Distributed services and a warehouse as an ecosystem on science and higher education – extended abstract

Emil Podwysocki, Łukasz Błaszczyk, Marta Niemczyk, Jarosław Protasiewicz, Marek Michajłowicz, Sylwia Rosiak, Iwona Kucharska

National Information Processing Institute, Warsaw, Poland

Keywords
warehouse; services for science; open data; web services

1. Summary

In this study, we demonstrate the concept of a new information ecosystem on science and higher education in Poland. We focus on a distributed architecture of services and a data integrating warehouse. The data warehouse is tailored to be well-suited to micro-services, the modern information system architecture. In addition, we show examples of descriptive and predictive analytics in the data warehouse, which assist policymakers in making decisions. The primary outcome of the study is a new architecture, which relies on combining the data warehouse and micro-services into one ecosystem of distributed services for science and higher education. We hope that our experiences and concepts will be beneficial to those who face challenges in redesigning of existing information systems.

2. DISTRIBUTED INFORMATION ECOSYSTEM OF SERVICES

Data warehouse solutions have been present in higher education and science for nearly two decades. They act both as data integrators and analytical tools. The validity of the study relies on combining a data warehouse and micro-services in the ecosystem of distributed services for science and higher education. According to our best knowledge, it is the first time such architecture has been proposed.

The information about Polish science and higher education has been collected since 2011 in an integrated and centralised information system. The system is tightly compliant with the legal requirements. Therefore, a new bill, the Constitution for Science, introduced in Poland last year, which has deeply altered the regulations for science and higher education, must result in profound and numerous changes within the system. We must develop new components so as to ensure smooth data migration to the new registers. Such an approach gives the opportunity to use modern technologies and optimise business processes. Amongst several improvements, we are planning to abandon the monolithic, multi-layered architecture with a central database, in favour of a distributed architecture of services. It will be composed of micro-services grouped into application, integration, and analytical layers (see Figure 1).

One possible application of this approach is event sourcing, in which changes in the state of business entities will be stored as a sequence of event objects. When a service detects a data change, it will publish an event object. Each published domain event will be available and ready-to-use by all interested modules (micro-services).

The data warehouse will integrate the data from various transactional systems without replacing it. The data will be converted into a form which allows optimal responsiveness of user interferences or web services, as well as generating various analyses and reports (Figure 2).

Every single event generated by POL-on 2.0 has to be transformed to Oracle Data Manipulation Language command. That approach provides near real time replications. Events after the transformation fill an application read models, and the next steps will be performed in conventional Oracle Data Integrator jobs.

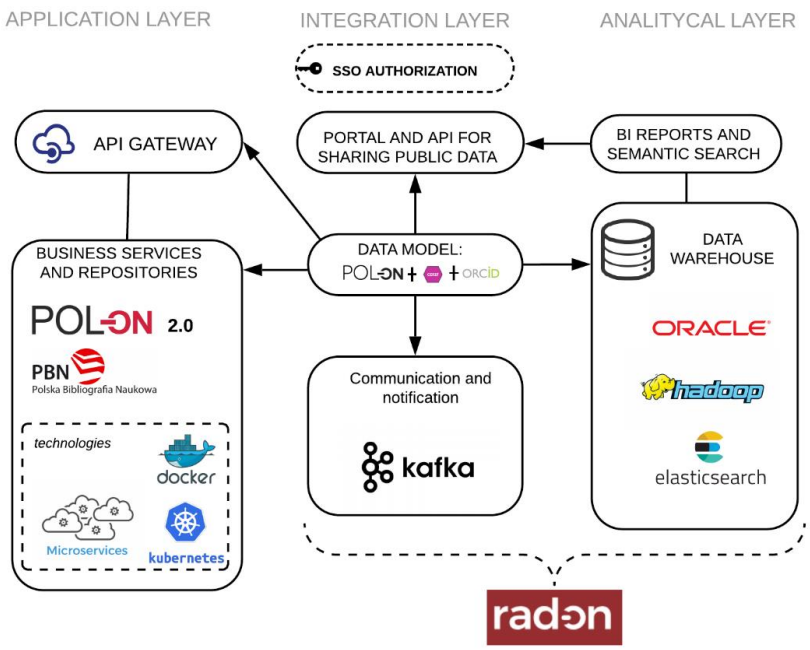

Figure 1. Vision of target architecture of the distributed system

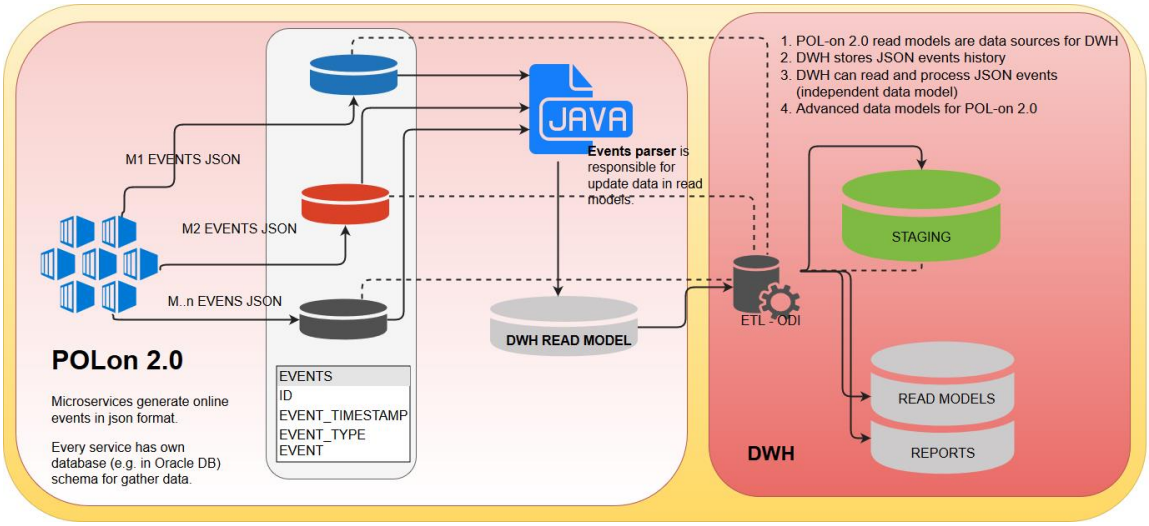

Figure 8. Events processing process.

3. CONCLUSION

We have outlined an ecosystem of distributed services for science and higher education in Poland. It includes a data warehouse, which provides us with data marts for these services to make them faster. Moreover, business intelligence tools offer descriptive and predictive analytics, which helps policymakers make proper decisions.

Currently, we are developing the designed architecture. Its deployment will verify the assumptions we have made and presented in this study. We hope that in the future work we will be able to provide quantitative data demonstrating the ecosystem's performance, and share our experiences.

4. AUTHORS' BIOGRAPHIES

Emil Podwysocki (MSc) received a Master degree in Telecommunications Systems at the Technical University of Lodz. He has 10 years of professional experience related to ETL/ELT, data warehouses and Business Intelligence. His areas of interest include Oracle technology, Big Data, Business Intelligence and data visualizations. Currently, he works as a Database and Business Intelligence Team Leader in the National Information Processing Institute.
Email: emil.podwysocki@opi.org.pl
Profile: www.linkedin.com/in/emil-podwysocki

Łukasz Błaszczyk (MSc) is a product owner and business analyst at the National Information Processing Institute. He earned his Masters degree in Forest Information Technology at the University of Applied Sciences in Eberswalde and a master's degree in Forestry at Warsaw University of Life Sciences – SGGW. He finished postgraduate studies in business analysis at SGH Warsaw School of Economics. He has 11 years of experience in business and systems analysis.
Email: lukasz.blaszczyk@opi.org.pl

Marta Niemczyk (MSc) received a Masters degree in Mathematics at the University of Warsaw. She has many years of professional experience as a business and systems analyst, endorsed by certificates in methods of IT systems modelling. Currently she works as a software designer and business analyst in the National Information Processing Institute.
E-mail: marta.niemczyk@opi.org.pl

Jarosław Protasiewicz (PhD) is an assistant professor at the National Information Processing Institute and the head of the Laboratory of Intelligent Systems. He received a Ph.D. in computer science at the Systems Research Institute of the Polish Academy of Sciences. His areas of interest include agile project management, software design and development, big data, machine learning, and bio-inspired algorithms.
Email: jaroslaw.protasiewicz@opi.org.pl

Marek Michajłowicz (MSc) is the Deputy Head of the Laboratory of Intelligent Systems. He received a Master degree in Sociology at Cardinal Stefan Wyszyński University in Warsaw and a Bachelor of Engineering (B.E.) in computer science at Warsaw School of Information Technology under the auspices of the Polish Academy of Sciences. He has several years of professional experience related to business and systems analysis. Since 2014 he has worked as the project manager of the POL-on system in the National Information Processing Institute. His areas of interest include agile project management, software design and development, big data and warehouses.
E- mail:marek.michajlowicz@opi.org.pl;
Profile: www.linkedin.com/in/marek-michajlowicz

Sylwia Rosiak (MSc) is a business analyst at the National Information Processing Institute. She received a Masters degree in Management at the University of Warsaw and completed postgraduate studies in designing information systems at the Warsaw University of Technology.
Email: sylwia.rosiak@opi.org.pl

Iwona Kucharska (MSc) received a Masters degree in Computer Science and Econometrics at the University of Computer Science and Economics in Olsztyn. She completed postgraduate studies in Business Processes Engineering and Business Intelligence at the Warsaw University of Technology. She has many years of professional experience as a software tester and as a business analyst, endorsed by certificates. Currently she works as a software designer and business analyst in the National Information Processing Institute.
E-mail: iwona.kucharska@opi.org.pl