Contents

[**Data Warehouse** 2](#_Toc131945112)

[Research-(Prastut) 2](#_Toc131945113)

[Cloud vs On Premise based Warehouse 2](#_Toc131945114)

[Architecture 3](#_Toc131945115)

[Alteryx 3](#_Toc131945116)

[MSSQL 3](#_Toc131945117)

# 

# **Data Warehouse**

## Research-(Prastut)

### Cloud vs On Premise based Warehouse

* Cloud has limited database security option therefore making it a risky environment for data storage.
* Cloud environment does not need a database administrator which exposes into more vulnerabilities.
* Cloud services are less customizable and may not cope with complex developments.
* On premise applications are to be more reliable, secure; offering complete ownership and control.

#### References

* Zhang, Q., Cheng, L. & Boutaba, R. Cloud computing: state-of-the-art and research challenges. *J Internet Serv Appl* **1**, 7–18 (2010). <https://doi.org/10.1007/s13174-010-0007-6>
* S. Bibi, D. Katsaros and P. Bozanis, "Business Application Acquisition: On-Premise or SaaS-Based Solutions?," in IEEE Software, vol. 29, no. 3, pp. 86-93, May-June 2012, doi: 10.1109/MS.2011.119.
* <https://www.cleo.com/blog/knowledge-base-on-premise-vs-cloud>
* https://www.xperience-group.com/news-item/cloud-vs-on-premise-software/

### Architecture

* For the architecture of the data warehouse consists of four main database storage systems, raw, staging, reporting and master database.
* The consumers will consume the data through staging or reporting with various data marts.
* The ETL tool will be used to create these databases from extracting the fit file data to loading the data into various data marts.

#### References

* <https://www.javatpoint.com/data-warehouse-architecture>
* W. Zhijuan, W. Hongchang and W. Xuefang, "A Data Warehouse Design Method," 2012 International Conference on Computer Science and Service System, Nanjing, China, 2012, pp. 2063-2066, doi: 10.1109/CSSS.2012.513.

### Alteryx

* Open-source data model and workflow with no coding hassle with drag and drop feature.
* An easy ETL process with very little chance of error.
* Handle large files proficiently with faster runtime.
* It has huge community support that will help to get expert answers.
* Can be connected to various databases either cloud or on-premises.
* Pentaho has been identified as a backup ETL tool incase if some of the features are not integrated through alteryx.

#### References

* <https://insightoriel.com/what-is-alteryx-advantages-of-alteryx/>
* <https://capitalizeconsulting.com/what-is-alteryx/#:~:text=Alteryx%20makes%20accessing%2C%20cleansing%2C%20testing,prone%2C%20and%20less%20risky%20way>.
* H. Singh, G. S. Matharu, A. K. Dardi, and J. S. Matharu, "Empirical Investigation of Big Data Analytical Tools: Comparative Analysis," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 1264-1269, doi: 10.1109/ICOEI.2019.8862739.

### MSSQL

* MS SQL Server has some useful built-in functionality to do analysis, automated reports, and machine learning with Python and R.
* SQL Server also offers good data protection and security.
* Snowflake schema will be used as a dimensional model as it is simple and most widely used.

#### References

* <https://blog.skyvia.com/sql-server-data-warehouse-the-easy-and-practical-guide/>
* <https://learn.microsoft.com/en-us/sql/relational-databases/data-collection/management-data-warehouse?view=sql-server-ver16>