An Indexing Approach of Historical States on Hyperledger Fabric

Khin Su Su Wai [1], Nwe Nwe Myint Thein [2] and Dim En Nyaung [3][+]

[1][2][3] University of Information Technology, Yangon, Myanmar

Abstract. A blockchain is a shared database that saves an unalterable ledger of the transaction. A public blockchain is the simplest to define and everyone can join the most active applications. It can only operate a limited number of transactions and storage size. The processing time of the transaction on the block is also slow. A consortium blockchain supports many of the same advantages with private blockchain but operates under the leadership of a group instead of one organization. The transaction speed of a private blockchain is quicker than public blockchain. The transaction fee in private blockchain is not very costly. Hyperledger Fabric is private blockchain and permissioned network. Although Hyperledger Fabric is a modular architecture, great performance, and scalability, it cannot provide efficient indexing on file-system. In this paper, a new indexing approach is proposed to overcome the issue of indexing in the past states. A user-defined time for each transaction is added as an additional parameter to access into blockchain. This approach will provide the performance for Hyperledger Fabric.

Keywords: blockchain, database, Hyperledger Fabric, file-system, user-defined time.

1. Introduction

Blockchain has become a growing popular and centre of the technology discussion by its network construction. Blockchain technology is one of the most popular in the current year. It is a new technology and can be considered an important building block. It is developing in many applications such as medical treatment, IOT, education, insurance, and financial market. Hyperledger Fabric default supports the ledger states as key-value pairs. The state database of blockchain ledger can be modified. The endorsing peers will result in changes to the world state. If the state is modified every time, the version number of a state is increased.

A block comprises block header, block data, and block metadata. Each block header consists of the block number, current block hash, and previous block hash. The block number is initiated as integer 0 in the genesis block and increased by 1 for every new block appended to the blockchain. The current block hash is the hash of all the transactions held in the current block. The hash from the previous block in the blockchain is copied in the previous block hash. Block Data contains a list of transactions sequences. Each transaction involves header, cryptographic signature, proposal, response, and endorsement. Block metadata consists of the time when the block is created, as well as the certificate, public key, and signature of the block writer. The block committer also adds a valid/invalid indicator for every transaction [1].

The rest of the paper is organized as follows. Section 2 is an illustration of related works with the proposed topic. The type of blockchain is presented in section 3. The Hyperledger Fabric architecture is discussed in section 4. State database is expressed in section 5. The blockchain is exposed in section 6. The proposed new indexing approach is discussed in section 7. Finally, this paper is concluded in section 8.

[+] Corresponding author. Tel.: +959-428127002; fax: +951-664250.
E-mail address: khinsusuwai@uit.edu.mm.

2. Related Works

J.J. Bambara and et al. [2] described a particular aspect of blockchain technology and its use cases. This is coupled with comprehensive treatment for getting started as a designer and developer of blockchain applications. This includes the Ethereum technology stack, code, and deployment techniques.

D. Puthal and et al. [3] discussed a comprehensive review of the blockchain by highlighting the working model of the blockchain. Consensus algorithms are described with different applications and use cases. Finally, this is concluded with presenting different security challenges.

E. Ademi [4] took a look at the current research on the challenges and limitations of the blockchain, and how these challenges affect the user. Recommendations on future research directions are provided for researchers. The proposed idea is to optimize the storage of the blockchain.

H. Gupta and et al. [5] proposed the two indexing models to improve the performance of temporal queries on Fabric. The first model is created by collecting each event inserted by a Fabric transaction and saving temporally close events together on Fabric. The Index construction time is 34% of the data ingestion time. This model is not clearly scalable. So, they discussed the next indexing model M2 which does not have separate indexing and hence does not have these short-comings. The second model is developed to keep the event count intact. It tags some metadata to each event being inserted on Fabric s.t. temporally close events share the same metadata. They [6] also discussed variants on these two models. They showed that which model is the best to handle the skew present in Fabric data.

3. Type of Blockchain

Blockchain systems can be classified into three types as public, consortium and private. The public blockchain is open and everyone in the various organizations can join the network. It is also called permission-less networks. It doesn't have privacy for transactions and only provides a weak concept of security. Proof-of-work (PoW) or Proof-of-stake (PoS) consensus mechanism is processed to write the public blockchain. Bitcoin is a type of public blockchain [4]. Consortium blockchain permits everyone or only participants to obtain a hybrid access method. Consortium blockchain platforms have many of the same advantages as a private blockchain. It doesn't ensure immutability and irreversibility. The type of consortium blockchain is Ethereum [7]. Private blockchain supports only predefined organization of users to write data to the blockchain. Everyone can't access this type of blockchain. It runs a fully centralized network. It copies the ledger and corporate to reach a consensus for updating. It can easily update the rules of a blockchain, revert transactions. Hyperledger Fabric is the type of private blockchain [7]. The information released from existing estimation is shown in Table. 1 [8].

Table. 1: Comparison of blockchain

Name	Transaction rate	Transaction fee	Network type	Network Access
Bitcoin	3(avg;), 7(max;)	Yes	Public	Permission- less
Ethereum	Depend on eth	Yes	Public (or) Private	Permission- less
Hyperledger fabric	>10k	No	Private	Permissioned

4. Hyperledger Fabric Architecture

Hyperledger Fabric is a private blockchain infrastructure. It saves CPU and network utilization. Hyperledger was established under the Linux Foundation. When considering a permission network, blockchain use case needs to comply with data protection regulations. Hyperledger Fabric smart contracts are called chaincode. Chaincode is software that defines assets and related transactions. Chaincode is called when an application needs to interact with the ledger. Chaincode can be written in Golang or Node.js. Hyperledger Fabric is designed with a modular architecture. It allows components, such as ledger database, consensus mechanism, and membership services, to be plug-and-play. Assets are represented in Hyperledger Fabric as a collection of key-value pairs. The performance of the execution layer can be measured by the observed transaction latency. Hyperledger Fabric applications can be decided using the Hyperledger Composer tool.

The client can create a transaction and request it to the endorsing peer by using the SDK API. The endorsing peer executes chaincode to simulate proposal in a subset of peers. It also generates ReadWriteSet based on keys that were query state database for reading and written. The Endorsing peers send back proposal response including response payload and ReadWriteSet to a client application. If the transaction is endorsed, the client submits the transaction to the ordering service. Otherwise, the transaction is cancelled. The validation logic is run in the committing peers to check the endorsement policy like validation system chaincode (VSCC). The committing peers also validate ReadSet versions like multi-version concurrency control (MVCC) in State database. Transactions are indicated in a block which is valid or invalid. Blocks are committed to blockchain on the file system. The valid transactions within the block are committed to the state database. The transaction life cycle of Hyperledger Fabric is shown in Figure. 1 [1].
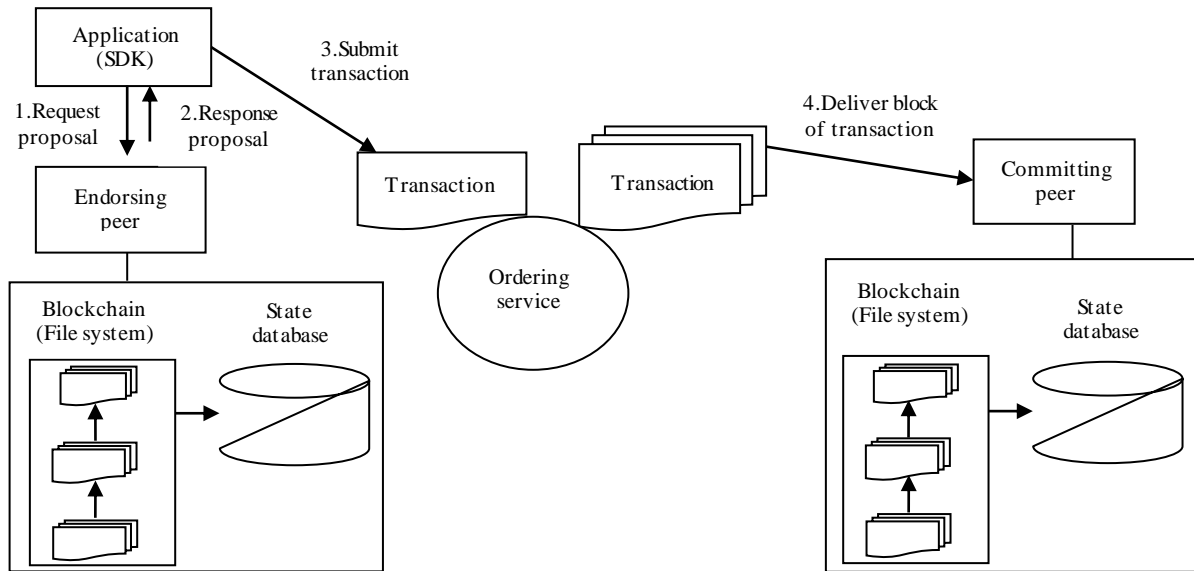

Fig. 1: Transaction lifecycle of Hyperledger Fabric

5. State Database

The current state of latest key-value pair is stored in the state database. The state database currently involves LevelDB and CouchDB. LevelDB is the default state database and supports very fast key-value store for a process. The basic operations are Put (key, value), Get (key), Delete (key). Although it can only support simple queries, it provides excellent performance. It does not support SQL queries and indexes. It can allow only a single process at a time. There is no client-server support to the library. The key-value store cannot be accessed by another process in a consistent way. It cannot support replication and map/reduce operation. There is no online backup and Web interface [9].

CouchDB is open source database software that emphases on ease of use and having a scalable architecture. It provides dynamic queries. It has a document-oriented database structure and is written by Erlang. Erlang is a general-purpose concurrent programming language and runtime system. CouchDB uses JSON file for storage and queries. It does not save data and relationships in tables like a relational database. A collection of independent documents is supported by each database. It provides ACID semantics. It constructs a form of multi-version concurrency control (MVCC) so can provide the set of concurrent readers and writers without conflict. It can replicate to devices that can go offline and support data synchronization to retain back online. The ordered lists and associative arrays can also be used. The field values may be simple things like strings, numbers, or dates. It can save any type of data together. CouchDB takes additional space for overhead compared to other databases [9].

6. Blockchain

The blockchain is a transaction log, structured as interlinked blocks. A blockchain ledger is often described as decentralized because it is replicated across many network participants. Each block involves an ordered of transactions which references the changes of the state database. It is sometimes called a history

database. The blockchain is physically constructed as a file-system. It allows only appending to the end of the blockchain. It is the source of data and immutable. The primary operation of the blockchain is to store the transaction. So, the retrieval of the transaction is currently infrequent operation [1].

Hyperledger Fabric cannot support efficient temporal indexing on file-system of the blockchain. The default Application Program Interface (API) of Hyperledger Fabric does not provide any temporal index. There is no inclusive call to query of those historical values for a key at exactly transaction time (t). GetHistoryForKey (K) (GHFK) is called to retrieve the past states of key k in the history file-system of the blockchain. It needs to operate a GHFK call for scanning all-state for key k between valid time (Vs, Ve) [6]. The valid time is the time range including start and end time. The temporal query is not known which blocks' time intervals (t1, t2) exactly involve between valid times (Vs, Ve) on Hyperledger Fabric. It always accesses between valid times (Vs, Ve) for any time intervals (t1, t2). So, the performance of the temporal query on Hyperledger Fabric is a worse case.

7. The Proposed New Indexing Approach

In a key-value database, the content of LevelDB can only be queried by key. It does not meet chaincode requirements for many use cases. The CouchDB is document-database which stores the content as JSON format and can complex queries. Capabilities for is enabled by the world state database called CouchDB. Hyperledger Fabric provides persistence data of any byte stream. This needs the stream to be in a JSON format. This supports dates to compare as strings. A timestamp can be used to prove that the events and transactions happened. It can be received from current server time that is kept synchronized with precise timeservers. If the time is in different time zones, querying will not work. All dates will be converted to ISO date format. The time is text format instead of a date type in SQL databases. The date type should be persisted using a strict format. All multiple peer servers accept the transaction from the client. They must be needed to support the trusted timestamp.

The original accessing of history state on Hyperledger Fabric is GetHistoryForKey (k). It only depends on key k to query all the past states in the blockchain. It cannot be effective for any temporal indexes. In this proposed system, the persistence user-defined time (t) will be stored for each transaction into both state database and blockchain with same strict length. The user-defined time (t) is a temporal attribute and must be only one for each transaction. The key k and persistence user-defined time (t) are used together to retrieve the history state of key k on blockchain like GetHistoryForKey (k, t). It will be used to execute the temporal query as an indexing technique on blockchain file-system. The value of a valid transaction given key and time can be found.

The proposed structure of the state database is shown in the figure. 2. The state database represents the current values of all ledger states. A ledger state such as, ample key (k), value (v) and user-defined time (t) is stored as version 0. The user-defined time (t) is considered as an additional argument to aid temporal index. The proposed block structure of blockchain is described in the figure. 3. The sample Transaction (T2) in block-data (D1) of a block (B1) involves transaction header (H2), a transaction signature (S2), a transaction proposal (P2), a transaction response (R2), and a list of endorsements (E2) and user-defined time (t2).
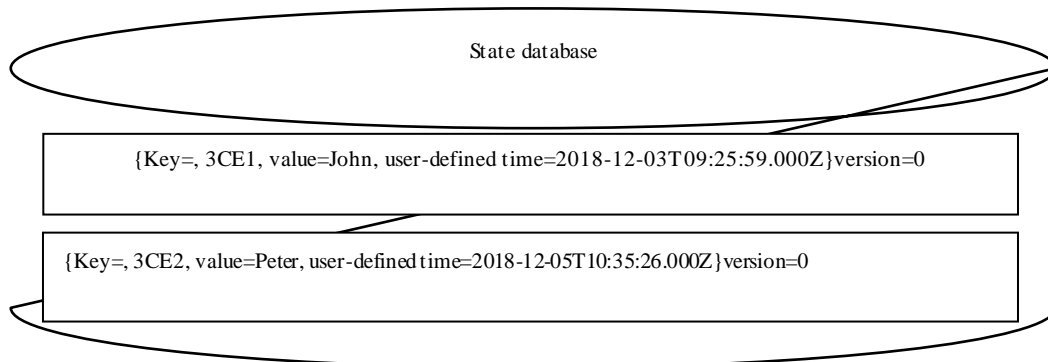

{Key=, 3CE1, value=John, user-defined time=2018-12-03T09:25:59.000Z}version=0

{Key=, 3CE2, value=Peter, user-defined time=2018-12-05T10:35:26.000Z}version=0
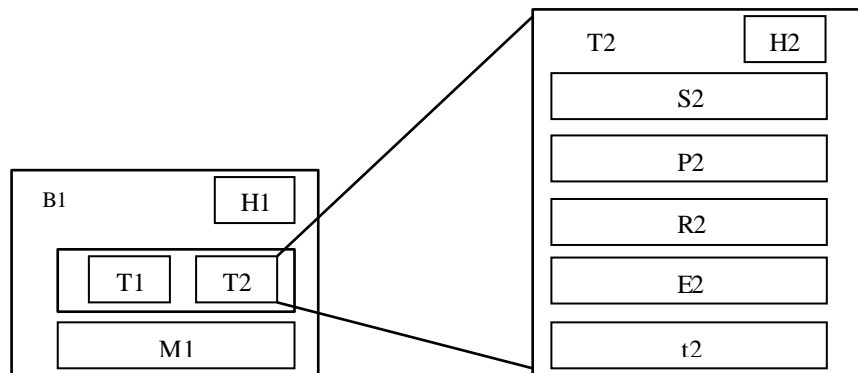
Fig. 2: Proposed structure of state database

Fig. 3: Proposed block structure of blockchain

8. Conclusion

In this paper, the temporal indexing of Hyperledger Fabric is focused. A user-defined time is proposed to support the indexing on the history states of the blockchain. The performance of the proposed system can be better than the original temporal query of Hyperledger Fabric. As a future work, many evaluation and experiments have to be done in order to get the efficiency of the proposed idea on Hyperledger Fabric.

9. Acknowledgements

10. References

[1] https://hyperledger-fabric.readthedocs.io/en/release-1.3/ledger/ledger.html.

[2] J.J. Bambara, P.R. Allen, K. Iyer, S. Lederer, R. Madsen and M. Wuehler. Blockchain: A practical guide to developing business, law, and technology solutions, 2018.

[3] D. Puth al, N. Malik, SP. Mohanty and E. Kougianos. Everything You Wanted to Know About the Blockchain: Its Promise, Components, Processes, and Problems. In IEEE Consumer Electronics Magazine 7(4). 2018, pp. 6-14.

[4] E. Ademi. A Comprehensive Study on the Scalability Challenges of the Blockchain Technology, 2018.

[5] H. Gupta, S. Hans, K. Aggarwal, S. Mehta, B. Chatterjee, and P. Jayachandran. Efficiently processing temporal queries on hyperledger fabric. In IEEE 34th International Conference on Data Engineering (ICDE). 2018, pp. 1489-1494.

[6] H. Gupta, S. Hans, K. Aggarwal, S. Mehta, B. Chatterjee, and P. Jayachandran. On Building Efficient Temporal Indexes on Hyperledger Fabric. In IEEE 11th International Conference on Cloud Computing (CLOUD). 2018, pp. 294-301.

[7] https://coinsutra.com/different-types-blockchains.

[8] Z. Hintzman. Comparing Blockchain Implementations –NCTA Technical Papers. In Technical Forum SCTE-ISBN, NCTA, CABLELABS, 2017.

[9] http://vschart.com/compare/leveldb/vs/couchdb.

Analysis of Students' Academic Performance, Behavior and Personality in Big Data

Gant Gaw Wutt Mhon[1][+] and Nang Saing Moon Kham[2]

[1]University of Computer Studies, Yangon, Myanmar

[2]Faculty of Information Science, University of Computer Studies, Yangon, Myanmar

[1]gantgawwuttmhon@ucsy.edu.mm, [2]moonkham@ucsy.edu.mm

Abstract. Today's every educational institution generate lots of data related to the admitted student from various electronic system due to the key aspects of a wide spectrum of age of Information Technology. Moreover, the lots of student data are accumulated and if that data is not analysis properly for management purpose and student achievement then all afford is going to be wasted. Notwithstanding, considering the correlations between students' academic performance, behavior and personality traits has more accurately revealed the presence of intriguing links rather than the focusing on their characteristics patterns, previous marks and other features. To detect these correlations, the system firstly attempts to explore the cluster analysis results as a characteristics of students in different groups from students' academic assessments, behaviors and demographics data. In this system, we adopt the performance of the effectiveness of electronic "Student Information System (SIS)" to collect students' demographics and assessments data and also explore the students' academic behaviors through social media platform, Facebook which is used as a teaching purposes because of similarly and dissimilarly acted in social interactions and interests of individuals in social network. This system uses "Big-Five Personality Model" to achieve students' personality results to analyze the correlations between the clustered results and personality traits which are firmly and consistently associated.

Keywords: Academic Behaviors, SIS, Clustsering, Big-Five Personality Model.

1. Introduction

With the development of new and emerging technologies, which present new opportunities to improve or redefine the university experience and academic achievement through activities including teaching, learning and management system. This means that being aware of new trends in emerging technologies and having the ability to rapidly harness their potential to drive improved outcomes will become a key differentiator within Higher Education. Therefore, by the use of an exploration of the effectiveness of electronic SIS system in managing student performance in education, it has important strategic side for analysis of students' performance. Because this system provides a simple interface for the maintenance of student information and so the records of students' performance can be collected easily without time consuming. This paper achieves the students' academic assessments and demographics data from SIS system to analyze their academic performance better to know about them. The main contribution of this paper is that the accurate analysis of individual differences in academic performance, behaviors and personality traits which can be easily to understand individual's strength and weakness and to fulfill their needs in time for their educational life effectively. Understanding and identifying the reasons for individual differences in behavior and personality have always been a concern of academic achievements.

With the growing popularity of social networking, different social behaviors are emerged through publicly available information on the analysis of social media domain. Social Networking sites, most notably Facebook have become so popular today which is the leading words on the lips of every student. This paper

---

[+] Corresponding author. Tel.: +9501610655; fax: +9501610633

E-mail address: gantgawwuttmhon@ucsy.edu.mm

conducts the findings of students' academic behaviors and interests of individuals in social network as a course-specific Facebook group created by the instructor was used to share additional material. Given most of the models are only considered such aspects (large number of friends and followers) from users of social media sites as accurate picture of personality.

In this paper, for the students' personality traits, Big-Five personality dimensions is played as a role of determiners based on its factors to achieve students' personality results. Big-Five personality traits are also known as Five Factor Model (FFM) which is a taxonomy for personality traits. When factor analysis is applied to personality survey data, some words used to describe aspects of personality are often applied to the same person. To our knowledge, we have learned that most of the analysis model of academic performance adopted this model as determiners to predict academic achievement based personality traits by utilizing academic performance and then quantify them into personality traits on top of the model. But learning and considering about the correlations are consistently associated or not between students' academic performances, behaviors and personality traits which may explore more accurately revealed the picture of the needs of academic success.

After every passing day, more and more students data are generated in the educational process and for finding ways to enhance students' motivation and improve their achievements, considering students' multidimensional data from various aspects are needed. For this problem, big data make this work much simpler and hassle-free for the accuracy of data and confidentiality. According to today's definition of big data, the concepts of big data is not only means the size but also refers to the use of data and how you are going to do that. [5] Also, the size always plays an important role when we talk about data but the technology world is progressing and no doubt the need for such options is highly on demand. Apparently, in order to get the right standard value from big data, we always need to look for quality and quantity in the corresponding raw data and we can gain enormous insight if that raw data is quality data even though small quantities of data. But earning a high standard of data quality isn't easy and we have to think about the waste and value from the raw data that we have collected.

The remainder of this paper has been arranged in different sections. Section 2 describes the related researchers in this area. Then section 3 implement background theory of this research work. In this section 4, we briefly explain our proposed method and conclusion is made in the final section.

2. Background Theory

2.1. Student Information System (SIS)

Today, the utilization of SIS system has been growing at a steady pace owing to increase in the application of analytics in the educational sector. As an implementation of state-of-the-art SIS, the data from this system helps in formulating policies and corrective measures to enhance students' academic performance. Due to the age of Information Technology (IT), SIS is popular among administrators and educators to gain accessibility of data to evaluate the performance of students based on the goal and mission of the institutions. Therefore, it is very time saving and effective way to observe all students' related data and many suggestions and idea can get from these data which are used to investigate the academic growth.

2.2. Big-Five Personality Model

There are huge amount for analysing and identifying academic success based on the usefulness of Big-Five Personality model due to a strong significance for the quality of educational work. Understanding the individual differences in levels of students' achievement based on their personality traits which will be concerned with identifying who will perform well, and who will perform poorly in academic programs. Many empirical studies have investigated that there are increasing focus on personality traits of students as a predictors of academic achievements. In the Big-Five Personality Model, each of the five factors represents a range of possible personality types and the definitions of each traits is described in Table. 1. Each person has five traits which are scored on a continuum from high to low and questionnaires which are used a Likert scale. But in some questions often contain some items which are to be reversed scored before we run a factor analysis on scale items.

Table. 1: Definitions of Each Traits

Preferences	Definitions
Openness (O)	Indicates how broad-minded and curious a person is.
Conscientiousness (C)	Indicates how hardworking and dependable a person is.
Extraversion (E)	Indicates how out-going and social a person is.
Agreeableness (A)	Indicates how warm, friendly and trustful a person is.
Neuroticism (N)	Indicates a person's ability to remain stable and balanced.

Reverse scoring means that the numerical scoring scale runs in the opposite direction and so we need to reverse-score all negatively-keyed items. For recode these items, we need to subtract the values from a constant one value higher than the highest value on the scale. According to calculated result, score description of each trait is as an example in extraversion, high scorers tend to be very social while low scorers prefer to work on their projects alone.

2.3. Using Facebook for Academic Behavior

With the advancements in technology, today's communication is easy as walking into a neighbor's residence to deliver a piece of information or vice versa through the use of social media. Due to Facebook usage, students' public data which includes interests and behaviors based on internal data but also external data are also emerged. [6] From their study, they recognize the major advantages of social media usage by students in higher education which included: enhancing relationship, improving learning motivation, offering personalized course material, and developing collaborative abilities. Apparently, Facebook is one of the social media platform which can be used for teaching purposes. Due to the Facebook usage with this purpose, students' academic behaviors through social network platform are conducted as effective way of one of the collection of academic performance.

[1] For an engagement index (EI) of the diverse array of possible units of engagement on Facebook, a measure of engagement depth was calculated for each student since the metrics of engagement on Facebook represent different levels of engagement based on Likes (seeing and/or reading, and critically thinking) can be considered a lower depth of engagement than comments (seeing, reading, and critically thinking), Student posts (finding/seeing, reading, critically thinking, and sharing with class) can be considered the highest level of engagement. The researchers also proposed a formula to calculate a particular students' overall level of engagement by taking into account the level and frequency of engagement and so EI was calculated for each engaged student as:

$$EI = [(1*\#likes) + (2* \#comments) + (3* \# posts)] / Total \# Facebook posts$$

where "likes" is an individual students' total number of likes on instructor and student posts, "comments" is an individual students' total number of comments on instructor and student posts and "posts" is the total number of posts that an individual shared on the Facebook page; "total number of Facebook posts" is the total number of instructor and student posts combined. According to our knowledge, to enhance for evaluating in students' academic behaviors using Facebook: likes, comments, student posts, and face-to-face interactions are also included.

2.4. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) Algorithm

[3] DBSCAN is a clustering method that is used in machine leaning to separate clusters of high density from clusters of low density. It is one of the Density-Based clustering algorithm and scans the whole dataset only one time and needs to calculate the distance of any pair of objects in the dataset. This clustering method can discover arbitrarily shaped clusters, find cluster completely surrounded by different clusters, robust towards outlier detection and require just two points which are very insensitive to the ordering of the points in the database. Due to the advantages of DBSCAN, the system attempts to focus the occurrence and necessity of formation of DBSCAN in a problem solving activity within multidimensional data to get a significant efficiency and reduce in computational complexity for analysis of student academic performance

with various features from different aspects. [4] The pseudocode of the DBSCAN algorithm is described in below:

Input (*eps, minpts,D*) , Output (a set of clusters)	11. mark p' as visited
1. initialize all points as unvisited	12. let N' be the eps−neighborhood of p'
2. for each unvisted point $p \in D$ do	13. if the size of N' is >= *minpts* then
3. mark p as visited	14. add those points to N
4. Let N be eps−neighborhood of p	15. endif
5. if the size of N < *minpts* points then	16. endif
6. mark p as noise	17. if p' is not yet a member of any cluster
7. else	18. add p' to C
8. create a new cluster C, and add p to C	19. endif
9. for each point $p' \in N$	20. endfor
10. if p' in unvisited then	21. endif
	22. endfor

3. Related Work

High quality education is an important fact and every educational organization working hard to achieve this from various aspects. Uddin et al [7] proposed utilizing relevant academic and personality features from big unstructured data to identify good and bad fit students by using logistictic regression, bayesian network and stochastic probability distribution methods. PERFECT algorithm engine is also used to infer academic based personality. Their combination of students' academic features and social networking features approach improved the prediction accuracy. In this approach, good fit candidate attributes are need to already import in prediction engine to evaluate the next student attribute for prediction of likeliness of success or failure.

Dong Din et al [2] evaluated student behavior clustering method based on campus big data with density based clustering method which is parallelized on the Spark platform and applied to subdivide student behavior into different group using historical data including student campus card consumption data, library book borrowing record data, attendance record data of campus, library access control records and achievement data are as a data source in digital campus shared database. This proposed approach is to study for universities to know students well and manage them reasonably and improved algorithm is also effective. But processing student behavior characteristic segmentation clustering model may not strongly accurate lack of personality due to individual differences in characteristics patterns.
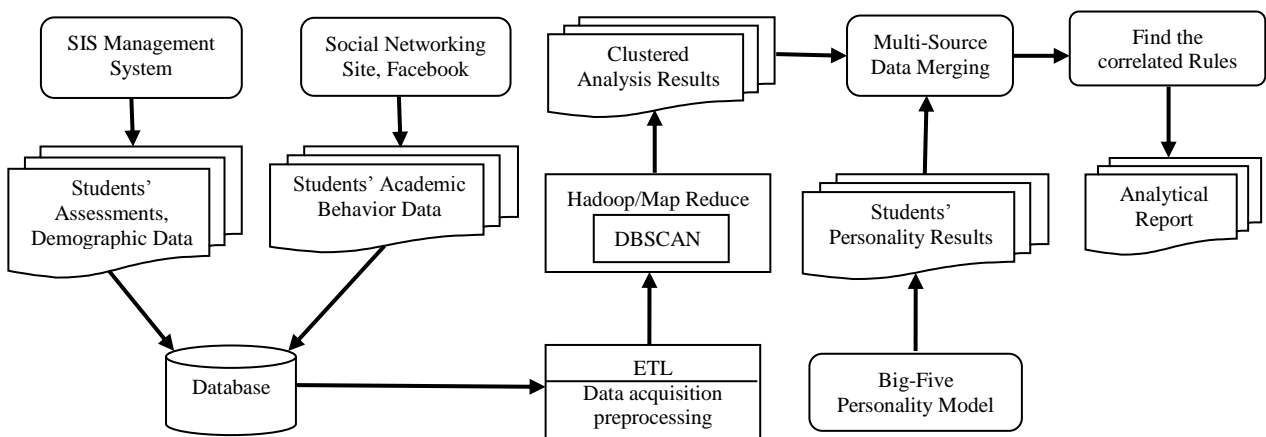
4. Proposed System

Fig. 1: Analysis Approach of Correlation Between Academic Performance, Behavior and Personality Traits

To our knowledge, there are various research works target to achieve in academic performance but there is no existing work of considering with this proposed approach idea for the correlations between students' academic performance, behavior and personality traits which reveal the presence of efficient way rather than the focusing on their characteristics patterns, previous marks and other features. Accurately, the system explores the students' academic assessments and demographic data by using the performance of SIS as a data source to help in formulating policies and corrective measures to enhance students' academic

performance. Additionally, the variety of open source student information management system software has better insight to collect lots of students' data with accurate ways without time consuming. The system uses Facebook for teaching purposes to provide students' academic behaviors because of the individual interests in their courses. Combining with open source framework Apache Hadoop, DBSCAN algorithm use the programming model of MapReduce that will reduce time of accessible algorithms and data set from the dissimilar location will work simultaneously from single node and find the appropriate outcome in distributed environment, cluster analysis result are explored as a characteristics of students in different groups based on students' data from the SIS and Facebook. To achieve students personality results, students need to take a personality test with fifty questions and each statement represents ten questions respectively to analyze their personality traits. The main idea of the proposed approach is that students' personality results and the clustered results are merged with their student identification number and then analyze with Apriori association rule algorithm, the most classical and important algorithm for mining frequent item sets, to conduct the correlated rules between them which are firmly and consistently associated. As an example, the possibility rules are included like:

- If a student is perform well in academic performance and behaviors and then his personality result is Openness;
- If a student is perform neutral in academic performance and behaviors and then his personality result is Openness;
- If a student is perform bad in academic performance and behaviors and then his personality result is Openness;

Therefore, to report the analytical result, final resulted correlated rules are emerged from these possibility by analyzing with Apriori algorithm.

5. Conclusion

In conclusion, this study can be clearly realized that considering the correlations between students' academic performance, behavior and personality traits has more accurately revealed the presence of intriguing links rather than the focusing on their characteristics patterns, previous marks and other features. This paper also tries to contribute in the occurrence and necessity of formation of DBSCAN in a problem solving activity within multidimensional data to get a significant efficiency for analysis of student academic performance with various features from different aspects by reducing in computational complexity. There are still many further works that are needed to be done for better analysis method. Considering the other students' features and using other statistical algorithms will be our future works.

6. Reference

[1] Clements, Jeff C; "Using Facebook to enhance independent student engagement: A case study of first-year undergraduates", Higher Education Studies, Volumn 5, Number 4, Pages 131, 2015.

[2] Ding, Dong; Li, Junhuai; Wang, Huaijun; Liang, Zhu; "Student Behavior Clustering Method Based on Campus Big Data"; IEEE, 2017.

[3] Evan Lutins, "DBSCAN: What is it? When to Use it? How to use it.", Sept 5, 2017.

[4] Han, Dianwei; Agrawal, Ankit; Liao, Wei-Keng; Choudhary, Alok; "A novel scalable DBSCAN algorithm with spark"; Parallel and Distributed Processing Symposium Workshops, 2016 IEEE International, Pages 1393-1402, 2016, IEEE.

[5] James Warner, "Big Data VS Traditional Data: What to know When it Comes to Defines Big Data?", March 18, 2018.

[6] Pečjak, Sonja; Puklek Levpušček, Melita; Valenčič Zuljan, Milena; Kalin, Jana; Peklaj, Cirila;"Students' social behaviour in relation to their academic achievement in primary and secondary school: Teacher's perspective", Psihologijske teme, Volumn 18, Number 1, Pages 55-74, 2009, Filozofski fakultet u Rijeci.

[7] Uddin, Muhammad Fahim; Lee, Jeongkyu; "Utilizing Relevant Academic and Personality Features from Big Unstructured Data to Identify Good and Bad Fit Students", Elsevier, 2016.