Tutorial: Blockchains and Databases

C. Mohan
IBM Almaden Research Center
650 Harry Road
San Jose, CA 95120, USA
cmohan@us.ibm.com

1. Overview

In the last few years, blockchain (also known as distributed ledger), the underlying technology of the permissionless or public Bitcoin network, has become very popular for use in private or permissioned environments. Computer companies like IBM and Microsoft, and many key players in different vertical industry segments have recognized the utility of blockchains for securely managing assets (physical/digital) other than cryptocurrencies. IBM did some pioneering work by architecting and implementing a private blockchain system, and then open sourcing it. That system, which has since then been named Fabric, is being enhanced via the Hyperledger Consortium set up under the auspices of the Linux Foundation. Other efforts in the industry include Enterprise Ethereum, R3 Corda and BigchainDB.

While currently there is no standard in the blockchain space, all the ongoing efforts involve some combination of database, transaction, encryption, virtualization, consensus and other distributed systems technologies. Some of the application areas in which blockchain pilots are being carried out are: smart contracts, supply chain management, Know Your Customer (KYC), derivatives processing and provenance management. A couple of production deployments are also in place now.

In this tutorial, I survey some of the ongoing projects with respect to their architectures in general and their approaches to some specific technical areas. Specifically, I focus on how the functionality of traditional and modern data stores are being utilized or not utilized in the different blockchain projects.

Because of the attention the world is paying to blockchain technologies, it is important for the database community to become more aware of the underlying technologies and other developments in this area. Then, the community could try to influence the approaches taken and, in particular, how database technologies could be better utilized or enhanced for blockchains. Since most of the blockchain efforts are still in a nascent state, the time is right for database researchers and practitioners to get more deeply involved!

Proceedings of the VLDB Endowment, Vol. 10, No. 12
Copyright 2017 VLDB Endowment 2150-8097/17/08.

2. Target Audience and Assumed Background

The tutorial is targeted at anyone who is curious about all the attention that blockchain technology and applications of it are getting currently not just from technical/industry publications and people but also from society at large (e.g., New York Times readers).

I assume minimal knowledge of database and distributed systems on the part of the attendees. This tutorial is more introductory in nature with respect to the usual attendees of VLDB: researchers, students and practitioners. The goal is to educate the attendees about the hype and the reality of blockchain technology and its intended usage in non-cryptocurrency use cases.

3. Scope and Structure

The scope of the tutorial is general in nature without getting into the nitty gritty of, e.g., cryptographic algorithms or the distributed consensus protocols. I provide introductory material on the basic aspects of blockchains and discuss the different projects that are in existence. Due to my personal involvement in the Hyperledger Fabric project related work at IBM and because this effort is more mature than other projects (e.g., Enterprise Ethereum and R3 Corda), I describe in greater detail Fabric V1 architecture, components and algorithms. I also try to give the audience a sense of where lot more work needs to be done, what the open issues are, and where there is scope for innovations and optimizations. I also emphasize scaling, performance and failure handling problems. I do not get into cryptocurrencies, mining and proof of work related concepts. I only discuss permissioned/private blockchains and not permissionless ones.

4. Speaker Biography

Dr. C. Mohan has been an IBM researcher for 35 years in the database area, impacting numerous IBM and non-IBM products, the research and academic communities, and standards, especially with his invention of the ARIES family of database locking and recovery algorithms, and the Presumed Abort commit protocol. This IBM (1997), and ACM/IEEE (2002) Fellow has also served as the IBM India Chief Scientist for 3 years (2006-2009). In addition to receiving the ACM SIGMOD Innovations Award (1996), the VLDB 10 Year Best Paper Award (1999) and numerous IBM awards, Mohan was elected to the US and Indian National Academies of Engineering (2009), and was named an IBM Master Inventor (1997). This Distinguished Alumnus of IIT Madras (1977) received his PhD at the University of Texas at

Austin (1981). He is an inventor of 50 patents. He is currently focused on Big Data, HTAP and Blockchain technologies (**http://bit.ly/CMbcDB**, **http://bit.ly/CMgMDS**). Since 2016, he has been a Distinguished Visiting Professor of China's prestigious Tsinghua University. He has served on the advisory board of IEEE Spectrum, and on numerous conference and journal boards. Mohan is a frequent speaker in North America, Europe and India, and has given talks in 40 countries. He is very active on social media and has a huge following. More information could be found in the Wikipedia page at **http://bit.ly/CMwIkP**

5. REFERENCES

[1] ASTRI. *Whitepaper on Distributed Ledger Technology*, November 2016. http://www.hkma.gov.hk/media/eng/doc/key-functions/finanical-infrastructure/Whitepaper_On_Distributed_Ledger_Technology.pdf

[2] *BLOCKBENCH: A Framework for Analyzing Private Blockchains*, National University of Singapore, Singapore. http://www.comp.nus.edu.sg/~dbsystem/blockbench/

[3] Brown, R., Carlyle, J., Grigg, I. and Hearn, M. *Corda: An Introduction*, Whitepaper, August 2016. https://www.r3cev.com/s/corda-introductory-whitepaper-final.pdf

[4] Cocco, S., Singh, G. Top 6 Technical Advantages of Hyperledger Fabric v1.0 for Blockchain Networks, IBM developerWorks, March 2017. https://www.ibm.com/developerworks/cloud/library/cl-top-technical-advantages-of-hyperledger-fabric-for-blockchain-networks/

[5] Digital Asset. *The Global Synchronization Log*, November 2016. https://digitalasset.com/static/documents/The_Global_Synchronization_Log.pdf

[6] Dih, A., Wang, J., Chen, G., Liu, R., Ooi, B.C., Tan, K.-L. *BLOCKBENCH: A Framework for Analyzing Private Blockchains*, Proc. ACM SIGMOD International Conference on Management of Data, Chicago, USA, June 2017. http://www.comp.nus.edu.sg/~ooibc/blockbench.pdf

[7] Gray, M. *Microsoft's Blockchain Architecture Overview*, Project Bletchley Whitepaper, September 2016. https://github.com/Azure/azure-blockchain-projects/blob/master/bletchley/bletchley-whitepaper.md

[8] Hearn, M. Corda: A distributed ledger, November 2016. https://docs.corda.net/_static/corda-technical-whitepaper.pdf

[9] Iansiti, M., Lakhani, K.R. *The Truth About Blockchain*, Harvard Business Review, January-February 2017. https://hbr.org/2017/01/the-truth-about-blockchain

[10] Interledger. *Interledger Architecture*, https://interledger.org/rfcs/0001-interledger-architecture/

[11] JP Morgan, *Quorum: A Permissioned Implementation of Ethereum Supporting Data Privacy*, September 2016. https://github.com/jpmorganchase/quorum-docs/raw/master/Blockchain_QuorumHyperledger_20160922.pdf

[12] McConaghy, T. et al. *BigchainDB: A Scalable Blockchain Database*, Whitepaper, June 2016. https://www.bigchaindb.com/whitepaper/bigchaindb-whitepaper.pdf

[13] McConaghy, T. *BigchainDB 2017 Roadmap*, 23 March 2017. https://blog.bigchaindb.com/bigchaindb-2017-roadmap-d2e7123f9874

[14] Meunier, S. *Blockchain technology—a very special kind of Distributed Database*, Medium Article, December 2016. https://medium.com/@sbmeunier/blockchain-technology-a-very-special-kind-of-distributed-database-e63d00781118#.way70psdu

[15] Murthy, C. *Blockchain DB-unked*, Presentation Slides, August 2016. https://ripple.com/files/db-unked.pdf

[16] Popper, N., Lohr, S. *Blockchain: A Better Way to Track Pork Chops, Bonds, Bad Peanut Butter?* New York Times, 4 March 2017, https://www.nytimes.com/2017/03/04/business/dealbook/blockchain-ibm-bitcoin.html?

[17] Quorum Whitepaper, Version 0.1., November 2016. https://github.com/jpmorganchase/quorum-docs/raw/master/Quorum%20Whitepaper%20v0.1.pdf

[18] Ripple. *Product Overview*, March 2017. https://ripple.com/files/ripple_product_overview.pdf

[19] Santo, A. et al. *Applicability of Distributed Ledger Technology to Capital Market Infrastructure*, JPX, August 2016. http://www.jpx.co.jp/english/corporate/research-study/working-paper/b5b4pj000000i468-att/E_JPX_working_paper_No15.pdf

[20] Sawtooth Lake, https://intelledger.github.io/introduction.html

[21] Summerwill, B., Khatchadourian, S. *Enterprise Ethereum Alliance Technical Roadmap*, 28 February 2017. https://bobsummerwill.files.wordpress.com/2017/02/enterprise-ethereum-technical-roadmap-slides-final.pdf

[22] Tapscott, D., Tapscott, A. *Blockchain Revolution: How the Technology Behind Bitcoin Is Changing Money, Business, and the World*, ISBN 978-1101980132.

[23] Vukolic, M. Rethinking Permissioned Blockchains, 1st ACM Workshop on Blockchain, Cryptocurrencies and Contracts, Abu Dhabi, UAE, April 2017. http://vukolic.com/rethinking-permissioned-blockchains-BCC2017.pdf

[24] Walch, A. *Open Source Operational Risk: Should Public Blockchains Serve as Financial Market Infrastructure?* Chapter in Handbook of Digital Banking and Internet Finance, Vol. 2, D. Lee, K. Chuen, R. Deng (Eds.), https://ssrn.com/abstract=2879239

[25] *Welcome to Fabric.* Fabric Documentation, March 2017. http://hyperledger-fabric.readthedocs.io/en/latest/

[26] Wikipedia. *Blockchain (database)*. https://en.wikipedia.org/wiki/Blockchain_(database)