Cloud Application Development Continual Assessment (2021)

Cloud data storage and visualisation: Microsoft vs Amazon – A comparison

Colm Gallagher

Kerry, Ireland

Abstract

Cloud services have become increasingly popular among businesses over the past decade or more. The opportunity to save costs on expensive in-house servers and physical drives is extremely appealing and so the bigger technology companies such as, Amazon, Microsoft, Google, and others have put vast amounts of time and money into being the market leaders in the battle for cloud superiority. As seen in the past year alone, more and more companies are having to move their services to the cloud as employees are increasingly working from home, whether through personal choice or the global pandemic.

Data storage and data visualisation are the two areas within the cloud realm that will be focussed on throughout this paper. Specifically, database warehouses and the ability to connect the data from them into other cloud services in which they can be visualised to give users a more asthetically pleasing way of receiving and understanding the data. The purpose of this paper is to simply compare the attributes of Amazon’s DynamoDB and AWS Quicksight against the features of Microsoft’s Azure Cosmos DB and Power BI. The study will conclude with the personal recommendations of the author.

*Keywords:* Microsoft; Azure; Cosmos DB; Power BI; Amazon; AWS; DynamoDB; S3; Quicksight; Cloud; Visualisation; Data

1. Introduction

As part of the final semester of a four-year *Computing with Media Development* honours degree, this author was set the task of producing a written document to supplement the technical aspect of research into a specific area of cloud application services. There are numerous areas with which one could focus on including, but not limited to, the following.

* Developing on the Cloud
* Cloud Data Storage
* Administration and Security
* Cloud Management Tools
* Networking and Content Delivery

The topic chosen as a preference for this paper was the field of *Cloud Data Storage* and, as an extension, the ability to produce visualisations from the said data. I have very much enjoyed the data analytics aspect of the degree course and have a keen interest in the way that media, in particular, use the form of visuals to enhance the experience of presenting data to their audience.

Upon discussing this with my tutors, it was agreed that it would be of interest to compare rival companies’ usage of these mediums. Amazon’s AWS cloud services include DynamoDB for storing databases, and Quicksight to produce data visualisations. Microsoft’s alternative cloud services to those mentioned would be Azure Cosmos DB, and Power BI.

1. Literature Review
   1. Cloud Database Storage

Amazon’s DynamoDB was first released in 2012 as a schema-free, noSQL database supporting such programming languages as.

* .Net
* Python
* Ruby
* Java
* PHP

Microsoft’s Azure Cosmos DB was released two-years later, in 2014. It also shares the similar noSQL, schema-free concept but supports some alternate languages.

* C#
* JavaScript
* Node.js
* Java

The well-known website, db-engines.com, ranks DynamoDB at number 16, nine places higher than Cosmos DB at number 25, in their current database rankings (as of May 2021) (DB-Engines Ranking, 2021).

* 1. Data Visualisation

Researching some comparison websites online, capterra.com shows that Power BI holds a 78% approval rating over Quicksight’s 62.5% (Amazon QuickSight vs Microsoft Power BI 2021 - Feature and Pricing Comparison on Capterra, 2021). (Sharma, 2018) writes on analyticsindiamag.com that both are industry standard business intelligence tools and have their advantages and disadvantages. Quicksight being the simpler to use of the two, Power BI containing stronger features.

* 1. Architecture

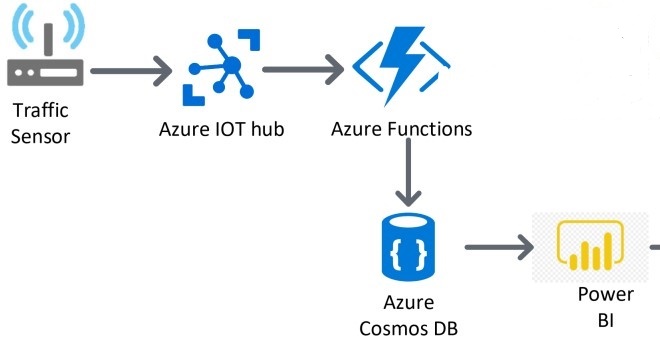


Figure 1 - Microsoft Cloud Services architecture

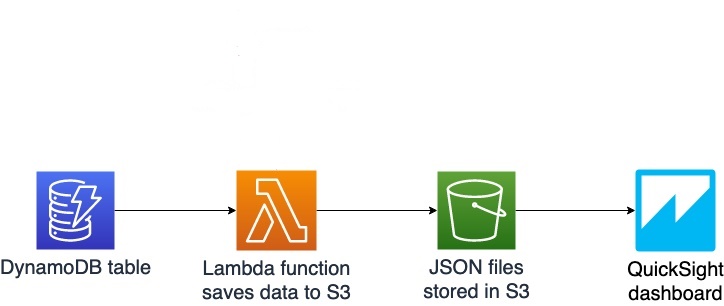


Figure 2 - Amazon Cloud Services architecture

1. Methodology

The Microsoft products were the first to be tested by the author. Barring a brief, and uncomfortable, project conducted in Azure Dev Ops, I had not used any of the other Azure services before attempting this task. I was less daunted by the prospect of using Cosmos DB though, as I have enjoyed other database related modules. A very short amount of time was also spent working in Microsoft Power BI, so I was not going into this completely cold.

Although feeling a little bit of pressure over time constraints, it was decided that the most practical approach was to follow an online tutorial (Visualize Azure Cosmos DB data by using the Power BI connector, 2019) with sample data before moving on with the data collected by myself. The prerequisites for connecting the two of these programs was to download the very latest version of Power BI (which would appear to be updated on a very regular basis), create an Azure Cosmos DB account, and to retrieve the sample data from a GitHub repository (azure-cosmos-db-sample-data/VolcanoData.json at main · Azure-Samples/azure-cosmos-db-sample-data, 2021).

There were some issues with deploying the Azure account at first as an error message stated that there was high demand in the original region I had selected. This culminated in me somehow losing my free tier account which took longer to resolve than I had wished. Fortunately, things became to move quite swiftly from that point onwards and I found a tool which would help me upload my csv data file to Cosmos DB with relative ease. Once again, I had to download this from a GitHub repository (Azure/azure-documentdb-datamigrationtool, 2021), and the simply named, Azure Cosmos DB Data Migration Tool (Figure 1), was also found to be extremely simple to use. A very clean interface with clearly designated selection boxes, I honestly could not fault anything about this clever tool. Once I had familiarised myself adequately with Cosmos DB, this cloud service also seemed to be straightforward to navigate around.

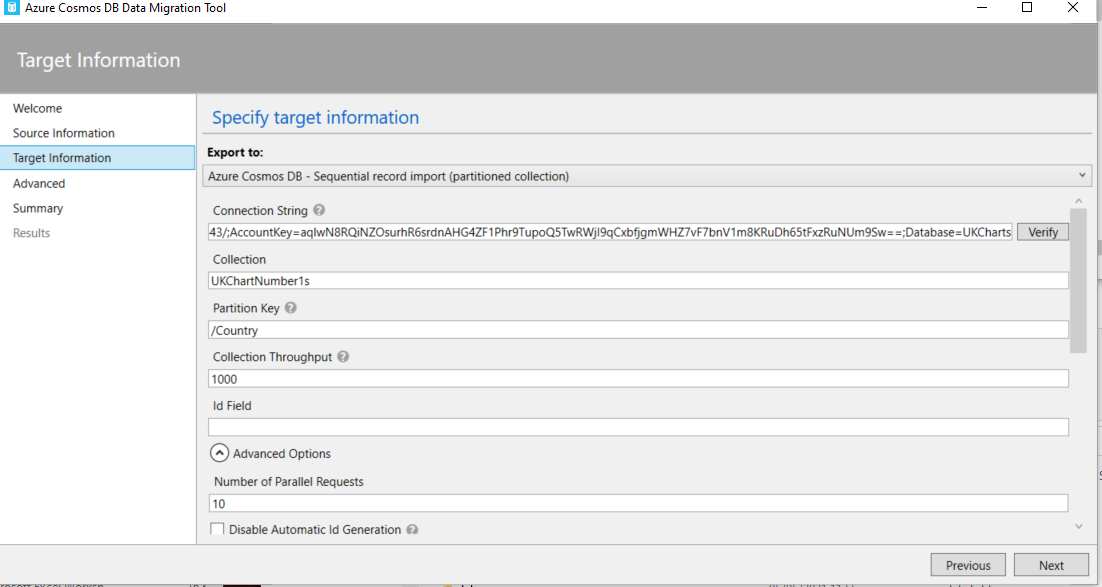


Figure 3 - User Interface of Data Migration Tool.

With my csv file up in the cloud and displaying no obvious issues, it was time to switch over to Microsoft’s business intelligence program, Power BI. As stated earlier, I had a little previous experience using this tool, but had not actually pulled any data from a database prior to this project. As with the other two programs used, it was pleasing to note the simplicity of the procedure. After clicking on just three tabs, all that needed to be done was to enter some security keys found in Azure Cosmos DB and the data was entered into Power BI.

As with any new program it takes a bit of time to get used to the layout and functionality. The most difficult feature of Power BI, it could be argued, is the transformation of data. I wanted the year column to be solely focused on just the year which meant using a delimiter tool to remove days and months attributes (Figure 2). There were also many null values which needed to be removed to ensure the accuracy of the data moving forwards into the visualization section. Also deleted were any duplicate entries which would have skewed the data.

Table

Description automatically generated

Figure 4 - Year column before (a) and after (b) using delimiter removal tool.

Similarly, to the data transformation, the task of visualising the data takes some time to master and there were many errors at first in relation to correct formats being used and in which order fields were to be displayed. But once the initial confusion passed, it became clear that Power BI can be an impressive program producing engaging and informative visual interpretations of the data it is provided with.

The main dashboards and reports are configured in the desktop version of Power BI before being pushed to the cloud version by publishing the report on completion (DMD (PowerBI - CA) (Colm Gallagher T00170514) (200421) - Power BI, 2021) [[1]](#endnote-1).

Next, it was the turn of the Amazon cloud services to be tested. The AWS alternative to Microsoft’s Cosmos DB is DynamoDB, also a noSQL database system. There would appear to be a plethora of tutorials and documentation on how to use AWS tools, including those provided by our tutors, which meant that it was easy to grasp the process of creating a table in DynamoDB. Just name the table, and then provide both a partition key and a sort key to help with the filtering of searches once inside the database.

Uploading the same csv file as used in the Microsoft tests though, was seemingly where this task became a whole lot more challenging. Unlike the connection between Azure’s data migration tool and Cosmos DB, trying to find a way to upload the csv file to DynamoDB was, at least for this author, increasingly a cause of frustration. An S3 bucket was created to store the csv file, which was fine, but finding a solution to transfer the data into DynamoDB was causing a major dilemma. The abundance of AWS documentation would appear to have had a negative impact in this situation as seemingly too many options to achieve the end aim were available, meaning many routes were explored only to culminate in a dead end on each occasion. Each of the following were attempted unsuccessfully to varying degrees.

* Redshift cluster
* AWS Lambda
* Zapier integration
* Panoply integration
* Stitch integration

After a considerable amount of time spent on trying to find a solution, the decision was made to manually enter some data into the DynamoDB table so that the next step of the project could be achieved for the purpose of making this report as accurate as possible in order to retain its integrity. Finally, an account was created in Quicksight to test the functionality of the program in comparison to Power BI. Ironically, I found that csv files can be directly uploaded to Quicksight, but that is irrelevant to this paper as the aim was always to test the database cloud services anyway. As time was against me, I did not spend as long as I would have hoped for using Quicksight and so feel that it would be unfair to review it either too positively or negatively at this stage.

1. Findings

The major finding of this process was the contrast in the ability to connect data between the respective companies’ services. Power BI’s ability to pull data directly from Azure Cosmos DB proved to be such a trump card in the research.

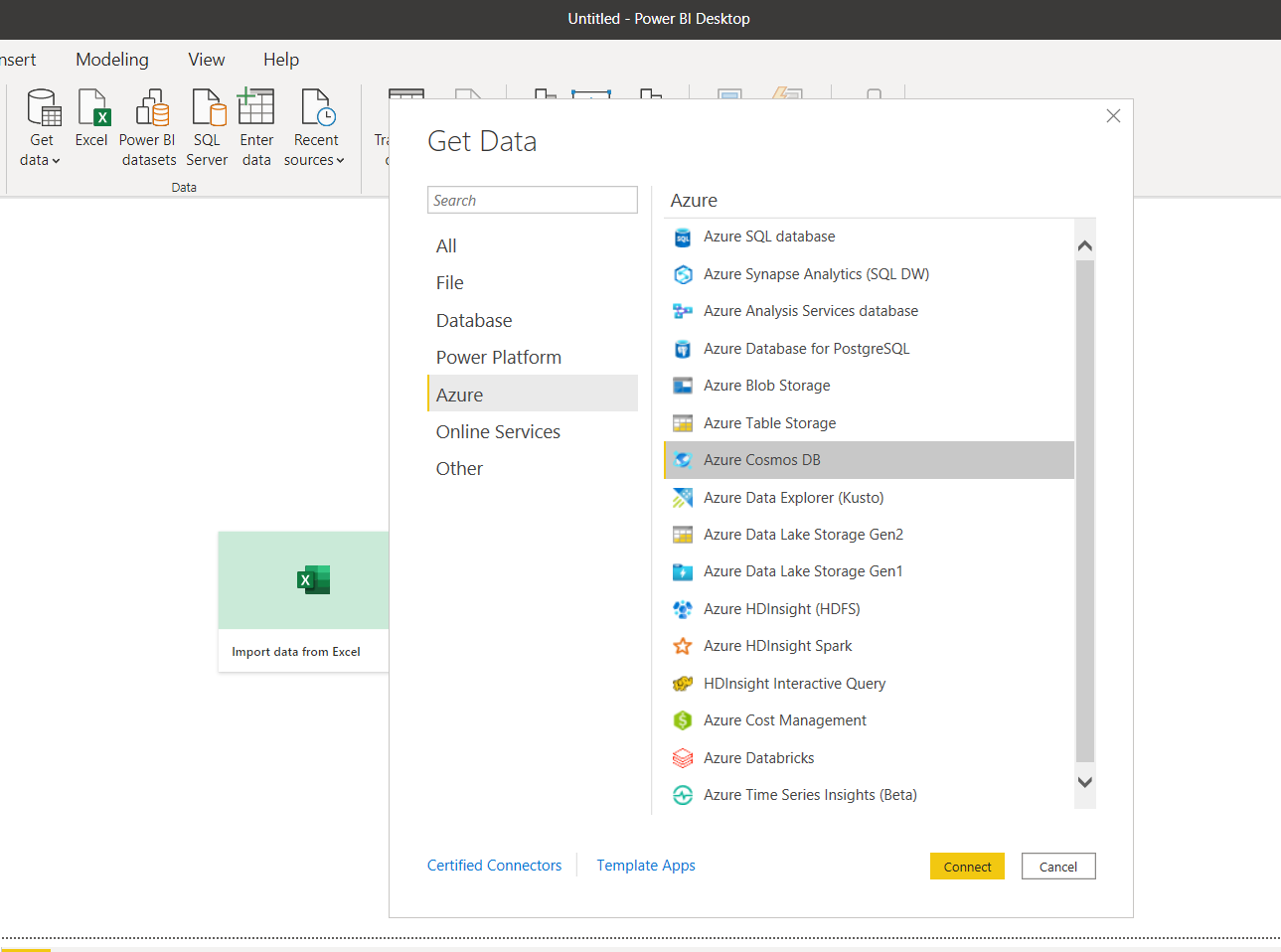


Figure 5 - Power BI connecting directly to Azure Cosmos DB

This proved impossible to do for this author (probably due to their own misgivings) and instead of pulling the data from AWS cloud services, it was eventually uploaded from the desktop.

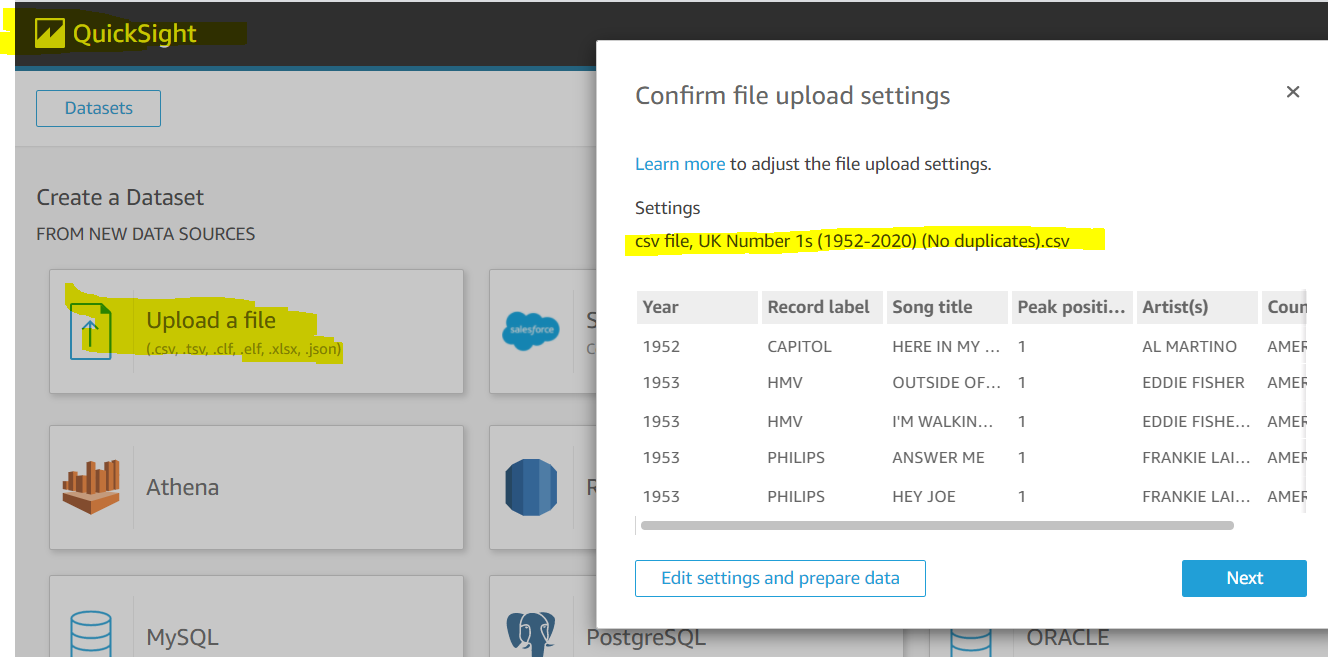


Figure 6 - An unsuccessful attempt at loading data into Quicksight from Amazon S3

Another feature with which Power BI appeared to eclipse Quicksight was with the usage of visualising map data, as seen below. Power BI produced a user friendly visual whilst Quicksight seemed to struggle with interpreting dual country attributes.

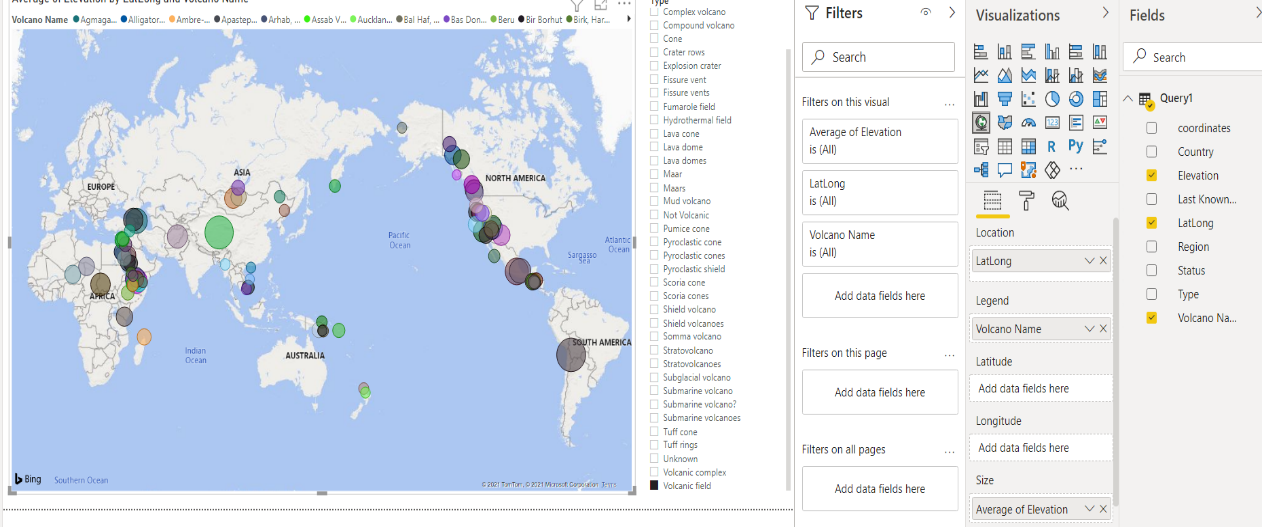


Figure 7 - Power BI's map function works well.

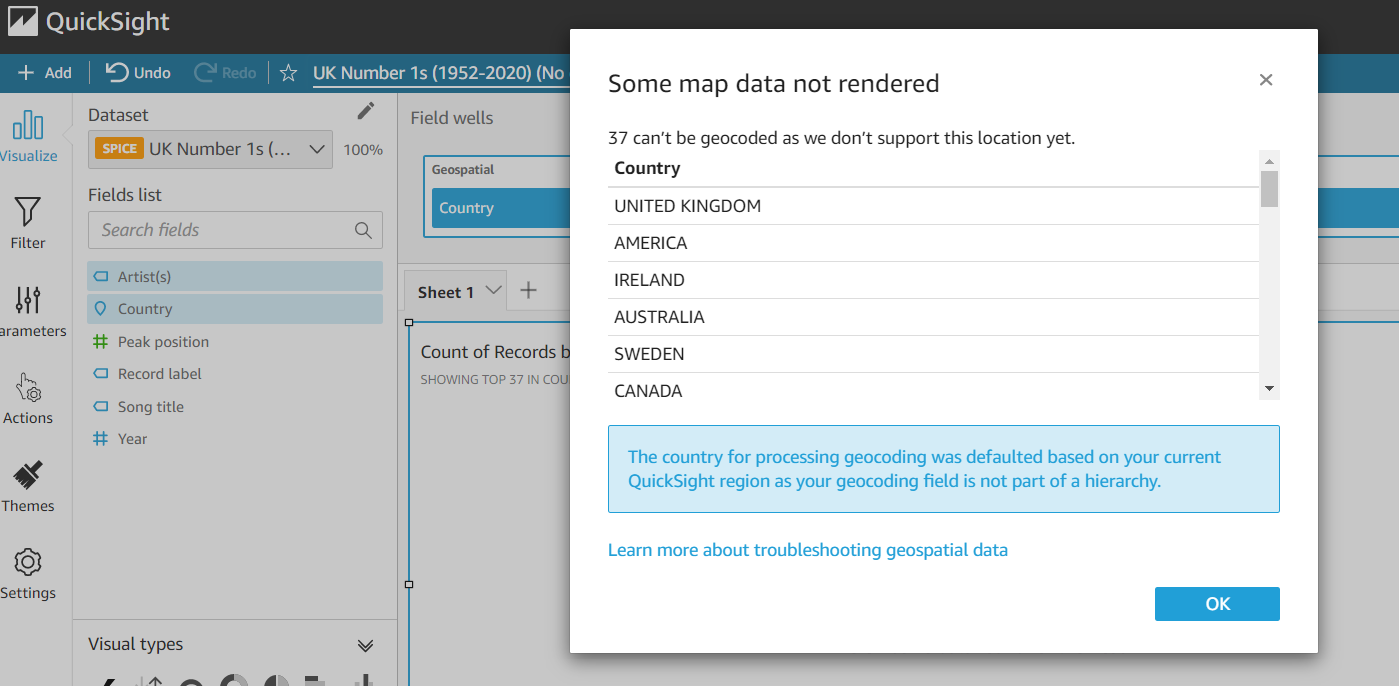


Figure 8 - Quicksight's map function appears not to support dual country attributes.

1. Discussion

As much as I would like this report to be as objective as possible, I must disclose that I had a very high preference for one of the contenders over the other, and that is the Microsoft suite of products. But I would definitely caveat this by saying that I feel my opinion is skewed by the fact that, even though miniscule in the grander scheme, I did have *some* previous experience with using both Azure and Power BI. This meant that the two programs certainly had an advantage in the sense that I had a certain level of knowledge and usability with regards to their functions, particularly in Power BI’s case. Therefore, I do have to take this into account when attempting to give a fair evaluation of the tests.

I found that Microsoft’s products just worked far better for the purpose of this paper, and I only set out to compare the specific task of pulling data from a cloud-based database into a data visualization program. The ease of how this was able to be executed between Cosmos DB and Power BI was significant in comparison to the Amazon services. And the simple data migration tool must take much of the credit for this. I was able to upload the csv, pull it into Power BI, and produce some impressive data visualisations within a matter of a few hours.

Conversely, I spent a couple of days trying to achieve the same result in the Amazon sphere but seemed to be blocked on every occasion in which I found myself, or so I believed, on the verge of making a breakthrough. It was early on that I figured the need to use an S3 bucket to store the csv file, and there was no sense of panic at this point. But from there onwards, I felt like I was crawling into an endless amount of rabbit holes as I tried to find a way of transferring the data from the S3 bucket into DynamoDB.

Reading through tutorials on Redshift and Lambda, I found myself confused as to how I was supposed to carry out the tasks suggested. This is an area where I feel that I can be totally objectional as I had not read any documentation on any of the services from either company before, and I can state that in my opinion, the layout, guidance, and instructions of the Microsoft tutorials and documentation are far superior to their Amazon alternative.

This meant moving on to try and find a third-party tool with which I could attempt to integrate S3 with DynamoDB. The first of those, Zapier, made me feel as though I was finally getting somewhere for the first time. It found a connection between both the S3 and DynamoDB services, but I just could not seem to complete the final step of bringing the correct data over. A similar scenario occurred with Panoply.

The final attempt was made through Stitch and, again, I was able to secure a connection between the two services. But it was looking for a .JSON type instead of the csv I was using and even though I felt I was very close; I eventually had to admit defeat and move onto making some sort of progress in Quicksight.

I pulled the csv file in directly which was a bonus by this stage as it meant I could finally produce some data visualization to compare to Power BI. It has still left me absolutely baffled as to why there is no simple native integration between Quicksight and either of the S3 or DynamoDB services, in the same way as there is between Power BI and Azure. It seems like it would provide a far better experience for the user and I do not even see the difficulty in implementing such functionality.

1. Conclusion

In conclusion, I would obviously recommend Microsoft’s Azure Cosmos DB and Power BI combination over the alternative Amazon offerings of DynamoDB and Quicksight. But in defence of the AWS services, I would like to once again clarify that my prior experience with the Microsoft tools gave them a big advantage from the beginning. This is to say that if I had spent an equal amount of time learning both services without any previous knowledge, there is certainly an argument to be made that this opinion would have had a different outcome. Many of the issues which I encountered with the Amazon services are, in all probability, due to my skills rather than any sort of slight on them.

References

Anon 2019. *Visualize Azure Cosmos DB data by using the Power BI connector*. Available at: <https://docs.microsoft.com/en-us/azure/cosmos-db/powerbi-visualize>.

Anon 2021. *Amazon QuickSight vs Microsoft Power BI 2021 - Feature and Pricing Comparison on Capterra*. [online] Available at: <https://www.capterra.com/business-intelligence-software/compare/169810-176586/Quicksight-vs-Power-BI> [Accessed 8 May 2021].

Anon 2021. *Azure/azure-documentdb-datamigrationtool*. [C#] Microsoft Azure. Available at: <https://github.com/Azure/azure-documentdb-datamigrationtool> [Accessed 7 May 2021].

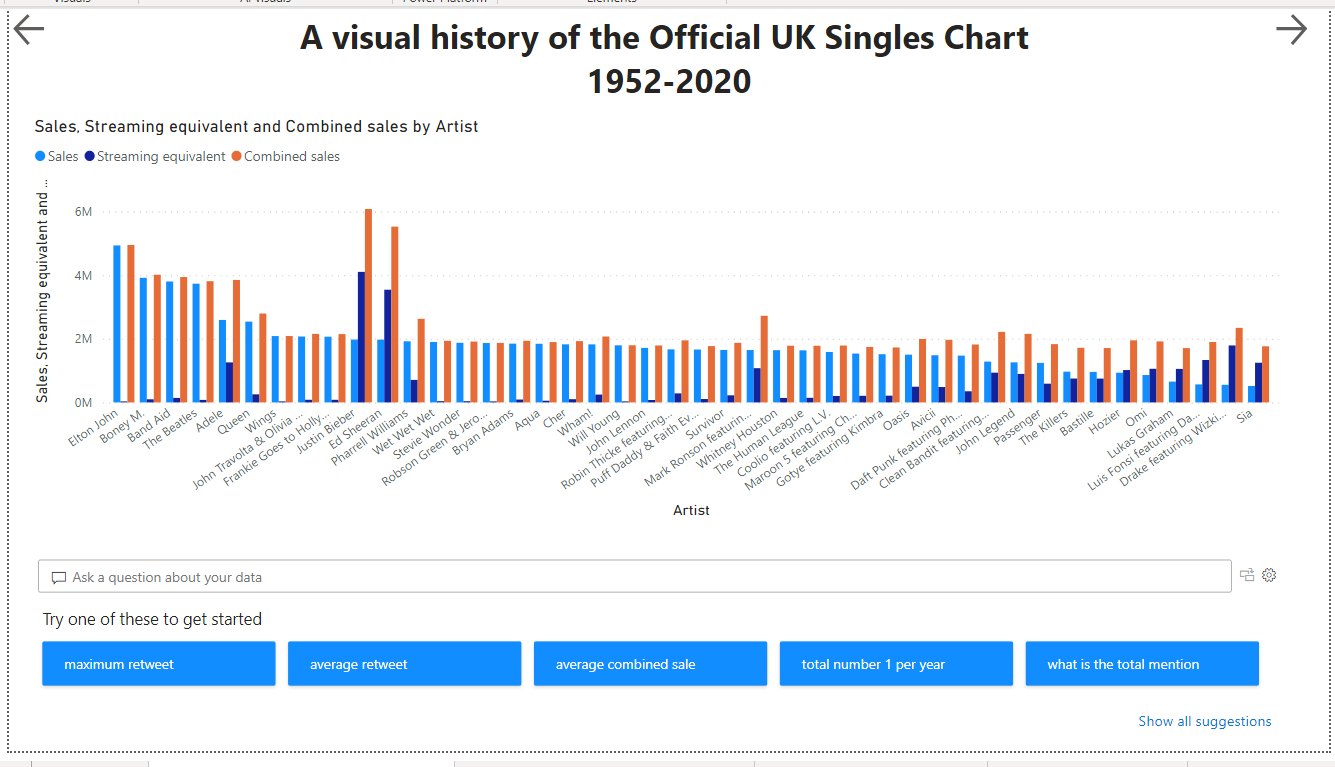
Anon 2021. *azure-cosmos-db-sample-data/VolcanoData.json at main · Azure-Samples/azure-cosmos-db-sample-data*. [online] Available at: <https://github.com/Azure-Samples/azure-cosmos-db-sample-data/blob/main/SampleData/VolcanoData.json> [Accessed 7 May 2021].

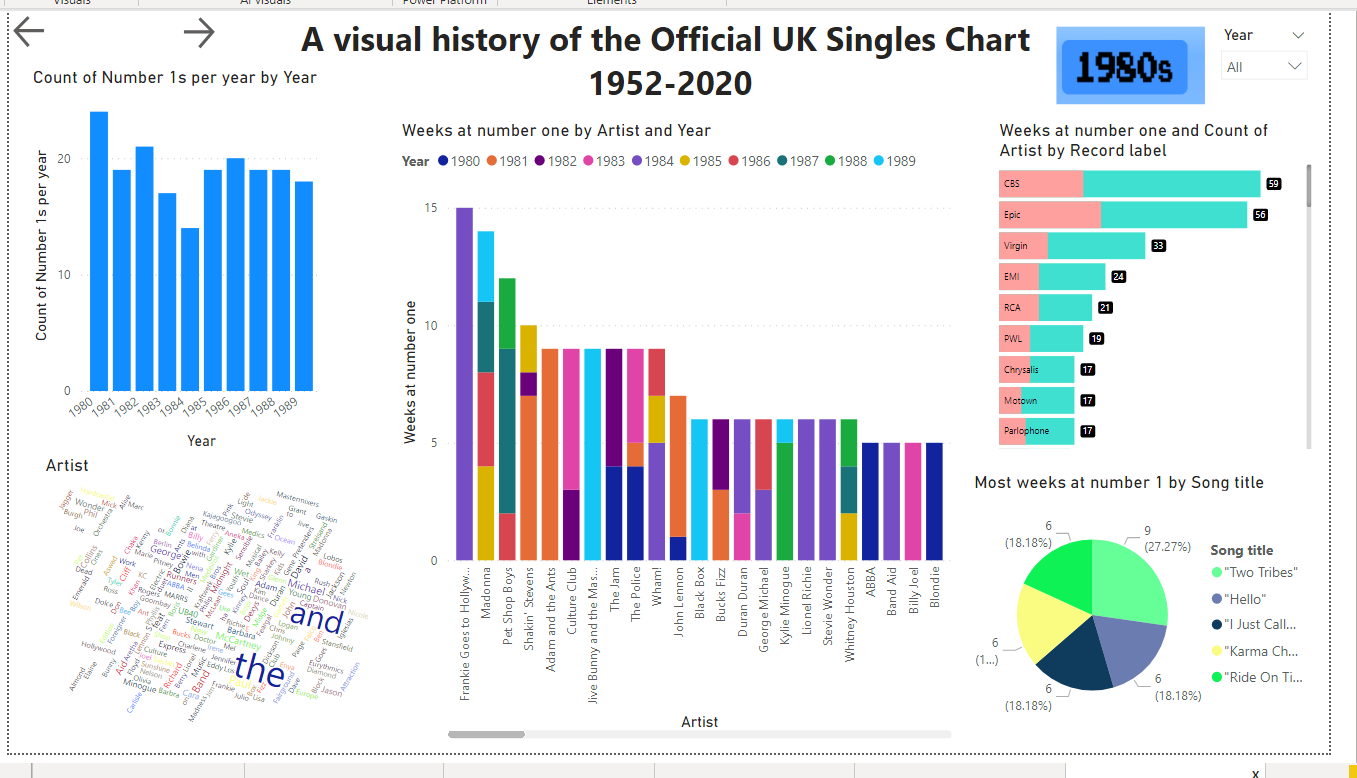
Anon 2021. *DB-Engines Ranking*. [online] DB-Engines. Available at: <https://db-engines.com/en/ranking> [Accessed 6 May 2021].

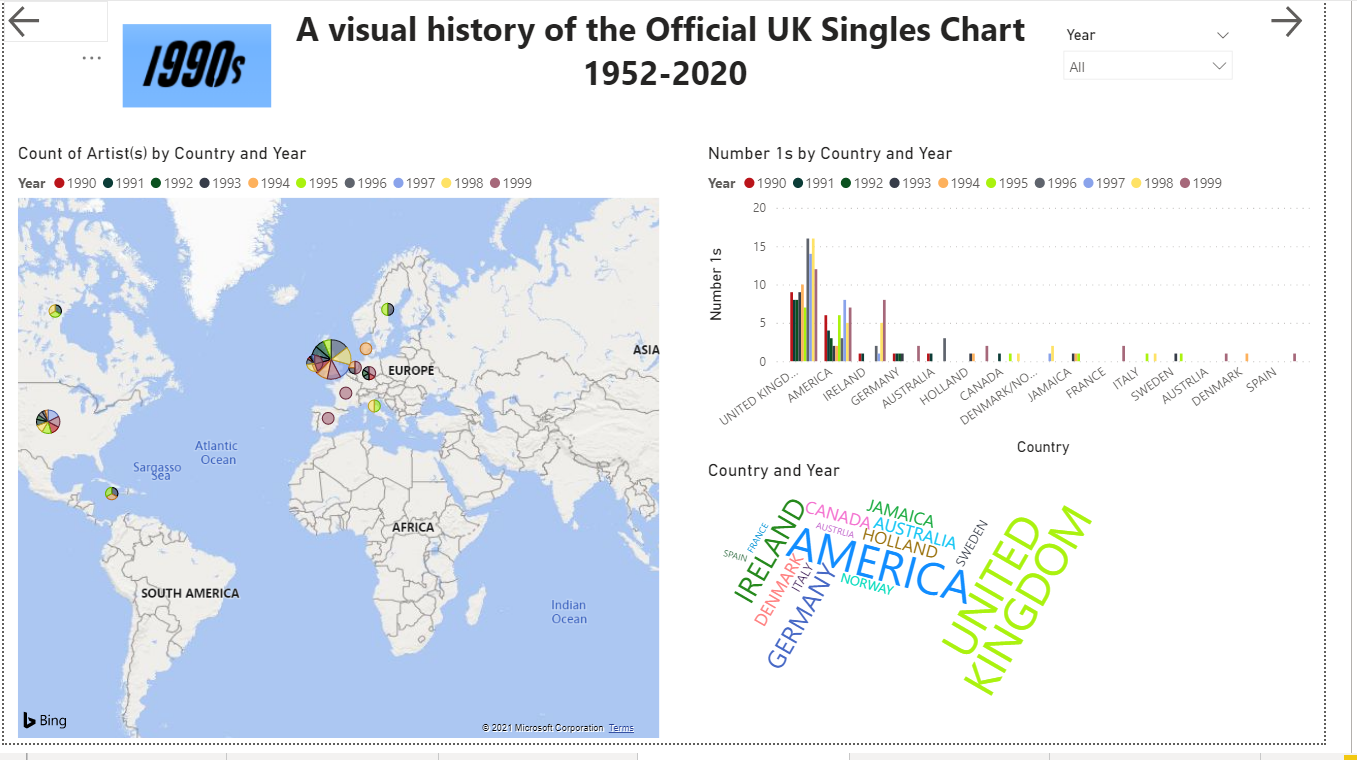
Anon 2021. *DMD (PowerBI - CA) (Colm Gallagher T00170514) (200421) - Power BI*. [online] Available at: <https://app.powerbi.com/groups/me/reports/bad2a075-b36d-4c1a-8c72-62683f701fb2/ReportSection> [Accessed 8 May 2021].

Sharma, A., 2018. Battle Of BI Tools: AWS QuickSight Vs Power BI. *Analytics India Magazine*. Available at: <https://analyticsindiamag.com/battle-of-bi-tools-aws-quicksight-vs-power-bi/> [Accessed 8 May 2021].

1. Sample images from the completed Power BI report stored on powerbi.com <https://app.powerbi.com/groups/me/reports/bad2a075-b36d-4c1a-8c72-62683f701fb2/ReportSection>

    [↑](#endnote-ref-1)