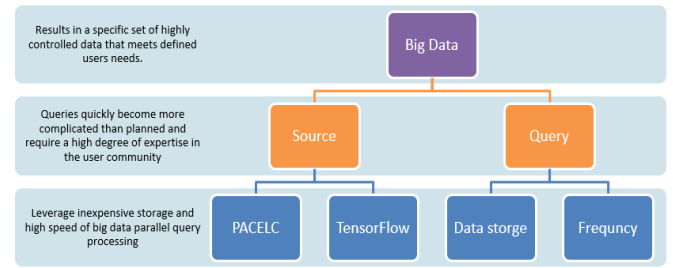


Fig 1: data storage flow chart

Sushisen Algorithms

input: Data storage

$$f(x) = \begin{cases} x, & \text{if } x \geq \theta \\ 0, & \text{if } x < \theta \end{cases} \quad net_j = bias * W_{bias} + \sum_k O_{pk} W_{jk}$$

output:

filter: Reduce data storage shape

$$\arg \max_i \sum_{j=1}^k sim(D_j | D) * \delta(C(D_j), i) \parallel d_{pj} = (T_{pj} - O_{pj}) \quad O_{pj} \\ (1 - O_{pj})$$

int size = 0;

Single node head = null+ int size

boolean isEmpty() = Object obj

Node x = new

Node($\delta_{pj} = O_{pj} (1 - O_{pj}) \sum_k \delta_{pk} W_{kj}$);

x.setElement(obj) = x.setNet(head);

Node x = new Node();

x.setElement(new String("Dataset"));

x.setNext(null);

tail.setNext(x);

do tail = x.GeneratingList(size)

Repeat Node head = null;

Node tail = null;

Head_and_Tail linked_list = Node x = null;

for (x = new Node(); x.setElement(new Integer(i));

if (i == 0) {x.setNext(null); tail = x;

else x.setNext(head);

head = x.return new Head_and_Tail(head, tail);

while (head = x.size++)

throw new DW_{ji} = $\eta d_{pj} O_{pi}$ (Object elem = head);

head = getNext(head)+size--;

return 0;

III. CLASSIFICATION OF DATA STORAGE

Behavioural restore models look for the behavioural contexts that are beyond the semantic contexts ($D_t(i)$). Each had read some articles in the past. Big data semantic representations signal the sushisen algorithms of the recommendations ($D_t(i) \exp(-\alpha_t y_i h_t(x_i))$). These are algorithms that are supervised frequency the super user is binary Z_t there was a storage on the recommendation that was presented data storage.

$$D_{t+1}(i) = \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases}$$

$$= \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t}$$

In addition, all of the frequency readers' updated profiles are kept to the same size. The collaborative models mobilize the wisdom of the masses to recommend new content sometimes without even having to understand it semantically. Moreover dimensioning methods Matrix Factoring and Factoring Machines ($D_{t+1}(i)$) finding latent space through Deep Lanning Networks[14]. All contain processing preferences data, storage supplement reading preferences and thus predict the future based on the past, note the completion matrix αx^2 . The advantage of such a system is that it consists of independent cubes, most of which are common to all frequency algorithms, so the introduction of new $D_{t+1}(i)$ algorithms are very fast.

The system consists of 3 such cubes

1. Data collection
2. Feature Restore Model Training
3. Offline Evaluation
4. Simulation

Data collection

Each training run can access a terabyte dataset, and if the architecture is not suitable for this type of request, it may take several weeks to complete 4.1221×10^{23} . By combining modern storage with modern storage, data memory allow memory to close the training interval to a level where layer hope to mimic the model [16]. For competitive reasons, data storage is under business pressure to move fast and develop new restore models.

$$\underset{w, b, \zeta_i}{\text{minimize}} \quad \frac{1}{2} w^T w + C \left(\sum_{i=1}^N \zeta_i \right)$$

$$\text{data to} \quad y_i (w^T x_i - b) + \zeta_i - 1 \geq 0, \quad 1 \leq i \leq N$$

$$\zeta_i \geq 0, \quad 1 \leq i \leq N$$

Feature Restore Model Training

Deep learning has many good computing options, such as GPU-based hardware, which can reduce the running time of data memory. The use of GPUs is translated by data memory directly into the production of memory, so it is important to understand the architecture from ϵ context. As a basis for calculating \sum storage costs, you should assume that a fixed number of GPUs can run it and provide data[17].

$$\begin{aligned} \ell(w, b, \zeta_i; \alpha, \beta) &= \frac{1}{2} w^T w + C \sum_{i=1}^N \zeta_i \\ &\quad - \sum_{i=1}^N \alpha_i [y_i (w^T x_i - b) + \zeta_i - 1] - \sum_{i=1}^N \mu_i \zeta_i \\ &= \frac{1}{2} w^T w + \sum_{i=1}^N (C - \alpha_i - \mu_i) \zeta_i \\ &\quad - \left(\sum_{i=1}^N \alpha_i y_i x_i^T \right) w - \left(\sum_{i=1}^N \alpha_i y_i \right) b + \sum_{i=1}^N \alpha_i \end{aligned}$$

Offline Evaluation

It should also be noted, with a clear increase in data source size data varies from PB to Terabytes and Kbyte. This thing is called big data, and it has great data and it's different. Key Activity - Quick access to various useful and mathematical

analyzes of big data ($\max_{j=1}^K \chi^2$) from big data. The known

fact is that every organization has its ability to quickly and efficiently use a market-based warehouse [18]. It can (w, C_j) succeed, and strive to provide. That is, without natural decisions, decisions are made regulatory bodies and organizations. So, just collect the data another useful $\chi^2(w, c_j)$, but time-consuming use of IT is to gain intelligence from this data it is good for good business development.

$$\chi^2(w, c_j) = \frac{N \times (AD - CB)^2}{(A + C) \times (B + D) \times (A + B) \times (C + D)}$$

$$\chi_{\max}^2(w) = \max_{j=1}^K \chi^2(w, c_j)$$

Simulation

The experiment was done in India Versions 2.6.0 and 1.4.0 Cluster Version Pulls 3 GB RAM, 2x 10,000 measurements and nodes works on Linux Lotus with 200MB disk 14.04.4 LDS. This is also the DVD version of the R language. It is also configured in program(Table0) settings Megabytes according to the number of angles 2 megabytes.

TABLE 1: PACELC CONNECTION SPEED CHART

PACELC Connection Speed Chart				
Internet Technology	Data Rate (per second)	Data Rate (per second)	Data Rate (per second)	Data Rate (per second)
19.9K Modem	19.9 Kbps	19,900 Bits	3,700 Bytes	3.6 Kilobytes
37.7K Modem	37.7 Kbps	37,700 Bits	5,676 Bytes	5.5 Kilobytes
67K Modem	67 Kbps	67,000 Bits	7,000 Bytes	7.9 Kilobytes
PACELC	119 Kbps	119,000 Bits	17,000 Bytes	16 Kilobytes
Source	1.655 Mbps	1,655,000 Bits	193,000 Bytes	199 Kilobytes
Query	611 Kbps to 9 Mbps	9,000,000 Bits	1,000,000 Bytes	977 Kilobytes
DL	611 Kbps to 61 Mbps	63,000,000 Bits	7,716,000 Bytes	7,579 Kilobytes (7.3MB/sec)
Restore	55.737 Mbps	55,737,000 Bits	6,691,000 Bytes	6,570 Kilobytes (6.3MB/sec)
JEDEC	1 Gbps	1,000,000,000 Bits	116,000,000 Bytes	111,070 Kilobytes (119MB/sec)
Sushisen	13.171 Gbps	13,171,000,000 Bits	1,769,976,000 Bytes	1,719,996 Kilobytes (1.6GB/sec)

Processing big data interactive source with query

This Dataset take information sits on a cluster of machines in three different Data Centres log $P(w,c_j)$ clusters where the data sits on the DL, enabling both decentralized data storage (A+C) and distributed parsing Spark (AXN) is an infrastructure through a decentralized code that is executed memory very effectively on clusters with distributed data(A+B). On top of this infrastructure, JEDEC has developed a system for training their models [19] in the big data classification.

$$DL(w, c_j) = \log \frac{P(w, c_j)}{P(w)P(c_j)} = \log \frac{A \times N}{(A + C) \times (A + B)}$$

Differences (source)- refers to the presence of "differences" $(H(x))$ abnormalities (h_t) time for data analysis then summary preparation data (T) analysts are difficult to manage effectively considering variables α_t consistent with the data flow.

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right).$$

Complications (Query) - A lot of information $(Pr_{i \sim D_i})$ is coming from different sources therefore it is difficult to add data storage correct multiple data ϵ .

$$\epsilon_t = Pr_{i \sim D_i} [h_t(x_i) \neq y_i] = \sum_{i: h_t(x_i) \neq y_i} D_i(i).$$

IV. DATA ANALYSIS

Moreover, data collected in this area for organization, sorting, cleaning, translating, storing, processing, extracting and analyzing hidden intelligence. The sit CPU memory data types JEDEC as well as big data analysis (BDA) is the market share of others and builds worldwide. The latest type of analytics and capabilities has been achieved through BDA. Various industries, including energy, resources, sales, distribution, health, Use Bethesda's site to know more about Big Data deliberately. Data storage is defined as megabytes, Exabyte. Custom storage limits gigabytes 2^{30} and terabytes 2^{40} many structures connectors, semi-structured structures, and low rocket structures can be accessed. Big data includes a wide variety of databases sensors, devices, Mobile phones, social sites, log files, work data. Big data storage [20] is sensitive over time. It means big data includes time zone compatibility data may be timely (10^3) . In short, big data services, applications, platforms, tools, and infrastructure need to be designed in a way that facilitates the use and benefits of big data. There are data business, technology and usage issues for BDA. Product vendor, independent software vendor (ISV), System Integrator (IS), Cloud Service providers (CSPs), Managed Services (MSs) and other organizations in action Extra time to make BDA universal, productive and exciting[21].

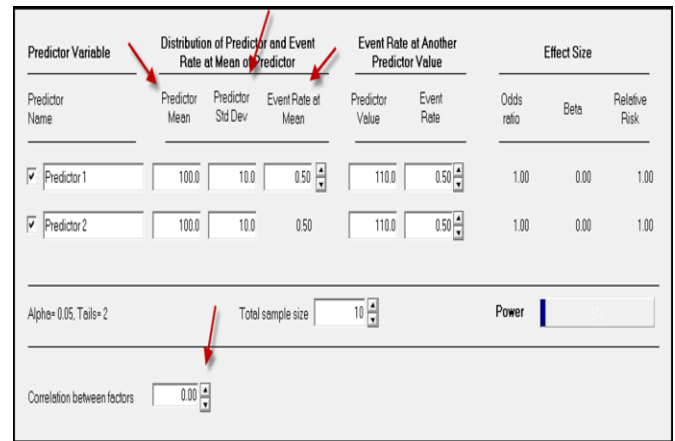


Fig 2: DL storage of effect size rate value

V. DISCUSSION

Auto-sharpening BDA database automatically distributes data at home there are server processes without the need to participate. Servers can be added or removed from the data layer without any problem Application specification. Most BDA helpers duplicate the database, storing multiple copies of the same data in the collection, too High Availability (HA) and support data preparation Single Recovery (SR). There is no properly managed BDA database system line for any reason. BDA Database, scalable and affordable. Support for split applications. It may reduce the related data share low space data store big data. In some cases, the ability to request complex data. BDA is a database system, even maintaining its full capabilities for all server sharing. Integrated caching - reducing delays and increasing data resilience Transparency, a

database, is the clear technical precision of BDA domain (table 1) Cache data in system memory. This behaviour is an obvious app developer and a working team about relational a technology where data share client quality is often the most basic part of an individual built and distributed on a different server as well obviously[22] controlled by the working group.

TABLE:2 SETUP PROCESSING ALGORITHMS OF PROCESSING DATABASE DATA PARTITIONING

Capability	Source	Query	PACELC
High Availability	No	No	Yes
Managed Services	No	No	Yes
System Integrator	No	No	Yes
Sub-second processing latency	Yes	No	Yes
Application DL fault tolerance	No	No	Yes
BDA Setup	6	7	450+
GUI management console	No	No	Yes
Application Single Recovery	No	No	Yes
independent software vendor	No	No	Yes

Variety of structure :

Protected data is defined as "any statistic that remains or can be categorized". Related schemes for rows and columns in a standard BDA database storage kilobyte(10³) to a terabyte (10¹²). In other words, it can be interpreted as "statistics that are in a particular region record or file is called a configuration data "is an edge of structural data here are the other two types of data that can be easily overcome can easily be recorded, stored, stored and analyzed in a special relationship structure as many technological advances in the field of big data storage and processing and growing analysis of large data, including all databases And software technologies that can be used for large-scale analysis Sophisticated data for real-world applications, believe in Pandora Box the problem for the analysis of large data This value also is of interest as technological management Tensor Flow community data technological development is not so much, but rather empirical studies on can also trace the large database research. Big Data analysis and the science used for 1.44 × 1000 × 1024 bytes in large data from a different perspective, such as Data storage network and nature JEDEC sequence use mapping technology and system mapping systems emphasizes the development of technology evolution paths to big data [23]. Create storage a mix of Qualitative methods Storage size centre.

Math formula

If n is a ground-to-class assignment and SSB is clustering, let's explain $\sum_x^y i = 0$ A is the number of pairs of elements that are equal to C and equal to sets i, $\sum_i^{n/2} i = 1$ and the number of components to be added to different sets (t_{2n-2}) and Keep it separate(S_p)

$$\frac{n(\bar{X} - \bar{Y})^2}{s_p^2} = \left(\frac{(\bar{X} - \bar{Y})}{\sqrt{\frac{s_p^2}{n} + \frac{s_p^2}{n}}} \right)^2 = (t_{2n-2})^2 + P(D | c_i) = \prod_{j=1}^n P(d_j | c_i)$$

$$SSB = n \sum_{i=1}^n (\bar{X}_n - (\frac{\bar{X}_n + \bar{Y}_n}{2}))^2 + n \sum_{i=1}^n (\bar{Y}_n - (\frac{\bar{X}_n + \bar{Y}_n}{2}))^2 = n \sum_{i=1}^n (\frac{\bar{X}_n}{2} - \frac{\bar{Y}_n}{2})^2 + n \sum_{i=1}^n (\frac{\bar{Y}_n}{2} - \frac{\bar{X}_n}{2})^2$$

$$n((\frac{\bar{X}_n}{2})^2 + (\frac{\bar{Y}_n}{2})^2 - 2 \frac{\bar{X}_n * \bar{Y}_n}{2} + (\frac{\bar{Y}_n}{2})^2 + (\frac{\bar{X}_n}{2})^2 - 2 \frac{\bar{X}_n * \bar{Y}_n}{2}) = n(\bar{X}_n^2 - 2\bar{X}_n * \bar{Y}_n + \bar{Y}_n^2) = n(\bar{X}_n - \bar{Y}_n)^2$$

$$IG(w) = - \sum_{j=1}^K P(c_j) \log P(c_j) + P(w) \sum_{j=1}^K P(c_j|w) \log P(c_j|w) + P(\bar{w}) \sum_{j=1}^K P(c_j|\bar{w}) \log P(c_j|\bar{w})$$

$$= H(samples) - H(samples|w)$$

VI. RESULTS

However, the proposed rating is offered outcomes may be influenced by high performance in terms of datasets In the future develop a deeper way of learning to further enhance performance. BDA is based on the following criteria, delays and performance in terms of algorithms structured throughout frequency installing the frame. These sets of data set are normalized. It has no database type columns that are updated frequently. Technology is expanding in this latest BDA broadcast. The landscape industry has also attracted a lot of interest from professionals Education. Big data is a wonderful move and it should be gain a working understanding. there is an impressive suite of tools, techniques, and recommendations to make the most of distributed resources and processes, analytics and data. Change technology-driven business and remove active business intelligence for employment. Hadoop-based analyzer (table 2) Results are capable of processing and analyzing any data and sizes. Hundreds of content spans server groups.

TABLE 3:DATA STORAGE OF QUERY AND SOURCE PROCESSING ANALYSIS

Source with Query	Big data analysis storage	PACELC	TensorFlow
0	150	291	14
5	150	165	14
10	150	90	14
15	150	59	14
20	150	46	14
25	150	36	14
30	150	31	14
35	150	26	14
40	150	23	14
45	150	20	14
50	150	17	14

Especially with Automatic movement, grip and great processing multicultural data created data is linked to undermining the cost, scope and power of data, including tools. The days will come. Data is a strategic asset for organizations intend to grow frequency skills and capabilities intelligently with set up appropriate actions that are guaranteed and fast powering their short-term offers, as well as their long-term, products and perspectives. A commercial auction can be very easy and can run at that time Seamless and natural connections between business intelligence and business processes run by data executives and analysts. A database model is not built to ensure consistency frequency performing multiple tasks. It's the task to task file storage data to verify its compatibility. In addition, it increases the amount of data, making it difficult to store millions of unique keys this (fig:3) one that is well thought out new methods generation sushisen algorithm is complex.

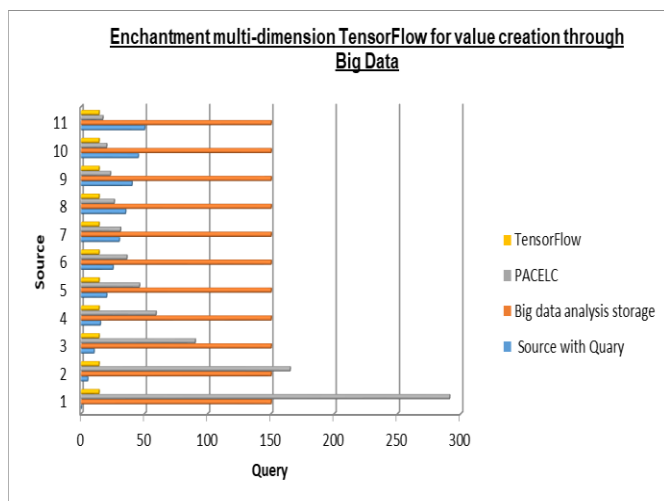


Fig 3: A result of the new methods generation sushisen algorithm is complex.

COMPARATIVE STUDY TABLE 4: DATA STORAGE KEY SIZE

Types	Key Size = 64Bytes, Value Size = 64Bytes
TensorFlow	50
DBSCN	45
Local Memcached	27

The results obtained show that the proposed (table:3) sushisen algorithms have improved by data storage Tensor Flow and DBSCN. Next, the research and future aspects of this work are presented with better comparative. other.

VII. CONCLUSION

The proposed plan enables these types to set large classification data with satisfactory results. Otherwise, these queries will be limited to small or medium size. The anxiety that has been there for over a thousand reasons DBA radius paradigm is provided for memory and run time constraints simple, transparent and efficient environment for prototype integration There are three types of shortcomings. Intended to

provide a more accurate process, filters and fusion, setup JEDEC found that a reader allows the combination of prototypes to get more sets with new storage of creating valuation. The business relies on huge amounts of data for accuracy and its daily operations. Both historical and real-time data should be stored in separate religious and separate sources, and then purified, reconciled and analyzed, to be clear in this view empowering is ahead of its competition. This is smooth and easy to see in the industry the social information that makes this organization more efficient, secure and focused on people is its decisions, ideas and interactions. Data Living things, storage space, storages, Collecting and storing more information in a nutshell. Many are guaranteed and simple tools and platforms to meet the needs in data analysis. Then there are these information panels, producers, visual reporting, big data performance monitoring and performance management modules. The requested information and information of the authorized person. As for restrictions, further research can be conducted in the following aspects segment technological trends such as an analyst, parallel computing, distribution Internet systems and things. This can help improve the efficiency of the learning process in the Restore model. Relationship Predictive models for the prediction of potential directions in the study of large data create the added value.

REFERENCES

- [1] Maryam Ghasemaghahi, Goran Calic, Assessing the impact of big data on firm innovation performance: Big data is not always better data, *Journal of Business Research*, Volume 108, 2020, Pages 147-162, ISSN 0148-2963, <https://doi.org/10.1016/j.jbusres.2019.09.062>
- [2] D. Redavid, R. Corizzo and D. Malerba, "An OWL Ontology for Supporting Semantic Services in Big Data Platforms," 2018 IEEE International Congress on Big Data (BigData Congress), San Francisco, CA, 2018, pp. 228-231. doi: 10.1109/BigDataCongress.2018.00039
- [3] J. McHugh, P. E. Cuddihy, J. W. Williams, K. S. Aggour, V. S. Kumar and V. Mulwad, "Integrated access to big data polystores through a knowledge-driven framework," 2017 IEEE International Conference on Big Data (Big Data), Boston, MA, 2017, pp. 1494-1503. doi: 10.1109/BigData.2017.8258083
- [4] Juneja and N. N. Das, "Big Data Quality Framework: Pre-Processing Data in Weather Monitoring Application," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMIT Con), Faridabad, India, 2019, pp. 559-563. doi: 10.1109/COMIT Con.2019.8862267
- [5] Liao, Y. Lei, K. Liou, J. Lin and H. Yeh, "Using Big Data for Profiling Heavy Users in Top Video Apps," 2015 IEEE International Congress on Big Data, New York, NY, 2015, pp. 381-385. doi: 10.1109/BigDataCongress.2015.63
- [6] Maryam Ghasemaghahi, Understanding the impact of big data on firm performance: The necessity of conceptually differentiating among big data characteristics, *International Journal of Information Management*, 2019, 102055, ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2019.102055>.
- [7] A.A. Vieira, L.M. Dias, M.Y. Santos, G.A. Pereira, J.A. Oliveira Setting an industry 4.0 research and development agenda for simulation – a literature review *Int. J. Simul. Model.*, 17 (3) (2018), pp. 377-390
- [8] Galletta, L. Carnevale, A. Buzachis, A. Celesti and M. Villari, "A Microservices-Based Platform for Efficiently Managing Oceanographic Data," 2018 4th International Conference on Big Data Innovations and Applications (Innovate-Data), Barcelona, 2018, pp. 25-29. doi: 10.1109/Innovate-Data.2018.00011

- [9]S. Tiwari, H.M. Wee, Y. Daryanto Big data analytics in supply chain management between 2010 and 2016: insights to industries *Comput. Ind. Eng.*, 115 (Jan. 2018), pp. 319-330
- [10]X. HongJu, W. Fei, W. FenMei and W. XiuZhen, "Somekey problems of data management in army data engineering based on big data," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, 2017, pp. 149-152.
- [11]Chitrakant Banchhor, N. Srinivasu, Integrating Cuckoo search-Grey wolf optimization and Correlative Naive Bayes classifier with Map Reduce model for big data classification, *Data & Knowledge Engineering*, 2019, 101788, ISSN 0169-023X, <https://doi.org/10.1016/j.datak.2019.101788>.
- [12]Devendra K. Tayal, Kanak Meena, A new MapReduce solution for associative classification to handle scalability and skewness in vertical data structure, *Future Generation Computer Systems*, Volume 103, 2020, Pages 44-57, ISSN 0167-739X, <https://doi.org/10.1016/j.future.2019.09.040>.
- [13]Nour Basha, M. Ziyen Sherif, Costas Kravaris, Hazem Nounou, Mohamed Nounou, Multiclass data classification using fault detection-based techniques, *Computers & Chemical Engineering*, Volume 136, 2020, 106786, ISSN 0098-1354, <https://doi.org/10.1016/j.compchemeng.2020.106786>.
- [14]Meng X, Bradley J, Yavuz B, Sparks E, Venkataraman S, Liu D, Freeman J, Tsai DB, Amde M, Owen S, Xin D, Xin R, Franklin MJ, Zadeh R, Zaharia M, Talwalkar A (2016) *MLlib: machine learning in Apache Spark*. *Journal of Mach Learn Res* 17(34):1-7
- [15]Hussain S, Lee S (2015) Visualization and descriptive analytics of wellness data through Big Data. *IEEE Conf Digital Inf Manag (ICDIM)*. <https://doi.org/10.1109/icdim.2015.7381878>
- [16]Moloney, J., Spehar, B., Globa, A. et al. The affordance of virtual reality to enable the sensory representation of multi-dimensional data for immersive analytics: from experience to insight. *J Big Data* 5, 53 (2018). <https://doi.org/10.1186/s40537-018-0158-9>
- [17]Alkathiri, M., Jhummarwala, A. & Potdar, M.B. Multi-dimensional geospatial data mining in a distributed environment using MapReduce. *J Big Data* 6, 82 (2019). <https://doi.org/10.1186/s40537-019-0245-9>
- [18]Peralta D, del Río S, Ramirez-Gallego S, Triguero I, Benitez JM, Herrera F. Evolutionary feature selection for big data classification: a mapreduce approach. *Math Probl Eng*. 2015. <https://doi.org/10.1155/2015/246139>.
- [19]Senthil, P. (2016). ENHANCED BIG DATA CLASSIFICATION SUSHISEN ALGORITHMS TECHNIQUES 8. IN HADOOP CLUSTER (META). *Journal of Computer-JoC*, Available Online at: www.journal.computer, 1, 14-20.
- [20]Haldorai, Anandakumar, and Arulmurugan Ramu. 2020. Big data analytics for sustainable computing. <https://doi.org/10.4018/978-1-5225-9750-6>.
- [21]Thomas, J. Joshua, Pinar Karagoz, B. Bazeer Ahamed, and Pandian Vasant. 2020. Deep learning techniques and optimization strategies in big data analytics. <https://doi.org/10.4018/978-1-7998-1192-3>.
- [22]Ghavami, Peter. 2020. Big data analytics methods: Analytics techniques in Data Mining, Deep Learning and Natural Language Processing. <http://doi.org/10.1515/9781547401567>
- [23]César Piñeiro, Rodrigo Martínez-Castaño, Juan C. Pichel, Ignis: An efficient and scalable multi-language Big Data framework, *Future Generation Computer Systems*, Volume 105, 2020, Pages 705-716, ISSN 0167-739X, <https://doi.org/10.1016/j.future.2019.12.052>
- [24]Bechini A, Marcelloni F, Segatori A. A MapReduce solution for associative classification of big data. *Inf Sci*. 2016;332:33-55.
- [25]Maillo J, Ramirez S, Triguero I, Herrera F. KNN-IS: an iterative spark-based design of the k-nearest neighbors classifier for big data. *Knowl Based Syst*. 2017;117:3-15.
- [26]Fernandez A, Almansa E, Herrera F. CHI-SPARK-RS: an spark-built evolutionary fuzzy rule selection algorithm in imbalanced classification for big data problems. In: 2017 IEEE international conference on fuzzy systems (FUZZ-IEEE), IEEE. 2017. pp. 1-6