Extension of the mediator-wrapper architecture for heterogeneous data source integration by adding a mask

Juraj Dončević, mag. ing.
(juraj.doncevic@fer.hr)
mentor: prof. dr. sc. Krešimir Fertalj,
University of Zagreb Faculty of Electrical Engineering and Computing

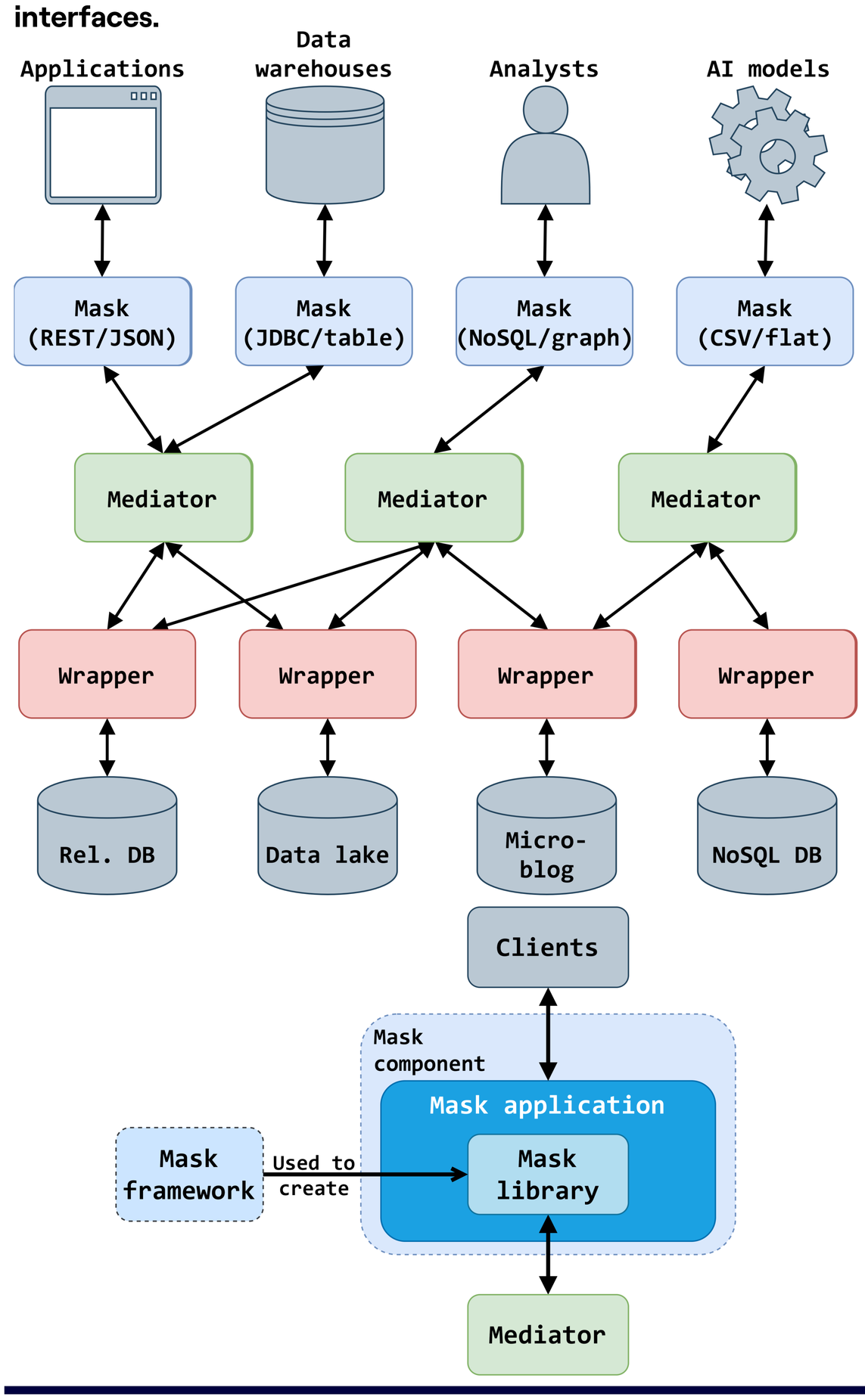
1. Introduction

The goal of this research is the extension of the mediator-wrapper architecture for heterogeneous data source integration with a new component type for representation of data, schemas and queries, called a **mask**. The purpose of the extension is to improve software flexibility in relation to the mediator-wrapper architecture in scenarios requiring the addition of different kinds of user access interfaces to a heterogeneous data integration system. The improvement of software flexibility would be proven through a quantitative analysis of software flexibility. The new architecture is called the **mask-mediator-wrapper architecture**.

2. Problem Description

The scientific community has determined that there are currently **no quality readily-available systems** for integrating data sources. Recent research in the field of data storage and integration emphasizes the importance of data representation to facilitate data preparation for data-oriented systems, which has also been neglected in previous research.

There is a need for a modern open-source data integration system that allows the addition of a variety of access interfaces.



3. Methodology

A quantitative shift-cost software flexibility analysis proves that the mask-mediator-wrapper architecture is an improvement in scenarios that require the addition of different kinds of user access interfaces to the system, as well as proving that there is no deterioration in other scenarios in comparison with the original mediator-wrapper architecture.

The following scenarios were elaborated for the mediator-wrapper architecture with two layers of mediators (2LMW) and the mask-mediator-wrapper architecture (MMW):

1. *Addition of a new representation type*
2. *Addition of a new representation*
3. *Addition of a mediator*
4. *Addition of a wrapper to an existing mediator*

The removal scenarios for these features are analogous to the scenarios above, due to the equal quantitative shift in the software.

During the analysis, the following assumptions were made about the implementation and deployment costs of a mediator component (Me'), a mediator with built-in representational functionalities (Me) and a mask (Ma):

(1)	Me' has a lower implementation cost, because it has no additional representational functionalities.	$C_{Me}^{impl} > C_{Me'}^{impl}$
(2)	Me' has a lower deployment cost, because fewer interfaces are exposed to the environment.	$C_{Me}^{depl} > C_{Me'}^{depl}$
(3)	Ma has a lower deployment cost, because fewer interfaces are exposed to the environment.	$C_{Me}^{depl} > C_{Ma}^{depl}$

4. Results

From the obtained software shift-cost estimates, it is possible to conclude that a system implemented in the mask-mediator-wrapper architecture has **improved flexibility** in scenarios 1, 2 and 3. In scenario 4, there is **no degradation**.

Sc	2LMW	MMW
1.	$C_{Me}^{impl} + C_{Me}^{depl} + C_{Conn}^{set}$	$C_{Ma}^{impl} + C_{Ma}^{depl} + C_{Conn}^{set}$
2.	$C_{Me}^{depl} + C_{Conn}^{set}$	$C_{Ma}^{depl} + C_{Conn}^{set}$
3.	$2 \times C_{Me}^{depl} + (N+1) \times C_{Conn}^{set}$	$C_{Me'}^{depl} + C_{Ma}^{depl} + (N+1) \times C_{Conn}^{set}$
4.	$C_{W}^{depl} + C_{Conn}^{set}$	$C_{W}^{depl} + C_{Conn}^{set}$

5. Conclusion

The analysis proves that the **addition of the mask is an enhancing extension** of the mediator-wrapper architecture. The extension allows for more flexible management of data, schema and query representation.

By separating the functionality of representation into a mask, it is possible to further reason about the standardized development of mask kinds. A mask framework allows the creation of specialized mask libraries, which are built into mask applications. The application is deployed into an existing system as a component.

See it all in action with **Janus** ☺