Leveraging big-data for business process analytics

Purpose

Business process improvement can drastically influence in the profit of corporations and helps them to remain viable. However, the use of traditional Business Intelligence systems are not sufficient to meet today's business needs. They normally are business domain specific and have not been sufficiently process-aware to support the needs of process improvement type activities, especially on large and complex supply chains, where it entails integrating, monitoring and analysing a vast amount of dispersed event logs, with no structure, and produced on a variety of heterogeneous environments. The aim of this paper is to present a solution to this variability by means of Big Data technology.

Design/methodology/approach

Authors present a cloud-based solution that leverages big data technology to provide essential insight into business process improvement. The proposed solution is aimed at measuring and improving overall business performance, especially in very large and complex cross-organizational business processes, where this type of visibility is hard to achieve across heterogeneous systems.

Findings

Three different Big Data approaches have been undertaken based on Hadoop and HBase. We introduced first, a map-reduce approach that it is suitable for batch processing and presents a very high scalability. Secondly, we have described an alternative solution by integrating the proposed system with Impala. This approach has significant improvements in respect with map reduce as it is focused on performing real-time queries over HBase. Finally, the use of secondary indexes has been also proposed with the aim of enabling immediate access to event instances for correlation in detriment of high duplication storage and synchronization issues. This approach has produced remarkable results in two real functional environments presented in the paper.

Originality/value

The value of the contribution relies on the comparison and integration of software packages towards an integrated solution that is aimed to be adopted by industry. Apart from that, in this paper authors illustrate the deployment of the architecture in two different settings.

Keywords

Big data, Business process analytics, Business intelligence, Business performance management, Event modelling, Process mining, Cloud computing.

1 Introduction

Big Data (BD) is an emerging phenomenon in IT. There are many definitions on the term, however, the predominate one seems to be that BD comprises datasets that have become too large to handle with the traditional or given computing environment (Costello & & Prohaska, 2013). Within the phenomenon of dramatic increase in information, Kraska (2013) points out that BD can be seen as two different issues: big throughput and big analytics; the former includes the problems associated with storing and manipulating large amounts of data and the latter those concerned with transforming this data into knowledge. Focusing on the analytics, BD analytics is a workflow that distils Terabytes of low-value data down to, in some cases, a single bit of high-value data with the goal is to see the big picture from the minutia (Fisher et al., 2012). This new discipline requires new approaches to obtain insights from highly detailed, contextualized, and rich contents that may require complex math operations, such as machine learning or clustering (Chiang, Goes, & Stohr, 2012). This diversity of tools and techniques for BD analytics–driven systems, makes the process not trivial requiring specific research on the topic.

BD analytics has become increasingly important also in the business community given its various fields of application. One of these fields is business process (BP) analysis. BD provides new prospects for BP management research given that organizations are recording large amounts of event data and this is great opportunity to promote "evidence-based BPM" (Van der Aalst, 2012). However, although large organizations appreciate the value of BD, they rarely connect event data to process models and process mining represents the missing link between analysis of BD and BP management (Van der Aalst, 2012). Taking this into account, this paper presents an architecture that aims to provide BD analytics for BP events in a distributed environment in order to provide organizations with a solution to analyse process events. Furthermore, this solution is deployed in two different settings.

2 Background

Traditional enterprise information systems were not process-aware 10 years ago, and corporations have had to evolve towards a process-oriented (PO) model due to a growing demand for an effective management of BPs, where these have to be fast, flexible, low-cost and deliver high quality consistently. As a result, a process-centric approach has been adopted by enterprises in recent years, and their organizational structures have been redesigned around a PO model in order to succeed in a global economy (Hammer & Champy, 2003).

A process orientation offers advantages to organizations. It leads companies to achieve an improved transparency and understanding of their BPs. This provides a better identification of the organizational problems and their root causes in order to respond to non-compliant situations very quickly. Moreover, PO systems provide a clear distinction of responsibilities, an increase in efficiency and productivity rates, and a significant improvement of product quality (Kohlbacher, 2009).

PO is gaining importance since there exists an imperious need for an effectively management of BPs in order to enable a proper decision support. The collection and reconciliation of operational data related to BPs, enables the measurement of process throughput and helps to improve business performance, aimed at identifying and discovering business opportunities (Seufert & Schiefer, 2005).

Business intelligence (BI) has become a key tool for business users in decision making. The integration of BI into business performance management (BPM), within a process improvement context, can be a powerful asset to business users in order to gain an insight into BP performance. However they are complex, expensive, require considerable resources and time to implement (Molloy & Sheridan, 2010), and most of them are business domain specific. Furthermore, they are not able to provide a mature data mining background (Van der Aalst, 2011). Currently, there is a noticeable disconnection between BPs and their actual event-data as they are focused on local decision making rather than end-to-end processes. Furthermore, their outputs tend to be unreliable since they are based on idealized models of

reality rather than on observed facts (Van der Aalst, 2012). According to Van der Aalst (2011), process mining enables event-based process analysis, where research outcomes must be fact-based, and therefore empirically evaluated with real data which lead to trustworthy analysis results.

Process mining can fill the gap between BI and PO systems by combining event-data and process models (Van der Aalst, 2012), and this can be leveraged by next generation of BI systems for providing insight into BP throughput, key intelligence in initiatives aimed at measuring and improving overall business performance.

Both event-data and process models together are essential to infer knowledge about process improvement. Process models by themselves provide a concrete understanding and representation of what needs to be monitored, measured and analysed, but a purely structural representation of process instances is not enough to construct a solid understanding of what needs to be improved. It is also required to capture and represent the behavioural state of these processes in order to be able to identify bottlenecks, detect non-compliant situations or discovering new business opportunities. Whereby, the measurement of process performance is key in BP improvement initiatives, and it is equally a critical factor.

The latest advances in technology make possible to organizations cooperate, whereby the integration of widespread business information systems is commonly present in large and complex supply chain scenarios. This leads to the management of non-trivial operational processes, where service technology and cloud computing (CC) have become more widespread while tending to produce cross-functional event logs that are beyond the company (and increasingly software) boundaries. This has experienced an incredible growth of event data in corporations that need to be merged for analysis (Van der Aalst, 2011).

According to Van der Aalst (2011), the isolated analysis within one of these organizations is insufficient to analyse and improve end-to-end processes, and thus "events need to be correlated across organizational boundaries", and this is a very challenging task. Very often, enterprises' business data are handled by heterogeneous systems which run on different technological platforms, and even use incompatible standards. Those systems are usually critical for corporations' efficiency which frequently refuses to replace them or redesign them. Furthermore, the continuous execution of distributed BP produces a vast amount of event data that cannot be efficiently managed by the use of traditional systems which are not adequate to manage event data of the order of hundreds of millions of linked records (Vera-Baquero, Colomo-Palacios, & Molloy, 2013). Thus, research contributions must be addressed to provide a fully distributed solution that leverages BD technology to support timely BP analytics on very complex and highly distributed supply chains.

3 Technological Solution

The solution presented herein is based on previous efforts (Vera-Baquero, Colomo-Palacios, & Molloy, 2013; Vera-Baquero & Molloy, 2013; Vera-Baquero, Colomo-Palacios, & Molloy, 2014; Vera-Baquero et al., 2015). In this paper, authors propose an extension to the aforementioned framework by using a cloud-based infrastructure complemented with a federative approach in terms of data warehousing and distributed query processing. The proposed architecture provides capabilities for enabling analysts to measure the performance of cross-functional BPs that are extended beyond the boundaries of organizations. One of the main challenges of this approach relies on the integration of event data from operational systems whose BPs flow through a diverse of heterogeneous systems such as BP execution language (BPEL) engines, ERP systems, etc. as well as storing very large volumes of data in a global and distributed BP execution repository through the use of BD technology. The monitoring of cross-organizational BPs is achieved by listening (in or near real-time) state changes and business actions from operational systems. This is achieved by collecting, unifying and storing the execution data outcomes across a collaborative network, where each node represents a participant organization within the global distributed BP. This will drive the analysis of these event logs to provide business users with

a powerful understanding of what happened in the past, evaluate what happens at present and predict the behaviour of process instances in the future (Zur Muehlen & Shapiro, 2010). Additionally, the structured data may serve as an input to simulation engines that will enable business users to anticipate actions by reproducing what-if scenarios, as well as performing predictive analysis or applying pattern behaviour recognition.

Finally, a CC service façade will be the core point for providing analytical services to third-party applications. This will empower next generation of BI systems to support advanced business performance analytics on any business domain and also, extending "the decision-making process beyond the company boundaries thanks to cooperation and data sharing with other companies and organizations" (Rizzi, 2012).

3.1 Event-based model

An event model is essential to provide the framework of a concrete understanding and representation of what needs to be monitored, measured and analysed. In the context of this work, it represents a relevant action occurred during the execution of a determined BP. Therefore, our model provides the information required to enable the system to perform analytical processes over them, as well as representing any measurable action performed during the execution of a whatever BP flow. Likewise, the model has been designed as generic enough as to accommodate whatever event data is produced on heterogeneous environments independently from the source system that originated the event.

The model is based on the BPAF standard (WfMC, 2012) published by the Workflow Management Coalition, combined with some important features of the iWISE model (Molloy & Sheridan, 2010) to support interrelated process correlation. Likewise, the source format has been modified slightly to support distributed storage by deriving into a model that represents operational business data supplied from heterogeneous environments where processes cross both organizational and software boundaries while maintaining fully agnostic to any specific business domain.

The Extended BPAF model

BPAF is a flexible XML-based data format for interchange audit data that flows across heterogeneous business process management systems, however the proposed solution pursues to record and analyse event information which are originated not only from business process management systems, but also from others whose operations are part of an upper global business process within a supply chain. Besides, the event store is aimed to be distributed across different organizations along the collaborative network. Hence, some extra information in the original BPAF structure is required in order to meet the goals of the proposed solution.

The proposed event model extends the BPAF standard and combines iWise features for enable heterogeneous systems to produce interchangeable event information regardless of the underlying concerns of the source systems. Likewise, the audit data can provide the behavioural information about business process executions, and thus it enables the measurement of processes performance. Finally, based on the event timing information and the BPAF state model described in (Zur Muehlen & Shapiro, 2010) and (Zur Muehlen & Swenson, 2011), it is possible to analyse the behaviour of completed processes or sub-activities at any nested level. Zur Muehlen & Shapiro (2010) propose to leverage the state change records in the life cycle of business process to determine metrics such as the turnaround, wait time, processing time and suspending time. This information will be key for the generation of key performance indicators and metrics which will be accessible through a specific-purpose SQL-like language to query business process performance data. This language will be introduced in further sections.

3.2 Event correlation

Event correlation refers to the determination of the sequence of events produced by the execution of interrelated and consecutive BPs or activities, and this is essential to generate metrics per process

instance or activity (Costello & Molloy, 2009), and consequently it is a key-driven at identifying exceptional situations and potential improvement opportunities.

The proposed solution uses an event correlation mechanism based on the shared data between business processes during their execution. This information makes reference to the payload of event messages, and it can be used to identify the start and end event data for a particular process instance or activity.

Figure 1 illustrates the event-capturing side of the proposed IT solution. It is a CC event-driven platform that relies on BD technology to handle very large volumes of event data with the aim of supporting timely business performance analysis. The continuous execution of distributed processes during the business lifetime produces a vast amount of unstructured event data that may occur in a variety of source systems which are run on different heterogeneous platforms. Based on the event model described herein, the solution has the capabilities to timely collect the enterprise events, correlate the data along their inter-related processes, and measure their throughput.

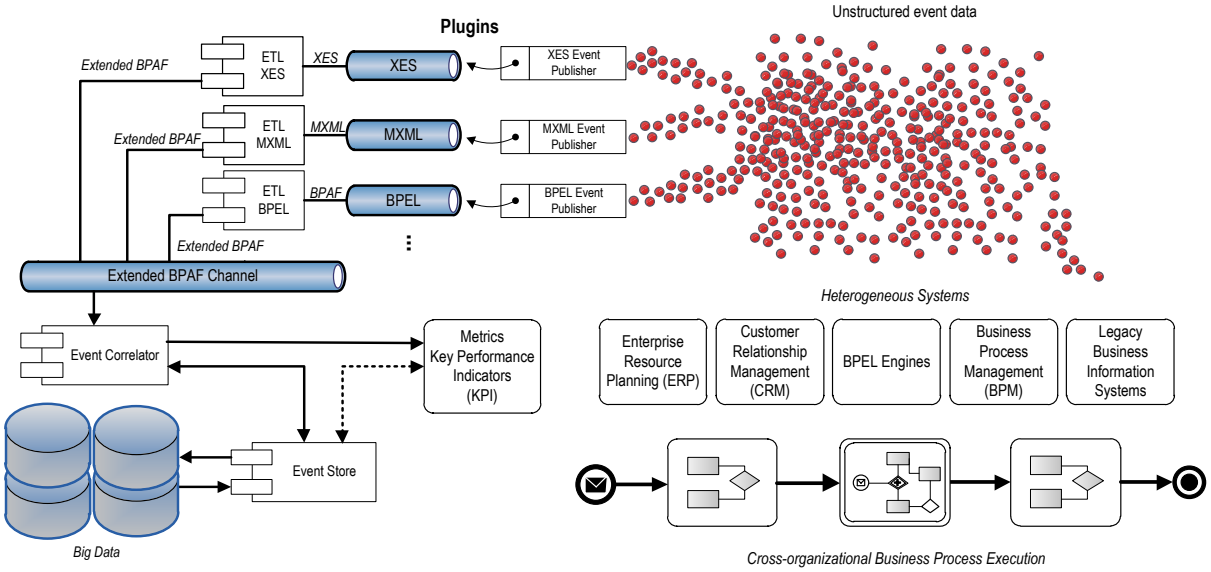

Fig. 1 – Contextual diagram of business event correlation

In this context, the listening software is responsible for specifying which part of the message payload will be used to correlate the events of instances associated to a specific model. This module is responsible for capturing the events from legacy business information systems and publishing them to the network throughout the ActiveMQ message broker. The legacy listener software emits the event information to different endpoints depending on the message format provided. Currently, the platform supports a variety of widely adopted formats for representing event logs such as XES, MXML or even raw BPAF. Consequently, a different set of plugins are available per supported event format, and in turn, each plugin incorporates specific ETL (Extract, Transform and Load) functions to convert source event streams into the proposed extended BPAF, which will enable the correlation of instances. Furthermore, the proposed platform also provides a plugin that directly mines BPEL databases for a specific vendor. This plugin has been specifically designed for Apache ODE engine, and has the capabilities to transform event logs into XML messages already structured in BPAF.

We leverage existing works around audit interchangeable event formats discussed in (Zur Muehlen & Swenson, 2011) and (Becker, Matzner, Müller, & Walter, 2012) to enable the transformation of multiple enterprise event formats into extended BPAF. Once the events are transformed, then they are forwarded to a specific channel for processing. The event correlation module is subscribed to this channel listening

continuously for new incoming events. Thereby, the enterprise events are correlated as they arrive by querying the event repository for previous instances. This is achieved by fetching the existence of a process instance associated with the correlation data provided. If no data is returned, it means that a new process has been created at the source system; thereby a new process instance is generated at destination. In such a case, a new identifier to the process instance is assigned that will later be used to correlate the subsequent events as they arrive.

In source systems that have the ability to generate and manage identifiers on their source instances, such as BPEL engines, it is not necessary to provide any correlation information on the event message. In such cases, the instance identifier is provided instead, and in turn, this is used to correlate the subsequent events.

As already stated, the retrieval of previously stored information is needed for the correlation mechanism. Hence, the timely access to this information at this stage is critical to provide timely business performance information, and thus the latency for querying BD tables must be minimal (Vera-Baquero, Colomo-Palacios, & Molloy, 2013). For this purpose, a BD approach has been used to overcome this significant pitfall. The underlying technology of the event store module is based on Apache HBase, whereby the event repository is implemented as BD tables.

3.3 BD event data store (HBase)

The event-based model introduced previously permits the representation of the structural and behavioural information of highly distributed BPs, but this is not enough to provide analysts with timely process performance information. The correct sequence of instances must be identified before any measurement can be applied, and thus, this action must be performed in a very-low latency (Vera-Baquero, Colomo-Palacios & Molloy, 2014).

As previously discussed, the event correlation relies on finding the associated process instance or activity in the event repository. This can be achieved by using the triplet - source, model and event correlation, where the event correlation can either be a set of key-value pairs, or a process or activity instance identifier. According to the data model described below, the join between event and event-correlation tables is needed for correlating processes that run on non-BPEL systems. However, on BPEL-like source platforms, it is only necessary to perform a simple scan on the process-instance or activity-instance tables over the source instance identifier attribute. Consequently, the measurement of process executions that are run on BPEL systems is much faster than those ones that are executed on legacy systems.

It worth to mention that the data model below has been denormalized due to performance reasons. Since join operations entail a very high performance cost on very large tables, the model attribute has been duplicated at the event level, and thus reducing the joins to just two tables. Furthermore, the event-correlation table cannot be denormalized as the correlation information contained in the message payload may have multiple items, so the relationship must be one-to-many in either case, otherwise the systems would be restricted to some specific BPs. Likewise, the correlation data can neither be implemented as a map attribute within the event table because the key-value pairs must be scanned and filtered. Therefore, the join operation is hard to be removed from the event-based model presented here.

4 Business Performance Analysis Scenarios

In order to provide timely business performance analysis we leverage current BD technology through the use of three different emerging approaches. In this paper we propose to write map-reduce jobs, use Cloudera Impala product or implement secondary indexes over HBase.

4.1 Map-reduce

Map-reduce is a distributed data-intensive processing framework introduced by Google that exploits cluster capabilities to support large distributed jobs in a cloud infrastructure. Due to their excellent fault

tolerance and scalability fatures, Map-reduce has now become the choice of data-intensive analysis in the cloud (Nurain, Sarwar, Sajjad, & Mostakim, 2012). We designed a map-reduce job for implementing the event correlation mechanism described above. Figure 2 depicts how the correlation process work.

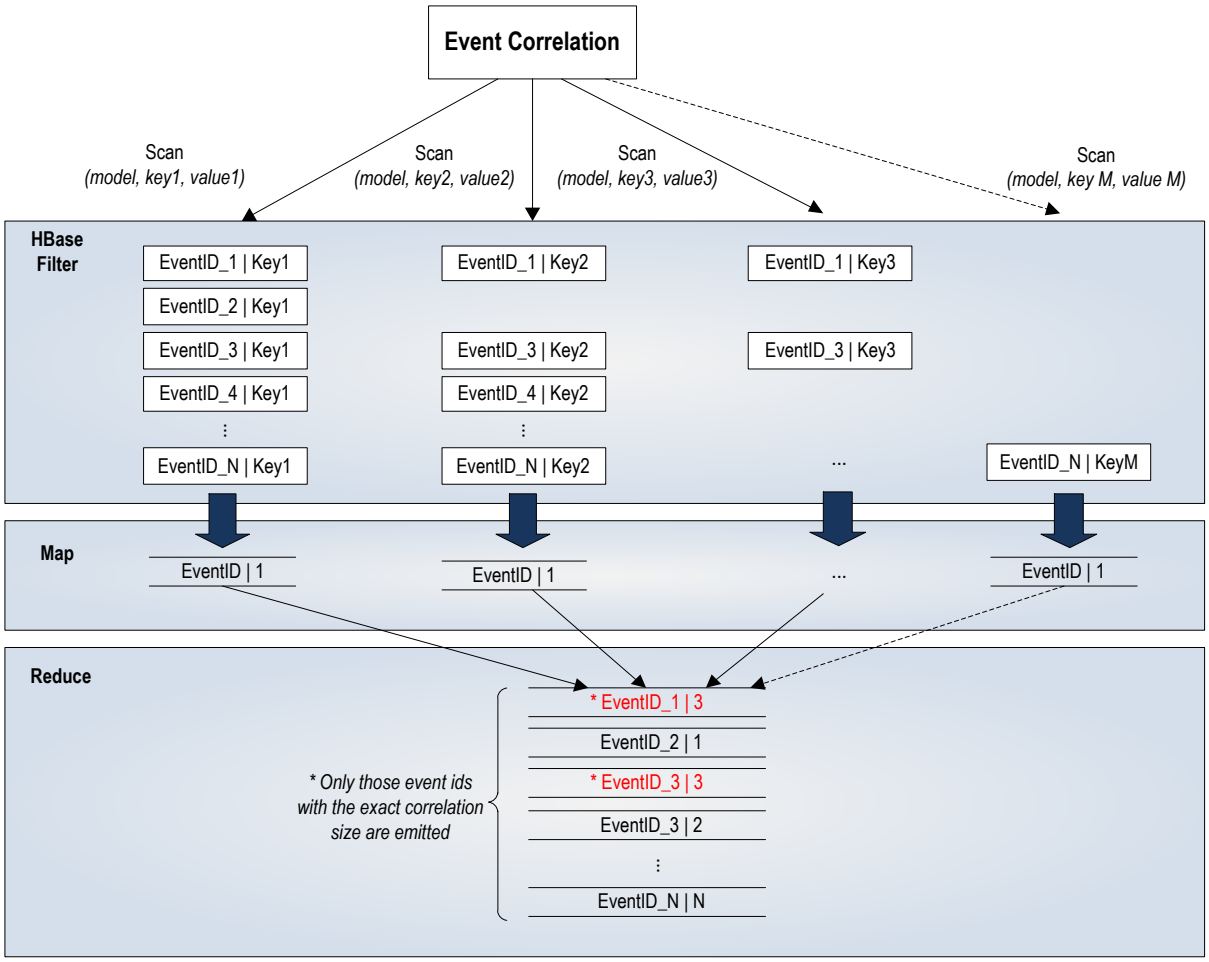

Fig. 2 – Event correlation based on map-reduce

During the map stage, the events are filtered by model and key-value in isolation per correlation set. Thereafter, every event identifier is emitted as key with value "1" representing the number of occurrences of the event within the correlation set. At the reducer stage, only those event identifiers that meet the exact correlation size are emitted. Thereby, we identify those events with the specific key-values combination thereof. The process or activity instance used to correlate subsequent events are retrieved by querying directly the event table by rowkey (event identifier). The performance of this approach will highly depend on the nature of business model and the clustering capabilities of the IT infrastructure.

4.2 Cloudera Impala

Cloudera Impala is an open source initiative from Cloudera Inc. inspired in the Google Dremel work (Melnik, et al., 2011) aiming at supporting real-time query capabilities on Apache Hadoop and HBase and complementing conventional map-reduce batch processing. It provides fast, interactive SQL queries directly on HDFS or HBase and uses the same metadata and SQL syntax (Hive SQL) as Apache Hive. This enables a unified platform for providing real-time or batch-oriented queries (Cloudera, 2013).

Furthermore, Impala is a Massively Parallel Processing (MPP) query engine that runs natively on Hadoop featuring scalable and parallel data-intensive technology for querying BD. This tool enables

end-users, or third-party systems, to issue low-latency SQL queries to data stored in HDFS and Apache HBase (Cloudera, 2013). And thus, it is a powerful technology for providing event-driven business performance analysis in real-time.

We have leveraged Impala capabilities to speed up the correlation of event data, and thus the availability of business performance measures to end-users. In our tests we have experienced a correlation performance improvement by a factor of 5 in respect with the map-reduce approach for a specific environment and with a fixed volume of data – around 5 million of linked event records. In the context of this work, a more considered study on determining the grade of improvement of this approach in comparison to map-reduce is currently on-going. Further research work is aimed at determining the relationship between the grade of improvement and the factors that directly affect to the throughput of these both approaches, namely the volume of data and clustering capabilities. This will give us a better insight on determining the trade-off between hardware investments and acceptable response time, as this strongly varies different among business cases with very specific demands.

Once the results are found, the set of event identifiers fetched should be very small and manageable, and thus the model can be easily filtered afterwards without causing any performance issue. Likewise, it is also possible to slightly modify the above statement to include a model filter by making distinctions per processes or activity cases.

4.3 Secondary Indexes

This last approach consists in using secondary indexes in HBase over the event correlation table in order to achieve immediate access to event identifiers that meet the set of key-value pair condition. According to (Vera-Baquero, Colomo-Palacios, & Molloy, 2013), the read operations over the rowkey in the event table are performed in the order of milliseconds; whereas these operations present a very-low latency, we can leverage this feature to enable real-time analysis. The idea behind this approach is to create an alternate HBase table that will be used as a secondary index for the event correlation table. The rowkey is established as a byte stream sequence of strictly ordered key-value pairs for every event as following:

RowKey: Key1Value1Key2Value2KeyN…ValueNSourceModel cf: “event_correlation” {eventIds}

Similarly as previous approaches, the correct process or activity instance can be identifying seamlessly by querying the event table per event identifier.

The downside of this approach is that it requires additional cluster space and processing, similarly as happens in RDBMS where alternate indexes require extra space and cost processing cycles to update. Furthermore, reliable synchronization measures must be implemented as the indexes could potentially be out of sync with the main data table. Even though RDBMS products have more advanced capabilities to handle alternate index management, HBase scales better at larger data volumes (Apache Software Foundation, 2013).

5 Big Data in motion: concerns for organizations

BD is more than just another buzzword that, like some others, it appears in the IT scenario. Indeed, BD is a powerful technology that is emerging at nowadays for providing data-intensive processing on large scale datasets. Thus, it is becoming a key enabler to organizations at meeting their business goals. Corporations are undergoing a growing demand for BD-based business analytics tool, and many of the top IT vendors have adopted BD as a cornerstone for their development: Fujitsu presents BD as the centre of its “Human-Centric Intelligent Society”, Oracle underlines that “Big data is the electricity of the 21st century—a new kind of power that transforms everything it touches in business, government, and private life”, Google launched in March 2014 BigQuery a service that lets businesses analyse data sets in the cloud, Microsoft recently launched Windows Azure HDInsight (built on Apache Hadoop) to integrate their traditional solutions and user familiar tools in a user-friendly solution… and this is just to name some of the most relevant IT vendors that presented their solutions in recent days. The amount

of solutions built on Hadoop, one of the the-facto standards in the scenario, is that important that Forrester issued in February 2014 “The Forrester Wave™: Big Data Hadoop Solutions” analysing nine solutions, namely: Amazon Web Services, Cloudera, Hortonworks, IBM, Intel, MapR Technologies, Microsoft, Pivotal Software, and Teradata (Gualtieri & Yuhanna, 2014).

However and apart from the commercial side, BD wave reached organizations too. It is true that a search in any of the scientific databases available retrieves normally more papers on the expectations, opportunities, future uses or strategies than on real industry projects, but it is also a true that there a lot of basic and applied research published in scientific journals on the topic in the last years. And this architecture is one of the examples of real experiments with BD. Thus Vera-Baquero, Colomo-Palacios, and Molloy (2014) present a case study on the architecture consisting in the analysis and improvement of the efficiency and security of the roads network in England. Execution results, measures and KPIs did not present any statistical significance in respect with the official values publicly available. On the other hand, Vera-Baquero et al. (2015) present another case study is focused on the improvement of the service delivery process for call centres to enhance productivity while maintaining effective customer relationships. One more time, results show a remarkable performance and exceptional cases such as bottlenecks, overloads and failure rates (abandons) were properly identified and detected by the system.

6 Summary and Outlook

Authors have presented a cloud-based platform that leverages BD technology to provide business analysts with visibility on process and business performance in a timely manner. The approach discussed is focused on providing timely correlation of event instances that are produced on a variety of heterogeneous PO systems. This is essential to make metrics and key performance indicators available to business users in an acceptable response time basis.

The integration and correlation of cross-organizational BP instances is a very challenging task, especially when they are run across distributed heterogeneous business information systems. To overcome this shortcoming, an event-based model is constructed with the aimed of interpret and process event streams that are part of a cross-functional BP. This model has the ability to represent both structural and behavioural aspects of BPs, thereby featuring a complete view for inferring knowledge from the collected information.

In order to provide the system with the capabilities to deal with very large amount of datasets, we have leveraged current BD technology aimed to generate timely business performance information. In this regard, three different approaches have been undertaken based on Hadoop and HBase. We introduced first, a map-reduce approach that it is suitable for batch processing and presents a very high scalability. Secondly, we have described an alternative solution by integrating the proposed system with Impala. This approach has significant improvements in respect with map reduce as it is focused on performing real-time queries over HBase. Finally, the use of secondary indexes has been also proposed with the aimed of enabling immediate access to event instances for correlation in detriment of high duplication storage and synchronization issues. A more considered analysis on these three approaches is ongoing and is part of future research work, as the choice will depend on specific business demands, and finding the balance between powerful hardware investments and the outcomes latency on business performance analysis.

Future work in the organizational side includes the deployment in more environments and the measurement of potential and actual consequences on the use of BD environments in capturing business value. Thus, it is aimed to conduct studies in functional areas inside organizations like Customer Relationship Management, Supply Chain Management and Financial Management in order to study differences in behaviour and performance taking into account the different qualities and amount of inputs these areas present. Finally, it is also aimed to conduct comparative studies on the deployment of the solution into different sectors like banking, education, automotive or tourism with regards to the deployment of the approaches presented in this paper.

REFERENCES

Apache Software Foundation. (2013). Apache HBase. Retrieved April 5, 2014, from: http://hbase.apache.org

Becker, J., Matzner, M., Müller, O., & Walter, M. (2012). A Review of Event Formats as Enablers of Event-Driven BPM. In F. Daniel, K. Barkaoui, & S. Dustdar (Eds.), Business Process Management Workshops (Vol. 99, pp. 433-445). Springer Verlag.

Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Discipline. ACM Transactions on Management Information Systems, 3(3), 12:1–12:13.

Cloudera. (2013). Cloudera Impala. Retrieved February 5, 2015, from http://www.cloudera.com

Costello, C., & Molloy, O. (2009). A Process Model to Support Automated Measurement and Detection of out-of-bounds events in a Hospital Laboratory Process. Journal of Theoretical and Applied Electronic Commerce Research, 4(2), 31-54.

Costello, T., & & Prohaska, B. (2013). Trends and Strategies. IT Professional, 15(1), 64–64.

Fisher, D., DeLine, R., Czerwinski, M., & Drucker, S. (2012). Interactions with big data analytics. interactions, 19(3), 50-59.

Gualtieri, M., & Yuhanna, N. (2014). The Forrester Wave™: Big Data Hadoop Solutions, Q1 2014. Available at http://www.forrester.com/The+Forrester+Wave+Big+Data+Hadoop+Solutions+Q1+2014/fulltext/-/E-RES112461

Hammer, M., & Champy, J. (2003). Reengineering the corporation - A manifesto for business revolution. HarperCollins.

Kohlbacher, M. (2009). The Perceived Effects of Business Process Management. Science and Technology for Humanity (TIC-STH). IEEE Toronto International Conference, (pp. 399-402). Toronto.

Kraska, T. (2013). Finding the Needle in the Big Data Systems Haystack. IEEE Internet Computing, 17(1), 84–86.

Melnik, S., Gubarev, A., Long, J., Romer, G., Shivakumar, S., Tolton, M., & Vassilakis, T. (2011). Dremel: Interactive Analysis of Web-Scale Datasets. Communications of the ACM, 54(6), 114-123.

Molloy, O., & Sheridan, C. (2010). A Framework for the use of Business Activity Monitoring in Process Improvement. In E. Alkhalifa (Ed.), E-Strategies for Resource Management Systems: Planning and Implementation. IGI Global.

Nurain, N., Sarwar, H., Sajjad, M., & Mostakim, M. (2012). An In-depth Study of Map Reduce in Cloud Environment. Advanced Computer Science Applications and Technologies (ACSAT), (pp. 263 - 268).

Rizzi, S. (2012). Collaborative Business Intelligence. In M. Afaure, & E. Zimanyi (Ed.), First European Summer School (eBISS 2011) (pp. 186-205). Paris: Springer.

Seufert, A., & Schiefer, J. (2005). Enhanced Business Intelligence - Supporting Business Processes with Real-Time Business Analytics. Database and Expert Systems Applications, 2005. Proc. Sixteenth International Workshop, 919-925.

Van der Aalst, W. M. (2011). Process Mining: Making Knowledge Discovery Process Centric. SIGKDD Explorations, 13, 45-49.

Van der Aalst, W. M. (2012). A Decade of Business Process Management Conferences: Personal Reflections on a Developing Discipline. In A. Barros, A. Gal, & E. Kindler (Eds.), Business Process Management (pp. 1-16). Springer Berlin Heidelberg.

Van der Aalst, W. M. (2012). Process mining. Communications of the ACM, 55(8), 76–83.

Vera-Baquero, A., & Molloy, O. (2013). A Framework to Support Business Process Analytics. Proc. of the International Conference on Knowledge Management and Information Sharing, (pp. 321-332). Barcelona.

Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2013). Business process analytics using a big data approach. IEEE IT Professional, 15(6), 29-35.

Vera-Baquero, A., Colomo-Palacios, R., & Molloy, O. (2014). Towards a process to guide Big Data based Decision Support Systems for Business Processes. In Proceedings of CENTERIS 2014 - Conference on ENTERprise Information Systems, Procedia Technology, 16, pp. 11-21.

Vera-Baquero, A., Colomo-Palacios, R., Molloy, O. & Elbattah, M. (2015). Business process improvement by means of Big Data based Decision Support Systems: a case study on Call Centers. International Journal of Information Systems and Project Management, 3 (1), 1-22.

WfMC. (2012, February). Retrieved from Workflow Management Coalition - Business Process Analytics Format Specification: http://www.wfmc.org/Download-document/Business-Process-Analytics-Format-R1.html

Zur Muehlen, M., & Shapiro, R. (2010). Business Process Analytics. In J. vom Brocke, & M. Rosemann (Eds.), Handbook on Business Process Management 2 (pp. 137-157). Springer Berlin Heidelberg.

Zur Muehlen, M., & Swenson, K. D. (2011). BPAF: A Standard for the Interchange of Process Analytics Data. In M. zur Muehlen, & J. Su (Eds.), Business Process Management Workshops (pp. 170-181). Springer Berlin Heidelberg.