

**Migrating MongoDB databases to Azure CosmosDB**

Project Report for Azure Deep Dive



February 10, 2018

Zhenglin Hu

**Migrating MongoDB databases to Azure CosmosDB**

(Executive Summary)

**Project Topic: Azure Database Migration Service**

**Project Topic Description**: Typically is focused on the migration of database schema and data from one database format to a specific database technology in the cloud.

**Problem Description**: We have a few MongoDB instances on-premise or on-virtual machines in the cloud. We would like to take advantage of Azure CosmosDB cloud service for DR, replication, scalability, and high availability. There is a need here to migrate of MongoDB databases to Azure CosmosDB database service.

**Sample Data:**

1. Northwind csv files from <https://github.com/tmcnab/northwind-mongo/archive/master.zip>. Northwind sample database is from Microsoft Access product. There are 11 tables with 3308 rows.
2. mongodb\_cases.json from <https://community.jaspersoft.com/wiki/mongodb-city-cases-example>. There are 10,000 json documents in this dataset. It is about cases in a city.

**Hardware and Software Used:**

1. Windows 7 and MongoDB Enterprise Server 3.6.1.

**High-Level Steps**

1. Install MongoDB, clean the data and load sample data into databases
2. Backup the databases in MongoDB which are ready for loading to CosmosDB
3. Create a CosmosDB account on portal.azure.com
4. Determine restore parameters for batch size and number of workers.
5. Restore the backups to CosmosDB using MongoDB API. Pre-create CosmosDB databases and collections if the default throughput 1000 RU/s is too slow.

**Issues and Lessons learned**

1. Uploading JSON documents to CosmosDB is slow. For a simple document, need about 0.072s which is equivalent of 14 documents/s. For a million documents, the rough estimate is that we need about one day.
2. We could increase the throughput above 10,000 RU/s to spead up the upload, but we need to implement sharding. When the collection size is larger than 10GB, sharding is also required. As JSON documents are schema free and no required constraint for a partition key as required by sharding. Thus, the application needs to ensure that there are candidate partition keys in the document.

**YourTube URLs**

2 min Video: 15 min Video:

**Migrating MongoDB databases to Azure CosmosDB**

**Project Topic: Azure Database Migration Service**

Project Topic Description: Typically is focused on the migration of database schema and data from one database format to a specific database technology in the cloud.

**Problem Description**: We have a few MongoDB instances on-premise or on-virtual machines in the cloud. We would like to take advantage of Azure CosmosDB cloud service for DR, replication, scalability, and high availability. There is a need here to migrate of MongoDB databases to Azure CosmosDB database service.

**Sample Data:**

1. Northwind csv files from <https://github.com/tmcnab/northwind-mongo/archive/master.zip>

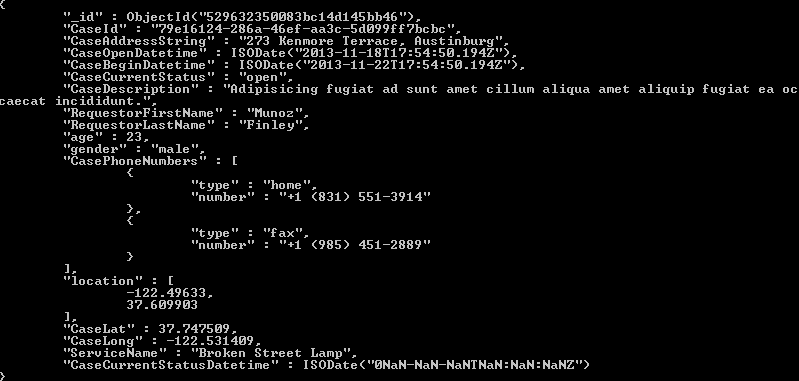
Northwind sample database is from Microsoft Access product. There are 11 tables with 3308 rows. Here are the details:

|  |  |  |
| --- | --- | --- |
| Table Name | Number of Rows | CSV File Name |
| categories | 8 | categories.csv |
| customers | 91 | customers.csv |
| employee-territories | 49 | employee-territories.csv |
| employees | 9 | employees.csv |
| order-details | 2155 | order-details.csv |
| orders | 830 | orders.csv |
| products | 77 | products.csv |
| regions | 4 | regions.csv |
| shippers | 3 | shippers.csv |
| suppliers | 29 | suppliers.csv |
| territories | 53 | territories.csv |

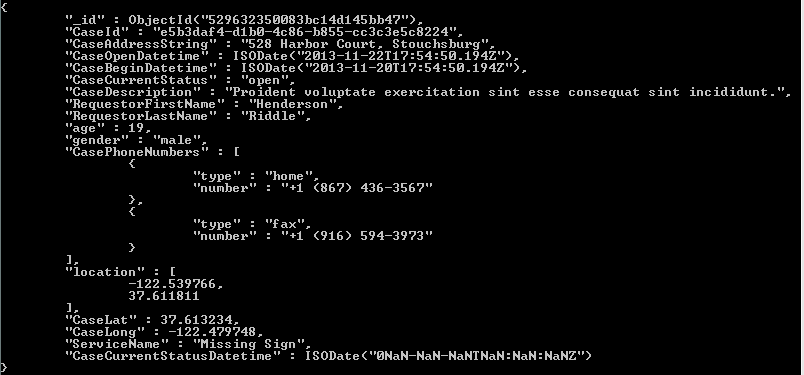
1. mongodb\_cases.json from <https://community.jaspersoft.com/wiki/mongodb-city-cases-example>

There are 10,000 json documents in this dataset. It is about service cases handled in a city. Here are some examples:

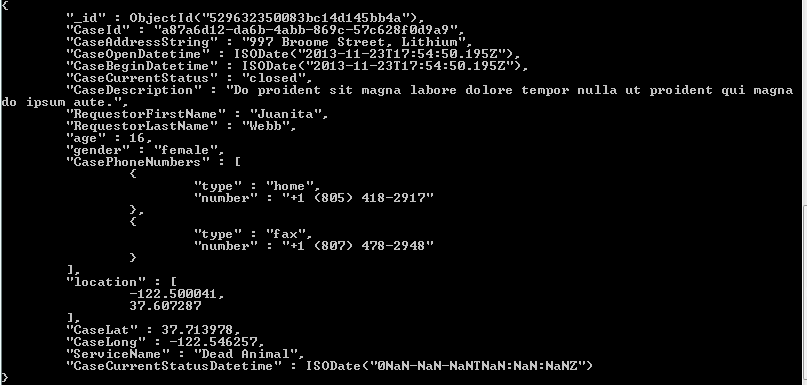
Broken Street Lamp:



Missing Sign



Dead Animal:



**Hardware and Software Used:**

1. Windows 7
2. MongoDB Enterprise Server download: https://www.mongodb.com/lp/download/mongodb-enterprise?jmp=nav

mongodb-win32-x86\_64-enterprise-windows-64-3.6.1-signed.msi

**Prepare the source MongoDB system**

1. Installation
   1. run mongodb msi file and install mongoDB. Created default database folder: C:\data\db
   2. Modify environment variable path to add the binary path for MongoDB

C:\Program Files\MongoDB\Server\3.6\bin;C:\Program Files\Microsoft VS Code\bin;C:\Users\elqtrlx\AppData\Roaming\npm;C:\MinGW\msys\1.0\bin;C:\Program Files\Java\jdk1.8.0\_151\bin

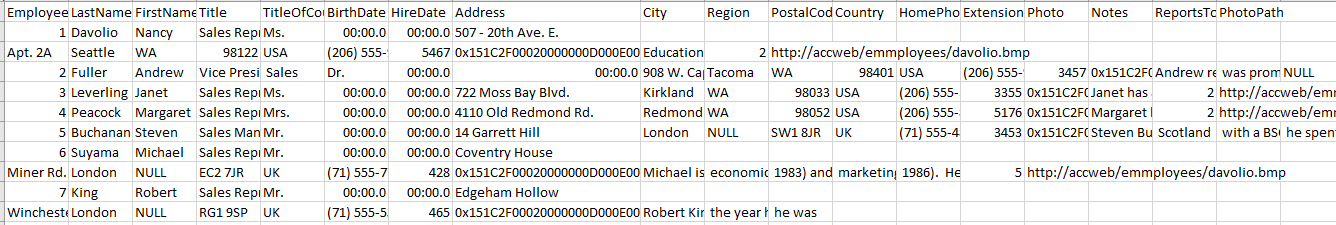
* 1. running mongodb

mongod to start the database to run from localhost with default port of 27017.

mongo to connect to the database as a client

1. Data cleansing

Address and Note fields contain coma in the employees CSV file. Create problems for loading into MongoDB. Added double quote for the field to resolve the loading issue.



mongoimport -d Northwind -c employees --type csv --file employees.csv --headerline

2018-01-29T15:45:00.213-0600 connected to: localhost

2018-01-29T15:45:00.216-0600 Failed: read error on entry #2: line 3, column 404: bare " in non-qu

oted-field

2018-01-29T15:45:00.216-0600 imported 1 document

1. Data load

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>run\_import.bat

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c categories --type csv --file categories.csv --headerline

2018-01-29T16:55:38.614-0600 connected to: localhost

2018-01-29T16:55:38.643-0600 imported 8 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c customers --type csv --file customers.csv --headerline

2018-01-29T16:55:38.872-0600 connected to: localhost

2018-01-29T16:55:38.910-0600 imported 91 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c employeeterritories --type csv --file employee-territories.csv --headerline

2018-01-29T16:55:39.119-0600 connected to: localhost

2018-01-29T16:55:39.151-0600 imported 49 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c employees --type csv --file employees.csv --headerline

2018-01-29T16:55:39.350-0600 connected to: localhost

2018-01-29T16:55:39.382-0600 imported 9 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c orderdetails --type csv --file order-details.csv --headerline

2018-01-29T16:55:39.608-0600 connected to: localhost

2018-01-29T16:55:39.676-0600 imported 2155 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c orders --type csv --file orders.csv --headerline

2018-01-29T16:55:39.911-0600 connected to: localhost

2018-01-29T16:55:39.964-0600 imported 830 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c products --type csv --file products.csv --headerline

2018-01-29T16:55:40.181-0600 connected to: localhost

2018-01-29T16:55:40.217-0600 imported 77 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c regions --type csv --file regions.csv --headerline

2018-01-29T16:55:40.430-0600 connected to: localhost

2018-01-29T16:55:40.468-0600 imported 4 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c shippers --type csv --file shippers.csv --headerline

2018-01-29T16:55:40.717-0600 connected to: localhost

2018-01-29T16:55:40.746-0600 imported 3 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c suppliers --type csv --file suppliers.csv --headerline

2018-01-29T16:55:40.966-0600 connected to: localhost

2018-01-29T16:55:40.998-0600 imported 29 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongoimport -d northwind -c territories --type csv --file territories.csv --headerline

2018-01-29T16:55:41.219-0600 connected to: localhost

2018-01-29T16:55:41.258-0600 imported 53 documents

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17\northwind-mongo-master\northwind-mongo-master>mongo

MongoDB Enterprise > use northwind

switched to db northwind

MongoDB Enterprise > show collections

categories

customers

employees

employeeterritories

orderdetails

orders

products

regions

shippers

suppliers

territories

MongoDB Enterprise > db.categories.count()

8

MongoDB Enterprise > db.customers.count()

91

MongoDB Enterprise > db.employees.count()

9

MongoDB Enterprise > db.employeeterritories.count()

49

MongoDB Enterprise > db.orderdetails.count()

2155

MongoDB Enterprise > db.orders.count()

830

MongoDB Enterprise > db.products.count()

77

MongoDB Enterprise > db.regions.count()

4

MongoDB Enterprise > db.shippers.count()

3

MongoDB Enterprise > db.suppliers.count()

29

MongoDB Enterprise > db.territories.count()

53

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>mongoimport --host localhost --db test --collection cases < mongodb\_cases.json

2018-01-29T17:52:38.994-0600 connected to: localhost

2018-01-29T17:52:39.566-0600 imported 10000 documents

MongoDB Enterprise > use test

switched to db test

MongoDB Enterprise > show collections

cases

MongoDB Enterprise > db.cases.find().count()

10000

MongoDB Enterprise >

**Migration Data to CosmosDB**

1. Take dumps from source system for databases: northwind and test

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>mongodump --db northwind --out northwind\_dump

2018-01-31T12:52:17.829-0600 writing northwind.orderdetails to

2018-01-31T12:52:17.830-0600 writing northwind.orders to

2018-01-31T12:52:17.830-0600 writing northwind.customers to

2018-01-31T12:52:17.830-0600 writing northwind.products to

2018-01-31T12:52:17.833-0600 done dumping northwind.customers (91 documents)

2018-01-31T12:52:17.833-0600 writing northwind.territories to

2018-01-31T12:52:17.840-0600 done dumping northwind.orders (830 documents)

2018-01-31T12:52:17.841-0600 writing northwind.employeeterritories to

2018-01-31T12:52:17.842-0600 done dumping northwind.orderdetails (2155 documents)

2018-01-31T12:52:17.842-0600 writing northwind.suppliers to

2018-01-31T12:52:17.843-0600 done dumping northwind.products (77 documents)

2018-01-31T12:52:17.844-0600 writing northwind.employees to

2018-01-31T12:52:17.844-0600 done dumping northwind.territories (53 documents)

2018-01-31T12:52:17.844-0600 writing northwind.categories to

2018-01-31T12:52:17.846-0600 done dumping northwind.employeeterritories (49 documents)

2018-01-31T12:52:17.846-0600 writing northwind.regions to

2018-01-31T12:52:17.846-0600 done dumping northwind.suppliers (29 documents)

2018-01-31T12:52:17.847-0600 writing northwind.shippers to

2018-01-31T12:52:17.848-0600 done dumping northwind.categories (8 documents)

2018-01-31T12:52:17.849-0600 done dumping northwind.employees (9 documents)

2018-01-31T12:52:17.849-0600 done dumping northwind.regions (4 documents)

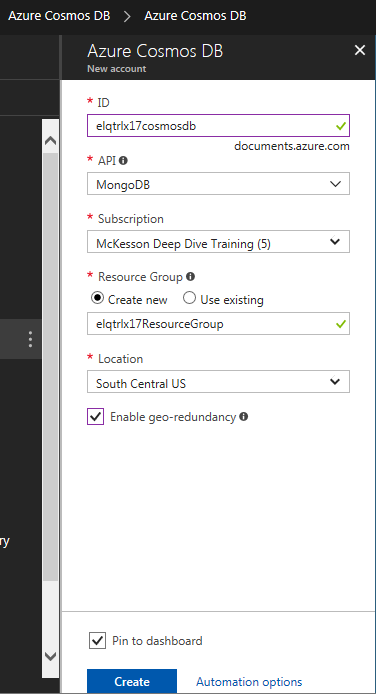
2018-01-31T12:52:17.850-0600 done dumping northwind.shippers (3 documents)

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>mongodump --db test --out test\_dump

2018-01-31T12:51:51.244-0600 writing test.cases to

2018-01-31T12:51:51.340-0600 done dumping test.cases (10000 documents)

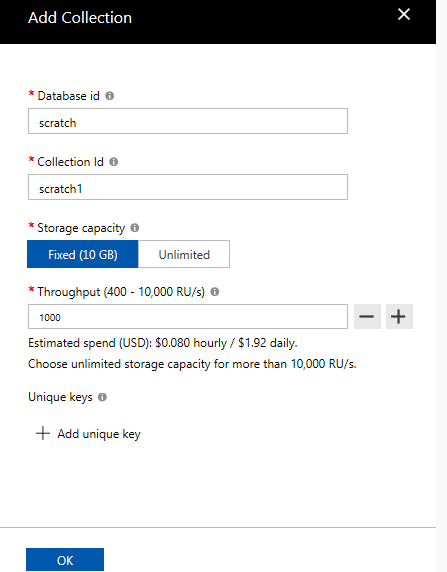
1. Create an Azure account for CosmosDB



1. Determine data restore parameters
   1. connect to cosmosdb via mongodb api (mongo\_shell\_cosmosdb.bat)

mongo --username <USERNAME for cosmosdb account> --password <PRIMARY PASSWORD> --host <USERNAME>.documents.azure.com --port 10255 --ssl

* 1. create a scratch database and scratch1 collection



* 1. run a simple insert into scratch database, scratch1 collection

globaldb:PRIMARY> use scratch

switched to db scratch

globaldb:PRIMARY> show collections

scratch1

globaldb:PRIMARY> db.scratch1.insert({ "playerId": "a067ff", "hashedid": "bb0091", "countryCode": "h

k" })

WriteResult({ "nInserted" : 1 })

* 1. get stats about Request Charge

globaldb:PRIMARY> db.runCommand({getLastRequestStatistics: 1})

{

"\_t" : "GetRequestStatisticsResponse",

"ok" : 1,

"CommandName" : "insert",

"RequestCharge" : 8.76,

"RequestDurationInMilliSeconds" : NumberLong(17)

}

3.5 Determine the latency from my machine to the Azure CosmosDB cloud service

globaldb:PRIMARY> setVerboseShell(true)

globaldb:PRIMARY> db.scratch1.find().limit(1)

{ "\_id" : ObjectId("5a721f72dab943d858c1a5cd"), "playerId" : "a067ff", "hashedid" : "bb0091", "count

ryCode" : "hk" }

Fetched 1 record(s) in 72ms

This is equivalent to 1/(0.001\*72)=1000/72=13.89 RU/s

* 1. determine batchsize and numInsertWorkders

Provisioned Throughput: 10000 RU/s.

BatchSize: 10000/13.89 = 720. Use 24 as batchsize per <https://docs.microsoft.com/en-us/azure/cosmos-db/mongodb-migrate>

Property Value

batchSize 24

RUs provisioned 10000

Latency 0.072 s

RU charged for 1 doc write 8.76 RUs

numInsertionWorkers (10000 RUs x 0.072 s) / (24 x 8.76 RUs) = 3.425 (use 4)

1. Upload data to CosmosDB

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>run\_restore.northwind.bat

2018-01-31T15:28:33.342-0600 preparing collections to restore from

2018-01-31T15:28:33.415-0600 reading metadata for northwind.orders from northwind\_dump\northwind\orders.metadata.json

2018-01-31T15:28:33.482-0600 reading metadata for northwind.orderdetails from northwind\_dump\northwind\orderdetails.metadata.json

2018-01-31T15:28:33.549-0600 reading metadata for northwind.customers from northwind\_dump\northwind\customers.metadata.json

2018-01-31T15:28:33.613-0600 reading metadata for northwind.products from northwind\_dump\northwind\products.metadata.json

2018-01-31T15:28:33.678-0600 reading metadata for northwind.suppliers from northwind\_dump\northwind\suppliers.metadata.json

2018-01-31T15:28:33.744-0600 reading metadata for northwind.employees from northwind\_dump\northwind\employees.metadata.json

2018-01-31T15:28:33.825-0600 reading metadata for northwind.territories from northwind\_dump\northwind\territories.metadata.json

2018-01-31T15:28:33.890-0600 reading metadata for northwind.categories from northwind\_dump\northwind\categories.metadata.json

2018-01-31T15:28:33.956-0600 reading metadata for northwind.employeeterritories from northwind\_dump\northwind\employeeterritories.metadata.json

2018-01-31T15:28:34.023-0600 reading metadata for northwind.shippers from northwind\_dump\northwind\shippers.metadata.json

2018-01-31T15:28:34.089-0600 reading metadata for northwind.regions from northwind\_dump\northwind\regions.metadata.json

2018-01-31T15:28:35.085-0600 restoring northwind.orderdetails from northwind\_dump\northwind\orderdetails.bson

2018-01-31T15:28:35.143-0600 restoring northwind.customers from northwind\_dump\northwind\customers.bson

2018-01-31T15:28:35.180-0600 [########################] northwind.orderdetails 207KB/207KB (100.0%)

2018-01-31T15:28:35.181-0600 [########################] northwind.customers 26.8KB/26.8KB (100.0%)

2018-01-31T15:28:35.181-0600

2018-01-31T15:28:35.254-0600 restoring northwind.suppliers from northwind\_dump\northwind\suppliers.bson

2018-01-31T15:28:35.258-0600 restoring northwind.regions from northwind\_dump\northwind\regions.bson

2018-01-31T15:28:35.331-0600 restoring northwind.products from northwind\_dump\northwind\products.bson

2018-01-31T15:28:35.378-0600 no indexes to restore

2018-01-31T15:28:35.379-0600 finished restoring northwind.regions (4 documents)

2018-01-31T15:28:35.403-0600 restoring northwind.employees from northwind\_dump\northwind\employees.bson

2018-01-31T15:28:35.428-0600 restoring northwind.categories from northwind\_dump\northwind\categories.bson

2018-01-31T15:28:35.463-0600 restoring northwind.territories from northwind\_dump\northwind\territories.bson

2018-01-31T15:28:35.581-0600 no indexes to restore

2018-01-31T15:28:35.581-0600 finished restoring northwind.employees (9 documents)

2018-01-31T15:28:35.598-0600 restoring northwind.shippers from northwind\_dump\northwind\shippers.bson

2018-01-31T15:28:35.642-0600 restoring northwind.orders from northwind\_dump\northwind\orders.bson

2018-01-31T15:28:35.716-0600 no indexes to restore

2018-01-31T15:28:35.717-0600 finished restoring northwind.territories (53 documents)

2018-01-31T15:28:35.737-0600 [########################] northwind.customers 26.8KB/26.8KB (100.0%)

2018-01-31T15:28:35.737-0600 no indexes to restore

2018-01-31T15:28:35.738-0600 finished restoring northwind.customers (91 documents)

2018-01-31T15:28:35.753-0600 no indexes to restore

2018-01-31T15:28:35.753-0600 finished restoring northwind.shippers (3 documents)

2018-01-31T15:28:35.845-0600 no indexes to restore

2018-01-31T15:28:35.846-0600 finished restoring northwind.suppliers (29 documents)

2018-01-31T15:28:35.908-0600 no indexes to restore

2018-01-31T15:28:35.908-0600 finished restoring northwind.products (77 documents)

2018-01-31T15:28:35.921-0600 no indexes to restore

2018-01-31T15:28:35.921-0600 finished restoring northwind.categories (8 documents)

2018-01-31T15:28:36.171-0600 restoring northwind.employeeterritories from northwind\_dump\northwind\employeeterritories.bson

2018-01-31T15:28:36.444-0600 no indexes to restore

2018-01-31T15:28:36.445-0600 finished restoring northwind.employeeterritories (49 documents)

2018-01-31T15:28:38.180-0600 [########################] northwind.orderdetails 207KB/207KB (100.0%)

2018-01-31T15:28:38.180-0600 [########################] northwind.orders 313KB/313KB (100.0%)

2018-01-31T15:28:38.180-0600

2018-01-31T15:28:41.180-0600 [########################] northwind.orderdetails 207KB/207KB (100.0%)

2018-01-31T15:28:41.180-0600 [########################] northwind.orders 313KB/313KB (100.0%)

2018-01-31T15:28:41.181-0600

2018-01-31T15:28:43.197-0600 [########################] northwind.orders 313KB/313KB (100.0%)

2018-01-31T15:28:43.198-0600 no indexes to restore

2018-01-31T15:28:43.198-0600 finished restoring northwind.orders (830 documents)

2018-01-31T15:28:44.180-0600 [########################] northwind.orderdetails 207KB/207KB (100.0%)

2018-01-31T15:28:46.744-0600 [########################] northwind.orderdetails 207KB/207KB (100.0%)

2018-01-31T15:28:46.745-0600 no indexes to restore

2018-01-31T15:28:46.745-0600 finished restoring northwind.orderdetails (2155 documents)

2018-01-31T15:28:46.745-0600 done

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>mongo\_shell\_cosmosdb.bat <northwind\_col\_count.js

MongoDB shell version v3.6.1

connecting to: mongodb://elqtrlx17cosmosdb.documents.azure.com:10255/

MongoDB server version: 3.2.0

WARNING: shell and server versions do not match

switched to db northwind

categories

customers

employees

employeeterritories

orderdetails

orders

products

regions

shippers

suppliers

territories

categories

NumberLong(8)

customers

NumberLong(91)

employees

NumberLong(9)

employeeterritories

NumberLong(49)

orderdetails

NumberLong(2155)

orders

NumberLong(830)

products

NumberLong(77)

regions

NumberLong(4)

shippers

NumberLong(3)

suppliers

NumberLong(29)

territories

NumberLong(53)

bye

Pre-create db test and collection cases on portal with throughput 10,000 RU/s

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>run\_restore.test.bat

Using --numInsertionWorkersPerCollection 4 --numParallelCollections 24

2018-01-31T14:25:44.555-0600 preparing collections to restore from

2018-01-31T14:25:44.625-0600 reading metadata for test.cases from test\_dump\test\cases.metadata.json

2018-01-31T14:25:44.626-0600 restoring test.cases from test\_dump\test\cases.bson

2018-01-31T14:25:46.672-0600 [#########...............] test.cases 2.43MB/6.07MB (40.0%)

2018-01-31T14:25:49.672-0600 [###################.....] test.cases 4.86MB/6.07MB (80.0%)

2018-01-31T14:25:52.672-0600 [###################.....] test.cases 4.86MB/6.07MB (80.0%)

2018-01-31T14:25:55.672-0600 [###################.....] test.cases 4.86MB/6.07MB (80.0%)

2018-01-31T14:25:58.672-0600 [#####################...] test.cases 5.48MB/6.07MB (90.2%)

2018-01-31T14:26:01.672-0600 [########################] test.cases 6.07MB/6.07MB (100.0%)

2018-01-31T14:26:04.524-0600 error: Message: The request rate is too large. Please retry after so

metime. ActivityId: f174511b-0000-0000-0000-000000000000, Request URI: /apps/97eda9fb-4d31-43ec-816e-e341bfd2b031/services/c31c35b8-bd10-40c7-a371-0a2ce3bf2c61/partitions/cb85e3cb-9d40-4c43-a842-293eb1fb8c32/replicas/131619034262500996p, RequestStats: , SDK: Microsoft.Azure.Documents.Common/1.19.162.2

2018-01-31T14:26:04.672-0600 [########################] test.cases 6.07MB/6.07MB (100.0%)

2018-01-31T14:26:06.462-0600 [########################] test.cases 6.07MB/6.07MB (100.0%)

2018-01-31T14:26:06.462-0600 restoring indexes for collection test.cases from metadata

2018-01-31T14:26:06.716-0600 finished restoring test.cases (10000 documents)

2018-01-31T14:26:06.716-0600 done

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>mongo\_shell\_cosmosdb.bat

MongoDB shell version v3.6.1

connecting to: mongodb://elqtrlx17cosmosdb.documents.azure.com:10255/

MongoDB server version: 3.2.0

WARNING: shell and server versions do not match

globaldb:PRIMARY> use test

switched to db test

globaldb:PRIMARY> show collections

cases

globaldb:PRIMARY> db.cases.find().count()

NumberLong(9000)

globaldb:PRIMARY>

(1000 documents were not loaded)

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>run\_restore.test.bat

Using --numInsertionWorkersPerCollection 1 --numParallelCollections 6

2018-01-31T16:11:39.125-0600 preparing collections to restore from

2018-01-31T16:11:39.195-0600 reading metadata for test.cases from test\_dump\test\cases.metadata.json

2018-01-31T16:11:39.196-0600 restoring test.cases from test\_dump\test\cases.bson

2018-01-31T16:11:41.359-0600 [##......................] test.cases 624KB/6.07MB (10.0%)

2018-01-31T16:11:44.359-0600 [####....................] test.cases 1.21MB/6.07MB (20.0%)

2018-01-31T16:11:47.359-0600 [#######.................] test.cases 1.83MB/6.07MB (30.1%)

2018-01-31T16:11:50.359-0600 [###########.............] test.cases 3.03MB/6.07MB (50.0%)

2018-01-31T16:11:53.359-0600 [##############..........] test.cases 3.65MB/6.07MB (60.1%)

2018-01-31T16:11:56.359-0600 [################........] test.cases 4.25MB/6.07MB (69.9%)

2018-01-31T16:11:59.359-0600 [###################.....] test.cases 4.86MB/6.07MB (80.0%)

2018-01-31T16:12:02.359-0600 [########################] test.cases 6.07MB/6.07MB (100.0%)

2018-01-31T16:12:04.712-0600 [########################] test.cases 6.07MB/6.07MB (100.0%)

2018-01-31T16:12:04.712-0600 restoring indexes for collection test.cases from metadata

2018-01-31T16:12:04.964-0600 finished restoring test.cases (10000 documents)

2018-01-31T16:12:04.965-0600 done

(need significantly lower the parallel workers and batchsize)

C:\Users\elqtrlx\Documents\Azure-Deep-Dive\Lesson 17>mongo\_shell\_cosmosdb.bat

MongoDB shell version v3.6.1

connecting to: mongodb://elqtrlx17cosmosdb.documents.azure.com:10255/

MongoDB server version: 3.2.0

WARNING: shell and server versions do not match

globaldb:PRIMARY> use test

switched to db test

globaldb:PRIMARY> show collections

cases

globaldb:PRIMARY> db.cases.find().count()

NumberLong(10000)

globaldb:PRIMARY>

**Issues and Lessons learned**

1. Uploading JSON documents to CosmosDB is slow. For a simple document, need about 0.072s which is equivalent of 14 documents/s. For a million documents, the rough estimate is that we need about one day.
2. We could increase the throughput above 10,000 RU/s to spead up the upload, but we need to implement sharding. When the collection size is larger than 10GB, sharding is also required. As JSON documents are schema free and no required constraint for a partition key as required by sharding. Thus, the application needs to ensure that there are candidate partition keys in the document.

**References**

1. <https://docs.microsoft.com/en-us/azure/cosmos-db/mongodb-migrate>
2. <https://docs.mongodb.com/manual/administration/>

**YourTube URLs**

2 min Video: <https://youtu.be/9vDRBgPeT50> 15 min Video: https://youtu.be/KHFuEFOMIH4

**GitHub Repository**: https://github.com/linnhu3210/week17