Project 2 - BI/DW Solution

ISDS 556

Bitcoin.org

Presented to
Dr. Turel

Spring 2017

Table of Contents

I. Overview:

Bitcoin is a payment system and which is also known as a cryptocurrency. Bitcoin is a system which enables a peer-to-peer transaction between users by taking out the middleman. The transaction is very cheap as the processing fee is negligible for the user. Bitcoin allows worldwide payment transfer with maintaining user privacy. The proof of concept for bitcoin was first published in 2009.

There are many companies working on this system. Bitcoin was invented by a group of programmers which is known as Satoshi Nakamoto. Bitcoin is an open source concept. Bitcoin.org was first registered by two developers Satoshi Nakamoto and Martti Malmi. Later, Nakamoto left the group and gave the ownership to other people to spread responsibility and to prevent the specific group from taking the ownership of the concept. All the bitcoin users control bitcoin all around the world.

The following are the missions of Bitcoin:

- Inform users to protect them from common mistakes.
- Give an accurate description of Bitcoin properties, potential uses, and limitations.
- Display transparent alerts and events regarding the Bitcoin network.
- Invite talented humans to help with Bitcoin development at many levels.
- Provide visibility to the large-scale Bitcoin ecosystem.
- Improve Bitcoin worldwide accessibility with internationalization.
- Remain a neutral informative resource about Bitcoin.

II. Description of the Organization and Need for BI:

Though the Bitcoin is a hot topic in the current market, Bitcoin is struggling to maintain its market share. There are some deficiencies in the Bitcoin model which are responsible for these difficult times. Recently, Bitcoin was banned from China as to make a deposit, Chinese government required the identification of the user for the account. User anonymity is a key feature of Bitcoin, and by this decision, the user base of bitcoin has been decreasing. Recently, the SEC declined the proposal of an ETF setup for Bitcoin users. The entrepreneurs Tyler and Cameron Winklevoss introduced this proposal. The incidence of SEC rejecting ETF resulted in the decrease in share value of Bitcoin in the market.

We identified the factors which are obstacles to Bitcoin business growth and design a dimensional data warehouse model which will help to build various BI systems to identify and resolve the problems which are causing the reduce in share value for bitcoin users.

Below is a representation of the dimensions which were identified to create sales dashboard.

Dimensions

Business Process	Customers	Companies	Date	Bitcoin	Location	Transaction	Time	Currency
Sales	X	X	X	X	X	X	X	X

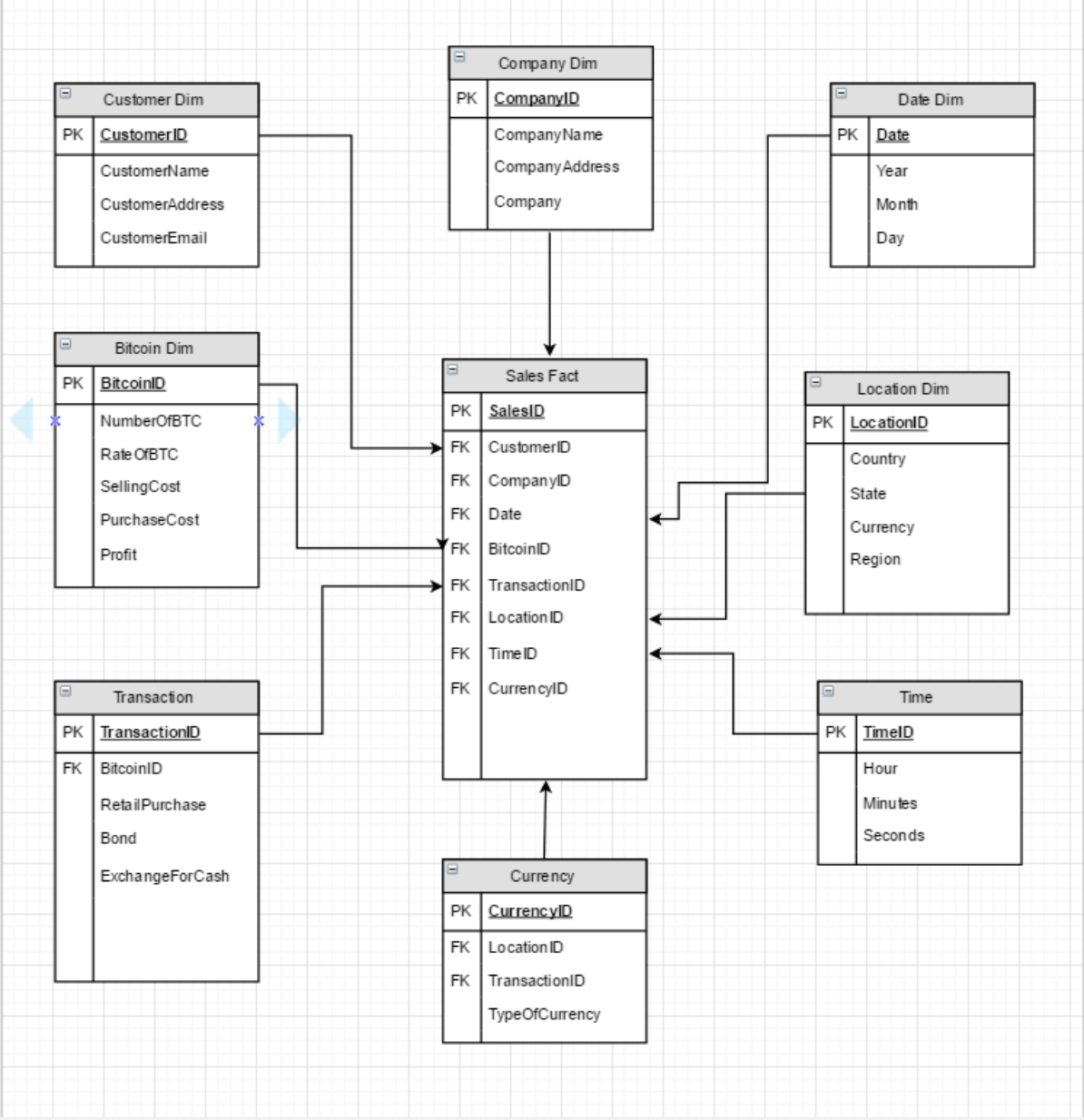
III. Description of the Functionality and Features of the Required BI system:

We are focusing on Sales Processing. Because we assume that this process is the important process in the value chain. In addition, the KPI of the BI system would be number of bitcoins, quantity purchased and cost of bitcoins.
Many types of transactional as well as summarized reports can be generated out of the system. Both summarized reports and transaction level reports would be used by upper management while the staff would be given access on transaction level reports. The reports could be run at any time. The BI application would fetch the data for presentation/analysis.

IV. Data Model:

We have developed a star schema for our data model, in which there is one fact table and 8 dimension tables. In the Customer and bitcoin dimension table all the attributes are being treated as Type-2, while in other dimension tables, all the attributes are being treated as Type-1. The attribute Sales ID is being treated as degenerate dimension in the fact table.

Note: The Date Dimension is a role-playing dimension.



V. Data Analysis:

In our data model, we evaluate the number of BitCoin transactions based upon their location, volumes purchased, currency and country. We also track the transaction records and thus it helps in eliminating the possibilities of a fraudulent transaction. The data is appropriate in nature and thus does not require additional data cleaning methods.

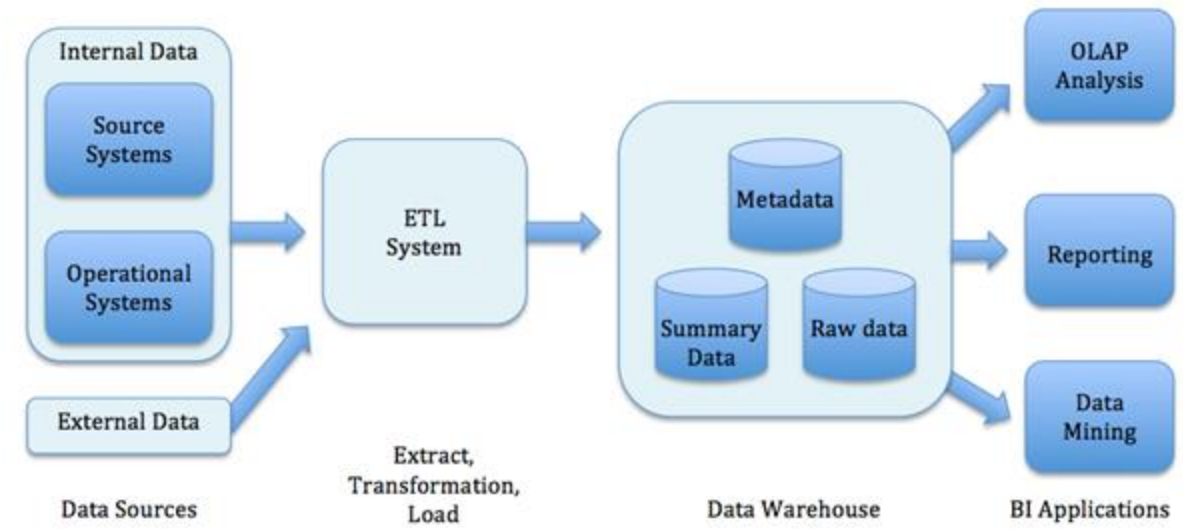
1. Total number of transactions with respect to time, location, and currency?
2. Maximum inter-country transaction by day/week/month?
3. Estimation of total revenue based on your BitcoinID and Type of Currency?
4. Most number of transactions that occurred in a single country?

5. Evaluating the KPI's based on the Number of Bit Coins, Rate of the purchase and Selling Cost and the profit that is generated?
6. Estimation of one of the frequently used services of Bit Coin and thus evaluating the performance regionally?
7. Who are the best customers in terms of the amount of purchase by day/month/year?
9. Customers from which state used our services by day/month/year?
Like that many other reports can be generated using this data model.

VI. Extraction, Transformation, and Loading Issues:

Data extraction might be an issue even though the data model is quite simple and contains minimal errors and incomplete data. Extracting the data takes a very large amount of time as it contains plenty of user parameters and has to take into consideration the security measures that needs to be followed. Applying validation check to the user interface (where users input the data) could at least reduce the formatting problem in the data, thus faster extraction.
Since all the data comes from customer database, it should be mostly consistent. We might want to have a validity check on customer phone number because the formatting could be different for each customer. Some might put parenthesis around area code or dash between chunks of numbers. E.g. (xxx)xxx-xxxx while some might put only just numbers. Each country has its own set of pin codes and each region has its different area code. Thus, this aspect must be taken into consideration as it is a very important phenomenon. Aggregations are also needed because we need to show many summary stats like the currency which is used most of the times or the countries that use Bitcoin majorly. Also at the end, all data will be sorted.
Extracting/Loading the data majorly depends upon the customer information as the customer data is very diverse in nature and consists of plenty of variations and redundancies. In general, we need real time loading as the data is sensitive in nature as it contains customer's demographic data as well as its financial data. The data if not updated on the regular basis, then it will cause fluctuations in the data.

VII. The Recommended Architecture:



The data source contains internal data and external data. The internal data includes the data from source systems and operational systems, such as OLTP system and ER system from different stores. And the external data includes the data from Bitcoin.org.

Next, the ETL system will extract, transform, and load all the data into data staging area. ETL system will integrate and process all the data from multiple systems in different stores. For example, looking up the Transaction status in an area while updating the Transaction status in other locations, sorting the new sales file while removing the duplicates on another file, or loading the historical sales data while creating the new sales data.
Metadata, summary data, and raw data describe stage in presentation area. Metadata summarizes basic information about data, which can make finding and working with particular instances of data easier. For example, retrieving the data of monthly Transaction of location and summarizing sales quantity of Bitcoin.
In the end, the end users can use BI applications, such as OLAP analysis, reporting, and data mining skills for getting the transaction and revenue reports from each country or area, finding potential patterns of sales, predicting the spending behavior of each customer, summarizing the average time the Bitcoin.org takes to transfer the bitcoin from one source to another.
Therefore, Bitcoin.org could use this application architecture model for improving its transaction processing procedure. First, it could analyze multidimensional data interactively from multiple perspectives by OLAP analysis. Second, it could get detailed sales summary reports and specific reports on each order process by reporting BI application. Third, it could use data mining skills to develop other data analysis. For example, classifying frequent Customers, predicting future transactions, finding association rules on its products, and creating multidimensional visualization charts for upper managers
In the end, the BI applications could benefit Bitcoin.org by decreasing transaction processing time, increasing sales for its company, creating value on its products, and further expanding its reputation.

VIII. Preliminary Budget Plan:

According to a research conducted by Cambridge University in 2017, there are between 2.9 million and 5.8 million unique users actively using a cryptocurrency wallet, most of them using bitcoin. It is not sufficient considering the population of the world. To improve the revenue, we have recommended a BI solution to Bitcoin.org. This provides a way to utilize the data existing customers and new customers to improve the sales based on various factors suggested above. The package costs about $40,000 per year.
Program Manager is required to communicate between the sponsor and all other teams. Core Project Team consisting of Business Analyst who converts the requirements into technical architecture. Data Architect for developing the overall data architecture strategy of the DW/BI system, ensuring reusability, integration, and optimization. The ETL developer is responsible for extract, transform and load data from related operational department to the BI system. He/she will also work for other backroom tasks. The BI developers are responsible for overall front room in the system. Data Mining Specialist develops data mining models and works with the BI application developers to design operational applications that use mining capabilities. The recruitment for these new employees could cause a salary increase of about $4, 00,000 per year.
In order to build a DW/BI system, the organization will have to invest on the creating a group exclusively for DW/BI tasks. Bitcoin is a cryptocurrency and an electronic payment system and

is a worldwide accepted currency system. This organization will definitely require a considerably bigger sized DW/BI team considering the increasing number of users every year. We concentrate mainly on Customer Acquisition, Sales and Revenue KPIs.

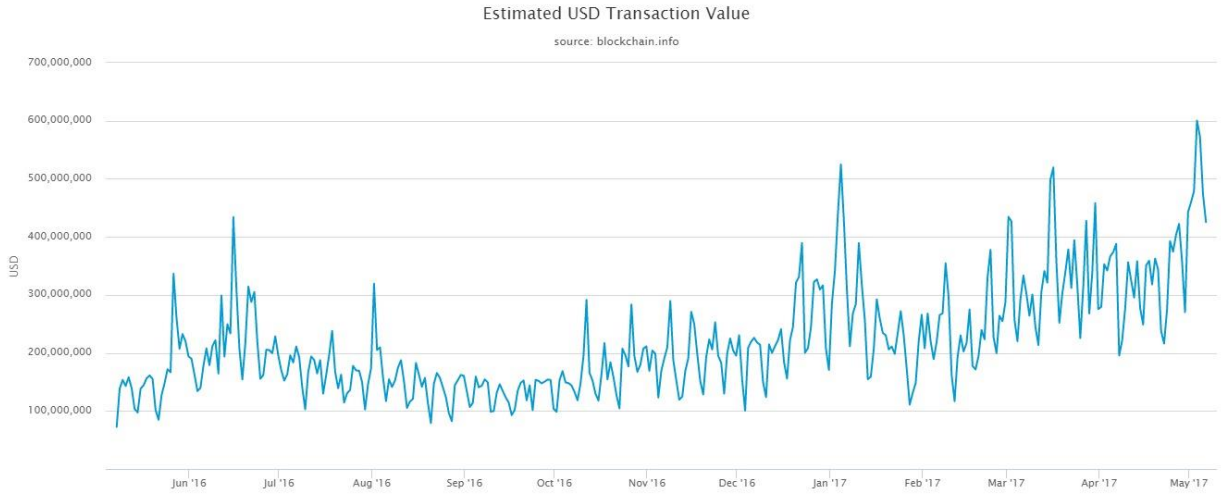
Even if the organization spends around 450000$ per year including the cost for different marketing strategies, they can definitely improve the customers. Firstly, the BI reports gives the company insights about the Bitcoin usage in different domains. The highest used domain like exchange of bitcoins can be ignored and lower used domain such as bonds can be improved. It basically gives the trends on the Bitcoin usage.

Secondly, the system help the company to get insight into customer behavior, it is unnecessary to hire an external consultant to do this kind of research any more. Moreover, since BI generates key business reports when and where you need them, it saves money to hire someone else to generate the reports. The total savings are expected to be at least $500000 including the increased number of customers, improved security and transaction, increased revenue for every transaction etc.

After introducing DW/BI, the company could have a yearly saving of about $50,000. Although it is a small number, it could not be neglected considering the fact that the investment is a long term one. It is already proven that a company based on Data Analytics tends to make a better ROI compared to the company that doesn't. As to the benefits it will bring to the business, the company is expected to generate higher revenue in the following years with an increasing number of customers and also providing a secure transactions.

Processes	Expenses
BI Products	$40,000
Salary	$400,000
Marketing	$60,000
Total Expenses	$500,000


Confirmed Transactions Per Day
source: blockchain.info
400,000
350,000
300,000
250,000
200,000
Transactions
Jun '16
Jul '16
Aug '16
Sep '16
Oct '16
Nov '16
Dec '16
Jan '17
Feb '17
Mar '17
Apr '17
May '17


Estimated USD Transaction Value
source: blockchain.info
700,000,000
600,000,000
500,000,000
400,000,000
300,000,000
200,000,000
100,000,000
USD
Jun '16
Jul '16
Aug '16
Sep '16
Oct '16
Nov '16
Dec '16
Jan '17
Feb '17
Mar '17
Apr '17
May '17

IX. Conclusion:

Faster Time to Value is the word that can be used when thought of BI solution. Every company aims at improving their return on investment by implementing different strategies. The best strategy is to implement a DW/BI system. As BI has become an increasingly popular and accepted technology for improving business performance. The new BI system can help the company get faster answers to business questions, get key business metrics reports (transactional and summarized) and also get insight into customer behavior. It is estimated that after introducing the BI system, the company could achieve a significant amount of yearly saving.

Every successful process starts with right planning before it is deployed. To make sure all the development efforts don't go in vain, the team's efforts need to be coordinated. We have included the KPI's that we have considered to build our BI solution. We have also included the multiple issues that were addressed like extraction, transformation and loading. We have provided a basic budget plan for our BI solution.

We would next develop front room and backroom provided in our architecture, deploy the DW/BI system. The developed system will be tested against different input conditions to make sure it generates valuable reports for such conditions along with documentation. In addition, documentation and user training will be required. The system will be built in robust way to handle different upcoming business issues. We would also maintain all the components of the system.