International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-8 Issue-4, April 2019

A Survey on Machine Learning Model for Building Efficient Agriculture Management System in Cloud Computing

Namita Naganur, Kuldeep Sambrekar, Vijay S. Rajpurohit

Abstract*: Agriculture is considered as a backbone of major countries economic such as India, China, and United states. In these countries good agriculture productivity will aid in attaining better industrial growth. Timely information need to be passed to farmers for obtaining better agriculture productivity. Thus, the geographic information system (GIS) and remote sensing forecasting (RSF) technology deployed different kind of sensor to collect information such as wind, temperature, humidity level and so on and are stored in cloud computing platform. These data can later be analyzed by applying machine learning (ML) technique for forecasting weather related information to farmers. This paper presents a deep rooted survey of various existing agriculture management and ML based model for enhancing agriculture management system (AMS). From survey it can be seen existing access control and machine learning model incurs computation overhead for processing and storing information on cloud platform. Thus, this paper gives a research direction for modelling scalable and cost-effective agriculture management system using ML and cloud computing environment..

Index Terms*: Access control mechanism, Agriculture management system, BigData, Cloud computing, Machine learning.*

I. INTRODUCTION

Farming is the spine of Indian economy that contributes around 40 percent towards Gross National Product (GNP) and 70 percent of the public rely upon it. Along these lines, for a nation that is essentially subject to agro industries requires exact and convenient data of information, for example, crop developed, crop yield, and crop growth development condition. This data is vital factor in meeting nation's sustenance security and dispersion framework. Pre--harvest appraisals of yield generation are prerequisite for acquiring ideal procedures for arranging, appropriation, minimum support price fixing, transportation and capacity of basic agriculture items [1]. The space borne remote forecasting and global information sensing organization are giving data to farmers or subscriber to different farming applications. They give data, for example, minor harvest developed in divided land possessions, measurement of its impact on crop yield and discovery of crop damage because of supplements and infection. Productive agro-information access and storing method are required for improving crop [2].

With the colossal development of data innovation, for example, sensor technology, cloud computing (CC) environment, Internet of Things, and so on. A tremendous quantum of information is constantly created by different associations for different applications purposes. A proficient storing, processing (analysis) and access control techniques of extensive scale information with negligible expense is generally required. Number of access control mechanism is presented which are survey in next section [3], [4], [5], and [6]. From analysis it can be seen these model suffers from role computing and storage overhead. Further, the agro data collected are generally stored in cloud computing environment for providing scalability [13], [14]. Along with, data mining approach is applied on agro data stored in cloud computing environment [7], [8], [9], [10], and [11], From survey conducted it can be seen these model suffers in bringing trade-off between processing time and memory [12]. Along with the state-of-art model is designed considering single dimension data and no prior work has considered analyzing agro data. For overcoming research challenges, this paper conduct deep rooted survey and identify research problem and gives a possible research direction for building cost effective agriculture management system.

The research contribution of this work is as follows

- Firstly, this work carried out extensive survey of various access control mechanism, machine learning algorithm for performing analysis of high and multi-dimensional data, and also data placement problem for storing BigData.
- Presenting a novel solution for building scalable and cost efficient agriculture management system.
- Further, generated multi-dimensional data using different kind of sensor for performing analysis on agro data.
- Along with showed how sensor error will lead to false ripening forecasting of banana.

The manuscript is organized as follows. In section II, literature survey is presented. In section III research and industry gaps are identified and possible future research solution is given. Lastly, the research is concluded with future research direction of work.

Manuscript published on 30 April 2019.
***** Correspondence Author (s)
Namita Naganur, Department of Computer Science, KLS, Gogte Institute of Technology, Belagavi, Karnataka, India.
Kuldeep Sambrekar, Department of Computer Science, KLS, Gogte Institute of Technology, Belagavi, Karnataka, India.
Vijay S. Rajpurohit, Department of Computer Science, KLS, Gogte Institute of Technology, Belagavi, Karnataka, India.

II. Literature Survey

This section conducted experiment analysis of various state-of-art models for enhancing agriculture management system using cloud computing and machine learning algorithm.

a) Access control method:

In [3] presented access control mechanism, named RTBAC, which accomplish time-release, role using rule set, and secure information access in CC. Their model at the same time attain the accompanying properties such as fine-grained data accessing mechanism, any client with the higher access rule can get permission to use information and Post revocation the client can no longer have permission to access the cloud and data stored in cloud computing environment, time-release, the delicate information proprietor may determine a session period which implies that the recipient can decode the ciphertext until a session period key has been acquired, provable secureness, their model is secure under the MBDH and q-BDHIP presumptions. Altogether hypothetical investigation and execution assessment demonstrate the viability and productivity of their access control model. In [4], a complete ontology structure depicting the insensible distributed storage environments and access control mechanism using role set that ought to be set up to deal with this security issues. They have built up a method to possess cloud information utilizing insensible information structure characterized and have additionally executed the ObliviCloudManager (OCM) presentation that enables clients to deal with their information stored on CC environment by approving it before putting away it in an insensible information structure. Their model utilizes access control mechanism using role set and gathering is performed using archiving mechanism to possess and recover information productively.

In [5], they presented a trust models to give a better security for possessing information in distributed storage frameworks that utilizes cryptographic for modeling role based access control mechanism. The trust based method offers a way to deal with the proprietors and roles to decide the dependability of distinct role sets and clients in role based access control framework. Their security method considers role history/legacy and hierarchical structure/importance in the assessment of reliability of role defined.

In [6], presented a methodology that consolidates the B languages and UML for the determination and approval of role based access control rules at the procedure level. In [15], address this issue utilizing another framework with malicious activity identification and malicious user are revoked. To actualize this framework, they presented another threshold based cryptography mechanism that is hierarchical in nature, which executes the incomplete request key order, like role set features in hierarchical role based access control, asymmetric based cryptography model.

b) Machine learning model for performing analysis on high dimensional data:

In [7], [8], [9], [10], and [11] presented a model to bring a good performance tradeoff issues such as accuracy, memory and processing time. In [8], [9] addressed memory overhead issues considering high dimensional data. In [10], [11] overcome the processing time and accuracy issues but failed to address memory overhead issues in analyzing high dimensional data. In [12] they addressed processing time improvement and memory overhead issues by presenting a novel nearest neighbor classification algorithm. However, they considered only single dimension data. As a result, it cannot be applied to agro data which is multi-dimension in nature.

c) BigData Data placement problem:

In [13], evaluate challenges attained in geographically distributed CC information environment: which information to buy (information buying) and where to placement/duplication (replication) the information for data placement (DP). They demonstrated that the cumulated issue of information obtaining and DP inside a CC information environment can be seen as an storage area issue and is in this way NP-hard. Nonetheless, they provided a provable ideal computation for the instance of an information showcase obtained from solitary cloud computing server environment within data center (DC) and afterward sum up the structure from the solitary DC setup/configuration so as to build up a close ideal, polynomial-time calculation for a geographically distributed information environment. In [14], showed non-volatile environment, for example, solid state drives, will be an indispensable fragment of the developing storage chain of command on largescale high performance computing environmental frameworks. These gadgets can be on the computational hubs as a major aspect of a conveyed burst traffic administration. Further, it can be used externally also. Wherever they are situated in the chain of command, one basic structure issue is the solid state drives durability under substantial load of tasks at hand, for example, the checkpoint I/O for logical (i.e., both scientific and data intensive) algorithm. They presented another checkpoint arrangement advancement show which cooperatively uses both the burst traffic and the parallel storage framework to poses the checkpoints, with plan objectives of boosting calculation proficiency while ensuring the SSD perseverance necessities.

From survey it can be seen for building cost effective agriculture management system it is important to build an efficient access control model, construct data analysis technique utilizing machine algorithm, and efficient data placement technique on cloud computing environment.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

Table 1: Survey of various state-of-art access control, data placement and machine learning

Author name	Methodology used	Advantage	Disadvantage
F. Li et al., [3]	Presented access control mechanism, named RTBAC, which accomplish time-release, role using rule set, and secure information access in CC	Provides session based security	Does not provide dynamic role revocation
V. Narkhede et al., [14]	They have built up a method to possess cloud information utilizing insensible information structure	Very limited information is used for creating roles	It is not efficient as it does not provide session based security mechanism
L. Zhou et al., [5]	They presented a trust models to give a better security for possessing information in distributed storage frameworks that utilizes cryptographic for modeling role based access control mechanism	Offer higher security for role creation.	They incurs computation overhead. Thus increase cost of provisioning security.
S. Chehida et al ., [6]	They presented a methodology that consolidates the B languages and UML for the determination and approval of role based access control rules	Can be used in provisioning security for real-time environment.	It incurs storage overhead for storing role and access control information
Y. Zhu et al ., [15]	They presented another threshold based cryptography mechanism that is hierarchical in nature.	Provide dynamic revocation of user.	Does not provide session based security model.
L. Kuang [7], Y. Gong et al., [8], T. Ge et al., [9]	They presented a universal design for dimension reduction technique for BigData analysis	Addresses memory overhead issues considering high dimensional data	Processing time issues is not addressed.
L. Bao et al., [10], D. Cozzolino et al., [11].	Presented machine leaning model for edge detection patch matching and forgery detection.	Overcome the processing time and accuracy	Failed to address memory overhead issues in analyzing high dimensional data
L. Verdoliva et al., [12]	Presented a novel nearest neighbor classification algorithm.	Addresses processing time improvement and memory overhead	They considered only single dimension data. As a result, it cannot be applied to agro data which is multi-dimension in nature
X. Ren et al., [13]	Provided a provable ideal computation for the instance of an information placement obtained from solitary cloud computing server environment within data center (DC)	Attain close to ideal, polynomial-time calculation for a geographically distributed information environment.	Incurs high latency for data access.
Lipeng Wan et al., [14]	They presented another checkpoint arrangement advancement show which cooperatively uses both the burst traffic and the parallel storage framework	Reducing processing time for placing data.	User level quality of service requirement is not considered for placing data on cloud environment.

III. RESEARCH/INDUSTRIAL GAP WITH POSSIBLE FUTURE RESEARCH DIRECTION

This section identifies research/industrial gap and also provides future possible solution to design efficient agriculture management system. The research problem of existing model access control model they cannot efficiently perform concurrent transaction considering large number of stake holders. Further, it requires to maintain the responsibility of different stake holder. Thus increasing storage cost when adopting cloud computing storage infrastructure. Further, it incurs for evaluating role computation as it requires the roles to be reorganized and regrouping of roles and view creation process is not light weight in nature. Further, designing efficient machine learning for classifying agro data is challenging. The existing model aimed at addressing either processing time or memory efficiency. As agro data are high dimensional in nature addressing memory issues is challenging. To address tradeoff problems between processing time and memory [9] presented machine leaning model using k nearest neighbor algorithm. However, for improving processing time it requires additional memory. This is due to as using multiple random tress the memory size also increases linearly. Further, the existing models are designed to perform classification considering single dimension data. Designing efficient agriculture management system requires efficient data access mechanism for different stake holder. Further, to bring a good trade-off between accuracy, processing time and memory utilization an accurate and efficient data mining algorithm is required to performing classifying agriculture data into different level and store it in cloud computing storage. Lastly, it requires efficient data placement method for minimizing service cost. The envisioned architecture for building efficient agriculture management system is shown in Fig 1.

For building efficient access and control mechanism this work consider role computing using binarization technique rather than performing XML based computation. Further, role set are stored in a binary form rather than storing in XML format. Thus will aid in reducing storage overhead. Further, for performing classification considering multi-dimensional agro data a novel classification model will be designed using clustering method such as K-means and K-nearest neighbor classification algorithm or by using supervised method such as neural network.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

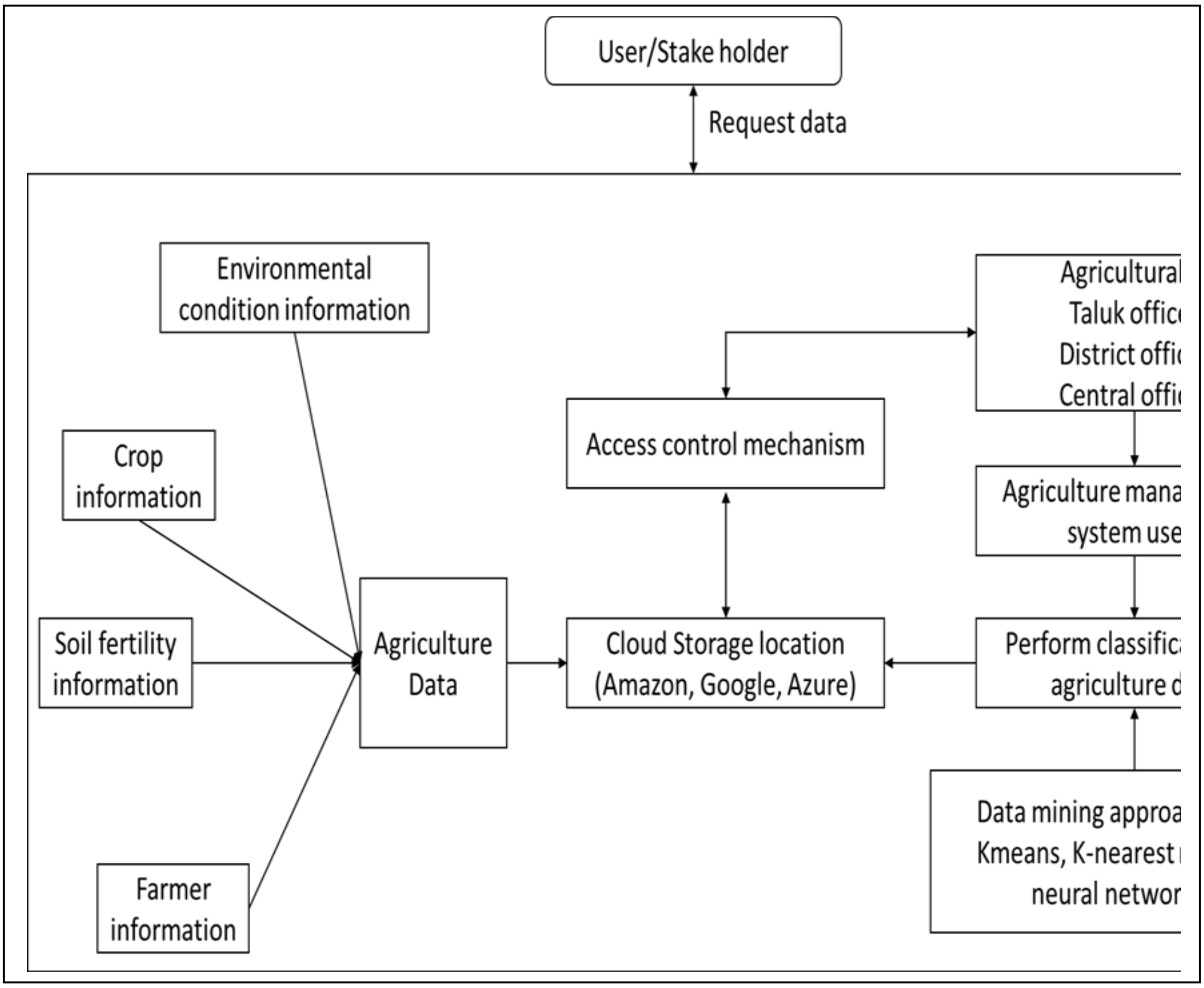

Fig. 1.Envisioned architecture of agriculture management system.

IV. MULTI DIMENSION AGRICULTURE DATA COLLECTION

This section describes the data collection process of agro data. For collecting agro data we considered using humidity, temperature sensor considering Sensirion SHT75 sensor. Along with used 4 MOX gas sensors. These sensor are used to monitor the background activity of banana. We considered different case study (i.e., every ten minutes the data is collected considering 12 hours). In this manner a time series of data is collected which composed of 7 days. The collected data can later be used to perform analysis using machine learning algorithm and see how external activity affecting banana ripening and other conditions. This work 4 MOX gas sensors to check how different sensor computing the background activity (i.e., identify variation in measurement). The sample multi-dimensional data collected is shown in Fig. 4 and Fig. 5. From Fig. 1 it can be seen how the measurement is varying considering different gas sensor. Sensor (S1) has measurement of 13.2041625, sensor (S2) has measurement of 10.5484325, sensor (S3) has measurement of 10.6071825, and sensor (S4) has measurement of 11.86528. Thus, improper measurement of sensor will lead to improper analysis in ripening forecasting of banana.

The Fig. 2, show average temperature (25.8752025) and humidity (55.60731) level of collected data.

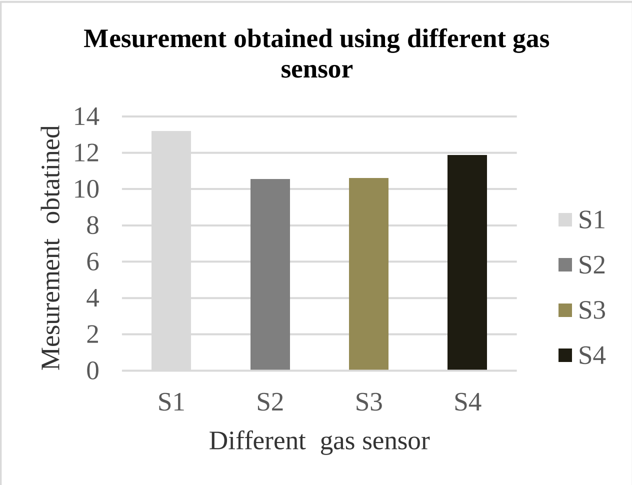

Fig. 2. Measurement obtained using different gas sensor.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)

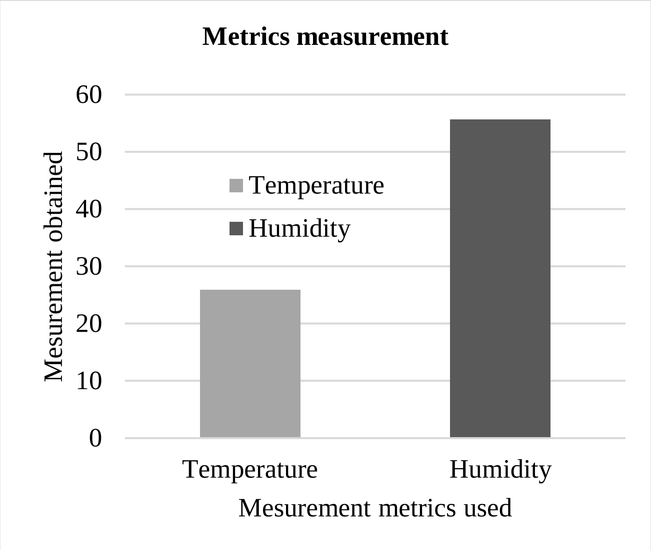

Fig. 3.Envisioned architecture of agriculture management system.

	id	time	R1	R2	R3	R4	Temp.	Humidity
2	0	-0.99975	12.8621	10.3683	10.4383	11.6699	26.2257	59.0528
3	0	-0.99947	12.8617	10.3682	10.4375	11.6697	26.2308	59.0299
4	0	-0.99919	12.8607	10.3686	10.437	11.6696	26.2365	59.0093
5	0	-0.99892	12.8602	10.3686	10.437	11.6697	26.2416	58.9905
6	0	-0.99863	12.8595	10.3688	10.4374	11.6699	26.2462	58.9736
7	0	-0.99835	12.8594	10.3691	10.4376	11.6702	26.2503	58.9583
8	0	-0.99807	12.8595	10.3695	10.4378	11.6702	26.2541	58.9445
9	0	-0.99779	12.8587	10.3696	10.4372	11.6698	26.2575	58.9319
10	0	-0.99752	12.8579	10.3693	10.4365	11.6694	26.2606	58.9206
11	0	-0.99723	12.8573	10.3692	10.4362	11.6693	26.2626	58.9103
12	0	-0.99695	12.8569	10.3695	10.4362	11.6692	26.2643	58.901
13	0	-0.99667	12.8565	10.3696	10.4363	11.6695	26.2658	58.8926
14	0	-0.99639	12.8569	10.3694	10.4362	11.6698	26.2671	58.885
15	0	-0.99611	12.8563	10.3695	10.4365	11.67	26.2683	58.8781
16	0	-0.99583	12.8558	10.3698	10.4368	11.6703	26.2695	58.8719
17	0	-0.99555	12.8553	10.3696	10.4367	11.6703	26.2705	58.8663
18	0	-0.99527	12.8549	10.3694	10.4366	11.6705	26.2708	58.8612
19	0	-0.99499	12.855	10.3698	10.437	11.6707	26.2707	58.8566
20	0	-0.99471	12.855	10.3699	10.4367	11.6704	26.2707	58.8524
21	0	-0.99443	12.8546	10.37	10.4366	11.6706	26.2706	58.8487
22	0	-0.99415	12.8543	10.3708	10.4359	11.6706	26.2706	58.8452
23	0	-0.99387	12.8539	10.3704	10.4357	11.6704	26.2705	58.8422

Fig.4.Sample multi-dimensional data obtained for performing analysis using machine learning method for first 12 hours.

	id	time	R1	R2	R3	R4	Temp.	Humidity
2	1	1.170528	13.1938	10.532	10.5945	11.8463	25.8593	55.7544
3	1	1.170806	13.1948	10.5338	10.5961	11.8487	25.8632	55.7345
4	1	1.171092	13.1952	10.5359	10.5974	11.8505	25.8667	55.7164
5	1	1.17137	13.196	10.5373	10.5982	11.8524	25.8688	55.7
6	1	1.171648	13.1973	10.5387	10.5993	11.854	25.8698	55.6852
7	1	1.171927	13.1984	10.5402	10.6005	11.8556	25.8708	55.6718
8	1	1.172205	13.1994	10.5413	10.601	11.8568	25.8717	55.6597
9	1	1.172494	13.1996	10.5426	10.6022	11.8585	25.8725	55.6487
10	1	1.172772	13.2005	10.5438	10.6035	11.8601	25.8732	55.6388
11	1	1.17305	13.2013	10.5447	10.6039	11.8611	25.8738	55.6298
12	1	1.173328	13.202	10.5453	10.604	11.8624	25.8744	55.6217
13	1	1.173606	13.2027	10.5463	10.6051	11.863	25.875	55.6143
14	1	1.173892	13.2033	10.5471	10.6058	11.8639	25.8754	55.6077
15	1	1.17417	13.2038	10.5479	10.6059	11.8648	25.8759	55.6017
16	1	1.174448	13.2043	10.5484	10.6061	11.8651	25.8763	55.5962
17	1	1.174726	13.2048	10.5491	10.6068	11.8658	25.8766	55.5913
18	1	1.175004	13.2052	10.5496	10.6069	11.8663	25.8769	55.5868
19	1	1.175293	13.2053	10.5502	10.6068	11.8663	25.8772	55.5828
20	1	1.175571	13.2056	10.5505	10.6071	11.8669	25.8775	55.5821
21	1	1.175849	13.2059	10.5509	10.6083	11.8673	25.8777	55.5815
22	1	1.176127	13.2062	10.5513	10.6088	11.8678	25.8779	55.581
23	1	1.176405	13.2065	10.5517	10.6094	11.8683	25.8781	55.5805

Fig.5.Sample multi-dimensional data obtained for performing analysis using machine learning method for second 12 hours.

V. CONCLUSION

This manuscript presented a deep rooted survey of BigData and machine learning model for building efficient agriculture management system. From extensive survey carried out it can be seen the efficient access control mechanism is required that reduces role computation overhead and also minimize roles and view (different stake holder) policies storage overhead. Thus, using binarization method for computing and storing could be an effective solution. Then, accurate and efficient classification method is required for considering multi dimension data is required. Using clustering method such as Kmeans and KNN could be effective solution when data obtained is not classified. Further, if agro data collected is labeled then using neural network based ML model could be an effective solution. Further, data placement is very crucial due to adoption of cloud computing platform. Thus, design multi-objective based placement strategies could be an effective solution. Lastly, this work collected multi-dimensional agriculture data. Meeting all these feature will aid in building scalable and cost-effective agriculture management system. Future work will consider implementing the envisioned agriculture management system using real-time agro data and cloud computing environment. Along with considering generating different king of fruits, vegetable, and crops.

REFERENCES

1. Bernard A (2003) A DSS is an Integration of Web-Based Programs. Geographic Information Systems (GIS) Capabilities and Databases, USA, pp: 484-495.
2. David H (2006) Web-Based Systems, Client can use any internet-Connected computer or web-enabled Mobile Phone or PDA to gain Real time Access to the data, USA.
3. F. Li, Y. Wang, J. Xiong and R. Xie, "Role and Time-Based Access Control with Efficient Revocation for Cloud Storage," 2016 IEEE First International Conference on Data Science in Cyberspace (DSC), Changsha, pp. 284-289, 2016.
4. V. Narkhede, K. Joshi, A. J. Aviv, S. G. Choi, D. S. Roche and T. Finin, "Managing Cloud Storage Obliviously," 2016 IEEE 9th International Conference on Cloud Computing (CLOUD), San Francisco, CA, pp. 990-993, 2016.
5. L. Zhou, V. Varadharajan and M. Hitchens, "Trust Enhanced Cryptographic Role-Based Access Control for Secure Cloud Data Storage," in IEEE Transactions on Information Forensics and Security, vol. 10, no. 11, pp. 2381-2395, Nov. 2015.
6. S. Chehida, A. Idani, Y. Ledru and M. Kamel Rahmouni, "Combining UML and B for the specification and validation of RBAC policies in business process activities," 2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS), Grenoble, pp. 1-12, 2016.
7. L. Kuang, L. T. Yang, J. Chen, F. Hao and C. Luo, "A Holistic Approach for Distributed Dimensionality Reduction of Big Data," in IEEE Transactions on Cloud Computing, vol. 6, no. 2, pp. 506-518, 2018.
8. Y. Gong, S. Lazebnik, A. Gordo, and F. Perronin, "Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 12, pp. 2916–2929, 2013.
9. T. Ge, K. He, Q. Ke, and J. Sun, "Optimized product quantization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 4, pp. 744–755, april 2014.
10. L. Bao, Q. Yang, and H. Jin, "Fast edge-preserving patchmatch for large displacement optical flow," IEEE Transactions on Image Processing, vol. 23, no. 12, pp. 4996–5006, 2014.
11. D.Cozzolino, G.Poggi, and L.Verdoliva, "Efficient dense-field copymove forgery detection," IEEE Transactions on Information Forensics and Security, vol. 10, no. 11, pp. 2284–2297, 2015.
12. L. Verdoliva, D. Cozzolino and G. Poggi, "A Reliable Order-Statistics-Based Approximate Nearest Neighbor Search Algorithm," in IEEE Transactions on Image Processing, vol. 26, no. 1, pp. 237-250, 2017.
13. X. Ren, P. London, J. Ziani and A. Wierman, "Datum: Managing Data Purchasing and Data Placement in a Geo-Distributed Data Market," in IEEE/ACM Transactions on Networking, vol. 26, no. 2, pp. 893-905, April 2018.
14. Lipeng Wan, Qing Cao, Feiyi Wang, Sarp Oral "Optimizing checkpoint data placement with guaranteed burst buffer endurance in large-scale hierarchical storage systems," Journal of Parallel and Distributed Computing, Volume 100, Pages 16-29, 2017.
15. Y. Zhu, G. Gan, G. Ruiqi and D. Huang, "PHE: An Efficient Traitor Tracing and Revocation for Encrypted File Syncing-and-Sharing in Cloud," in IEEE Transactions on Cloud Computing, vol. PP, no. 99, pp. 1-1, 2016.

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication (BEIESP)