Application of Microservices for Digital Transformation of Data-Intensive Business Processes

Jānis Grabis and Jānis Kampars
Information Technology Institute, Riga Technical University, Kalku 1, Riga, Latvia

Keywords: Business Process Redesign, Digital Transformation, Data Analysis Model, Microservices.

Abstract: Business processes are redesigned as a part of business process management lifecycle and data intensive activities such as image processing, prediction and classification are increasingly incorporated into business processes. Data intensive activities often involve usage of data analysis models. It is argued that successful development and execution of data intensive business processes requires synchronization of business process redesign and data analysis models development activities. The business process architecture integrating core business process with data analysis model setup and updating sub-processes is developed. Business process transformation stages for incorporating data-intensive activities are outlined. The process redesign and execution is supported by the technical architecture based on microservices. An example of business process redesign is discussed.

1 INTRODUCTION

Business processes are subject to continuous improvement not least because of new technologies becoming available. There have been multiple cycles of technology driven business process redesign and the most recent cycle is frequently referred as to digital transformation (Zimmermann et al., 2016). This cycle is characterized by increasing usage of solutions enabling business process execution autonomy what is often achieved by relying on advanced data processing capabilities (Roedder et al., 2016). Bringing these data processing capabilities into business processes requires a new set of development technologies. These technologies should support simultaneous business process redesign, development of complex data analysis models and process execution software as well as establishing appropriate infrastructure. These activities concern various domains and require domain specific knowledge not only during the process redesign phase but also during business process execution.

Business process management lifecycle models and methodologies cover various stages of business process development (De Morais et al., 2014). They provide process redesign patterns and guidelines. However, they provide limited guidance concerning data processing issues and implementation aspects are platform dependent. Traditional implementation platforms such as ERP systems and Business Process Management (BPM) suites are better equipped for transaction processing rather than analytical processing. Additionally, data analysis models incorporated in data-intensive business processes have their own lifecycle and often require extensive computational resources at their disposal. That requires integration of core business process activities with activities associated with data analysis and an ability to provision the required computational resources.

Business process architecture (Weske, 2007) is an approach allowing integration of various dimensions of business process redesign and execution. Cloud-computing and service-orientation are two key technologies providing infrastructural services for running computationally intensive operations (Moreno-Vozmediano et al., 2013).

The overall goal of the proposed research is to elaborate a framework for development of data-intensive business processes. The framework integrates core business process activities and activities associated with data analysis for enabling data-intensive activities (i.e., horizontal business process architecture integration) as well as considers development of appropriate computational infrastructure (i.e., vertical business process

Grabis, J. and Kampars, J.
Application of Microservices for Digital Transformation of Data-Intensive Business Processes.
DOI: 10.5220/0006805207360742
In *Proceedings of the 20th International Conference on Enterprise Information Systems (ICEIS 2018)*, pages 736-742
ISBN: 978-989-758-298-1

architecture integration). It focuses on process redesign and implementation stages of the business process management lifecycle.

The objective of this paper is to outline the proposed transformation approach and to introduce key components of business process architecture for transformation of data-intensive business processes. The data-intensive business process is a process requiring data of various types and from different sources as well as analytical data transformations to guide and automate business process execution. In this paper, conversion of paper-based documents (e.g., travel receipts) into meaningful business data is considered as an example of the data intensive business process. Optical characters recognition is used to convert document images into characters and neural networks are used to extract business data from the text.

The rest of the paper is structured as follows. Section two introduces technologies used in elaboration of the approach. The transformation approach is presented in Section 3. An application example is provided in Section 4. Section 5 concludes

2 BACKGROUND AND REQUIREMENTS

2.1 Business Process Redesign

Business process is a sequence of activities performed to achieve specific business goals. Business processes are continuously improved through their lifecycle (Van der Aalst et al., 2003). The improvement cycle includes activities of process design, modeling, implementation, monitoring and optimization (De Morais et al., 2014). Becker et al. (2011) provide guidelines for the design of business process. The guidelines include As-Is modeling, analysis of As-Is models, To-Be modeling and optimization. Process documentation, usage of reference models and benchmarking as well as measurement and simulation methods are used for these purposes.

The process improvement methods rely on a set of general principles (De Morais et al., 2014) and are often tailored to specific needs and organizations. Therefore, there is a large variety of methods. Reijers and Mansar (2005) define a set of heuristic rules for business process improvement. The qualitative evaluation of the heuristic rules is also provided which helps business analysts to justify their BP improvement decisions. Barros (2007) attempts to formalize business improvement guidelines as reusable patterns. These patterns can be used to construct business process from existing best practices. Damij et al. (2008) propose a Tabular Activities Development methodology, which addresses both business process improvement and implementation of information system supporting the improved business processes. Process simulation is used to determine the process cycle time and related measurements are one of the key parts of the methodology. Dumas et al. (2013) categorize various process improvement alternative and their potential impact on process improvement as well as advocates usage of quantitative models in business process design. The review of BP improvement approaches is provided by Zellner (2011). Sidorova and Isik (2010) present an extensive review of business process research. They have identified business process design and business process on-going management and control as two of the four cornerstones of core business process research. Decision Model and Notation allows representing decision-making logics in business processes (Biard et al. 2015; Mertens et al., 2017). Although the authors acknowledge importance of quantitative analysis and decision-making in business process design and execution, integration of quantitative models into business processes is rarely explore what is especially important since quantitative and analysis models have their own life-cycle, which should be synchronized with business process life-cycle. It is also desired that models should not be developed for every client using a particular data intensive business processes. The models are pre-packaged or provided as a service and only their configuration is required for particular clients.

The business process architecture is an approach allowing to combine various dimensions of business process redesign (Lapouchnian et al., 2015). Lapouchnian et al. (2017) show that business process architecture is suitable for design of business processes requiring cognitive capabilities. The set of related business processes include processes for analytical model creation, validation and business improvement.

Business processes are implemented and executed using various technologies. ERP systems provide monolithic business process development environment, which is efficient for stable processes while being difficult to modify and to implement custom requirements. Business process management (BPM) suites allow implementing custom processes and jointly with service-oriented architecture create an environment support flexible modification of the processes by selecting appropriate services. However, BPM and to some extent SOA are not well-suited for data analysis purposes, especially, requiring integration of various data processing technologies. Scalability is also achieved at the expense of efficiency. Business intelligence technologies are often used to develop data analysis models though their integration in business processes is often performed in off-line or ad-hoc manner (Chou et al., 2005).

Microservices (Dragoni et al., 2017) is a technology recently gaining prominence as self-contained light-weight containers of business services. They allow for high degree of modularity, containers are created on-demand quickly and with a little overhead as well as various data processing and storage technologies can be utilized.

2.2 Requirements

The proposed research focuses specifically on design and execution of data intensive business processes. The existing research and industrial experience suggest that the following requirements should be considered:

- Support for development and tuning of data analysis models - business process and data analysis models' life-cycles should be synchronized and model development should be a part of the overall business process redesign and execution;
- Presentation and understanding of data analysis models - user of data analysis model should have utmost understanding of models used in business process execution. That can be achieved by explaining levers available for controlling behavior of the models as well as by providing facilities for experimenting with models. The experimentation allows identification of appropriate model usage modes and understanding of model usage consequences;
- Modification - data analysis models are subject of frequent revision as more data become available and models are refined. The new models should be evaluated and integrated in the business processes without affecting the rest of the process;
- Scalability - data analysis models often require significant computational resources and decisions should be made during business process execution for potentially large number of users. That requires both vertically and horizontally scalable execution environment;
- Flexibility - data analysis models, data used in the models and model solving algorithms come in different forms and the development and execution environment should be flexible to support usage of the most appropriate data processing methods and technologies.

3 TRANSFORMATION APPROACH

The transformation approach consists of three main components: 1) business process architecture; 2) transformation process; and 3) technical architecture. The business process architecture defines an interrelated set of sub-processes required to design and run data-intensive business processes. The transformation process defines steps to be performed to redesign a traditional business process into a data-intensive business process. The technical architecture specifies technologies used to execute data-intensive business processes.

3.1 Business Process Architecture

The business process architecture consists of three sub-processes: 1) sub-process of setting up models of data analysis; 2) actual execution of data-intensive business process or operations; and 3) sub-process of updating the data analysis models once additional data are accumulated during the operations (Figure 1).

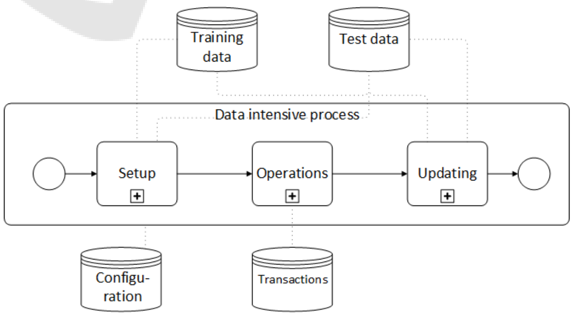

Figure 1: Data-intensive business process architecture.

The operations sub-process is the core business process providing the required business functions. Many activities of this process rely on data analysis models, which are invoked during the process execution. These models are setup prior to their usage. The setup sub-process might involve configuration of several models for all data-intensive activities in the business process. Assuming that a number of data analysis models for solving the decision-making problem are available, typical activities of the setup process are model selection (e.g., exponential smoothing or moving average for forecasting purposes), selection of structural parameters (e.g., input variables for a regression model or layers for neural networks), estimation of models' parameters (e.g., estimation of regression coefficients) and evaluation of modeling performance. The evaluation activities are of particular importance to ensure acceptance of data analysis results and their adoption for business process execution.

The setup sub-process requires training and test data to estimate and to evaluate models, respectively. The business process architecture represents these as coming from data stores, which physically could belong to different types of storage facilities. It also uses configuration data. The configuration data can be divided in three groups: 1) models' parameters; 2) performance indicators; and 3) design of experiments. The configuration data can be manipulated by business process users. All models' parameters are not included in the configuration data but specifically promoted parameters important for experimenting, tuning and understanding of the models. The performance indicators set targets for both modeling and process performance. That is important because in many situations data analysis models can be used only if reasonably accurate. The design of experiments is necessary for evaluation purposes. That allows business process users to gain understanding of models' behavior and implications of modeling results.

The operations sub-model is business problem specific. The data analysis models are invoked during execution of its activities. The execution results are stored as transactional records, which could be used for further refinement of the models. The refinement is handled by the update sub-process. The main issues represented in the update process are updating conditions (i.e., when to replace existing models with the updated ones) and handling of model life-cycle issues (e.g., model versioning for long-running processes).

3.2 Transformation Stages

The transformation process defines steps of process redesign into a data intensive business process and processes execution. The transformation process itself is a part of the overall business process management lifecycle though other stages of this lifecycle are not considered in this paper.

The following steps are performed to develop a data intensive business processes:

- Define business process - representation of the core business process model is developed assuming that process identification has been performed
- Identify data intensive activities - the process is vetted to discover opportunities for automation or improved decision-making due to data analysis. A data intensive activity requires data entities, which are not directly associated with the data record currently processed by the activity, and involves analytical transformations;
- Integrate data - data sources for training data are identified and data delivery channels are established;
- Develop data analysis models - data analysis models appropriate for a specific business area are developed. The models are developed as packaged products configured and tuned for a particular case;
- Deploy models as microservices - the models are packaged and deployed as microservices to achieve high scalability and modifyability;
- Redesign process around data intensive activities - introduction of data intensive activities might result in significant changes in the business process to achieve full benefits of digital transformation;
- Integrate models in business process - bind data intensive business process activities with corresponding microservices and development of data flows;
- Execute business process - the business process is executed including invocation of the microservices;
- Update data analysis models - as additional data are accumulated as the result of business process execution, the data analysis models are updated to account for the latest tendencies.

Some of these steps are performed in parallel, for example, data integration and development of data analysis models.

3.3 Technology

The technical architecture (Figure 2) consists of four main components: 1) operations execution component; 2) model setup component; 3) model execution component; 4) data storage. The operations execution module is a container for running data intensive activities and invoking microservices. It can be implemented using various technologies suitable for developing business processes. The model execution component is built using microservices as lightweight containers of executable data analysis models. The containers are created and disposed on-demand depending on computational requirements. The microservices are decoupled from operation execution component by a queue thus providing scalability and load balancing capabilities. The model setup component also can be implemented in various technologies. However, it is proposed that notebooks are suitable technology because they combine model development, execution, experimentation and documentation thus clearly explicating the data analysis model. The model setup often results in a model specification (e.g., PMML standard), which can be used to configure the microservices. Data storage technologies are selected depending on requirements and include traditional relational databases, object storages and document-oriented databases. All components access the storage component though isolated storage facilities usually are created for the setup and operations purposes.

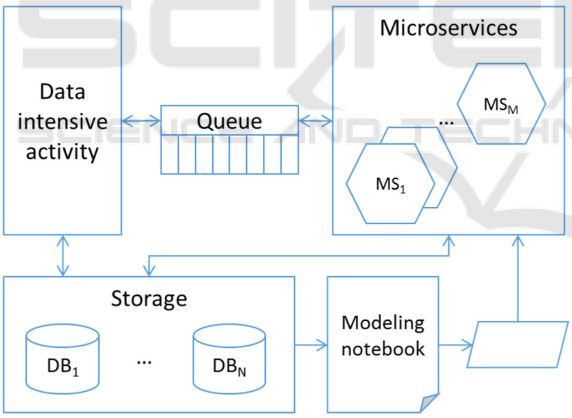

Figure 2: Main components used to create technical architecture.

4 EXAMPLE

The proposed approach is illustrated using the travel management processes as an example. Business travelers are required to report their expenses after returning from the trip what includes filling out expense forms and submitting travel receipts (Figure 3.a). The current business process captures images of travel receipts without much of valuable use of these data and requires manual filling of the expense report. Apparently, there are at least two data-intensive activities in the current process, namely, Gather receipts and Fill expense report. The images captured during the Gather receipts activity can be transformed into text making them available for further processing. The text extracted from the images can be analyzed to fill the expense report automatically. The redesigned process is shown in Figure 3.b. Character recognition (OCR) is used to extract text. Neural networks are used to recognize type of travel expenses, date and amount. The training data needed for these purposes are a data set containing images of travel receipts and actual expenses associated with these images. The extracted expenses still need to be verified by a human to ensure correctness.

The setup sub-process starts with preparation of input data including training and test data. These data are either previously accumulated images of travel receipts with known (manually filled) expense data items or specifically generated data sets. Assuming that neural networks are used to identify expense data items, a structure of the neural network is defined and training of the model is performed. The final activity is evaluation of the accuracy of expense data identification. The automated identification is enabled only if this accuracy is satisfactory as defined by performance indicators. Every new enterprise using the travel expenses reporting business process requires tailoring of the model and the setup sub-process provides clear

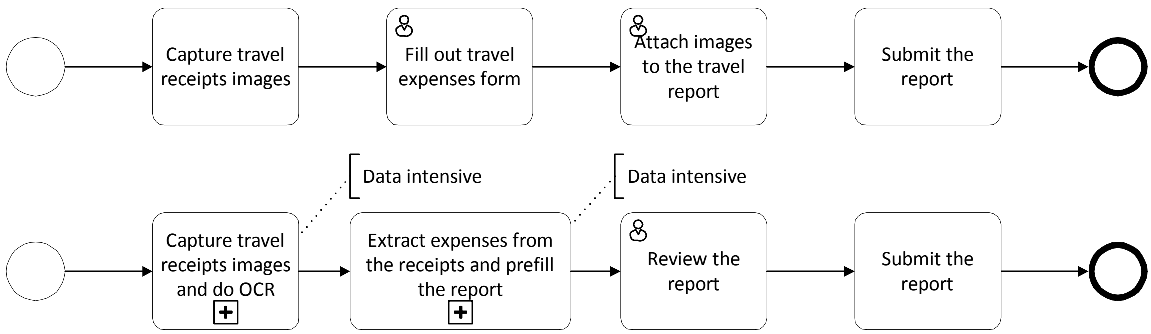

Figure 3: Travel expenses reporting: a) original process; and b) operations sub-process of the redesigned process.

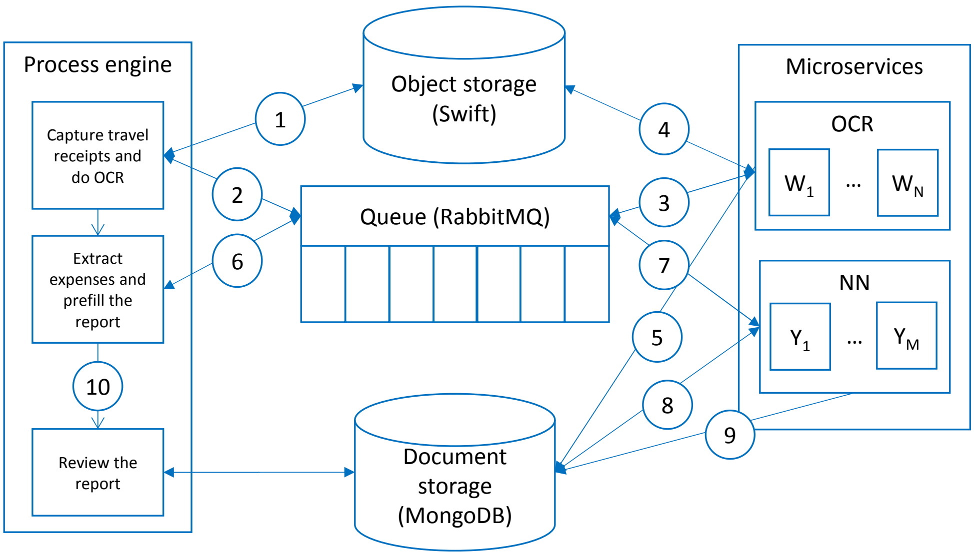

Figure 4: Technical architecture (the numbers indicate the sequence of interactions).

guidance for that. The configuration parameters and the design of experiments are of particular importance because they allow for understanding of the model and its behavior. In this case, the estimated expenses extraction accuracy threshold is one of the configuration parameters. If the estimated accuracy is below the threshold then the extracted data are not used and manual entry is requested. Another configuration parameter for travel receipt capture is the number of images taken of the single receipt to improve OCR precision.

At the infrastructure level, the redesign process is supported by a scalable data analysis solution (Figure 4). The image capture activity creates an image processing job and places it in the queue (implement using RabbitMQ[1]). The images captured are also stored in object storage implemented using Swift[2]. The image to text conversion is realized as a microservice using the Tesseract[3] library. Incoming image conversion messages are pulled from the queue by microservice instances and the instances read image data from the object storage. The recognized text is stored in the document-oriented database (implemented using MongoDB[4]) and the queue is notified that the text is available for further processing. That triggers the expense data extraction using neural networks (realized using TensorFlow[5] and dedicated neural networks (NN) are created for each expense item). The neural network was trained in the model setup component and production neural networks are again implemented as microservices. The extraction results are stored in the document-oriented database and the data intensive activity captures events indicating that extraction results are available for verification.

Following the business process architecture allows to redesign the business process as well as to establish supporting processes for model setup and update. The technical solution is highly scalable and the components are decoupled one from another. That allows to modify the components, to use specialized technological components as necessary and to integrate the solution in the enterprise's existing technological landscape. In this case, the expense data extraction algorithm is regularly modified to achieve better accuracy. Once the new version of the algorithm is approved for application it is deployed as a microservice without a delay.

5 CONCLUSION AND OUTLOOK

The paper proposes a framework for design execution of data-intensive business processes. Integration of setup, operations and update sub-processes allows to implement data intensive business processes as one package rather than implementing the key process and its supporting processes separately as it is often a case. At the current stage of the proposed research, a proof-of-

[1] https://www.rabbitmq.com/
[2] https://docs.openstack.org/swift/latest/
[3] https://github.com/tesseract-ocr
[4] https://www.mongodb.com/
[5] https://www.tensorflow.org/

concept has been developed and the further elaboration of framework is required.

The example described is currently implemented as a prototype and experiments are conducted with this prototype to tune data analysis models and to evaluate efficiency of the redesigned processes. Performance of the technical solution is also being evaluated with focus on scalability.

The paper has proposed the business process architecture intended for data intensive business processes. The further formalization of this architecture is required. That includes elaboration of transition from the setup sub-process to the operations sub-process both at the business process and infrastructure levels. From the transformation process perspective, guidance on identification of data intensive activities should be provided. Additionally, redesign around data-intensive activities often results in substantial changes in other business process activities. These changes are similar across business processes depending of data analysis technique used and patterns for process redesign could be formulated.

The prototype developed relies on the microservice architecture and described technological choices. These design decisions are also subject of further investigation.

REFERENCES

Barros, O., 2007.Business process patterns and frameworks: Reusing knowledge in process innovation. *Business Process Management Journal 13, 47-69*

Becker, J., Kugeler, M., Rosemann, M. 2011. *Process Management: A Guide for the Design of Business Process*, Springer, New York

Biard, T., Mauff, A.L., Bigand, M., Bourey, J.-. 2015. Separation of decision modeling from business process modeling using new decision model and notation (DMN) for automating operational decision-making. IFIP Advances in Information and Communication Technology 463, pp. 489-496.

Chou, D.C., Bindu Tripuramallu, H., Chou, A.Y. 2005. BI and ERP integration, Information Management, Computer Security, Vol. 13, No. 5, 340–349.

Damij, N., Damij, T., Grad, J., Jelenc, F. 2008. A methodology for BP improvement and IS development. Information and Software Technology 50, 1127-1141

De Morais, R. M., Kazan, S., de Pádua, S. I. D., Costa, A. L. 2014, An analysis of BPM lifecycles: From a literature review to a framework proposal, *Business Process Management Journal, vol. 20, no. 3, pp. 412-432.*

Dragoni N. et al. 2017. Microservices: Yesterday, Today, and Tomorrow. In: Mazzara M., Meyer B. eds. Present and Ulterior Software Engineering. *Springer*

Dumas M., La Rosa M., Mendling J., Reijers H.A., 2013. *Fundamentals of Business Process Management*. Springer, Berlin, Heidelberg

Lapouchnian, A., Yu, E., Sturm, A. 2015. Design dimensions for business process architecture. *Lecture Notes in Computer Science 9381, pp. 276-284.*

Lapouchnian, A., Babar, Z., Yu, E., Chan, A., Carbajales, S., 2017. Designing process architectures for user engagement with enterprise cognitive systems. *Lecture Notes in Business Information Processing, 305, 141-155.*

Mertens, S., Gailly, F., Poels, G. 2017. Towards a decision-aware declarative process modeling language for knowledge-intensive processes, *Expert Systems with Applications, vol. 87, pp. 316-334*.

Moreno-Vozmediano, R., Montero, R.S., Llorente, I.M. 2013. Key challenges in cloud computing: Enabling the future internet of services, *IEEE Internet Computing, vol. 17, no. 4, pp. 18-25.*

Reijers, HA, Liman Mansar, S. 2005. Best practices in BP redesign: An overview and qualitative evaluation of successful redesign heuristics. *Omega 33, 283-306*

Roedder, N., Dauer, D., Laubis, K., Karaenke, P., Weinhardt, C. 2016. The digital transformation and smart data analytics: An overview of enabling developments and application areas, *Proceedings - 2016 IEEE International Conference on Big Data, Big Data 2016, pp. 2795.*

Sidorova, A., Isik, O. 2010. BP research: A cross-disciplinary review, *Business Process Management Journal, vol. 16, no. 4, pp. 566-597.*

Van der Aalst, W.M.P., Ter Hofstede, A.H.M., Weske, M. 2003. Business process management: A survey. In: van der Aalst W.M.P., Weske M. (eds) Business Process Management. BPM 2003. *Lecture Notes in Computer Science, vol 2678. Springer, Berlin, Heidelberg.*

Weske, M. 2007. Business process management: Concepts, languages, architectures in Business Process Management: Concepts, Languages, Architectures, *Springer*.

Zellner, G., 2011. A structured evaluation of Business Process improvement approaches, *Business Process Management Journal, vol. 17, no. 2, pp. 203-237.*

Zimmermann, A., Schmidt, R., Jugel, D., Möhring, M. 2016. Adaptive enterprise architecture for digital transformation. Communications in *Computer and Information Science 567, pp. 308-319.*

Zimmermann, A., Schmidt, R., Jugel, D., Möhring, M., 2015. Evolving enterprise architectures for digital transformations, Lecture Notes in *Informatics LNI., Proceedings - Series of the Gesellschaft fur Informatik GI., pp. 183.*