Company level Big Data Management

Tamás Orosz
orosz.tamas@arek.uni-obuda.hu

István Orosz
orosz.istvan@arek.uni-obuda.hu

Abstract — Carefully Processing enormous data in Company-level evokes new tools and methods in Enterprise Data Management. The volume of information and data increases in an exponential way, which makes strategic issues not only for individuals but also in Company-level. Typical approaches are well known for these kinds of problems. Date is stored in Databases, Data warehouses and Cloud applications. There are various well-known and appropriate solutions are ready-to-use for any of the above listed applications, but they are rarely complete at Corporation levels. Therefore, Enterprises require temporary solutions. The current article discusses the main issues together with the complete and temporary solutions for typical Data management. Effectiveness, efficiency and costs are demonstrated in different point of views, e.g. Project and Knowledge Management.

Keywords: Big Data Management, ERP Systems, Cost-efficiency, change management

I. INTRODUCTION

Big Data is moving from a focus on individual projects to an influence on enterprises strategic information. Dealing with data volume, variety and complexity is leading organizations to abandon the concept of a single enterprise data warehouse[1][2][3] containing all information needed for decisions. Rather than, businesses need a focus on multiple systems, including content management and data centers combined with data services and metadata, acting as an “enterprise data warehouse,” according to Gartner. Big data combined with actionable analytics open up a whole new range of possibilities for business. These solutions will help businesses not only integrate and understand these new types of data, but will also get to better answers with suggested actions to enable people to calculate and/or recalculate business possibilities in real time.

All the systems and processes, which are necessary for achieving this is known as Big Data Management. The major parts and flow of data can be seen in Figure1. Although it is a popular terms nowadays, the meaning of Big Data varies from company to company. Manufacturing and companies dealing with supply chain [12] it is a new area. If a simple definition is needed, in that case it is huge volume of data which is hard to process.

There are several reasons behind which causes the processing issues:
1. The data structure can be complex
2. The data in unstructured
3. The data requires deed analysis

New tools are needed to solve these complex issues. The main reason why it is in focus is, that it can improve the efficiency of supply chain processes, while reducing costs. In the end, the company can become a data-driven one, and the management will have the chance to make real time decisions.

If we look at these companies, there are getting fairly low margin, so every little detail and advantage counts. Mistake and not effective business processes can have a huge negative impact on customer relationships and competitiveness. Big Data analysis [4][5] can be a key point to improve. The ability to use these kind of almost live data from a company’s warehouse management system (WMS) and sensors throughout the transportation chain combined with external information from customers and suppliers can make these kind of capacity management and allocation of resources more agile for a real-time decision making. Adopting a robust and complex data analytical system is a complicated task, but it doesn't have to be always like this. The first step is for
everyone to evaluate the company Big Data requirements and needs how they fit within a business intelligence (BI) environment in the future or in the now existing model. Most companies already have an implemented ERP system in place. The necessary and suitable business intelligence solution can be up and running above the ERP system, as another layer, by relying on extended integration platforms shown in Figure 2 and Figure 3, with a wide range of data sources, with standard and customized cubes, ready-to-use analyses, and business reports which will provide the real value of business intelligence.

Figure 2. setup of BI tools in Microsoft Dynamics AX2012

A vast range of systems are the on the platform for keeping a company’s business running. Extracting data from each can be very complicated and time consuming. Not only the system complexity is very high, but the amount of data which is being generated and stored is very much. That doesn't mean that this amount of data can't be processed and digested quickly and easily by business experts.

The top BI and analytics solutions [6][7] deliver all of the needed analytics tools in a single platform. These include data discovery tools, self-service analytics, reporting tools, and dashboards in an integrated solution.

So it is does not matter if a company extract data from a data warehouse instead of a standard database. These tools were designed to take the IT department out of the Big Data process and put it in the hands of the everyday business user.

To go even further, with the right business intelligence tool it's possible to set parameters for higher level, so that if a KPI falls outside a predicted range, the users are automatically notified when the company has achieved a goal or deviated from the norm. There's never a need to run through reports when the system can do all the work by automatically in batch processing. Users can even access company analyses on their mobile devices as well.

Figure 3. Big Data: volume, velocity and variety (based upon past presentations on Microsoft's Channel 9)

If it comes to volume it is already not about megabytes or gigabytes like with traditional ERP and CRM, rather than it is about terabytes, petabytes or even exabytes. As for velocity, it is about real-time data. And variety is not only about structured and unstructured data but vastly different types of data. The volume is quite incomprehensible but let's try to put some numbers to it; one petabyte is one quadrillion bytes and an exabyte is the thousand fold of this. It is estimated that Walmart network of shops located in the US gathers more than 2.5 petabytes of business and internal transactional data in every working hour.

II. MANAGING BIG DATA

The vision on BI is based on empowering the end user to interpret data-even Big Data. The domain of Big Data is ruled by open source vendors that are supplying excellent platform solutions to store and process the volume and variety of data associated with Big Data. Technology-wise there is no doubt that the open source solutions can be a framework used to manage data and applications on commodity hardware. Plus a very dynamic vendor community already exists to analyze and provide extra support on open source clusters, like Hadoop.

Microsoft is opening up options for the use of Big Data by end users through its HDInsights on standard infrastructure for On-Premise and in the cloud with HDInsights for Azure. HDInsights is a Hadoop distribution that is fully compatible with Apache Hadoop. If one is reading the analyst reports on the leaders in the BI and Data Warehouse field you will see that these kind of solutions are gaining favorable assessments of its approach, by repairing the right format of master data often means standardize the data.

As the first step the data structure has to be normalized. After the missing values have to be filled, e.g. default values. Next step is standardizing the values, e.g. convert all dimensions to metric, and convert all prices to a common currency. Mapping the attributes is the last step of standardization [8].

Self-service BI options are provided through SharePoint and Excel. Think about ad-hoc, animated data visualizations through Power View, self-service data mining through Excel and working with giant datasets in Excel through PowerPivot, as can be seen in Figure 4.
The good news is that all these options are also available when analyzing Big Data!

Getting access to the data sets in the HDInsights data warehouse is handled through Hive. This add-in lets you query data in the data warehouse and use the output of the query to further process the data. A representation on how all these components of the BI stack come together is available in the following Figure 5 [21]:

But currently the world of Big Data operates through analyzing data in longer running analytic jobs and coming up with data that is refined on a higher level and can be used to drill for details.

More options are already available for handling data at high velocity within the standard technology stack. StreamInsight is a product that is part of the SQL Server stack and is especially designed for handling complex event processing applications with low-latency. StreamInsight can be used to do the first analysis of data coming out of Big Data sources such as sensor networks or, for example, social media feeds. When using StreamInsight the data flows through StreamInsight and if it adheres to certain aspects, then it can be managed as a trigger for a follow-up event.

Many Big Data deployments using the standard platform are in progress [9]. In this case the technologies are used to sift through massive amounts of data that are coming in from the e.g. telephone network, support cases, and social media feeds. StreamInsight and HDInsights process and analyze the data and the overall sentiment is calculated based on the combined input. This sentiment score is used to determine the priority of support cases and to determine which ones might be linked based on association algorithms.

The solution helps provide more insight to customer service representatives on the priority of cases and background info on how cases might be linked to each other and to network disruptions. Think about what this handling of high velocity, high volume and wide variety of data could do for your business challenges.

Now that we know more about our customers it is time to determine how we are going to reach our customers based on what we know about them.

Increasing adoption of the cloud solutions in IT departments represents a clear next step for big data, according to IBM. IBM has found that [23]:

- 80% of the world's data is still unstructured, meaning
- Businesses are making critical decisions with only 20% of the knowledge they need, and
- 56% of business leaders feel overwhelmed at how much data they need to manage

III. DATA DRIVEN PERFORMANCE

The first step on the journey toward data-driven performance is a significant one: getting access to your data. With the right BI platform [10], a company can get up and running almost instantly with acceleration packages. Pre-built data stores come standard and users experience easy adoption with pre-defined reports, dashboards, and analyses.

This operational level of BI involves the incorporation of the data sources that are most critical to the business. Basic reports and analytics are pulled easily. Management has committed to a unified version of the truth based on approved data sources, and starts to dig into internal data to reveal new kinds of actionable insight.

At the beginning of this stage, analytics is often the realm of a few super users in IT or finance. As the organization becomes more mature in BI and analytics,
standard reports give way to specialized dashboards and reports. A cycle of better insight leading to better questions and answers begins. Users become fluent in creating multi-dimensional analytics.

Information is easy to disseminate throughout the organization with dynamic storyboards and mobility features so information remains consistent and up-to-date.

By the end of this stage, an increasing number of people in an increasing number of departments are embracing BI and analytics, shown in Figure 6. At a basic level, fact-based processes and decision cycles are on the rise.

As the organization grows in analytics maturity, data sets get broader. The volume of structured and unstructured data being analyzed increases dramatically to include sources outside the ERP system. This is the time when the organization need to take it as the link between learning and doing. And analytics is accelerating that learning process.

As data volume, variety, and velocity increase so does the organization's ability to use it [11]. Analytics resources and skills are no longer limited to a few departments. Everyone from sales to marketing and supply chain is leveraging the benefits. Users across the company become smarter about the entire value chain. Learn from what you're analyzing and improve your system.

Decision making is spread throughout the company because business intelligence is arming users with the information they need to make decisions and take actions quickly and with confidence.

Analyses are increasingly used to strategically plan for the future, not just examine the past. Automated notifications based on multi-dimensional analyses keep people informed of key developments within their KPIs without keeping them glued to the screen. Users can proactively make real-time operational adjustments. Predictive analytics learns from the past to provide educated guesses about probable futures.

Users at this stage have begun to take advantage of some of the more powerful BI tools. Ad-hoc analytics give instant answers to variables inside and outside of the data warehouse. And user-friendly dashboards help make creating and sharing comprehensive reports and analyses easy.

SHAPING FUTURE STRATEGY AND COMPETING ON ANALYTICS

This is the most strategic step. Companies operating at this level of maturity employ the most highly skilled analytics resources they can find. Analytics is used to formulate the future strategy, and new KPIs emerge and are put to the test. Current strategies are adjusted and are changing based on new knowledge found within analyses.

More and more external data is used strategically. There are a vast number of data sources that are outside of your control in terms of data availability and quality, but they are highly valuable. This is the so-called Big Data.

In many cases, that data imbalance means customers are forming their opinions about your brand before they even come in contact with you. In fact, 60-70 percent of customers say their buying decisions are made before even talking to a vendor.

Analytics makes it possible to know what customers are saying and searching for by providing an easily digestible snapshot of positive/neural/negative speech about your company, products, and competitors happening on Twitter, Facebook, and LinkedIn. This type of analysis is the fastest route to understanding what's working in the market. Only here can a company listen for the questions you didn't even know to ask.

At this level, there is a dynamic integration of BI. Organizations are contending against the competition based on their analytics. Internal and external data sources are at play. This is a true analytics-driven culture.

Companies don't progress from stage one to three without spending time in between. It's critical to strive for incremental, evolutionary improvements that are achievable, rather than revolutionary leaps that promise fantastic results but also require significant shifts in corporate culture.

Another lesson learned [16] is that moving from one stage to another requires organizations to change. And as we all know, change isn't easy. There are too many companies evolve from level one to level two only to see inertia regain control. Instead of continuing to push for greater analytics power throughout the organization, they level off in what becomes a new comfort zone.

Adaptation is the key to survival in the business jungle. The courage to act is also the courage to change old habits - and keep progressing on the path toward the data-driven business.

IV. SYNTHESIS OF THE SHOWN METHODS

As a high level review of Big Data, we have tried to focus on selecting and mixing the appropriate techniques when dealing it.

We still have a long work in front to select and adopt the most improved techniques to deal with company Big Data. The goal of this study was to show, that it is not enough to select only the ERP system to us in a company, but we have to be able to solve the Big Data processing problem as well.
A new methodology has to be settle to prepare up the companies for the proper handling of issue related to bid Data management. We were focusing on resolving not only reporting issues, but the transaction processing as well. With the aid of the OLTP and OLAP techniques, we can focus the processing of KPI figures on OLAP Cubes, which are a most efficient way than the traditional SQL based one. Also we have to focus on the parallel processing in this respect.

V. CONCLUSIONS

A high level review of ERP using big Data change was given in the last pages. On the business user side, every system can offer almost the same high level of functionality, but their ecosystems differ a lot. We have tried to select the most advanced techniques to deal with this tremendous amount of data, more or less regardless the current purpose of the selected ERP system.

According to Gartner, transformation remains a critically important phenomenon across all industries. Many industries will face intense challenges in 2014 and beyond, and will have no choice but to radically change their established business models. Last year saw many industry decision-makers focusing on adopting new technologies to improve business operations by addressing developments such as the Nexus of Forces, the convergence of social, mobile, cloud, and information. Today, by contrast, leaders are significantly shifting their business models and processes.”

We are heading a long work to make a new methodology for the proper handling of Big Data.

ACKNOWLEDGMENT

The author(s) gratefully acknowledge the grant provided by the project TAMOP-4.2.2/B-10/1-2010-0020, Support of the scientific training, workshops, and establish talent management system at the Óbuda University.

REFERENCES