Educational Administration: Theory and Practice
2024, 30(1), 3358-3366
ISSN: 2148-2403
https://kuey.net/

Research Article

A Detailed Study On Google Cloud Compute, Storage Capabilities And Its Usages In Investment Banking For Recomending Probable Clients

Shivendra Singh Chauhan[1*], Prof. (Dr). Hari Om Sharan[2]

[1*]Research Scholar, Department of Computer Science, Faculty of Engineering & Technology, Rama University, Uttar Pradesh Kanpur, Shiveny2k@gmail.com
[2]Professor, CSE, FET, RAMA University, Uttar Pradesh, Kanpur, deanengineering@ramauniversity.ac.in

Citation: Shivendra Singh Chauhan, et al, (2024), A Detailed Study On Google Cloud Compute, Storage Capabilities And Its Usages In Investment Banking For Recomending Probable Clients, *Educational Administration: Theory and Practice*, *30*(1), 3358-3366
Doi: 10.53555/kuey.v30i1.7178

ARTICLE INFO

ABSTRACT

At present era of innovation, cloud infrastructure through efficient computing power and huge storage potential has enabled big data and machine learning capabilities a lot. Storing huge amount of data, managing this, extracting required data in a click within few milliseconds is now possible with different innovation that had happened and continuing in cloud computing and data storage landscape. Based on humongous available data and its analysis different trading & sales desk can be supported to find potential investors for a particular debt / security / commodity / equity and take more efficient, tailored and quality decision.
This was not an easy journey and there is a wealthy history associated with this journey. This paper talks more about Google Cloud Infrastructure its Compute & Storage capabilities, its usages in investment banking industry.

Keywords: Big data, Google Cloud, Cloud Infrastructure, Cloud Compute, Cloud Storage Capabilities, Client's Recommendation, Investment Banking

INTRODUCTION:

Different big names like: - Amazon, Google etc., has been continuously working with data and artificial intelligence since decades. Google is working in this field since 1998 and in 2008 (after a decade) they launched Cloud Platform to provide secure and flexible computing and storage services.

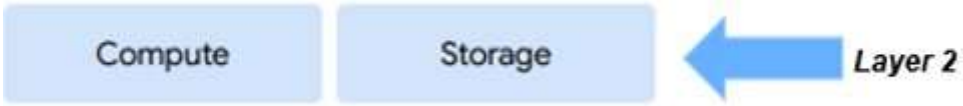
Google and its Cloud Infrastructure:

Google cloud infrastructure can be thought in form of three different layers: -

- At the base layer is networking and security, which lays the foundation to support all of Google's infrastructure and applications.
- On the next layer sit compute and storage. Google Cloud separates, or decouples, as it's technically called, compute and storage so they can scale independently based on need.
- And on the top layer sit the big data and machine learning products, which enable you to perform tasks to ingest, store, process, and deliver business insights, data pipelines, and ML models. And thanks to Google Cloud, these tasks can be accomplished without needing to manage and scale the underlying infrastructure.

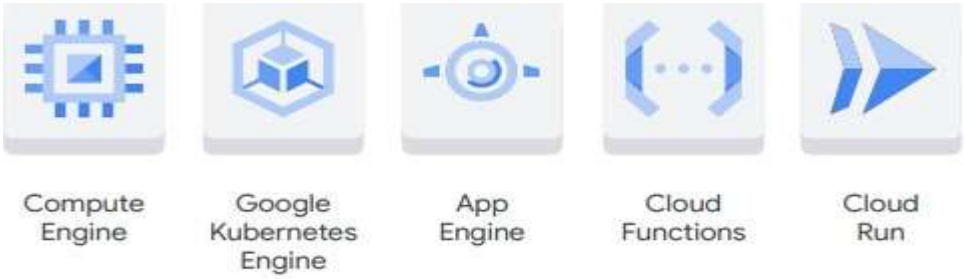
Surprisingly above listed tasks can be completed without needing to scale and manage the underlying

infrastructure.
In this paper our focus will be more on layer two so let's focus more on Compute and Storage capabilities of cloud.



Compute:
Google cloud provide following computing services.

- Compute Engine
- Google Kubernetes Engine
- App Engine
- Cloud Functions
- Cloud Run



Organizations like Investment Banking Industry, with growing data needs often require lots of compute power to run big data jobs, stored procedures, or queries. Almost for every organization compute power only grows so for meeting these ask Google Cloud Platform offers a range of compute services which are listed above, let's talk about individual component below.

Compute Engine:
Compute Engine is an Infrastructure as a service (IaaS), which provides compute, storage and network resources virtually and this is similar to physical data centers. We can use the virtual compute and storage resources in same way as we manage them locally.

So compute engine in summary can be captured by following key points:-

- IaaS Offering
- Compute
- Storage
- Network
- Maximum Flexibility

Kubernetes Engine:
Kubernetes Engine runs containerized applications in a cloud environment. A container represents code

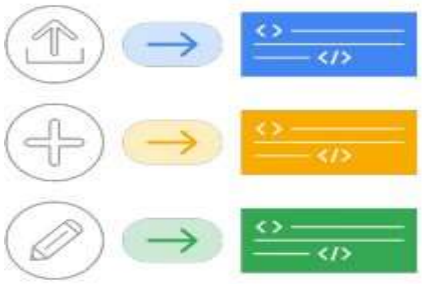
packaged up with all its dependencies.

App Engine:

The App Engine is *Platform as a Service (PaaS)* offering, this bind code to libraries that provide access to the infrastructure application needs. This capability allows more resource to be focused on application logic.

Cloud Functions:

The fourth is Cloud Functions, this is also referred as *Functions as a Service*, this feature executes code in response to events, like when a new file is uploaded to Cloud Storage. It's a completely serverless execution environment and in this environment, we don't need to install any application or software locally, to run the code we are free from provisioning and managing servers.



Cloud Run:

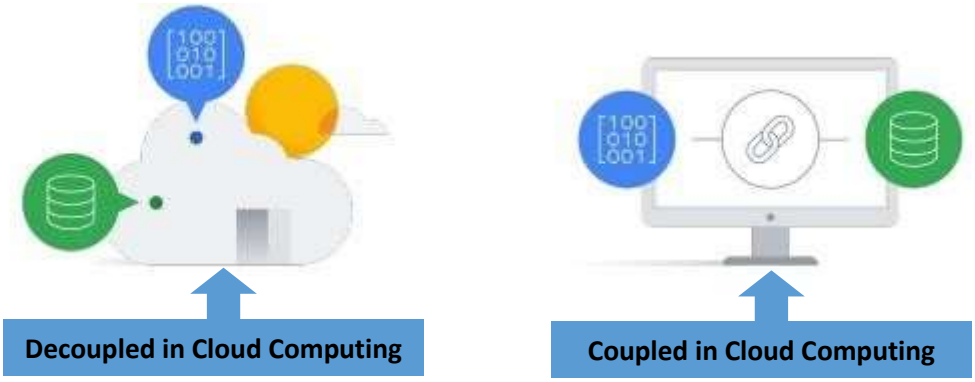
Cloud Run is a fully managed compute platform that enables us to run request or event- driven stateless workloads without having to worry about servers. This feature abstracts away all infrastructure management so we can focus more on writing code. It automatically scales up and down, so we also need to worry about scale configuration. Costing also very efficient with cloud run and we only need to pay for the resource that we have used.

So Run in summary can be captured by following key points: -

- Fully managed platform
- Let us focus on writing code.
- Auto Scale up and down
- Charges only for used resource.

Storage Capabilities:

One of the major differences between cloud computing and desktop computing is – in Cloud, Compute and Storage are decoupled and it provides better scaling capabilities



Google Cloud Database and Storage Services:

Google offers following database and storage services: -

- Cloud Storage
- Cloud Bigtable
- Cloud SQL
- Cloud Spanner
- Firestore

BigQuery The goal for above products is to reduce effort and time needed to store data. This means creating an elastic storage bucket directly in a web interface or through command line.

Storage Offerings:

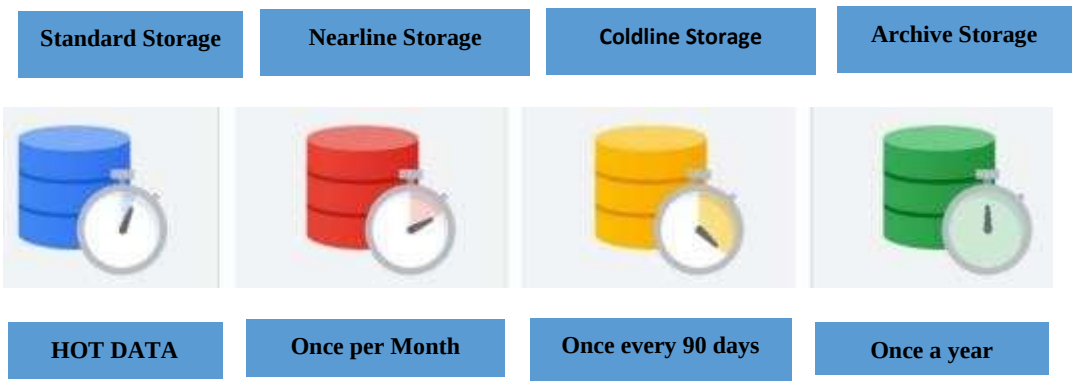
Google cloud platform offers following storage options.

- Relational databases
- Non-relational databases
- Worldwide object storage

Selecting right data base option to store and process depends on organizations and their business need.

Google Cloud Storage Primary Storage Classes:
It has four primary storage classes:

1. Standard Storage
2. Nearline Storage
3. Coldline Storage
4. Archive Storage

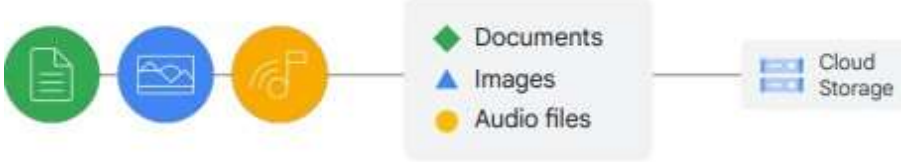


Standard Storage:
This is considered best for very frequently accessed or "hot" data. This is suitable for data that need to be stored for very short span of time.

Nearline Storage:
This is best for storing infrequently accessed data, like reading or modifying data once per month or less, on average.

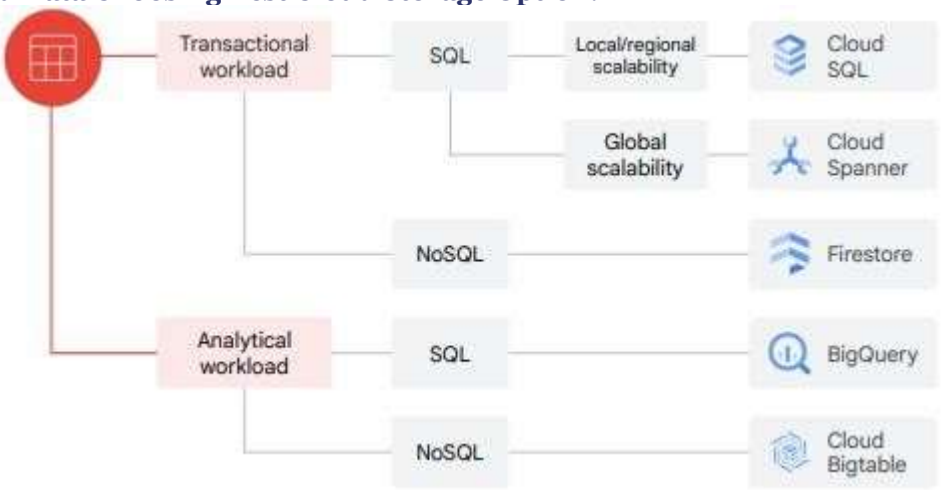
Coldline Storage: This is low-cost option for storing infrequently accessed data. If we talk about access frequency of coldline storage, then it come around 90 days.
Archive Storage: This is the lowest-cost option used ideally for data archiving, backup, disaster recovery. It is the best choice if we use to access data once in a year.

Cloud Storage: - Unstructured data
Unstructured data stored in non-tabular form such as documents, images, audio or video files. Unstructured data is usually suited to Cloud Storage, but BQ (Big Query) now offers the capability to store unstructured data as well.



Cloud Storage: - Structured data
This represents information stored in form of tables, rows, and columns.

In Structured Data Choosing Best Cloud Storage Option:



Structured data comes in following two types: -

1. Transactional workloads

This is used for online transaction processing systems which are used when fast data inserts and updates are required to build row-based records. This maintains a snapshot. This requires standardized queries that impact only a few records. So if your data is transactional and you need to access it using SQL then Cloud SQL and Cloud Spanner are two options.

- **Cloud SQL:** - Work best for local or regional scalability
- **Cloud Spanner:** - Best to scale Database globally.

Firestore: - Is a transactional No-SQL, document-oriented database.

2. Analytical workloads

This is used for Online Analytical Processing Systems, which are used when entire datasets need to be read. They often require complex queries, for example, aggregations.
So if ask is for analytical workload that require SQL commands then Big Query is best option. Alternatively Cloud Bigtable provides a scalable NoSQL solution for analytical workloads. Best for real-time, high throughput applications that require only millisecond latency.

Real Life Usages of Cloud Capabilities in Investment Banking: -

Leveraging huge client data stored in Google Cloud Platform to augment Sales desk Capability by analysing Clients / Customers who have previously traded in or displayed an interest in the type of debt / security / commodity / equity currently being offered by an Investment Institution. Any application / webapp powered by machine learning algorithms can generate list of potential clients / investors based on few key attributes of a new offering. Recommended solution will enable sales desk to place offerings in line with a particular Clients need.

Client Recommendation Application: - Wireframe

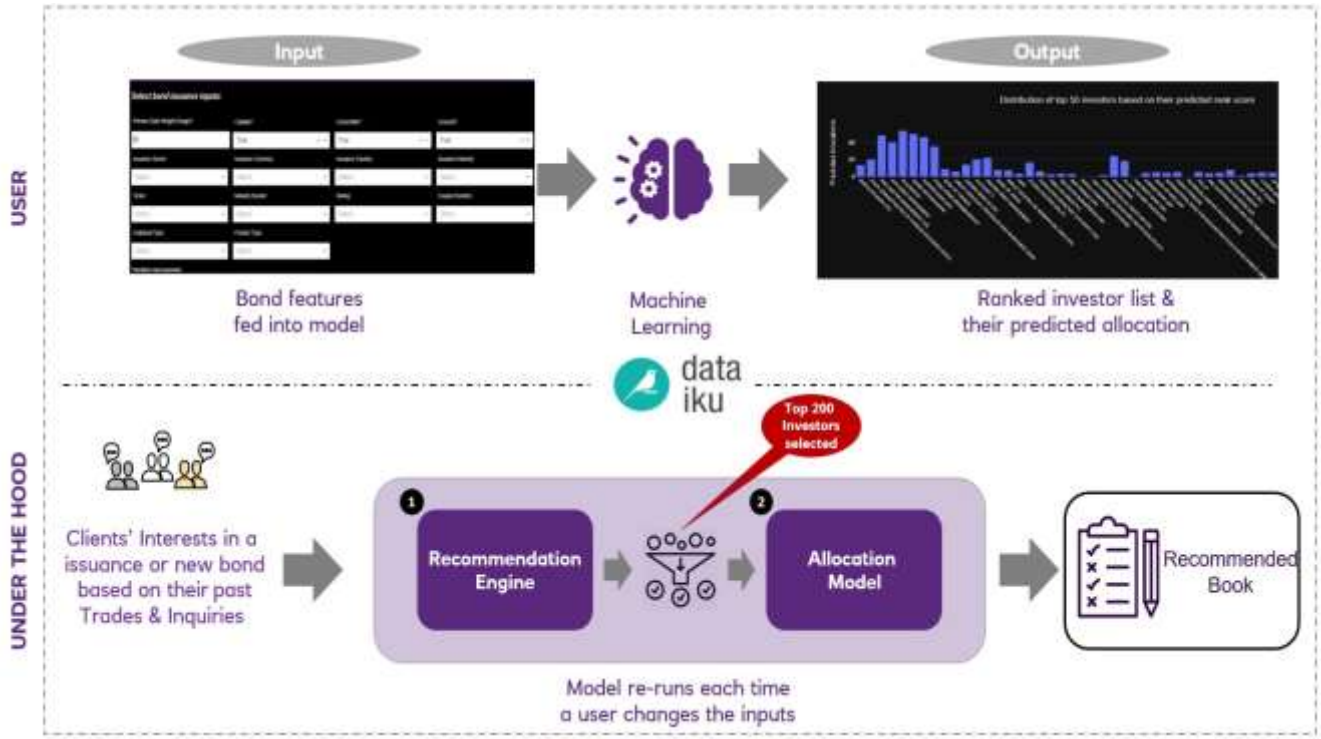
This consist of following three data sources: -

- Book Builder Data
- Client Enquiries Data from different markets / exchanges / sources
- Instrument Details Data from different markets / exchanges / reference data sets

What is defined as Client Interest: - Any client who has traded in Primary Market or on Secondary Market on same offering. E.g. France 20yr debt can be decoded as: -
A Client who has participated in a: -

- ❖ 20-year France Bank Debt
  - ○ Sector = Bank; Issue Country = France; Tenor = 20yrs
- ❖ Long dated Euro Bank Debt
  - ○ Sector = Bank; Currency = EURO; Tenor >= 20yrs

Similarly in secondary market if client has participated in a same / similar offering, it will be an eligible client for same use case.



Client Recommendation Application: - Working Flow

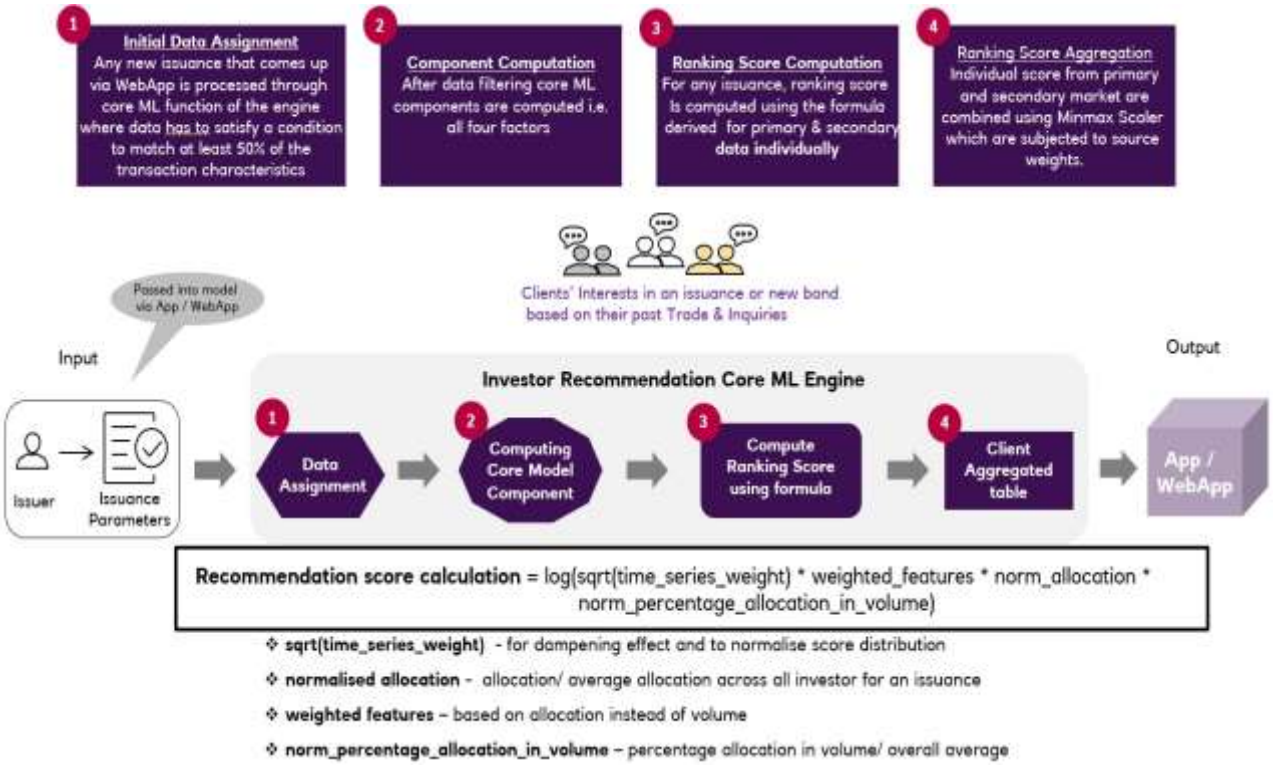
Different **Input** Fields Based on Which Recommendation Engine recommends probable clients are as follows: -

- ❖ Primary Data Weight (%age)*
- ❖ Callable*
- ❖ Convertable*
- ❖ Covered*
- ❖ Issuance Sector
- ❖ Issuance Currency
- ❖ Issuance Country
- ❖ Insurance Industry
- ❖ Ticker
- ❖ Maturity Bucket
- ❖ Rating
- ❖ Coupon Buckets
- ❖ Collateral Type
- ❖ Product Type

*Mandatory input parameters

Output based on above input parameters: -

- ❖ List of clients with their predicted score.



Probable Client Ranking Recommendation – Framework

In above use case we have seen how huge set of clients and their interest's data on different instruments, securities listed and traded in different markets are kept in Cloud platform for our analysis and recommending customers based on analysis result. Please note this data is stored from different exchanges / markets and it could be in Petabyte or higher units, querying such humongous data using Big Query and recommending probable clients to sales desk in milliseconds itself talks about Cloud Compute Power and its Storage Capability.

Other Usages / Use-cases in Investment Banking Industry and Benefits:

Cloud usage in Investment banking industry is growing rapidly. As per the research happened in 2021 Google Cloud share is 9% worldwide.

According to Stanford University's 2019 AI Index Report states, since 2012, the required computing power has been doubling approximately every three and a half months.

If we talk around Investment banks then they always need best rates (bid and ask prices, foreign exchange rates etc.) from different exchanges. Business expects to bring best deals at best rates, provide good currencies exchange rates (FX rates), on time regulatory reporting etc. so fulfilling these asks and storing huge historical transaction plus valuation data for reporting purpose they need excellent computing capabilities and huge data storage, fortunately Google Cloud Platform provide all this, hence in huge demand in banking industry.

Investment banks are using GCP for following benefits: -

- Quick decision making in Sales and Trading
- Enhanced Security
- Supports huge data storage and fast retrieval.
- Easier data analysis
- For enhanced quality controls
- Scalability
- Cost saving
- Improved performance
- Increasing computing power
- Risk Simulation
- Regulatory reporting

Reference:

1. https://cloud.google.com/
2. M. Beyer, ``Gartner says solving `Big Data' challenge involves more than just managing volumes of data," Tech. Rep., AaltoDoc, Aalto Univ., 2011.
3. R. Mital, J. Coughlin, and M. Canaday, ``Using big data technologies and analytics to predict sensor anomalies," in *Proc. Adv. Maui Opt. Space Surveill. Technol. Conf.*, Sep. 2014, p. 84.

4. Essentials of Cloud Computing: A Holistic Perspective by Chellammal Surianarayanan, Pethuru Raj Chelliah.
5. A Brief Guide to Cloud Computing: An Essential Introduction to the Next Revolution in Computing: An essential guide to the next computing revolution by Christopher Barnatt
6. Y. Wang, L. Kung, W. Y. C. Wang, and C. G. Cegielski, "An integrated big data analytics-enabled transformation model: Application to health care," *Inf. Manage.*, vol. 55, no. 1, pp. 64–79, Jan. 2018.
7. R. Khan, S. Khan, R. Zaheer & S. Khan, "Future Internet: The internet of things architecture, possible applications, and key challenges," In *Proceedings of international conference on frontiers of information technology*, pp. 275-260, 2012.
8. Ilapakurti, J. S. Vuppalapati, S. Kedari, S. Kedari, C. Chauhan, and C. Vuppalapati, "iDispenser #x2014; Big Data Enabled Intelligent Dispenser," in *2017 IEEE Third International Conference on Big Data Computing Service and Applications (BigDataService)*, pp. 124–130, 2017.
9. Y. Wang, L. Kung, and T. A. Byrd, "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations," *Technol. Forecast. Soc. Change*, vol. 126, pp. 3–13, Jan. 2018.
10. M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I.A.T. Hashem, A. Siddiqa, I. Yaqoob "Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges," *IEEE Access*, vol. 5, pp. 5247–5261, 2017.
11. E. Ahmed, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Imran Khan, Abdelmuttlib Ibrahim Abdalla Ahmed, Muhammad Imran, Athanasios V. Vasilakos, "The role of big data analytics in Internet of Things," *Computer Networks*, vol. 129, pp. 459–471, Dec. 2017.
12. Acharjya, D.P., Ahmed, K., "A survey on Big Data analytics: challenges, open research issues, and tools." in *Int. J. Adv. Comput. Sci. Appl*. Vol.7, issue 2, pp. No.- 511–518, 2016.
13. F. Constante Nicolalde, F. Silva, B. Herrera, and A. Pereira, "Big Data Analytics in IoT: Challenges, Open Research Issues and Tools," in *Trends and Advances in Information Systems and Technologies*, Cham, 2018, pp. 775–788.
14. S. Foundation: Spark 0.8.0: This document gives a short overview of how Spark runs on clusters, to make it easier to understand the components involved, 2014, https://spark. apache.org/docs/0.8.0/cluster-overview.html
15. V. Morabito, "Managing change for big data driven innovation," in *Big Data and Analytics*. Springer, 2015, pp. 125–153.
16. Bhardwaj, S. Bhattacherjee, A. Chavan, A. Deshpande, A. J. Elmore, S. Madden, and G. Parameswaran, "Datahub: Collaborative data science & dataset version management at scale," *arXiv preprint arXiv:1409.0798*, 2014.
17. F. Farber, S. K. Cha, J. Primsch, C. Bornh¨ovd, S. Sigg, and W. Lehner, "Sap hana database: data management for modern business applications," *ACM Sigmod Record*, vol. 40, no. 4, pp. 45–51, 2012.
18. Y. Zhuang, Y.Wang, J. Shao, L. Chen, W. Lu, J. Sun, B.Wei, and J. Wu, "D-ocean: an unstructured data management system for data ocean environment," *Frontiers of Computer Science*, vol. 10, no. 2, pp. 353–369, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11704-015-5045-6
19. Isard, M., Budiu, M., Yu, Y., Birrell, A., Fetterly, D. Dryad, "distributed data-parallel programs from sequential building blocks" in ACM SIGOPS Oper. Syst. Rev. 41, pp. No.- 59–72, 2007.
20. Kelly, J.: Apache Drill Brings SQL-Like, Ad Hoc Query Capabilities to Big Data (2013). http://wikibon.org/wiki/v/Apache_Drill_Brings_SQL-Like,_Ad_Hoc_Query_Capabilities_to_Big_Data
21. C.L.P., Chen, C.Y. Zhang, "Data-intensive applications, challenges, techniques, and technologies: a survey on Big Data." In *Inf. Sci.* 275, pp. no. -314–347, 2014.
22. G. Ingersoll, "Introducing apache mahout: Scalable, commercial-friendly machine learning for building intelligent applications," White Paper, IBM Developer Works, pp. no. - 1- 8, 2009.
23. C.-W. Tsai, C.-F. Lai and A. V. Vasilakos, "Future Internet of Things: open issues and challenges," *Wireless Netw*, vol. 20, no. 8, pp. 2201–2217, Nov. 2014, DOI: 10.1007/s11276-014-0731-0.
24. M. Chen, S. Mao, and Y. Liu, "Big Data: A Survey," *Mobile Netw Appl*, vol. 19, no. 2, pp. 171–209, Apr. 2014, DOI: 10.1007/s11036-013-0489-0.