REMS.PA: A Complex Framework for Supporting OLAP-based Big Data Analytics over Data-intensive Business Processes

Alfredo Cuzzocrea[1], Salvatore Cavalieri[2], Orazio Tomarchio[2], Giuseppe Di Modica[2], Concetta Cantone[3] and Angela Di Bilio[3]

[1]*DIA Dept., University of Trieste, Trieste, Italy*

[2]*DIEEI Dept., University of Catania, Catania, Italy*

[3]*Xenia Software Solutions, Catania, Italy*

Keywords: Big Data Analytics, Data-intensive Business Processes, OLAP-based Big Data Analytics and Complex Architectures and Systems.

Abstract: In this paper, we provide architecture and functionalities of REMS.PA, a complex framework for supporting OLAP-based big data analytics over data-intensive business processes, with particular regards to business processes of the Public Administration. The framework has been designed and developed in the context of a real-life project. In addition to the anatomy of the framework, we describe some case studies that contribute to highlight the benefits coming from our proposed framework.

1 INTRODUCTION

Nowadays, a relevant interest is growing around the issue of extracting useful knowledge from *big data* (e.g., (Cuzzocrea *et al.*, 2013a; Cuzzocrea *et al.*, 2013b; Li *et al.*, 2015; Zikopoulos and Eaton, 2011; Manyika *et al.*, 2011)), in order to derive *actionable knowledge* to use for decision making purposes. This trend has originated a rich research vein (e.g., (Barnaghi *et al.*, 2013; Depeige and Doyencourt, 2015; Illa and Padhi, 2018; Raghupathi *et al.*, 2018; Yan *et al.*, 2018; Oneto *et al.*, 2018)), which still attracts a great deal of attention from the research community.

The problem of supporting *big data analytics* (e.g., (Cuzzocrea *et al.*, 2011; Cuzzocrea, 2013; Cuzzocrea and Song, 2014; Russom, 2011; Raghupathi and Raghupathi, 2014; Hui *et al.*, 2018)) over so-called *data-intensive business processes* (e.g., (Ahmad *et al.*, 2017; Shchapov *et al.*, 2017; Grabis and Kampars, 2018; Sakr *et al.*, 2018;)) plays a relevant role. This because, on one hand, business processes still keep the most of the data, information and knowledge of very-large enterprises and organizations, and, on the other hand, perfectly marry with the emerging characteristics of big data.

Indeed, the first important issue to face-off is represented by the explicit *graph-like nature* of business processes, which casts for advanced solutions that share poor insights with traditional approaches developed in the context of business intelligence.

Another important aspect is represented by *performance issues*. In fact, managing graph-like data introduces two essential requirements: (*i*) efficient in-memory representation; (*ii*) powerful access and management primitives. Both these requirements must be considered when dealing with data-intensive business processes.

An important solution for supporting big data analytics concerns with applying fortunate *multidimensional metaphors and abstractions*, mainly falling in the well-known OLAP context, thus originating an evolving trend that can be safely recognized within the term "*OLAP-based big data analytics*" (e.g., (Cuzzocrea, 2017; Cuzzocrea *et al.*, 2016; Schuetz *et al.*, 2017)).

Inspired by this research context, in this paper we focus the attention on the problem of supporting OLAP-based big data analytics over data-intensive business processes, and we describe a real-life framework developed in the context of a real-life project, called REMS.PA, which has produced the corresponding framework, mainly designed on top of open-source technologies, and that, particularly, focuses on business processes of the Public administration.

The remaining part of this paper is organized as

Cuzzocrea, A., Cavalieri, S., Tomarchio, O., Di Modica, G., Cantone, C. and Di Bilio, A.
REMS.PA: A Complex Framework for Supporting OLAP-based Big Data Analytics over Data-intensive Business Processes.
DOI: 10.5220/0007737002230230
In *Proceedings of the 21st International Conference on Enterprise Information Systems (ICEIS 2019)*, pages 223-230
ISBN: 978-989-758-372-8

follows. In Section 2, we report on main research issues of supporting OLAP-based big data analytics over data-intensive business processes. In Section 3, we describe the proposed framework. In Section 4, we provide several case studies that show how the framework works in practice. Finally, in Section 5, we provide conclusions and future work for our research.

2 OLAP-BASED BIG DATA ANALYTICS OVER DATA-INTENSIVE BUSINESS PROCESSES: EMERGING RESEARCH ISSUES

OLAP-based big data analytics over data-intensive business processes opens the door to several emerging research issues, among which some noticeable ones are the following:

- computing multidimensional OLAP aggregations over data-intensive business processes;
- supporting OLAP querying, operators and operations over so-computed OLAP *cubes*;
- effective and efficient in-memory representation of business process cubes;
- supporting flexible big data prediction methodologies over so-computed OLAP cubes.

How to aggregate a collection of data-intensive business processes? This is a relevant question that has attracted the attention of several studies. Basically, classical OLAP aggregation algorithms cannot be applied as they are, but suitable adaptations must be devised. A possibility consists in considering the graph-like nature of business processes in this respect. Doing this, the *scalability* property, which is relevant for big data management and processing (e.g., (Wu *et al.*, 2018; Sun *et al.*, 2018; Yang *et al.*, 2014; Cuzzocrea *et al.*, 2013)), must be taken into account.

After computing aggregations, the support for OLAP querying, operators and operations must be ensured. Among queries, *range queries* are very significant in this context. In addition, supporting *roll-up* and *drill-down* operators is, for instance, a first-class problem in this respect. At the same, *slice* and *dice* operations are significant in order to provide a comprehensive support to ad-hoc big data analytics procedures.

Effectively and efficiently supporting in-memory representation of business process cubes conveys on several challenges to be faced-off. Indeed, so-computed OLAP cubes can achieve very large sizes when stored in suitable Cloud storage systems. Therefore, specialized approaches must be devised in order to tame such enormous sizes. Partition-based approaches seem a promise trend to this end.

Finally, another critical problem is represented by the issue of supporting flexible big data prediction methodologies over target OLAP cubes, as the final goal is that of discovering useful knowledge from data-intensive business processes (e.g., (Braun *et al.*, 2014; Wang *et al.*, 2018; Shen, 2018)). Again, multidimensional paradigms, such as *multidimensional clustering* (e.g., (Murtagh, 1985)), can be successfully applied to this end.

3 REMS.PA: AN INNOVATIVE FRAMEWORK FOR SUPPORTING OLAP-BASED BIG DATA ANALYTICS OVER DATA-INTENSIVE BUSINESS PROCESSES

The proposed framework aims at supporting OLAP-based big data analytics over data-intensive business processes. It combines two main assets: analysis and prediction of business processes, with focus on the case of business processes in the Public Administration, and intends to reach the definition of the framework for the automated management and optimization of business processes in the Public Administration. From a strictly technological point of view, the fundamental components of the framework are the following:

- tools to support multidimensional analysis of business process schemes using the OLAP paradigm;
- visual analytics tools for business processes based on multidimensional abstractions;
- tools to support the prediction of executions of business processes based on a data-driven approach.

The framework has been realized by using and integrating open-source software technologies for the support of business process management with the aim of speeding up and simplifying the management of the operational workflows of the Public Administration, via defining and building the management processes in a rigorous and reliable way, and finally monitor the real status of their execution. More generally, the proposed framework aims at optimizing and automating the management of Public Administration processes through their analysis and prediction of their executions. Business process analysis and prediction are therefore the two central themes of the proposed business process management framework. By recognizing in these two phases critical elements for the improvement of the management of Public Administration processes as well as the provision of services to the citizen, the framework aims at optimizing these critical components. Therefore, the resulting optimizations tend towards the general objective of achieving efficiency and flexibility of the Public Administration processes. To this end, the proposed framework includes two innovative components to support the analysis and prediction phases: (*i*) visual analytics on business processes, which focuses on the analysis of business processes (and their execution traces) using multidimensional abstractions for the support of OLAP analysis on business process schemes; (*ii*) execution prediction on business processes, which focuses on the prediction of business process executions, to support their optimization, through an innovative data-driven approach. In short, this approach aims to predict execution of Public Administration business processes by resorting to the analysis of the variations that business-processes' previous performances have produced on the data (focusing the attention, therefore, on the nature of the data distributions that characterize these variations). A software tool has been implemented, as to allow the Public Administration to optimize the management of internal processes, evaluate their effectiveness, and adopt the necessary corrections in order to make the service offered to the community efficient and transparent.

Indeed, the level of citizen satisfaction is a yardstick for the Public Administration with respect to public management. In this sense, the framework aims to ensure significant changes, including:

- improvement of administrative transparency (e.g., telematics desk for the citizen, and so forth);
- certainty of compliance with procedures and regulations and the traceability of activities;
- control and optimization of processes;
- reduction in the time required for administrative procedures;
- increase in "company" productivity;
- global reduction of associated costs;
- automation of the planned activities;
- accountability and monitoring of the people involved.

The innovative features introduced by the proposed framework are the following.

Feature #1 – Innovative Techniques and Tools for OLAP Analysis on Business Process Schemes: Although OLAP is a methodology applied to many data models (such as graphs, sequences, text, etc.), in literature, as well as in industry, there are no proposals that offer an "explicit" OLAP support on business processes (for example: multidimensional browsing and exploration of aggregated business process schemes, coverage of the most common OLAP operators and operations - such as roll-up, drill-down, pivoting, etc., and so forth), in spite of the embryonic tools for multidimensional analysis made available by some tools (e.g., *ProM* (van Dongen *et al.*, 2005)).

Feature #2 – Visual Analytics Tools and Techniques on BP that Exploit Multidimensional Abstractions: Even in this case, the visual analytics solution proposed by the framework directly exploit the power of multidimensional abstractions, for example thanks to multi-resolution analysis, which it is both powerful and very intuitive. It should be noted that, both in literature and in the field of industrial solutions, there are no approaches that propose this vision of visual analytics on business processes.

Feature #3 – Data-driven Process Mining: From a purely scientific and industrial point of view, the most valuable result that the framework introduces is represented by the innovative data-driven process mining methodology. This methodology is not only innovative in research (academic and industrial), but, despite its complexity, it effectively captures real-world application scenarios of business process management systems. Indeed, these systems are, in turn, characterized by a certain intrinsic complexity, hence the proposed methodology aims at using multidimensional abstractions in a very powerful and flexible manner via an innovative data-driven approach. Contrary to this, other approaches known in the state-of-the-art literature solve the difficult problem of monitoring and optimizing business processes through solution-driven approaches, which introduce little flexibility and extensibility, thus exposing important limitations for its application in real-life settings.

Summarizing, the main scientific and technical research issues addressed by the framework are the following:

- definition of methodologies, models and tools for supporting multidimensional analysis of business process schemes;
- effective and efficient representation of aggregated business process schemes in secondary storage;
- definition of paradigms for the support of OLAP functionalities and extensions on aggregated business process schemes;
- definition of methodologies, models and tools for supporting the multi-resolution OLAP analysis of business process schemes;
- optimization techniques for OLAP roll-up and drill-down operators on aggregated business pro-

cess schemes;

- definition of appropriate multidimensional metaphors for the support of visual analytics for business process using OLAP methodologies and paradigms;
- efficient and scalable solutions for the support of visual analytics for business processes;
- definition of the predictive analysis method of data-driven process mining;
- cumulative similarity techniques between discrete data distributions;
- techniques for optimizing procedures for processing and analyzing discrete distributions on big business process data.

4 REMS.PA: LOGICAL ARCHITECTURE

Figure 1 shows the logical architecture of the proposed framework for supporting OLAP-based big data analytics over data-intensive business processes.

As shown in Figure 1, the proposed framework introduces the following layers:

- *BPM Layer*: is it the layer where the input business processes are located and exploited to populate the big data layer of the framework;
- *OLAP Aggregation Layer*: it is the layer where business processes are aggregated into cubes in order to supporting OLAP-based big data analytics;
- *OLAP Analysis Layer*: it is the layer where the OLAP querying, operators and operations over business processes are implemented;
- *Application Layer*: it is the layer where the consumer applications are located, being visual analytics and prediction analytics the main functionalities supported.

4.1 The BPM Layer

The BPM Layer is the bottom layer of the platform. It collects data from the various BPM Systems, standardizes the process representation, and makes this information available at the top level.

The following components populate this layer (see Figure 2):

- *Adapters:* they are responsible for acquiring data from the various BPM Suites and standardizing data according to a common format. Specifically, the adapters have been designed to acquire both BPM schemes and process instances. The standardised format also allows for the graphical representation of BPM schemes to which the data refer. Adapters have been implemented for the following BPM Suites: *Camunda*, *Activiti* and *jBPM*.
- *Process Scheme & Instance:* this module is responsible for storing the process execution instances (adequately uniformed by the adapters) and the BPM schemes.

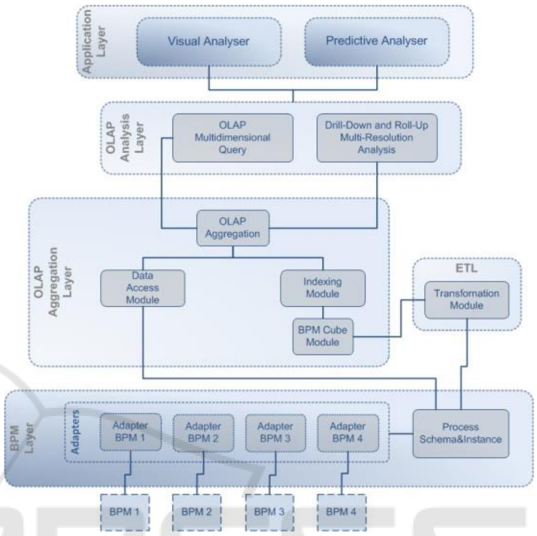

Figure 1: REMS.PA: Logical Architecture.

In the following, we provide a detailed description of the layers of the proposed architecture.

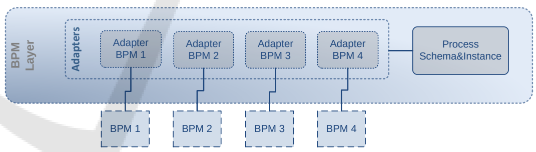

Figure 2: The BPM Layer.

4.2 The Olap Aggregation Layer

The OLAP Aggregation Layer processes the data coming from the BPM layer with the purpose of extracting useful information for the analysis. In particular, it deals with the definition of the relationships between the various process instances and the aggregation of the BPM Schemes (*BPM Scheme Aggregation*) through the definition of the BPM OLAP Cube.

As depicted in Figure 3, this layer consists of four components:

- *ETL - Extract, Transform, Load*. This module takes care of extracting, transforming and loading data into the data mart. This transformation aims to make the data homogeneous and ensure

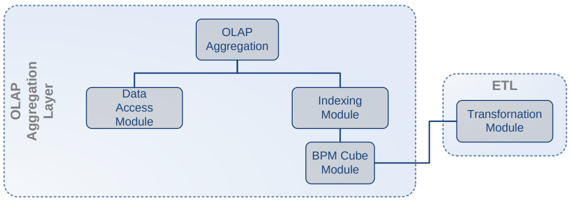

Figure 3: OLAP Aggregation Layer.

that they are congruent with the business logic of the analysis system for which it is developed. The data thus processed are then stored in the tables (load).

- *BPM Cube:* This module handles OLAP cubes (basic aggregation). OLAP cubes reside in primary memory and are defined by the administrator according to specific analytical objectives. OLAP cubes represent multidimensional partitioning and contain references to objects in secondary memory, that is, keys that uniquely identify the processes (graphs). Processes provide access to the Cubes to retrieve the key set via the indexing process, and access to the serialized objects (handled by the Process Schema & Instance Module) for each key extracted using the *Data Access Module*. For each multidimensional range, Cubes contain intervals of the OLAP aggregations on that range and collection of references in secondary memory of the objects logically contained by that range.
- *Indexing Module:* The Indexing Module is responsible for indexing the objects in the BPM Cube. This indexing system features the data space as organized into MBR (*Minimum Bounding Rectangles*), which are hierarchically organized and, when possible, overlapped.
- *Data Access Module:* The Data Access Module is responsible for accessing objects in the Process Schema & Instance Module. Because objects reside in secondary memory, the indexes are organized according to the Cluster-Index mode, which helps when building highly scalable systems.
- *OLAP Aggregation:* This module takes care of the actual aggregation of the BPM schemes through the use of specific algorithms. Two aggregation modes have been defined: a) "Inter-schema" aggregation mode, in which the data cube cells contain BPM schemes that follow the multidimensional aggregation - in other words, each cell in the data cube contains a partition of BPM schemes, and portions of duplicated BPM schemas between data cube cells are not eligible; b) "Infra-schema" aggregation mode, according to which the cells of the data cube contain portions of BPM schemes appropriately connected (based on the established semantics) - in other words, each cell of the data cube contains a BPM meta-schema built on portions of BPM schemes, and portions of duplicated BPM schemas between data cube cells are eligible.

4.3 The OLAP Analysis Layer

This OLAP Analysis layer, which is depicted in Figure 4, takes care of the actual OLAP analysis starting from the OLAP Cube and from the aggregations of the schemas generated by the underlying OLAP Aggregation layer.

Figure 4: OLAP Analysis layer.

Two macro modules populate this layer:

- *OLAP Multidimensional Query:* This module deals with the interrogation of the Cube in order to present the results of the analysis requested by the analyst. The OLAP quering process firstly executes the query on the reference Cube and then accesses, in secondary memory, the serialized objects contained in the Process Schema & Instance module. Queries can be performed in MDX, which is a query language for multidimensional databases.
- *Drill-Down and Roll-Up Multi-Resolution Analysis:* This module provides the algorithms and techniques to support OLAP (Drill-Down and Roll-Up Multi-Resolution Analysis) operations. In particular, it deals with the definition and application of advanced techniques on basic operations such as Drill-Down and Roll-Up. The Roll-Up operation is implemented by aggregating the graphs (BPM schemas and process instances). The output of this aggregation operation is a new graph. The Drill-Down is a disaggregating operation that allows data to be analyzed at a finer level of detail (e.g., the number of properties per province and the number of properties per municipality). It consists in disaggregating the graphs of a cell into smaller graphs within a set of cells. It should be noted that, unlike the roll-up operation, the transformation triggered by the drill-down operation is non-completely-reversible, in the sense that the way of disaggregating the graphs is not unambiguous.

4.4 The Application Layer

The Application Layer (see Figure 5)implements the front-end components that display the result of the data processing carried out in the OLAP Layer.

Figure 5: Application Layer.

This layer is composed of two macro-blocks:

- *Visual Analyzer:* It offers dashboards and graphical tools for the visualization of the data analysis performed at the OLAP Analysis Layer. While the classical BI tools aggregate on the data and show the result, the platform works on processes information, so it has to provide process-centric views that ease the analysts' work of optimizing the management of the PA processes. The proposed Visual Analytics solution exploits the power of multidimensional abstractions through the multi-resolution technique, which is both powerful and very intuitive.
- *Predictive Analyzer:* It is the tool that visualizes the result of the forecast analyses based on the data manipulated in the OLAP Analysis Layer.

5 REMS.PA: CASE STUDIES

A prototype of the REMS.PA framework has been implemented and tested by using example data coming from process logs of the Catania municipality.

In this Section, we present some of the main results (by means of screenshots of the platform's front end) obtained by applying the different techniques so far described: in particular we show the results of the Visual Analyzer and of the Predictive Analyzer components.

The set of analysed processes involves services related to the local taxes (on houses, properties, waste collection and so on) and to the cadastrial management (check-up, re-evaluation, etc).

In order to test the effectiveness of the Visual Analytics components, some KPI indicators of involved services have been defined. As shown in Figure 6, some preliminary steps to obtain KPI data to be visualized are performed by the platform behind the scene:

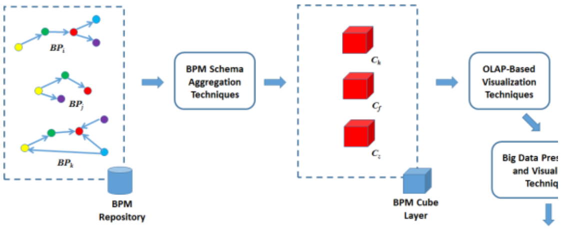

Figure 6: Preliminary Steps for Visual Analytics.

- identification of all the BPM schema involved by the municipality services under observation and collection of their logs;
- application of the process aggregation techniques in order to build the BPM cube layer;
- application of multiresolution analysis algorithms on the BPM cube layer in order to extract data to be visualized.

For space reason, we show only two of the defined KPIs, namely:

- a KPI showing the distribution of the status of citizens' administrative procedures; in Figure 7 three different statuses can be recognized: "running", "under verification", and "closed". The different colors refer to different districts of the city, while the size of the circle represents the number of procedures in that status.
- a KPI showing the average time to complete an administrative procedure. In Figure 8, the number of civil servants employed to manage the procedure and its mean execution time are showed: bars of different colors refer to different districts of the city.

The other component of the Application layer is the Predictive Analyzer. While in the Business Process Monitoring context many kind of predictions can be defined (e.g., time related prediction such as delays, deadline violation, etc, or prediction of process outcomes), in the REMS.PA project we are interested in a specific kind of path completion. More specifically, given a process scheme which includes a very large number of possible execution tracks (all potentially activable at runtime), at a certain time we want to determine which is the trace that, most likely, will be activated at the next execution. The information on which this predictive analysis is based includes all the logs related to the traces of past executions observed in a well-defined time interval. Such a kind of prediction is of paramount importance for Public Administration; such a knowledge could be used to optimize resources (people, infrastructures) required for the execution of the process. More generally, a Public Administration could implement proactive actions for a more efficient execution of that process.

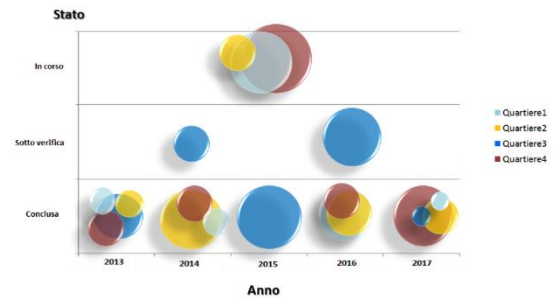

Figure 7: KPI Showing the Status of Citizens' Procedures.

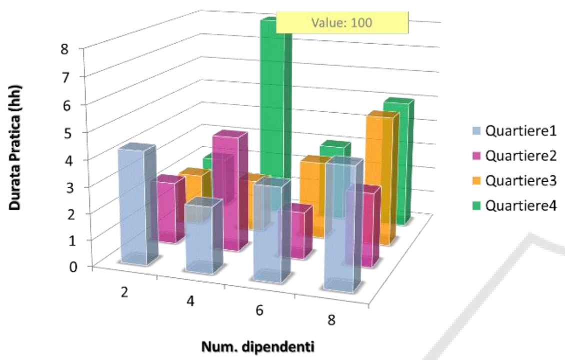

Figure 8: KPI Showing the Average Time to Complete An Administrative Procedure.

Without entering in details, the adopted approach is based on the analysis of changes in the data that the past executions of a given process originated: the basic idea is that executions of "similar" processes (by adopting a certain similarity threshold) give rise to changes in the data that follow the same distributions. This approach is suitable for "data- intensive" processes, such as those typical of public administrations.

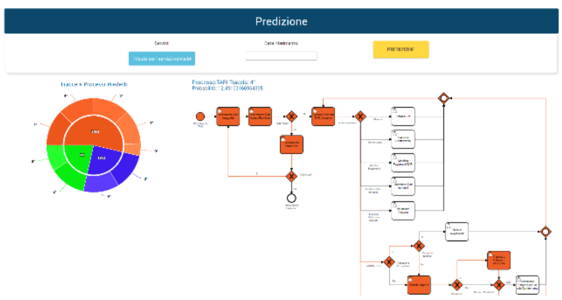

Figure 9: Predictive Analyzer Dashboard.

The results of the application of these techniques are presented in the platform on a dashboard, shown in figure 9. After a process and a reference time windows have been selected, the dashboard shows (by highlighting it in the process scheme) the trace which, most likely, will be activated at the next execution (on the left in Figure 9). On the right part of the windows, a pie chart reporting the information about the probability of execution of different traces is instead shown.

6 CONCLUSIONS AND FUTURE WORK

This paper has focused the attention on the problem of supporting big data analytics over so-called data-intensive business processes, i.e. business processes connected to big data sources. We explored issues, models and proposals in the field, and finally the architecture of a real-life framework developed in the context of a real-life project has been provided.

Future work is mainly oriented to enrich the proposed framework via innovative big data properties, such as: *privacy preservation* (e.g., (Cuzzocrea and Bertino, 2011; Cuzzocrea and Russo, 2009; Yang *et al.*, 2018)), *open big data predicates* (e.g., (Karau, 2017)), and *consistency checking* (e.g., (Tran *et al.*, 2015)).

ACKNOWLEDGEMENTS

This research has been developed in the context of the *MISE Horizon 2020 – PON 2014/2020* project: "*REMS.PA (Resource in Engineering Management for Software process automation in Public Administration)*".

REFERENCES

Saima Gulzar Ahmad, Chee Sun Liew, M. Mustafa Rafique, Ehsan Ullah Munir, "Optimization of data-intensive workflows in stream-based data processing models", *The Journal of Supercomputing 73(9)*, 2017

Payam M. Barnaghi, Amit P. Sheth, Cory A. Henson, "From Data to Actionable Knowledge: Big Data Challenges in the Web of Things", *IEEE Intelligent Systems 28(6)*, 2013

Peter Braun, Juan J. Cameron, Alfredo Cuzzocrea, Fan Jiang, Carson Kai-Sang Leung, "Effectively and Efficiently Mining Frequent Patterns from Dense Graph Streams on Disk", *Procedia Computer Science 35*, 2014

Alfredo Cuzzocrea, "Analytics over Big Data: Exploring the Convergence of Data Warehousing, OLAP and Data-Intensive Cloud Infrastructures", *IEEE COMPSAC*, 2013

Alfredo Cuzzocrea, “Scalable OLAP-Based Big Data Analytics over Cloud Infrastructures: Models, Issues, Algorithms”, *ACM ICCBDC*, 2017
Alfredo Cuzzocrea, Ladjel Bellatreche, Il-Yeol Song, “Data warehousing and OLAP over big data: current challenges and future research directions”, *ACM DOLAP*, 2013
Alfredo Cuzzocrea, Elisa Bertino, “Privacy Preserving OLAP over Distributed XML Data: A Theoretically-Sound Secure-Multiparty-Computation Approach”, *Journal of Computer and System Sciences 77(6)*, 2011
Alfredo Cuzzocrea, Carmen De Maio, Giuseppe Fenza, Vincenzo Loia, Mimmo Parente, “OLAP analysis of multidimensional tweet streams for supporting advanced analytics”, *ACM SAC*, 2016
Alfredo Cuzzocrea, Rim Moussa, Guandong Xu, “OLAP*: Effectively and Efficiently Supporting Parallel OLAP over Big Data”, *MEDI*, 2013
Alfredo Cuzzocrea, Vincenzo Russo, “Privacy Preserving OLAP and OLAP Security”, *Encyclopedia of Data Warehousing and Mining*, 2009
Alfredo Cuzzocrea, Domenico Saccà, Jeffrey D. Ullman, “Big data: a research agenda”, *ACM IDEAS*, 2013
Alfredo Cuzzocrea, Il-Yeol Song, “Big Graph Analytics: The State of the Art and Future Research Agenda”, *ACM DOLAP*, 2014
Alfredo Cuzzocrea, Il-Yeol Song, Karen C. Davis, “Analytics over large-scale multidimensional data: the big data revolution!”, *ACM DOLAP*, 2011
Audrey Depeige, Dimitri Doyencourt, “Actionable Knowledge As A Service (AKAAS): Leveraging big data analytics in cloud computing environments”, *Journal of Big Data 2(12)*, 2015
Boudewijn F. van Dongen, Ana Karla A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, Wil M. P. van der Aalst, “The ProM Framework: A New Era in Process Mining Tool Support”, *ICATPN*, 2005
Janis Grabis, Janis Kampars, “Application of Microservices for Digital Transformation of Data-Intensive Business Processes”, *ICEIS*, 2018
Pan Hui, Yong Li, Jörg Ott, Steve Uhlig, Bo Han, Kun Tan, “Mobile Big Data for Urban Analytics”, *IEEE Communications Magazine 56(11)*, 2018
Prasanna Kumar Illa, Nikhil Padhi, “Practical Guide to Smart Factory Transition Using IoT, Big Data and Edge Analytics”, *IEEE Access 6*, 2018
Holden Karau, “Unifying the open big data world: The possibilities∗ of apache BEAM”, *IEEE BigData*, 2017
Thanh Tran Thi Kim, Erhard Weiss, Christoph Ruhsam, Christoph Czepa, Huy Tran, Uwe Zdun, “Embracing Process Compliance and Flexibility Through Behavioral Consistency Checking in ACM - A Repair Service Management Case”, *BPMW*, 2015
Kuan-Ching Li, Hai Jiang, Laurence T. Yang, Alfredo Cuzzocrea, *“Big Data - Algorithms, Analytics, and Applications”*, Chapman and Hall/CRC, 2015
J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. Byers, “Big Data: The Next Frontier for Innovation, Competition, and Productivity”, *McKinsey Global Institute TR*, 2011
F. Murtagh, *“Multidimensional Clustering Algorithms”*, Physica-Verlag, 1985
Luca Oneto, Emanuele Fumeo, Giorgio Clerico, Renzo Canepa, Federico Papa, Carlo Dambra, Nadia Mazzino, Davide Anguita, “Train Delay Prediction Systems: A Big Data Analytics Perspective”, *Big Data Research 11*, 2018
Wullianallur Raghupathi, Viju Raghupathi, “Big data analytics in healthcare: promise and potential”, *Health Information Science and Systems 2(3)*, 2014
Viju Raghupathi, Yilu Zhou, Wullianallur Raghupathi, “Legal Decision Support: Exploring Big Data Analytics Approach to Modeling Pharma Patent Validity Cases”, *IEEE Access 6*, 2018
Philip Russom, “Big data analytics”, *TDWI best practices report*, fourth quarter, 2011
Sherif Sakr, Zakaria Maamar, Ahmed Awad, Boualem Benatallah, Wil M. P. van der Aalst, “Business Process Analytics and Big Data Systems: A Roadmap to Bridge the Gap”, *IEEE Access 6*, 2018
Christoph G. Schuetz, Simon Schausberger, Ilko Kovacic, Michael Schrefl, “Semantic OLAP Patterns: Elements of Reusable Business Analytics”, *OTM Conferences*, 2017
Vladislav A. Shchapov, Aleksei G. Masich, Grigorii F. Masich, “The technology of processing intensive structured dataflow on a supercomputer”, *Journal of Systems and Software 127*, 2017
Bin Shen, “Universal knowledge discovery from big data using combined dual-cycle”, *Int. J. Machine Learning & Cybernetics 9(1)*, 2018
Dawei Sun, Hongbin Yan, Shang Gao, Zhangbing Zhou, “Performance Evauation and Analysis of Multiple Scenarios of Big Data Stream Computing on Storm Platform”, *KSII Transactions on Internet and Information Systems 12(7)*, 2018
Xinyang Wang, Deyu Qi, Weiwei Lin, Mincong Yu, Zhishuo Zheng, Naqin Zhou, Pengguang Chen, A general framework for big data knowledge discovery and integration”, *Concurrency and Computation: Practice and Experience 30(13)*, 2018
Yulei Wu, Yang Xiang, Jingguo Ge, Peter Mueller, “High-Performance Computing for Big Data Processing”, *Future Generation Computer Systems 88*, 2018
Hehua Yan, Jiafu Wan, Chunhua Zhang, Shenglong Tang, Qingsong Hua, Zhongren Wang, “Industrial Big Data Analytics for Prediction of Remaining Useful Life Based on Deep Learning”, *IEEE Access 6*, 2018
Chao-Tung Yang, Jung-Chun Liu, Ching-Hsien Hsu, Wei-Li Chou, “On improvement of cloud virtual machine availability with virtualization fault tolerance mechanism”, *The Journal of Supercomputing 69(3)*, 2014
Yang Yang, Xianghan Zheng, Wenzhong Guo, Ximeng Liu, Victor I. Chang, “Privacy-preserving fusion of IoT and big data for e-health”, *Future Generation Computer Systems 86*, 2018
Paul Zikopoulos, Chris Eaton, *“Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data”*, McGraw-Hill Osborne Media, 2011