Fusion of blockchain and machine learning for secure analytics

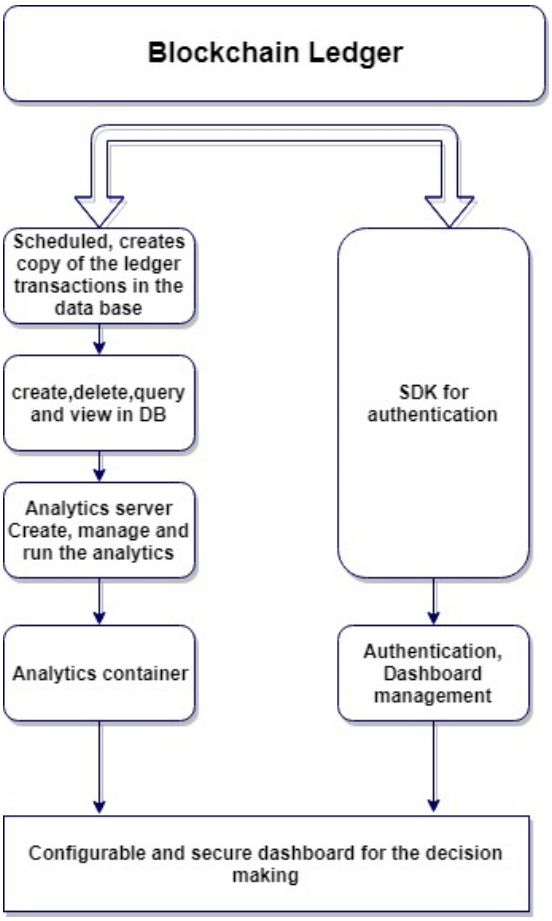
The blockchain technology is growing fast particularly for the logistic, financial services, identity verifications and asset tracking. Analytics on the stored data in the blockchain can provide useful information for predictive planning, regulatory compliance and range of other analysis. In this research, analytics engine will be connected to the blockchain to be able to present the analysis results on a configurable dashboard. These dashboards can be combined with the external data sources for further analysis.

The data stored on the enterprise blockchains are provide rich warehouse of information for the analytics. In many of the current analytics projects data cleaning is a major time-consuming element and could take up to 80% of the data science team. The stored data in the blockchain network is in a common format and organized which could also reduce the data cleansing part of the analytics.

Few studies are conducted on analytics of blockchain data. Currently, there are popular blockchain implementations available including but not limited to Fabric [1], Ethereum [2], and Parity [3]. At the current stage, the blockbench [4] study concluded that existing blockchain systems are not suited for the large-scale data processing workloads. The same issue exists for the public blockchain systems such as Bitcoin . The current research focuses mainly on temporal queries on the blockchain platforms. The temporal query has been a very well researched topic in the data science community and various studies conducted on the temporal queries subject[5]. The dashboard creation for the secure analytics services on blockchain platform and the visualization part is described in the following section. A rule engine is going to be integrated to check the blockchain data for the compliance.

The other section of the research belongs to the predictive models. This part of the research will focus on how the blockchain technology can help machine learning for creation of the trusted trained models. At the current stage, data scientist does not have access to the immutable records of data for machine learning training. If the trained data is biased or manipulated by a third party, the trained model provides predictions that are detrimental to the company using the models [6]

The following architecture will be used for the prototype implementation.



The architecture can be deployed by various providers and it is not specific to specific blockchain platform. The business analytics as described in the above architecture need to follow the control and security procedure in the underlying blockchain platform.

To sum up, the current research is focused on creating decision making dashboards that analysis the data records on the blockchain. The stored data could be related to the financial services, logistics, identity records, asset tracking, etc. The data retrieval from the graph nodes such as Tangle technology will not be the focus of the current research as a result of restricted APIs on available blockchain platforms [6, 7].

References:

[1] (2019). *Linux Foundation*. Available: https://www.hyperledger.org/projects/fabric

[2] (2019). *Ethereum Blockchain App Platform*. Available: https://www.ethereum.org

[3] (2019). *Ethcore. Parity: next generation ethereum browse.* Available: https://ethcore.io/parity.html

[4] T. T. A. Dinh, J. Wang, G. Chen, R. Liu, B. C. Ooi, and K.-L. Tan, "BLOCKBENCH: A Framework for Analyzing Private Blockchains," *arXiv e-prints,* Accessed on: March 01, 2017Available: https://ui.adsabs.harvard.edu/abs/2017arXiv170304057D

[5] S. Nakamoto, "Bitcoin: A Peer-to-Peer Electronic Cash System ", Available: https://bitcoin.org/bitcoin.pdf.

[6] L. Muñoz-González *et al.*, "Towards Poisoning of Deep Learning Algorithms with Back-gradient Optimization," *arXiv e-prints,* Accessed on: August 01, 2017Available: https://ui.adsabs.harvard.edu/abs/2017arXiv170808689M

[7] M. Jagielski, A. Oprea, B. Biggio, C. Liu, C. Nita-Rotaru, and B. Li, "Manipulating Machine Learning: Poisoning Attacks and Countermeasures for Regression Learning," *arXiv e-prints,* Accessed on: April 01, 2018Available: https://ui.adsabs.harvard.edu/abs/2018arXiv180400308J