Reviewing features of Data Warehouse Architectures: A Taxonomy

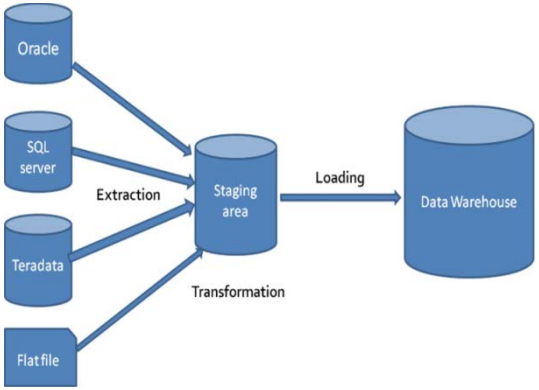
Qishan Yanga and Tai Mai
Linz, Austria
June 11, 2018

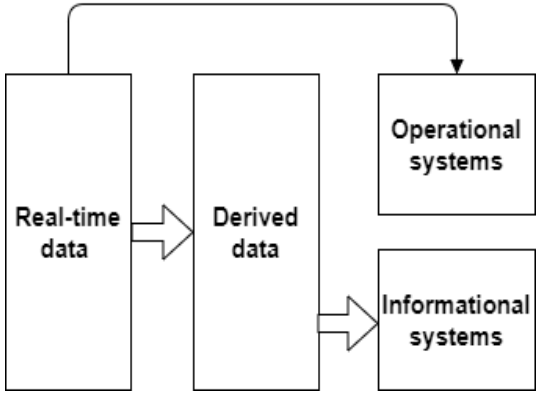
fh OBERÖSTERREICH
UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

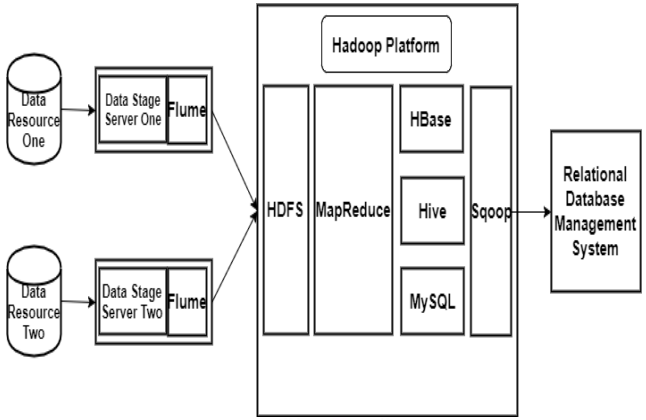
Introduction

A data warehouse (DWH) is a subject oriented, integrated, non-volatile and time-variant collection of data that supports the capability of the decision-making in organisations [1].

Through decades of development and innovation, their architectures have been extended to the variety of derivatives for achieving different either business or technical requirements of organisations.







Problem Statement

Confused!!!

BI

ETL

Staging Area

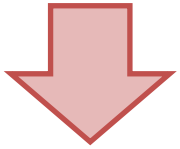
Data Flow

Data Storage

Refresh

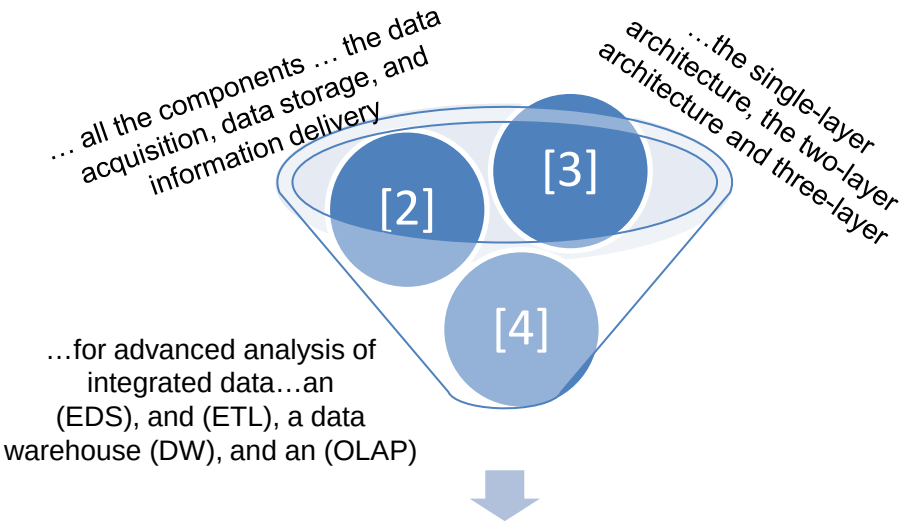
Motivation

To the best of our knowledge, there is no effort from the literature which collects and classifies features for data warehouse architecture evaluations generally and systematically.

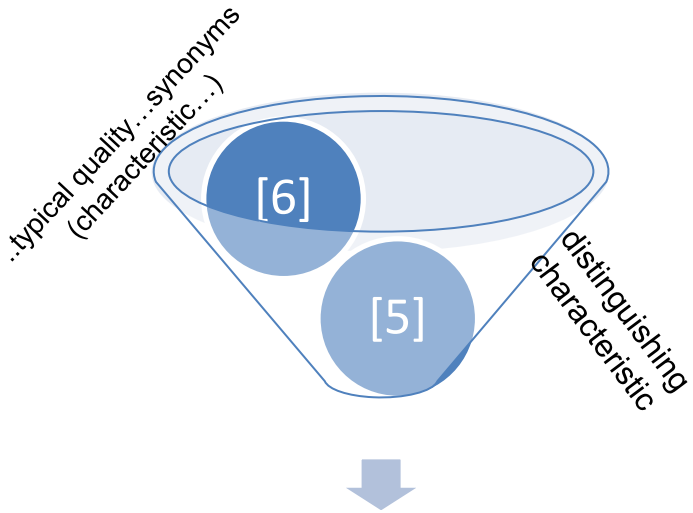


According to the problem and current situation, it motivates us to investigate DWHA features, which will benefit DWHA evaluations for different perspectives. This research conducts a systematical literature review for data warehouse architecture feature collection and classification.

Theoretical Background



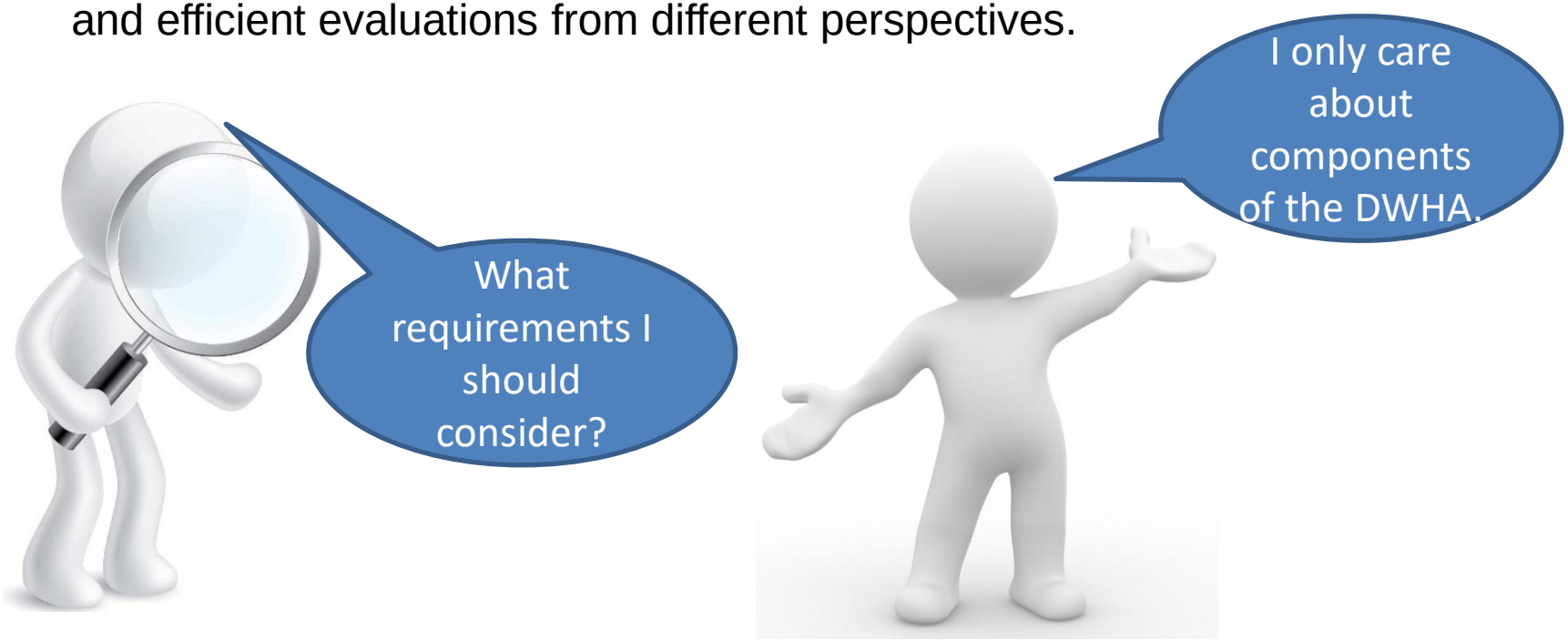
The **Data Warehouse Architecture** is a principled system with **fundamental properties** and **rational relationships** for data manipulations, storage and analyses, which includes the **Source Layer**, **ETL Layer**, **Data Warehouse Layer** and **Data Presentation Layer**.



A **feature** of the DWHA is a **distinguishing characteristic** to reflect an **attribute** or **component** to constitute an important portion of the DWHA

Hypothesis

Features of data warehouse architectures can be categorised into several classifications to provide guidelines for better requirement understanding and efficient evaluations from different perspectives.



Research Method

The systematic literature review is processed based on [7, 8].

1. Research question
2. Identify concepts for searching
3. Searching
   i. Database: IEEE, dblp, Google Scholar.
   ii. Year range: 2008 - 2018.
   iii. Keywords: “ Data Warehouse Architecture, feature and its synonyms”
   iv. Identification is made by reading the title and abstract to distinguish potentially eligible for inclusion.
4. Features extraction.
   i. Collect features mentioned in each selected paper
   ii. Remove duplicated features
5. Features classification.

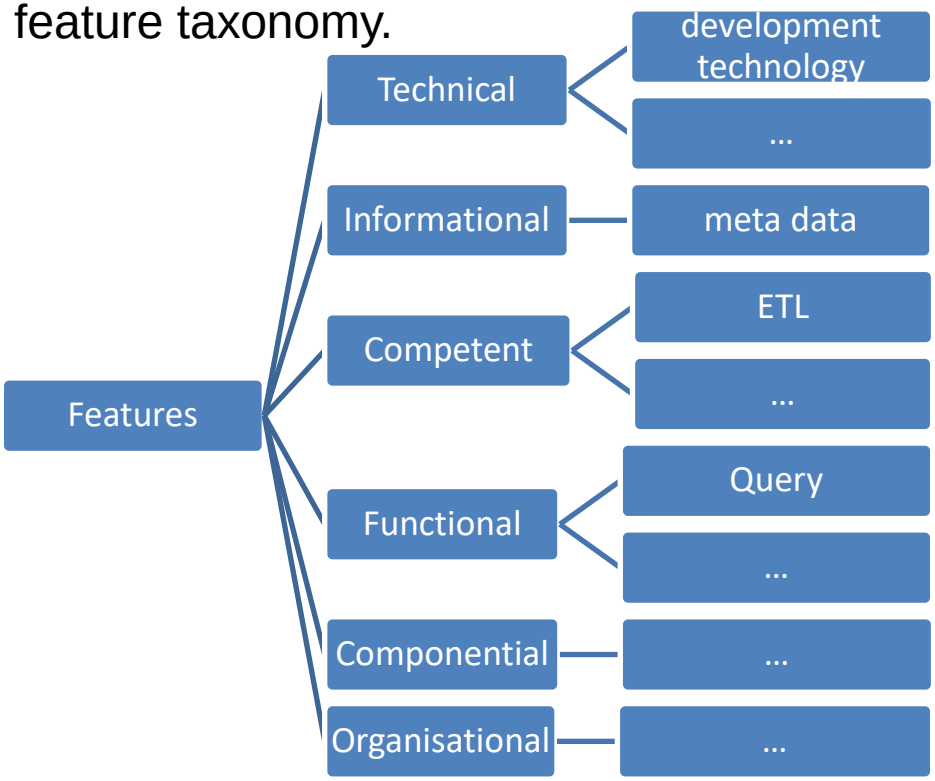
Initial Results

247 features are abstracted, collected. Then, duplicated features are deleted and 148 unique features are summarised. Some of them are tabulated blow.

user	meta data	ETL	management	cycle	OLAP	external data
operational data	data flow	schema	functionality	internal data	data mart	information interdependence
interface	security	urgency	cost	unstructured data	data age	strategic
respond time	team skill	integration area	view of the data warehouse	usability	data storage component	direct cost
information delivery component	efficiency	data access tool	distributed	operational data quality	compatibility with existing systems	size
business requirement	technical issues	reliability	availability	development skill	computing budget	portability
information systems	vendor stability	labor usage	vendor reputation	vendor experience	transformation tool	development approach
vendor support	DW administration	warehouse engine	data model	project team	data staging component	data quality check
indirect cost	lifecylce of data	operational system	operational data store	source of sponsorship	query	database support

Initial Results

These features are categoried into six groups as the data warehouse architecture feature taxonomy.



Conclusion

This research conducts a systematical literature review for data warehouse architecture feature collection and classification. The contributions are summarised below:

- It organises and provids features for people who want to investigate or evaluate data warehouse architectures.
- Various features are provided, which would benefit the requirement collection.
- A taxonomy of DWHA feature will be porposed as a guideline for further evaluation (To-Do).

References

[1] Inmon, W.H., 1997. What is a data warehouse?. Prism Tech. Topic 1(1).

[2] Laney D, Warehouse factors to address for success, HP Professional 14 (5) (2000).

[3] Crane G. MacGregor-I. Meyer S. Chan, A. and 2001. WWW-based Accelerator Data Warehouse. Sass, R.

[4] Wrembel, R., 2011. On handling the evolution of external data sources in a data warehouse architecture. In Integrations of data warehousing, data mining and database technologies: innovative approaches (pp. 106-147). IGI Global.

[5] ANSI\/IEEE STD 829-1983, 1983. Standard for Software Test Documentation. Institute of Electrical and Electronics Engineers, New York.

[6] Cambridge Dictionary, (2017) Meaning of “feature” in the English Dictionary. Available at: https://en.oxforddictionaries.com/definition/feature [Accessed 18 May 2017].

[7] Okoli, C. and Schabram, K., 2010. A guide to conducting a systematic literature review of information systems research.

[8] Webster, J. and Watson, R.T., 2002. Analyzing the past to prepare for the future: Writing a literature review. MIS quarterly, pp.xiii-xxiii.

JOHANNES KEPLER
UNIVERSITY LINZ

DCU

Thanks For Your Attention!

UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA