BIG DATA IN BANKING

Constantine DZIK

Utech LLC location – Madison, MS USA

Farooq MOHAMMED

Utech LLC location – Madison, MS USA

Dzmitry BALTUNOU

Utech LLC location – Madison, MS USA

Utech LLC location – Madison, MS USA

Big data is a hot topic in banking but yet it is a complex one,which is clearly in the early stages of evolution. Many organizations drive business using big data technologies are still challenged by multiple issues and rapidly shifting technologies. Expecting positive results on big data investment needs solution that focuses on purpose, people, and process prior to exploring data and technologies.Big Data opens up innovative ways for banking organizations to interact and leverage data. Banks should be flexible to changes, willing to be prudent and practical in their strategy for rollout. They should be willing to adapt to calculative risk strategy requiring to be open to failures, on a small scale if not large, resulting in large-scale success.As a result, organizations need to change their focus on their architecture, development, migrations, and maintaining big data solutions. In this presentation, we will discuss: traditional data management in banking, risks and impediments to using big data, success stories of banks using Big Data and use cases.

Big data is a hot topic in banking but yet it is a complex one, which is clearly in the early stages of evolution. Many organizations drive business using big data technologies are still challenged by multiple issues and rapidly shifting technologies. Expecting positive results on big data investment needs solution that focuses on purpose, people, and process prior to exploring data and technologies. Big Data opens up innovative ways for banking organizations to interact and leverage data.

Banks should be flexible to changes, willing to be prudent and practical in their strategy for rollout. They should be willing to adapt to calculative risk strategy requiring to be open to failures, on a small scale if not large, resulting in large-scale success. As a result, organizations need to change their focus on their architecture, development, migrations, and maintaining big data solutions.

Banking industry is greatly dependent on Information Technology now, than ever before to compete with constantly changing business methodologies and global economy.

Banks use data warehousing and data mining methodology to build long-term relationships with their customers. To make data useful, bank enterprises collect data from almost every platform and data is formatted, cleansed and transformed into information that users will understand and store the information in an efficient data warehouse struc-

ture. Data analysis provides ability to use the same information on different levels. This is where data mining comes in play to explore the information stored in data warehouse structure.

Below are the tree most prominent areas in banking sector, forcing banks to change traditional ways of handling data.

- Challenges with managing vast amount of data and accurately reporting financial positions to both regulatory agencies and the general public.

- Banks are failing to address the magnitude of the problems they face around data risk aggregation.

- Legacy infrastructure leading to incompatibility issues with advanced banking technologies.

Data is the source for every business decision making process, but typical data is not in the correct format to support decision-making. Decision makers must have the right information at the right time to make informed and intelligent decisions.

Information flows from various sources like Social Media, Machine Data, VoIP, Search Engines, & Emails.

Wonder what happens in an internet minute?

- 4.0 Million Google searches.
- 433,000 Twitter tweets
- 3.3 Million Facebook Posts
- 204 Million Emails
- 1.4 Million Minutes Voice Calls
- 50 Billion WhatsApp Messages

In present conditions, various financial organizations are waking up to new realities. It is time to rethink and re- innovate their data management strategies. Banking organizations are realizing that they can no longer depend on data management using a "data provider" mindset that is vaguely connected to downstream functionality.

Large Enterprise systems use a typical pattern of DATA WAREHOUSE:

Interactive DWH handles requests coming from the Web site or any other custom applications, data is then retrieved from a relational database and loaded into the data warehouse for further processing and archiving. Data is usually de-normalized in OLAP (online analytical processing).

Unfortunately, modern DWH systems cannot accommodate the enormous amount of data that is created in large companies and then there is a need to compromise, partial data copied to DWH or purged after a certain timeframe. There is no need for such trade-offs, if Hadoop is used as the intermediate layer between the online database and data warehouse.

Performance of data processing increases in proportion to the amount of storage, while in high-end servers with traditional DWH the increase in storage is easily achievable, but computing performance remains the same.

Cutting edge Hadoop technology increases processing performance by simply adding new nodes to the data warehouse.

DWH advantages:

- Handle complex transactions.
- Processes hundreds of thousands of queries per second.
- Real time results.
- Simple but effective query statements.

DWH Disadvantages:

- Defined schema required prior to data transfer.
- DWH capacity amounts up to hundreds of terabytes of storage.
- Limited amount of record results in a single query.
- Change in limitation results in performance issues.

Hadoop Advantages :

- Low-cost hardware and software
- Schema-free data staging
- Supports widely diverse data and file types
- Understanding consumer behavior via clickstreams
- Recognition of sales and market opportunities
- Fraud detection and Enhanced transactions

Hadoop Disadvantages :

- Demand and Supply ratio is too low for Hadoop Skillset
- Security for Hadoop data
- Excessive hand coding required of Hadoop
- Cost of implementing a new technology
- Challenges implementing architecture of big data analytic system
- Big data not usable for end users

Financial organizations might decide to collect data from various sources including social networking and collaborate it with data from the internal DWH to update customer's acquaintance. In this case, we can use Hadoop to quickly bank on that person's social influence and then data being relayed back to the DWH so that the event coordinator can view that person's influence and re-segment accordingly.

Hadoop isn't going to replace enterprise data warehouse. For example, Alasdair Anderson of HSBC gave great, concrete example: "There's no relationship between the EDW and Hadoop right now — they are going to be complimentary. It's NOT about rip and replace: we're not going to get rid of RDBMS or MPP, but instead use the right tool for right job — and that will very much be driven by price."

“Deutsche Bank has been working on a big data implementation since the beginning of 2012 in an attempt to analyze all of its unstructured data. However, problems have arisen while attempting to unravel the traditional systems – mainframes and databases, and trying to make big data tools work with these systems. The bank has been collecting data from the front end (trading data), the middle (operations data) and the back-end (finance data). Petabytes of this data are stored across 46 data warehouses, where there is 90% overlap of data. It is difficult to unravel these data warehouses that have been built over the last two to three decades. The data integration challenge and the significant investments made by the bank in traditional IT infrastructure pose a key question for the bank’s senior executives – what do they do now with their traditional system? They believe that big, unstructured and raw data analysis will provide important insights, mainly unknown to the bank. They need to extract this data, streamline it and build traceability and linkages from the traditional systems, which is an expensive proposition.”

Source: Computerworld UK, Deutsche Bank: Big data plans held back by legacy systems, February 2013. Web. 07 June 2015

“Bank of America is focusing on big data with an emphasis on an integrated approach to customers and internal operations. The key objective of its big data efforts is to understand the customer across all channels, interactions, and presenting consistent appealing offers to well-defined customer segments. For example, the bank utilizes transaction and propensity models to determine which of its primary relationship customers may have a credit card, or a mortgage loan that could benefit from refinancing. When the customer accesses the bank’s online channel, calls a call center, or visits a branch, that information is available to the online app or the sales associate to present the offer. The bank has launched a program called ‘BankAmeriDeals’, which provides cash-back offers to holders of the bank’s credit and debit cards based on analysis of where they have made payments in the past.

Banks are moving from shared-services data modeling environment to a dedicated ‘Grid Computing’ platform to drive operational efficiency by early detection of high-risk accounts. This initiative is benefiting in several ways, such as reducing its loan default calculation time for a mortgage book of more than 10 million loans from 96 hours to just four. Ad hoc processing times has improved three times when compared with previous environment.

Banks are modifying it's organizational structure in line with big data initiative by historically employing several quantitative analysts. In order to support its big data initiatives, banks consolidated dispersed analytics talent and also set up matrix reporting lines from its analytics teams to a central analytics group as well as business units”.

Source: Davenport, Thomas H., and Jill Dyché. "Big Data in Big Companies." SAS Institute Inc. N.p., May 2013. Web. 7 June 2015.

“Rabobank named big data as one of the 10 most important trends in their 2013 yearly report and started developing a strategy around it. They created a list of 67 possible big data use cases, divided them into four categories – fix organizational bottlenecks, improve efficiency in business processes, create new business opportunities and develop new business models. For each of these categories they measured IT impact, time required for implementation, and business value proposition. The bank moved ahead with big data applications for the improvement of business processes due to their low IT impact and the possibility of a positive ROI.

Rabobank started with a few proof-of-concepts using only internal data. Later, the bank extended the scope of its big data program to include web data (click behavior), social network data, public data from government sources and macro-trend data. The bank built small clusters using open-source technology to test and analyze unstructured data sets, which kept costs low and offered the scalability to expand. A dedicated multidisciplinary team was setup to implement big data use cases. The team experimented with small and short implementation cycles.

One of the use cases at Rabobank involved analyzing criminal activities at ATMs. Rabobank found that the proximity of highways, and the season and weather conditions increased the risk of criminal activities. The bank also used big data tools to analyze customer data to find the best locations for ATMs. Based on its initial success with big data analytics, Rabobank is now focusing on addressing more pressing big data issues around privacy concerns and data ownership.”

Source: Rijmenam, Mark Van. "Datafloq - Connecting Data and People." Datafloq Read RSS. Datafloq, 2 May 2013. Web. 07 June 2015.

Digital revolution dictates the future of companies: they have to become a "smarty" of the business. Now, thanks to Hadoop and other technologies so-called big data companies can treat unstructured data as a whole. ETL streams from various sources into the Big Data Lake of a bank allow them to analyze real time scenarios to identify fraud, shape its marketing strategy and make special offers based on the customer’s online behavior.

However it’s not just about the volume of data because this is what first comes to mind at the mention of big data. On the contrary, the main thing is that between this data regardless of the type and source there is a very important hidden relation, such as between the information from the call center, data on the use of the website and sales amounts. The ability to assess the data is much more important whether the flow of information directly from the Internet, or part of it leaked through the firewall, sensor data or information from public sources and then link all that into a single coherent picture.

So the youthfulness of big data is definitely not blocking, but the new turning point of data use. We can understand from the above examples that large organizations are

joining traditional analytics and big data together to create a synergistic effect for the business. It is no doubts, this is an evolutional process for the companies, but the sooner they start the transition to the new model the more favorable return of investments will be. Practicing and building proof of concepts plays a vital role, while architecture, technology, skills, organizational structure should also undergo the change process. The main power of big data is coming not from accumulating enormous amounts of raw data, but rather from sharp analysis, discovering hidden relations of the unstructured data.

References

1. Computerworld UK, Deutsche Bank: Big data plans held back by legacy systems, February 2013.

2. Davenport, Thomas H., and Jill Dyché. "Big Data in Big Companies." SAS Institute Inc. N.p., May 2013.

3. Rijmenam, Mark Van. "Datafloq - Connecting Data and People." Datafloq Read RSS. Datafloq, 2 May 2013.