# DATA VAULT BASED SYSTEM CATALOG FOR NOSQL STORE INTEGRATION IN THE ENTERPRISE DATA WAREHOUSE

Katerina Černjeka
kcernjeka@uniri.hr

Danijela Jakšić
dsubotic@uniri.hr

Patrizia Poščić
patrizia@uniri.hr

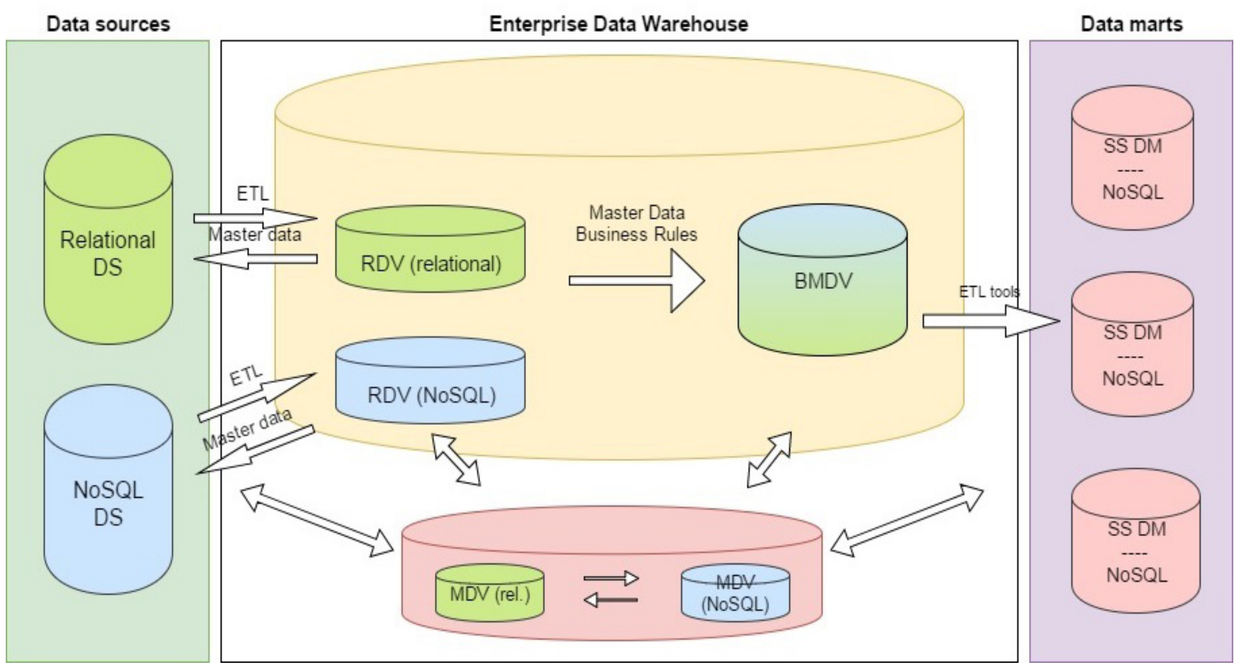
## 1 INTRODUCTION

In the context of data heterogeneity, our Data warehouse (DW) system [3] will provide not only the ability to query relational and NoSQL data but to integrate, store and preserve history of all the corporate data and their changes into a single system of records.
We aim to extend the current system catalog from [1] in order to integrate NoSQL data sources with relational ones in Data Vault (DV) [2] based Enterprise data warehouse (EDW) and by doing so track the origin and history of changes for both, data and metadata of NoSQL stores and relational databases, as well as their schemas.

## 2 RESEARCH QUESTION

*Can our system catalog, built upon the proposed DW architecture, be able to track and store changes of data (and metadata) from relational source and NoSQL source through the whole DW architecture, and by doing so serve as a basis for data auditing and data governance?*

## 3 PROPOSED DW ARCHITECTURE




EDW is partially oriented towards the data sources side as *raw data vaults*, and partially oriented towards reporting side as *business data vault*.

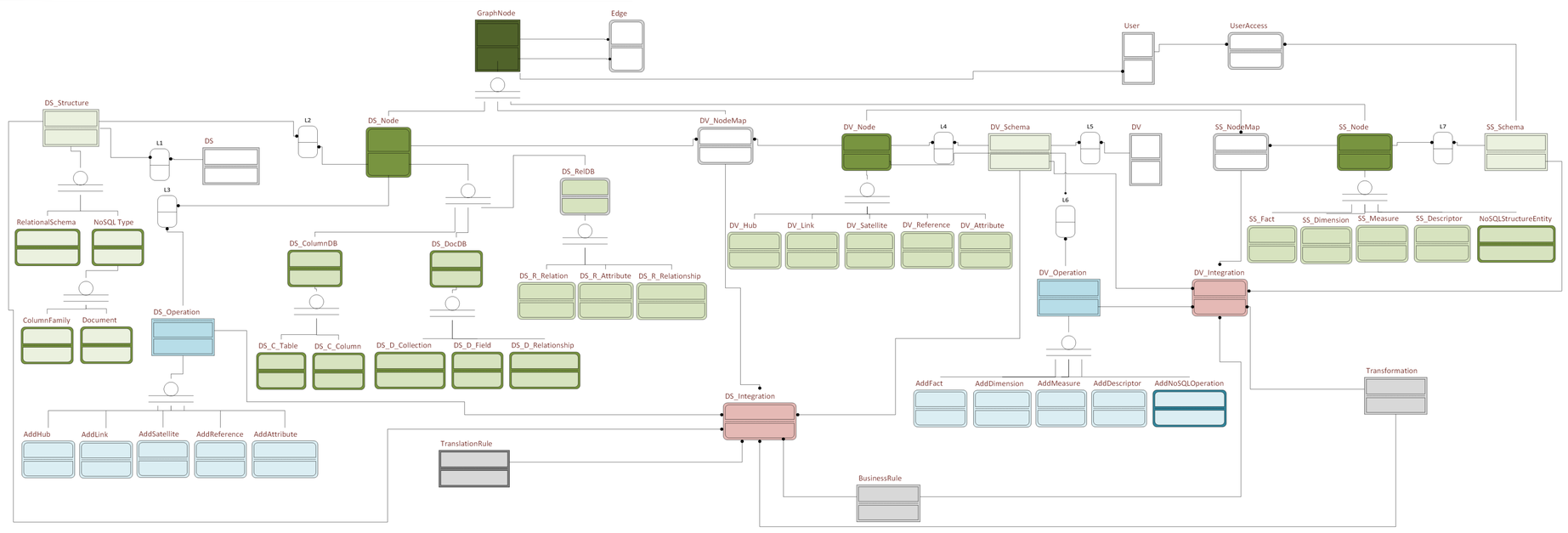
Raw data vaults (RDV) contain actual (unchanged) copies of the originals from the data sources.
Business data vault (BDV) is created by updating and consolidating this raw data with application of business rules and standardized master data. Since copies of the originals are permanently kept in raw data vaults there is no loss of information and the basis for audit process is created.

Because of the separation between raw data vault and business data vault we distinguish *reversible* and *irreversible transformations* which allow us to *track* data back to its source and reconstruct them, if necessary.

The meta-data vault (MDV) represents the metadata repository and it is the basis for our extended system catalog. It integrates meta-data from all layers of DW architecture with an emphasis on integration of raw data vaults and business data vault in the EDW system of records.

*This is the key idea for getting an integrated central EDW system of records and basis for data governance*

## 4 METADATA DATA VAULT MODEL




Data sources | EDW | Data Marts

4 main parts (nodes): *data sources* (DS_Node), *DV EDW* (DV_Node), *data marts* (SS_Node) and *user access* part (User and UserAccess).

## 5 CONTRIBUTIONS

**A** New integrated DW architecture

**B** Meta-data vault model of EDW system catalog for both, relational and NoSQL data sources

**C** An EDW system catalog prototype that stores metadata of both, relational and NoSQL data sources, as well as their schemas.

## 6 REFERENCES

[1] D. Jakšić, "Metadata repository model for data warehouse schema evolution and integration with master data management system," Sveučilište u Rijeci, 2016., doctoral thesis
[2] D. Linstedt and O. Michael, Building a Scalable Data Warehouse with Data Vault 2.0. Todd Green, 2016.
[3] R. Kimball, The Data Warehouse Lifecycle Toolkit. New York: Wiley, 2008.
[4] M. Golfarelli and S. Rizzi, Data Warehouse Design, Modern principles and methodologies. The McGraw-Hill Companies, S.r.l.-Publishing Group Italia, 2009.
[5] D. Subotic, V. Jovanović, and P. Poščić, "Data warehouse and master data management evolution – a meta data vault approach," Issues Inf. Syst., vol. 15, no. Ii, pp. 14–23, 2014.
[6] W. H. Inmon, Building the Data Warehouse. Wiley Computer Publishing, 1992.

DEPARTMENT OF INFORMATICS
UNIVERSITY OF RIJEKA