Towards Blockchain-enabled School Information Hub

Nelson Bore
IBM Research Africa
Nairobi, Kenya
nelsonbo@ke.ibm.com

Samuel Karumba
IBM Research Africa
Nairobi, Kenya
karumbaS@ke.ibm.com

Juliet Mutahi
IBM Research Africa
Nairobi, Kenya
julimuta@ke.ibm.com

Shelby Solomon Darnell
IBM Research Africa
Nairobi, Kenya
shelbyD@ke.ibm.com

Charity Wayua
IBM Research Africa
Nairobi, Kenya
charitywayua@ke.ibm.com

Komminist Weldemariam
IBM Research Africa
Nairobi, Kenya
k.weldemariam@ke.ibm.com

ABSTRACT

Several initiatives have been proposed to collect, report, and analyze data about school systems for supporting decision-making. These initiatives rely mostly on self-reported and summarized data collected irregularly and rarely. They also lack a single independent and systematic process to validate the collected data during its entire lifecycle. Furthermore, schools in developing countries still do not maintain complete and up-to-date school records. Due to these and other factors addressing the education challenges in those countries remains a high priority for local and international governments, donor and non-governmental agencies across the world. In this paper, we discuss our initial design, implementation, and evaluation of a blockchain-enabled School Information Hub (SIH) using Kenya's school system as a case study.

CCS CONCEPTS

• **Applied computing** → *Business process management systems*; *Cross-organizational business processes*;

KEYWORDS

ICTD, Education, Blockchain, Data Management

1 INTRODUCTION

In Sub-Saharan Africa, the use of school data for decision-making has been historically poor and lacking [12]. Currently, each school keeps track of data related to their school, by maintaining a repository of data relating to parents, teachers, students, facilities, and school assets. The stakeholders in the school manage these records by "checking in the records" in a centrally owned database [2, 4, 13].

Existing school management systems are useful in automating and simplifying the working of an individual school. However, it is increasingly difficult and complex when schools need to provide access to data and information to additional stakeholders. For example, for a ministry of education to obtain data on schools, they need to employ additional data collection efforts as they do not have access to school-level data directly from what schools have collected. Today, a student who is transferring from one school to another school is forced to recreate their record when they go to the new school. At best, they might be asked to bring a manual copy of their old school record that can perhaps be transferred to the new school in a manual file. As a result, the management of schools at a macro level is very challenging.

Policy makers charged with monitoring learning outcomes, tracking out of school children, designing intervention mechanisms, etc are unable to make accurate decisions because of the paucity of data [6, 8, 11]. In many cases, data they receive is fraught with data manipulation, fraudulent reports (e.g., over/under reporting), data duplication and incompleteness. As a result, decision-makers are looking for a unified solution that enables them to get visibility into what is happening at a school and student level so that they can make better decisions and design appropriate interventions.

Blockchain presents a novel concept for managing business data and transactions [5]. It has been applied in various use cases with the most popular being its application to digital currencies [3, 7, 9]. The algorithms and protocols of blockchain establish verifiable control points and manage compliance assurance (data), assure the integrity of transactionality (immutability), and assure that transaction are executed by identifiable entities (non-repudiation). For example, all entry of, and changes to, school records of entities can be entered in an immutable form, a contemporary means for assuring correct attribution of entries and changes, and immutability of these entries and changes using a consensus mechanism layered on top of known cryptographic algorithms. Thus, any alteration of school records can be verified against a particular instance of the record. This can be done by obtaining a historical block identifier from a school record historical blockchain (representative of historical record events) that have been conducted with respect to the particular instance of the record.

This paper proposes a novel use of blockchain for School Information Hub (SIH). Section 2 motivates the need for a blockchain-based SIH. We present our methodology and a conceptual framework of the proposed SIH system in Sections 3 and 4. An initial prototype implementation based on Kenya's school system is discussed in Section 5. We conclude the paper in Section 6.

ICTD '17, November 16–19, 2017, Lahore, Pakistan

ACM ISBN 978-1-4503-5277-2/17/11…$15.00
https://doi.org/10.1145/3136560.3136584

2 CHALLENGES WITH SCHOOL MANAGEMENT

School Census Hub (SCH) is one of the projects we have been working on in partnership with local and international education partners [4]. SCH leverages cloud-based technologies equipped with a multi-modal suite of applications enabled on tablet devices that facilitate the collection, validation, and visualisation of data in schools. The primary objective of the SCH was to test if a technology can improve the efficiency and effectiveness of school level data collection while ensuring transparency and accountability in school systems. This work has presented an opportunity to determine the specific challenges faced by policy makers at various levels with regard to managing school systems. These challenges have formed the basis for our proposed solution. They include:

Challenge C1. Public school systems obtain significant portions of their funding from national governments and usually based on the number of students. Today, this allocation is based on an unverifiable number of active students and teachers. As principals self-report the number of students per school, it is unclear that the numbers reported are accurate. This lack of data quality has significant implications on budget allocation especially in cases where budget requests are inflated. For example, an overestimate by an average of 10% per school can lead to an excess annual budget allocation of $50 million for non-existing students in Kenya alone. Similarly, while the teacher payroll allocation is centralized, teachers are assigned to specific schools. Hence, teachers that have left the service or do not show up for work may still get paid due to a lack of coordination between the responsible central office and the schools. Using the same example in Kenya, a 1% "ghost teachers" in payroll can amount to almost $20 million a year in losses.

Challenge C2. Once budget allocations are made, spending on assets is not transparent, and therefore not verifiable. The spending of the budget is tracked against invoices issued by 3rd party vendors and not against the actual assets. There are discrepancies between the invoices for purchase of learning materials compared to the actual materials in the classroom. The budget allocated for repairs in schools to improve learning environments may not be spent appropriately, but it is not currently possible to provide authoritative proof of work.

Challenge C3. A third issue concerns improving the learning environment. There is a limited insight into the statistical relationship between school effectiveness and demographic variables. To understand the performance of schools across districts and tie this performance to external conditions, such as the state of the facilities, resources available, and other exogenous factors such as neighbourhood income levels, local or in-school violence, health challenges in the area, etc., requires deep understanding of the context of a school. Currently, there is no direct way to correlate a school's performance to such things as teacher absenteeism let alone to such exogenous factors.

Challenge C4. Insufficient data on teachers and students to support personalized education. Personalized education is an approach to improving the learning experience of both teachers and students by fostering engaged teacher-student experiences. It requires an understanding of the learning environment and a detailed understanding of the student's learning history and their performance challenges. Outside of exam results, there is insufficient data to help support widespread teacher-student engagement.

Proper and reliable visibility into the number of students, teachers, school assets and the various transactions each of these entities are involved in throughout their lifecycle is therefore the foundation for proper data driven policy decisions to ensure learning outcomes are measured and appropriate interventions are designed.

3 METHODOLOGY

We propose a generic methodology (Figure 1) to guide the design and development of the proposed SIH, which shall facilitate the following core requirements:

- *Lock In Attribution*— help create a permanent and unbreakable link between the entity (or data owner) and the record, and that link, the record of ownership can be forever verified and tracked.
- *Secure Sharing*— help securely share one's digital content with others. Sharing behaviour or lifelong gains in skill is made as easy as sharing or copying a piece of data with blockchain based consent management logic.
- *Improve Visibility*— help trace when and how learning or performance improves, how school systems evolve/grow over time, what/how to prioritize resource needs based on single point of truth, etc.
- *Certificate of Authenticity*— each logged SIH events may come with a certificate of authenticity (COA), a built in unique cryptographic ID, and complete ownership history. The COA can be verified any time and printed out.

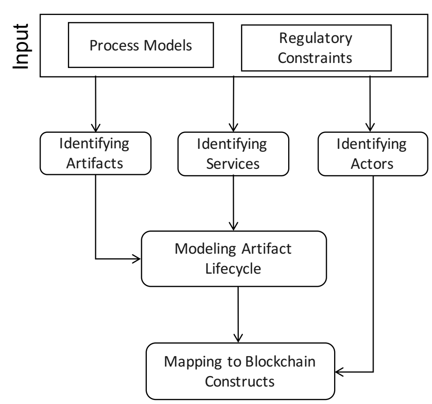

Figure 1: High-level methodology.

A business process specification and set of associated regulations are the basis for the methodology. Automated or semi-automated (i.e. "human in the loop") toolsets can be use to mine documented processes, in conjunction with documented regulatory requirements that pertain to that process, to extract actors, artifacts, and services associated with the process and regulatory matters.

The methodology then breaks the services into a series of transformations on the artifacts that are applied by the actors. The services, in general, will have some governing control logic (i.e. "do this if that"). It then takes these sets of transformations and associated control logic (one set per service) and incorporates them into blocks in a blockchain, together with associated smart contracts. Note that artifacts are typically documents of one type or another, i.e. either physical or electronic documents. In the case of physical documents, there would have to be some human in the loop, e.g., at least to facilitate the transforming of the physical document to electronic form to enable further electronic processing.

For example, consider capturing and scoring student examination where the exams are paper-based, one human-in-the-loop approach would be to have a human (student or proctor) scan the exam immediately after the student takes the exam. The series of human actions on these exams (taking and submitting the exam, scanning the exam, grading the exam (possibly by different human graders according to section), summing up the scores to give a total score, etc.) could be separately entered into the blockchain, or alternatively, the exam results could be read, say, by an optical mark reader, the recording of this reading then added to the blockchain, and then further processing could be handled in purely electronic fashion. Once the actors and artifacts are identified, the set of possible artifact states and state transitions are identified, along with the actors that have the ability to affect these state transitions. These are the transformations (or set of transformations, possibly including some amount of control logic) that constitute the services provided by the various actors.

Moreover, process completion is a key notion and is defined as a set of artifacts reaching a given end or goal state. There may additionally be a set of required intermediary states. It is always possible to add variables to the definition of the system so that the requirement of attaining intermediary states are encoded in the end/goal state. Finally, to formalize our methodology, we extended the methodology presented in [1, 14] based on our context.

4 CONCEPTUAL FRAMEWORK

Figure 2 shows a simplified view of the proposed SIH, comprising preprocessor module configured with a School Data Hub (SDH) framework, and also in communication with a blockchain network. The system further includes user interfaces and devices, and a decision support layer. The SDH is a suite of apps that collect, store and manage school records by facilitating transactions related to student, teacher and administrative staff and school.

A school record includes a student record, a teacher record, and an asset record. A student record consists of a profile, attendance, academic performance, etc. A teacher record consists of a profile, attendance, social network, career progression, etc. An asset record can be and are not limited to learning content, curriculums, instructional materials, etc., and any other record, such as school metadata, school residence type, category, demography, type, etc.

A transaction related to a student record can include: registering a new student into a school system, enrolling or re-enrolling in classes, updating user identity, such as biometric data, logging attendance of the student, promotion in grade, transferring the student, graduating student, termination, placement, etc. A transaction related to a teacher record can include: registering a new teacher into a school system, updating the profile of the teacher, assigning a course to a teacher, promotion, deployment and assignment of teacher, attendance and teaching related transactions such as recording the performance of teacher, etc. Similarly, transactions

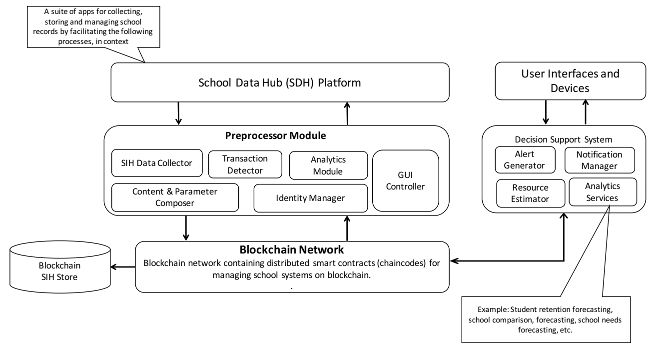

Figure 2: A simplified view of a blockchain enabled SIH.

related to assets, learning resources, classes, etc., include asset registration, uploading or updating curriculum, creating a new course, starting or closing of a course, etc.

The preprocessor module includes an SDH data collector, a transaction detector, a learning analytics module, a GUI controller, a parameter composer, and an identity manager. The SDH data collector further collects, prepares or generates transactions and data per student to be securely stored and managed on the student record. The transaction detector detects the addition of content to blockchain records. The parameter processor and composer compiles contents and parameters and then triggers the appending of record transactions associated with a give entity (e.g., student, teacher, school) into the chain of a historic entity's record blockchain, based on new content detected by the transaction detector. The analytics module is a collection of various analytics models related to risk assessment, transfer patterns, etc. based on data from the SDH, longitudinal school data, and historical blockchain data. The identity manager facilitates the creation of identity, by, e.g., using biometric data such as fingerprints and iris scan, demographic data, etc., during student registration, facilitates authentication of students, e.g., during attendance, and manages other identity related services.

The blockchain network contains distributed smart contracts ("chaincodes") for managing school systems on a blockchain, and is connected to a blockchain school lifelong store, which is a ledger that serves as a database for storing school records and chaincodes. A chaincode is an executable smart contract that is distributed across blockchain nodes in the network to enable interaction with that network's shared ledger. The school record is stored on the blockchain SIH store. Chaincodes are deployed at each node of the blockchain network for managing school systems. All the functionalities and services discussed are governed by the necessary smart contracts such as process chaincode ensures the corresponding processes are followed and agreed up on by all the nodes, etc. Transactions associated with a record instance are compiled into a chain of record transaction blocks. The chain can be considered a chronicle of a record path through time. When a transaction occurs, one or more corresponding parameters related to administrative events, such as registering a student, transferring a student, enrolling a student for a course, promoting a teacher, assigning a course for a teacher, etc., are recorded. The recorded parameters establish the validity of the transaction and generate a new block.

Once the new block has been generated, it can be appended to the historic school's record blockchain.

The decision support component is a collection of analytics services that can provide actionable insights about students, teachers, schools and resources for decision makers across various levels. These capabilities can be built on top of other analytics modules to enable services such as resource allocation, budget allocation, impact assessment, student drop-out pattern forecasting, in-take pattern analysis (i.e. "what types of students are most attracted to this school?"), transfer pattern, student similarity analysis, benchmark analysis, fraud prevention, teacher distribution, school comparison across different demographic variables, and decision support services. A sample decision support question that one might ask is "Given a certain amount of money to spend to improve student retention, what will have the greatest impact?" At the student level, these analytics help decision makers identify risks and needs unique to each learner and design personalized interventions that can shift learning trajectories towards desired outcomes.

5 PROTOTYPE IMPLEMENTATION

We implemented a prototype system using a permissioned blockchain technology —called the Hyperledger Fabric [10]. Our initial implementation and experiment were primarily focused on a registration use case, which is a very particular one as it creates the foundation for addressing Challenges C1, C2, and C3 (see Section 2). The system generates and stores (on blockchain ledger) digital artifacts that uniquely define individual records related to all the entities (e.g., students, teachers, etc) after a successful completion the registration process. The relevant SDH apps, middleware (i.e., preprocessor module), and smart contracts were implemented.

The SDH apps are implemented as a collection of mobile apps for registering, onboarding, and verifying a digital record of an entity (a human subject) and associated transactions. Using the SDH apps, the registrar collects demographic data, captures biometric data, and scans documents of the human subject.

The preprocessor module is implemented as a collection of microservices that provide separation of services for ease of maintenance and scalability across all the nodes. All the requests to the API are served via REST interfaces that receive a request from a mobile client and determine the appropriate middleware (e.g., transactions detector, analytics, identity) to handle the request. All the services in the backend communicate via a message bus in order to handle real-time data and requests from client applications. The middleware extracts relevant metadata from the collected data/document, computes hashing and encryption of the document, and submits the transaction for validation.

We implemented chaincodes to facilitate the registration use case. For instance, the document manager chaincode is responsible for manipulating all documents specific to an entity, both stored in blockchain and externally. Currently, we only store hash values of biometric data, documents, images, and any large multimedia data, are stored in the blockchain. Biometric data, documents and large multimedia data such as those captured by SDH apps are stored outside of the blockchain ledger. However, blockchain can be used for controlling the off-ledger data, in which confidentiality, authorization, and integrity of the off-ledger data can be ensured.

We tested our prototype implementation on blockchain network composed of four organizational nodes and created digital identities for internal human subjects. Our experiment primarily focused on testing the effectiveness of the system. More specifically, we collect and process biometric data (fingerprint, iris scan, and face), demographic data, and at least one document (birth certificate, nationally accepted id, or passport). For each data (and document), we compute the corresponding hash value and encrypt them. A transaction composed by the computed hashes, URLs (the locations where the encrypted documents are stored), and relevant events and metadata associated with each data (and document) is transmitted to one of the blockchain nodes for validity checks using deployed smart contracts. Then, after validity check has been performed, <hash, URL, events> tuples are stored on the shared ledger. Once the block commit has been received by the application layer, the encrypted versions (of the data and documents) are persisted on a secured off-ledger storage, where versioning and indexing are performed to enable operations, analytics, and reporting.

6 CONCLUSION

In this paper, we have discussed the conceptual design of a blockchain-enabled SIH using our proposed methodology. An initial prototype implementation of the subset of the proposed system has discussed. We believe that (once refined, extended, and optimized) the registration use case can be used to create a large-scale system for collecting and managing immutable school records for use in accurate decision-making and preventing fraudulent activities.

REFERENCES

[1] Kamal Bhattacharya, Cagdas Evren Gerede, Richard Hull, Rong Liu, and Jianwen Su. 2007. Towards Formal Analysis of Artifact-Centric Business Process Models. In *BPM*. 288–304.

[2] M. Fetaji, B. Fetaji, A. Ajredini, and M. Ebibi. 2013. Devising a model of electronic School Management System based on web services for secondary schools in Macedonia. In *Proceedings of the 35th International Conference on Information Technology Interfaces*. 187–192.

[3] Adiseshu Hari and T. V. Lakshman. 2016. The Internet Blockchain: A Distributed, Tamper-Resistant Transaction Framework for the Internet. In *HotNets '16*. ACM, New York, NY, USA, 204–210.

[4] Andrew Kinai, Juliet Mutahi, Nelson Bore, Samuel Karumba, Srihari Sridharan, Abdigani Diriye, and Komminist Weldemariam. 2017. Deploying large scale school census hub: an experience report. In *Proceedings of the 39th International Conference on ICSE: Software Engineering in Society Track*. IEEE Press, 47–56.

[5] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system,? http://bitcoin.org/bitcoin.pdf. (2008).

[6] Standard Media. 2016. EACC: How school heads steal education billions. (2016). https://goo.gl/KnZjJ9.

[7] Melanie Swan. 2015. *Blockchain: Blueprint for a New Economy* (1st ed.). O'Reilly Media, Inc.

[8] Simon Thuranira Taaliu. 2017. Corruption in Schools and Universities in Kenya. *Open Journal of Social Sciences* 5 (2017), 227–231.

[9] Don Tapscott and Alex Tapscott. 2016. *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*. Brilliance Audio.

[10] The Linux Foundation. 2017. The Hyperledger Project. (2017). https://www.hyperledger.org.

[11] Timothy Machasio. 2017. On Pseudo-Excellence and Corruption in the Kenyan Education System. (2017). https://goo.gl/yFaAna.

[12] UNESCO. 2016. Publications on various Working Papers on Education Policy. (2016). https://goo.gl/MqZ1Rv.

[13] A. Vladimirou and A. Kokkinaki. 2013. Design and development of school assets management system. In *2013 36th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 587–590.

[14] Wei Zhao, Barrett R. Bryant, Fei Cao, Rainer Hauser, Kamal Bhattacharya, and Tao Tao. 2007. Transforming business process models in the presence of irreducibility and concurrency. *IJBPIM* 2, 1 (2007), 37–48.