ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017
An International Conference on Recent Trends in IT Innovations - Tec'afe 2017
Organized by
Dept. of Computer Science, Garden City University, Bangalore-560049, India

Real Time Database Migration to Microsoft Azure Cloud

Vikash Kumar, Keerthana N

Assistant Professor, Dept. of Computer Science & IT, JAIN University, Bangalore, India

Student, Dept. of Master of Computer Applications, JAIN University, Bangalore, India

ABSTRACT: The Windows Azure is a cloud computing platform that provides users to store large amount of data with efficient storage. The infrastructure is created by the Microsoft Corporation and it manage applications and its products through a global network known as Microsoft Azure. One of the important advantage of using azure is that the data which is stored on azure is present for any duration of time and it can be easily accessible to user at any time and from any place. The Structured Query Language (SQL) is having some limitations as the data is not retrieved efficiently. SQL has an issue, on its flexibility. This paper purposes to migrate the SQL database to Azure cloud which uses NoSQL structure for storing the data. In azure cloud the data which is migrated, is stored in normalized form. Based on some experimental results, our mechanism of migration will help the current application system which uses SQL to improve its performance and maintains the security. If we do normalization on-premises, then we can reduce required computation time while migrating.

KEYWORDS: Cloud Storage, Windows Azure, SQL, Not Only SQL (NoSQL), Data migration.

I. INTRODUCTION

Windows Azure Cloud is a cloud computing platform which is provided by Microsoft Corporation. This is used in many application areas which includes customer managing system, retailers of shops, social networking sites and so on. The advantage of using azure is that it gives the customers very large amount of space to store their data and they also manage that data.

As we are using NoSQL, it provides many advantages over the SQL such as it does not provide tabular relationship of database rather than this they provide document which contains Key-Value pair. The data is efficiently accessed and it also represented in normalized form. The SQL support strong join operation after the table is normalized, when we are migrating it to NoSQL the relationship are often captured by de-normalizing data and presenting all data for an object in a single record, which gives high performance reducing any complexity that exist

As many industries are using SQL only it is efficient to them to use the Microsoft azure cloud storage which can give high performance. [4]

II. LITERATURE SURVEY

The conversion of schema mode to SQL Database to NoSQL is presented by the Gansen Zhao, Qiaoying Lin, Libo Li, Zijing Li in their research paper. This paper focuses on the schema conversion of SQL Database to the NoSQL. This paper gives the comparison of query performance between Nested solution and the Non-Nest solution. As we all know, if the SQL query statement involves multi-table join, the query need to access more than one table in the original relational database. The advantage of this paper is the correctness of the conversion algorithm has been conducted a rigorous proof. They also have stated that the future work about the subject is to minimize the spatial redundancy [1].

Another paper is about the schema denormalization and migration by using the study of various content management system. Content Management System are able to let people who have no technical skill to manage the website, rapidly create, edit or publish online content. So the paper purpose is that as the exploring growth of huge amount of data, the well-structured characteristics of SQL may limit to handle scalability of the data so the use of migration from SQL to NoSQL.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017
An International Conference on Recent Trends in IT Innovations - Tec'afe 2017
Organized by
Dept. of Computer Science, Garden City University, Bangalore-560049, India

In this paper the flow about the conversion is described with various steps. Advantage of this paper is that the migration useful for various Content Management System is described thoroughly [2]. Windows Azure Storage, a highly available cloud storage service with strong consistency, is another paper by Brad Calder, Ju Wang, Aaron ogus, Niranjan Nilakantan,

ArildSkjolsvold, Sam Mckelvie. The paper focuses on the Windows Azure storage and its various description. Currently, WAS storage coming in the form of Blobs (Files), Tables (structured storage), and Queues (message delivery). In this paper, they describe the WAS architecture, global namespace, and data model, as well as its resource provisioning, load balancing, and the replication systems [3].

III. PROPOSED SYSTEM

In our system we are accepting existing relational database, we need to normalize if necessary and identify the table joins and move the data to azure tables in denormalized form. So that the data redundancy can be eliminated. Provide a command line utility which can accept the old queries (SQL) and it gives the same result with data in Azure cloud.

Example: consider a simplified scenario that there are following tables in a business database-
Customer (cust_id, name, address)
Item (item_id, description, price)
Purchases (purchase_id, cust_id, item_id, date, amount)
When we move the purchases table to azure table in denormalized way it would probably look like the following
DenormalizedPurchase (purchase_id, cust_id, item_id, cust_name, cust_address, item_desciption, item_price, purchase_date, purchase_amount)
There would be a single denormalized azure table for all the purchases containing each record. [4]

Relational SQL- based database systems can perform queries to alter a table or database, or run update a number of rows at a time. NoSQL databases are schema-less. Thus a migration script is essential in performing bulk of operations.

Before doing and running anything it is advised to take backup of your data. If the script will run multiple times, it is a good idea to wrap it in an easy to use command-line-interface(CLI). Writing the script in such a way that it can run repeatedly. The best migration scripts can figure out whether a database has already been transformed or modified and will skip it in a subsequent run.

A. System Architecture

System consist of four main components Data Migrator, Query Parser, Normalization, Table Join. System takes the input as relational database and processes it and then it is converted into azure storage format i.e. the NoSQL.

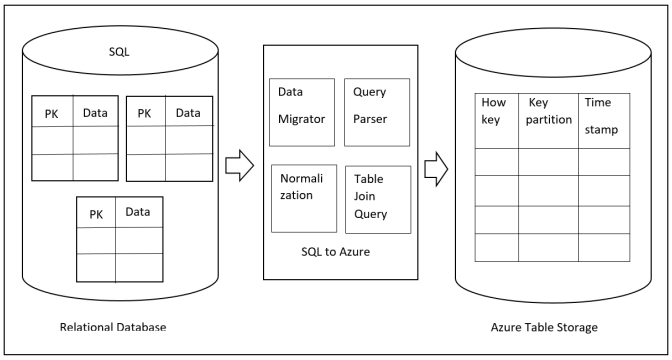

Fig: Relational Database Migration to Azure table

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017
An International Conference on Recent Trends in IT Innovations - Tec'afe 2017
Organized by
Dept. of Computer Science, Garden City University, Bangalore-560049, India

B. There are four modules in system

Data migrator: Data migrator it is used to move the data from SQL to the azure table storage. Data which is migrated from SQL to Azure table storage is in normalization form. Data in the SQL database is in the structured format and when data is migrated from SQL to Azure table storage it is in the structured NoSQL data in the cloud.

Query parser: To accept the existing SQL queries and dynamically translate to azure data fetch logic. Query Parser module provide a command line utility which can accept the old SQL queries and it gives the same results with data in the Azure cloud.

Normalization: Normalization module is use to reduce the data redundancy and improve data integrity. Normalization is the process of efficiently organizing data in a database.

Table join: Table join module identify the table joins in relational database. Tables and relations in entities to store entities in NoSQL databases.

List of Normal Forms

a. UNF – "Unnormalized Form"
b. 1NF – First Normal Form
c. 2NF –Second Normal Form
d. 3NF –Third Normal Form
e. EKNF – Elementary Key Normal Form
f. BCNF – Boyce-Codd Normal Form
g. 4NF –Fourth Normal Form
h. 5NF –Fifth Normal Form

Unnormalized Form – An unnormalized relation is a relation that contains the repeating values. An unnormalized relation can also contain relations nested within other relations, as well as all kinds of transitive dependencies.

First Normal Form - First Normal Form (1NF or Minimal Form) is a normal form used in database normalization. A relational database table that has 1NF is one that meets a certain minimum set of criteria. These criteria are basically concerned with ensuring that the table is a faithful representation of a relation and it is free of repeating groups. A table I in 1NF if and only if all the key attributes have been defined and it contains no repeating groups. A relation R is in 1NF if and only if it has only single-valued attributes (atomic values).

Second Normal Form – 2NF is the Second normal form used in relational databases. A table is second normal form (2NF) if and only if it is in the 1NF and every non-key attributes fully functionally dependent on the whole of the primary key (i.e. there are no partial dependencies). A non-prime attribute is one that does not belong to any candidate key.

Third Normal Form – 3NF is the Third normal form used in relational database normalization. According to the Codd's definition, a table is said to be in 3NF, if and only if, that table is in the second normal form (2NF), and every attribute in the table that do not belong to a candidate key, should directly depend on every candidate key of that table. The tables that comply with the 3NF generally do not contain anomalies that occur when inserting, deleting or updating records in the table.
Elementary Key Normal Form – EKNF is an enhancement of 3NF. The tables are also in third normal form. This happens when there is more than one unique composite key and they overlap. Such cases can cause redundant information in the overlapping columns.

BOYCE CODD Normal Form (BCNF) -BCNF is a normal form used in database normalization. It is a slightly stronger version of the third normal form (3NF). BCNF was developed in 1974 by Raymond F. Boyce and Edgar F. A

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017
An International Conference on Recent Trends in IT Innovations - Tec'afe 2017
Organized by
Dept. of Computer Science, Garden City University, Bangalore-560049, India

relation is in BCNF, if and only if, every determinant is a candidate key. A table complies with BCNF if it is in third normal form and for every functional dependency X -> Y, X should be the super key of the table.

Fourth Normal Form – The tables are said to be in fourth normal form when the table meets the criteria for third normal form. Situations where non-key attributes depend on the key column exclusive of other non-key columns are eliminated. The Fourth normal form removes all multi-valued dependencies.

Fifth Normal Form – A relation R is in 5NF also called projection-join normal form (PJ/NF) if and only if every join dependency in R is implied by the candidate keys of R. The tables are said to be in fifth normal form when the table meets the criteria for fourth normal form. The table consists of a key attribute and a non-key attribute only. [12]

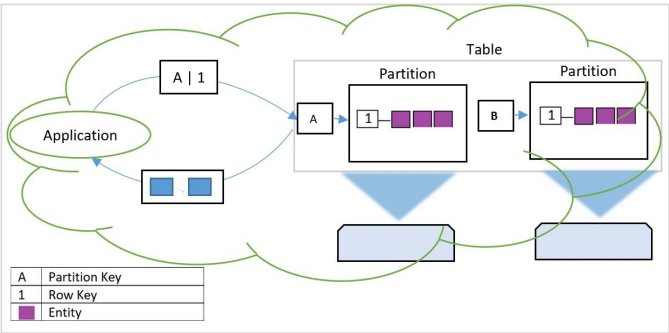

Fig : Windows Azure Storage Architecture

C. Storing and Retrieval of Key/Value Pair

The details will vary in different technologies, as how the components are named, but the basic idea of a key/value store remains the same. An application gives the store a unique key, from that we get back one or more values associated with that key. In the Windows Azure Table, let's consider an example, data is held in tables despite of their names nothing are like relational tables. Each table consists of one or more partitions. Each partition has some number of entities with each entity containing properties.

Each property has a type, such as integer or character string or any other, and each one holds a single value. There are two properties, one property is in each entity is designated as the partition key and it contains the same value for all entities in a particular partition. Second property is in each entity is designated as the row key, and it contains a value that's unique within its partition. Unlike a relational database (but like other key/value stores), Windows Azure Tables has no notion of schema. Each entity in a partition can contain different kinds of properties this makes the most sense for an application.

To retrieve an entity, an application provides both the partition key and the row key for that entity. What comes back is the entity this key pair identifies, includes some or all of its properties. It is also possible to do other kinds of queries, such as requesting properties whose keys fall within a specifies range of values. Windows Azure Tables has no support for the secondary indexes, however, which is typical of key/value stores. [11]

D. How to store data in Relational Database

Relational Databases use tables to store the information. The standard fields and record are represented as columns(fields) and rows(records) in table. With a relational database, you can quickly compare information because of the arrangement of data in columns. The relational database model takes advantage of this uniformity to build completely new tables out of required information from existing tables.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017
An International Conference on Recent Trends in IT Innovations - Tec'afe 2017
Organized by
Dept. of Computer Science, Garden City University, Bangalore-560049, India

E. How to store data in Azure Table Storage

Azure Table storage it is the service that stores structured NoSQL data in the cloud. Table storage is a key/attribute store with a schemaless design. Because the Table storage is schemaless, it is easy to adapt your data as the needs of your application evolve. Access to the data is fast and cost-effective for all kinds of the applications. Table storage is typically significantly lower in cost than the traditional SQL for the similar volumes of data.

You can use Table storage to store the flexible datasets, such as user data for the web applications, address book, device information, and any other type of metadata that your services. You van store any number of entities in a table, and a storage account contain any number of tables, up to the capacity limit of the storage account.

F. NoSQL Database here we are using Key Value Store database model and it consists of two features Redis and Riak

Table 1

Features	Redis	Riak
Query Language	API Calls	HTTP, JavaScript, REST, Erlang
Operating System	Linux, *NIX, Windows, Mac OSX	Cross-Platform
Programming language	C, C++	Erlang

G. Difference between SQL database table and Azure storage table

Table 2

SQL Database Table	Azure Storage Table
SQL database tables are designed to store the same type of data in each and every row.	Azure tables are capable of storing one type of data in a row and another type of data in an another row.
In relational database, a primary key is used to identify a unique row.	A unique row is identified by the combination pair of Partition key and Row key.
In relational database, we can create number of secondary indexes.	In azure tables we can create only index based on the Partition key and Row key properties.

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017
An International Conference on Recent Trends in IT Innovations - Tec'afe 2017
Organized by
Dept. of Computer Science, Garden City University, Bangalore-560049, India

H. Database Migration Toolkits

Table 3: Comparison Table

S. NO	Name	From	To
1	OSDM Toolkit	Oracle, Sybase, Informix, DB2, MS Access, MS SQL	PostgreSQL and MySQL
2	DB Migration	Oracle and MS SQL	PostgreSQL and MySQL
3	Mssql2 Pgsql	MS SQL	PostgreSQL
4	MySQL Migration Toolkit	MS Access and Oracle	MySQL
5	MySQL Migration Toolkit	MS Access, MS SQL, Dbase and Oracle	MySQL
6	Open DBcopy	Any RDB*	Any RDB*
7	Progression DB	MS SQL	PostgreSQL, MySQL and Ingres
8	Shift2Ingres	Oracle and DB2	Ingres
9	SQL Porter	Oracle, MS SQL DB2 and Sybase	MySQL
10	SQL Ways	All Relational Databases	PostgreSQL and MySQL
11	SwisSQL Data Migration Tool	Oracle, DB2, MS SQL, Sybase and MaxDB	MySQL
12	SwisSQL SQLOne Console	Oracle, DB2, MS SQL, Sybase and Informix	PostgreSQL and MySQL
13	MapForce	SQL Server, DB2, MS Access, MySQL and PostgreSQL	SQL Server, DB2, MS Access and Oracle
14	Centerprise Data Integrator	SQL Server, DB2, MS Access, MySQL and PostgreSQL	SQL Server, DB2, MS Access, MySQL and PostgreSQL
15	DBConvert	Oracle, DB2, SQLite, MS Access, MySQL and PostgreSQL and Foxpro	Oracle, DB2, SQLite, MS Access, MySQL and PostgreSQL and Foxpro

IV. CONCLUSION & FUTURE SCOPE

For the developers of cloud based applications the Windows Azure storage platform provides the essential services. It provides strong consistency, global partitioned namespace, and disaster recovery for large scale applications. In this paper, we propose SQL to Azure, intermediate system which will accept the relational database and we migrate those

ISSN(Online) : 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering
An ISO 3297: 2007 Certified Organization Vol.5, Special Issue 2, April 2017
An International Conference on Recent Trends in IT Innovations - Tec'afe 2017
Organized by
Dept. of Computer Science, Garden City University, Bangalore-560049, India

data to azure table storage which is of NoSQL for cloud based applications in normalized form. In future real time data migration from one cloud vendor to another can be explored.

REFERENCES

[1] Gansen Zhao QiaoyingLin Libo Li Zijing Li: Schema Conversion Model of SQL Database to NoSQL, Ninth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (2014).
[2] Chao-Hsien Lee, Yu-Lin Zheng:SQL-to-NoSQL Schema Denormalization and Migration: A Study on Content Management Systems, IEEE International Conference on Systems (2015).
[3] Brad Calder, Ju Wang, Aaron Ogus, Niranjan Nilakantan, ArildSkjolsvold, Sam McKelvie, YikangXu, Shashwat Srivastav, Jiesheng Wu, HuseyinSimitci, JaidevHaridas, Chakravarthy Uddaraju, Hemal Khatri, Andrew Edwards, Vaman Bedekar, Shane Mainali, Rafay Abbasi, Arpit Agarwal, Mian Fahimul Haq, Muhammad Ikramul Haq, Deepali Bhardwaj, Sowmya Dayanand, Anitha Adusumilli, Marvin McNett, Sriram Sankaran, Kavitha Manivannan, Leonidas Rigas :Windows Azure Storage: A Highly Available Cloud Storage Service with Strong Consistency,ACM Transactions on Computer Systems,(2011).
[4] Omkar Dhamale ,Rushikesh Birajdar ,Harshal Wadkar , Rishabh Singh ,S. G. Malvi , S. G. Patil, "Migration of relational Database to Microsoft Azure Cloud"
[5] DongzhaoLiang, Yunzhen Lin, Guiguang Ding: Mid-model Design Used in Model Transition and Data Migration between Relational Databases and NoSQL Databases, IEEE International Conference, (2015).
[6] Erik Meijer and Gavin Bierman, Microsoft Research: A co-Relational Model of Data for Large Shared Data Banks, ACM Transactions, (2011).
[7] Eric A. Brewer, "Towards Robust Distributed Systems. (Invited Talk)," in Principles of Distributed Computing, Portland, Oregon, (2000).
[8] F. Chang et al., "Bigtable: A Distributed Storage System for Structured Data," in OSDI, 2006.
[9] S. Lombardo, E. Di Nitto, and D. Ardagna, "Issues in Handling Complex Data Structures with NoSQL Databases," Proceedings of the 14 International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC),2012.
[10] https://docs.microsoft.com/en-us/azure/documentdb/documentdb-nosql- vs-sql
[11] David Chappell, "Understanding NoSQL Technologies on Windows Azure".
[12] Ashwini S Diwakar, "Database Management Systems".
[13] Bisbal, J.; Lawless, D.; Wu, B.; and Grimson, J. (1999). Legacy Information System Migration: A Brief Review of Problems, Solutions and Research Issue. IEEE Software, 6(5), 1-18.
[14] Feroz Alam, "DATA MIGRATION: RELATIONAL RDBMS TO NON-RELATIONAL NOSQL"
[15] Andreas Meier et al., (1994). 'Hierarchical to Relational Database Migration', IEEE Software, Vol. 11 Issue 3, pp 21-27