

Development of Visualization System for Relational Database Health

Ekkarach Sunanta¹ and Pruet Boonma²

¹ Data Science Program, Chiang Mai University, Chiang Mai, Thailand

² Faculty of Engineering, Chiang Mai University, Chiang Mai, Thailand

Ekkarach_Sunanta@cmu.ac.th

Abstract. The objective of this independent study was to develop a system to summarize the health status of multiple relational databases using ETL (Extract Transforms Load) techniques which extract data from various sources, transform and load it into a master database. Then, the health status data is analyzed and presented by using the business intelligence tool. This system provides a convenient and timesaving platform for observing the availability and identifying anomalies of each database. Furthermore, this system allows the administrators to detect problems that can occur, such as disk space full, which can cause the system to stop and, consequently, damage the system's reliability.

The proposed system is evaluated in two aspects. First, the time spent making a summary of the health of the database to check abnormal from these databases is compared between the existing system and the proposed system. Second, the accuracy and completeness of the information provided by the proposed approach are compared with that of the current system. The evaluation found that the proposed system can meet the expected time saving, contains complete information as expected, and has verified that the data is correct with the fact data in each database source.

Keywords: Data Integration, Extract Transforms Load, Relational Database, Data Visualizations.

1 Introduction

Nowadays, software development in various companies develops applications or services with connecting through API (Application Program Interface) to create new software to meet the requirements of customers or the needs of each organization. For this reason, many companies have databases that store on multiple servers. Sometimes, the database servers have various platforms, such as the SQL Server database system on virtual machines and on-premises servers, SQL server databases in virtual servers on the cloud, and cloud databases such as Azure SQL Database. As different platforms are

used, it is challenging to verify availability or prevent problems. From that problem, this research proposes gathering the detailed data of each database and summarized into a visualized report on the database's health by using the ETL (Extract Transforms Load) technique [5].

2 The Proposed system

The proposed system is separated into three parts. The first part is preparing data from each data source. The second process is the data integration system to transform and load the data from prepare data source process to one destination database. The last part is using Microsoft Power BI to visualize the data. The overview of the proposed system is illustrated in Fig. 1.

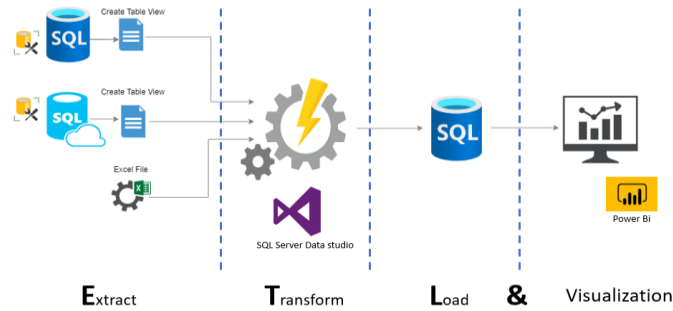


Fig. 1. The systems overview

2.1 Prepare Data and Destination Database Process

This step aims to get complete data as needed in the relational databases health definition [11][12] and take that data into the following process. The source data has the relational databases health definition, the number of databases that will be built into the system, and the source of the database includes the underlying data of that database. To meet the desired format and complete the definition of relational database health, the proposed system deploys a SQL view in each database to retrieve the details of each database. In addition, the proposed system prepares the master data of each database and the primary data of each database to use as the master data of each database. When all the data is ready, the tables used to store those data in the destination database will be designed according to the data format prepared at the beginning.

In particular, to create an SQL view at the source database, the view would differ between SQL Server and Azure SQL databases. For the database that is the SQL Server is Infrastructure as a service (IaaS), it is possible to get access to the system's Super Admin level. However, the Azure SQL Database is a Platform as a service (PaaS) which is impossible to access the super admin level, so it is required to view to get the necessary data. [3][8][10]

2.2 Develop Data Integration System Process

Once the source data has been prepared, the next step is creating a data integration system to retrieve data from the source database obtained from the data preparation process. Then, the system loads the data into the created destination database. The data integration system used SQL Server Integration Services (SSIS) in SQL Server Data Tools (SSDT) on Visual Studio. Next, it used SQL Server Management Studio (SSMS) to create stored procedures to transform data from temporary tables received from Integration Services into main tables. First, SSIS will separate database sources by control flow. Then, there is a data flow of each data source, consisting of data conversion and destination source. Finally, SQL Server Management Studio creates stored procedures to transform data from temporary tables to the main table and create job schedules for the execute SSIS Package.

2.3 Develop Data Visualization System Process

In this step, data visualization uses data from the master database and publish to web service. Microsoft Power BI is used to present visualized data from the dataset that was made before.

3 Result

The study results were divided into three parts in the Development of Visualization System for Relational Database Health. The first part is determined by the validity of the data extracted from the source database that load into the destination database it is correct according to realistic. The second part considers the data in the destination database, and the visualization system meets the database health definition. The last part considers the reduced time spent on staff work who maintains and summarizes the overview of the database in the organization that before there was this system and after there was the system.

Database Health Definition	Description
Sizing	Total database size must be not over than 80% of Disk Size or Attach Disk size
Last Backup / Have Backup	There should be regular backups. For very important databases such as customer databases
Last Access	How often is it used to check if it's still being accessed?
Status	Used to see if the database is online or not (there are some databases still online but not in used that wastes monthly expenses)
Cost	To see which databases can be to reduced or increased or optimized.
Critical Level	To show that each important level for each the database (Enterprise,Corporate,Product)
Type	To show what kind of database it is (Production , UAT, Development).

Table 1. Database Health Definition

Database Health Definition	Column In Destination View	Status
Sizing	Disksize,DiskLoc,AttachDisk , Rowsize, Logsize,	Completed
Last Backup / Have Backup	BackupSizeMB,LastFullDB,BackupDateBackupFileLocation	Completed
Last Access	LastAccess	Completed
Status	state_desc	Completed
Cost	CostUS	Completed
Critical Level	CustType , SVDBCATEGORY	Completed
Type	SVDBCATEGORY	Completed

Table 2. Database Health Definition compares with column in destination table.

For the evaluation of the reduction of working time of staff, this research experimented with making a summary data overview of the database without using the relational database health system. Found that it took about 13-15 minutes, but when using the Development of Visualization System for Relational Database Health, it does take about 10 seconds.

Process Type	Workflow	Working times(Min./All Database environment)										Avg.	Remark
		First round			Second round			Thirst round					
		AzureSQL	SQL Server	Total	AzureSQL	SQL Server	Total	AzureSQL	SQL Server	Total			
Traditional process	1. List master data from Azure web portal and record to excel	4.46	3.42	7.88	4.23	3.44	7.67	3.47	3.52	6.99	7.51	The timer will start record from opening the Azure portal web page and end on the data of all databases are saved in Excel already.	
	2. Execute view in each data and record to excel.	4.33	5.46	9.79	4.21	5.06	9.27	4.16	5.16	9.32	9.46	The timer will start record from start connect to each database and end on the data of all databases are saved in Excel already	
	3. Summarize report in Excel.	5.09	6.17	11.26	4.37	5.48	9.85	4.47	5.31	9.78	10.30	The timer will start record from start doing summary data in excel until save is done.	
	TOTAL Woking time	13.88	15.05	28.93	12.81	13.98	26.79	12.10	13.99	26.09	27.27	Total Avg. to used for Traditional process ~27 Min.	
Used The New System	Open system from URL link and waiting for system loading.	0.14			0.11			0.05			0.10	The timer will begin record time since open the URL link of system until system load already.	
								Total average difference			27.17	The new system can reduce working time average at 27.17 min	
Note :													
There are 5 database servers for SQL Server and 6 database servers for Azure SQL Database													
There are 128 databases for SQL Server and 7 databases for Azure SQL Database													

Table 3. Time results comparing the traditional operation with the operation after the system was established.

4 Conclusion

This independent research is the application of knowledge gained from graduate studies in data science. This research focuses on the data engineering approach to the organization's Development of Visualization System for Relational Database Health. To help as a tool to detect and monitor anomalies with relational databases in the organization and to help as a tool for considering resource management in the organization's database. In addition, the developed system also reduces the amount of time spent by staff in monitoring anomalies and summarizing the overall database within the organization. The results showed that Development of Visualization System for Relational Database Health was developed according to the objectives set is have accurate and complete information according to the definition of health of the database. Therefore, it can be used as a tool to help detect anomalies and a tool to help summarize the overview of all relational databases in the organization to consider the organization's database resource management. Additionally, it can reduce the work time of officers who must maintain the database and summarize an overview of the organization's database.

References

1. สารานุกรมไทยสำหรับเยาวชนฯ / เล่มที่ ๒๕ / เรื่องที่ ๒ ระบบฐานข้อมูล / การรวมกันของข้อมูล (Data Integration). <https://www.saranukromthai.or.th/>
2. Run Integration Services (SSIS) Packages. <https://docs.microsoft.com/en-us/sql/integration-services/packages/sql-server-agent-jobs-for-packages?view=sql-server-ver15#schedule>
3. T-SQL differences between SQL Server and Azure SQL Database., <https://docs.microsoft.com/en-us/azure/azure-sql/database/transact-sql-tsql-differences-sql-server>
4. พงษ์พันธ์ ศิวลิข. SQL Server 2014 ฉบับ DBA มืออาชีพ กรุงเทพฯ, บริษัท ซีอีดี เอ็ดดูเคชั่น จำกัด (มหาชน), 2559.
5. ETL. Open wide to ETL. <https://sites.google.com/site/ectlearning/home/etl-extract-transform-and-load>
6. ภาษา SQL (Structured Query Language) เบื้องต้น. last accessed 2021/02/03, <https://beprogrammer.tripod.com/WEBprogrammer/sql/sql2.html>
7. รูปแบบฐานข้อมูลเชิงสัมพันธ์. last accessed 2021/02/03, <http://elearning.psru.ac.th/courses/256/pdf>
8. There is no MSDB (system) database with Azure SQL Database offered as PaaS. <https://social.msdn.microsoft.com/Forums/azure/en-US/ce416e83-28ff-4712-b323-729776c32e2b/is-there-an-msdb-database-in-azure-sql?forum=ssdsgetstarted>
9. ณัฐวุฒิ สิริกุลสัมพันธ์. Extraction, Transformation and Loading Data to Data Warehouse in Employee, Student, Research and Curriculum Subject Areas of Chiang Mai University, 2552
10. Azure SQL Database Backup retention. <https://docs.microsoft.com/en-us/azure/azure-sql/database/automated-backups-overview?tabs=single-database>
11. Md. Harun Ar Rashid. DATABASE FAILURE | CAUSES OF DATABASE FAILURE. <https://limbd.org/database-failure-causes-of-database-failure/>
12. Cloud Secure Tech. What Are the Causes of Database Loss? <https://www.cloudsecuretech.com/what-are-the-causes-of-database-loss/>
13. SSIS Connection Managers: OLE DB vs ODBC vs ADO.NET, <https://www.sqlshack.com/ssis-connection-managers-ole-db-vs-odbc-vs-ado-net/>