Reviewing features of Data Warehouse Architectures: A Taxonomy

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

Staging Area

Data Storage

ETL

Data Flow

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.

What requirements I should consider?

I only care about components of the DWHA.
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.
   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.
247 features are abstracted, collected. Then, duplicated features are deleted and 148 unique features are summarised. Some of them are tabulated blow.

<table>
<thead>
<tr>
<th>operational data</th>
<th>ETL</th>
<th>management</th>
<th>cycle</th>
<th>OLAP</th>
<th>external data</th>
</tr>
</thead>
<tbody>
<tr>
<td>data flow</td>
<td>schema</td>
<td>functionality</td>
<td>internal data</td>
<td>data mart</td>
<td>information interdependence</td>
</tr>
<tr>
<td>interface</td>
<td>security</td>
<td>urgency</td>
<td>cost</td>
<td>unstructured data</td>
<td>data age</td>
</tr>
<tr>
<td>respond time</td>
<td>team skill</td>
<td>integration area</td>
<td>view of the data warehouse</td>
<td>usability</td>
<td>data storage component</td>
</tr>
<tr>
<td>information delivery component</td>
<td>efficiency</td>
<td>data access tool</td>
<td>distributed</td>
<td>operational data quality</td>
<td>compatibility with existing systems</td>
</tr>
<tr>
<td>business requirement</td>
<td>technical issues</td>
<td>reliability</td>
<td>availability</td>
<td>development skill</td>
<td>computing budget</td>
</tr>
<tr>
<td>information systems</td>
<td>vendor stability</td>
<td>labor usage</td>
<td>vendor reputation</td>
<td>vendor experience</td>
<td>transformation tool</td>
</tr>
<tr>
<td>vendor support</td>
<td>DW administration</td>
<td>warehouse engine</td>
<td>data model</td>
<td>project team</td>
<td>data staging component</td>
</tr>
<tr>
<td>indirect cost</td>
<td>lifecyle of data</td>
<td>operational system</td>
<td>operational data store</td>
<td>source of sponsorship</td>
<td>query</td>
</tr>
</tbody>
</table>
Initial Results

These features are categorized 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 provides 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 proposed as a guideline for further evaluation (To-Do).
References


Thanks For Your Attention!