Chapter 3
Extracting Insights From Bitcoin Transactions:
Data Warehouse Modeling and Analytical Questions

Rim Moussa
University of Carthage, Tunisia

Alfredo Cuzzocrea
University of Calabria, Italy

ABSTRACT

Bitcoin is the most well-known cryptocurrency. It was first released in 2009 by Satoshi Nakamoto. Bitcoin serves as a decentralized medium of digital exchange, with transactions verified and recorded in the blockchain. The latter is a public immutable distributed ledger that operates without the need of a trusted record keeping authority or a central intermediary. It provides OLTP capabilities with both atomic transactions and data durability guarantees for blockchain transactions. Blockchain ledgers were not designed to perform analytics questions. The availability of the entire bitcoin transaction history, stored in its public blockchain, offers interesting opportunities for analyzing the transactions to obtain insights on users/entities patterns and transactions patterns. For these purposes, the authors need to store and analyze cryptocurrency transactions in a data warehouse. In this chapter, they investigate public blockchain datasets, and they overview different data models for setting up a data warehouse appliance of cryptocurrencies.

INTRODUCTION

Blockchains use cases are emerging in the financial services, such as supply chain, media, and many highly digitized industries. Blockchains are being used for distributed value exchange, based on cryptographically signed, irrevocable transactional records shared by all participants in a network. Each record contains a timestamp and reference links to previous transactions. The Bitcoin blockchain in particular

DOI: 10.4018/978-1-7998-5839-3.ch003

aims to remedy financial industry flaws. As motivated by Satoshi Nakamoto (Nakamoto, 2008), it is the first truly crypto-currency which does not discriminate its users based on citizenship or location, is available all time, and is secure with very low fees. It manages the life cycle of digitalized assets and immutably records operations in a distributed ledger. A digitalized asset can be any valuable object (e.g. crypto-currencies, securities, patient health records). Users trade electronically and more anonymously than via traditional electronic transfers. Bitcoins design keeps all transactions in a *public immutable distributed ledger*.

The Blockchain guarantees three main features – *Accessibility*, *Security*, and *Accountability*. Blockchain, being shared by all parties, makes data accessible for everyone involved. The data is stored on every computer, so that it is both decentralized and distributed. This enables a high level of security because intruders would need to access and alter the data on all linked computers at the same time in order to change one transaction. As a single, and fixed cache of information, Blockchain ensures accountability by everyone in the network.

While blockchain ledgers provide OLTP capabilities namely atomic transactions and data durability for transactions, they don't support On-Line Analytical Processing workloads (OLAP). OLAP performs multidimensional analysis of business data and provides the capability for complex calculations, trend analysis, and sophisticated data modeling. The capability to regularly generate time-scale and ergonomic reports on specific or aggregated money flows stored in the ledger is very important. The inability to easily build reports from the blockchain can reduce transparency and increase the difficulty of price discovery of BTC versus fiat currencies (e.g. US$, euro,...), as well as other fundamental analytical questions such as transactions and entities' patterns. Consequently, blockchain data must be ingested into a data warehouse system to be queried efficiently. Typically, Data Warehouses are implemented on relational stores. Achieving scalability and elasticity is a huge challenge for relational database management systems. Relational databases were designed to run on a single server in order to maintain the integrity of the table mappings and avoid the problems of distributed computing. The scalability, fit-to-data model, denormalization, and schema flexibility makes NoSQL stores a viable alternative option. NoSQL stands for "Not Only SQL". The most common types of NoSQL databases are key-value (e.g. Redis, Amazon DynamoDB), document (e.g. BaseX, MongoDB, CouchDB, ElasticSearch), column (e.g. BigQuery, Apache Drill, Cassandra, Apache HBase), and graph databases (e.g. Neo4j, Apache ArangoDB, JanusGraph, RedisGraph). Graph compute engines can be used in online analytical processing (OLAP) for bulk analysis (Chen, 2008).

This chapter describes different data models for setting up a data warehouse appliance for crypto-currencies. For that purpose, we focus on the relational model, the nested-immutable model, and the graph model. For each model, we show typical queries which execute on the data warehouse.

Blockchain analytics specifically of Bitcoin blockchain should provide insight into a variety of economic indicators, illegal activities (e.g. ransoms, tracking sellers and buyers of illegal items, tracking laundering of large sums of money, gambling…).

The chapter is organized as follows, first we introduce key concepts of bitcoin transactions. Then, we present a sketch of Blockchain Relational Data Warehouse and detail integration workflows and typical business questions. After that, we present the nested-immutable model implemented by *Google* proposed as a cryptocurrency *warehouse* on *BigQuery.* We also present different graphs modeling and detail the insights they allow to extract. Finally, we conclude the chapter and present a research agenda.