Analytics in the database – Less world on Amazon Cloud

You must have heard the phrase "Data is the new Oil" coined by mathematician Clive Humby. Allow me to extend it to – "And so are the server-less services being created to query and magnify data insight."

Do all applications require a database? The answer was usually 100% yes, just a couple of years ago, and continues to be answered in the affirmative for most legacy business applications even now; Legacy here, means 3 & 4-year-old applications! In the new world of server-less data analytics services, data does not necessarily mandate having a database. There is a paradigm shift happening from the traditional methodology of data in a database toward an accelerated, integrated and innovative edge while engaging with data.

The database-less use cases always existed, however database in one form or the other was the chosen solution due to lack of alternate paths or tools to take hold of and query data not residing in the cool confines of a database. The very idea of creating a data lake without a database sounded risky, costly and far-fetched.

It resembled travelling by a city's fabulous roads (read databases – RDBMS/Data Warehouse etc.), with multiple modes to commute (read database flavours – enterprise/standard etc.) but no alternate commute paths; in spite of knowing that your travel time could be shortened. You could save on the cost too (read no license), but were not able to, as all modes of commute used the same or similar roads in one form or the other.

Imagine a world where we had the tools that would eliminate the need to use a database (No RDMS/Data warehouse) but we still had the ability to:

AWS cloud has empowered the developer community with some very powerful, highlyavailable, scalable & server-less data analytics services.

- Discover & store metadata

- Query data within flat files via SQL queries• Sort, filter & manipulate data

- Share queried data with other applications

- Create visualizations on queried data

This would have essentially led to reduced cost with the following elements and tasks becoming redundant:

- Database License

- Database Cluster

- Database Monitoring

- Planning/Staff for managing maintenance windows

AWS cloud has empowered the developer community with some very powerful, highly available, scalable & server-less data analytics services. These 'Pay As You Go' consumption model services help deliver and fulfil data query and data visualizations using a few clicks, which would otherwise require thousands of lines of code to draw the same conclusion.

In simple terms, databases and associated tools get redundant in the new scheme of server-less analytics by replacing:
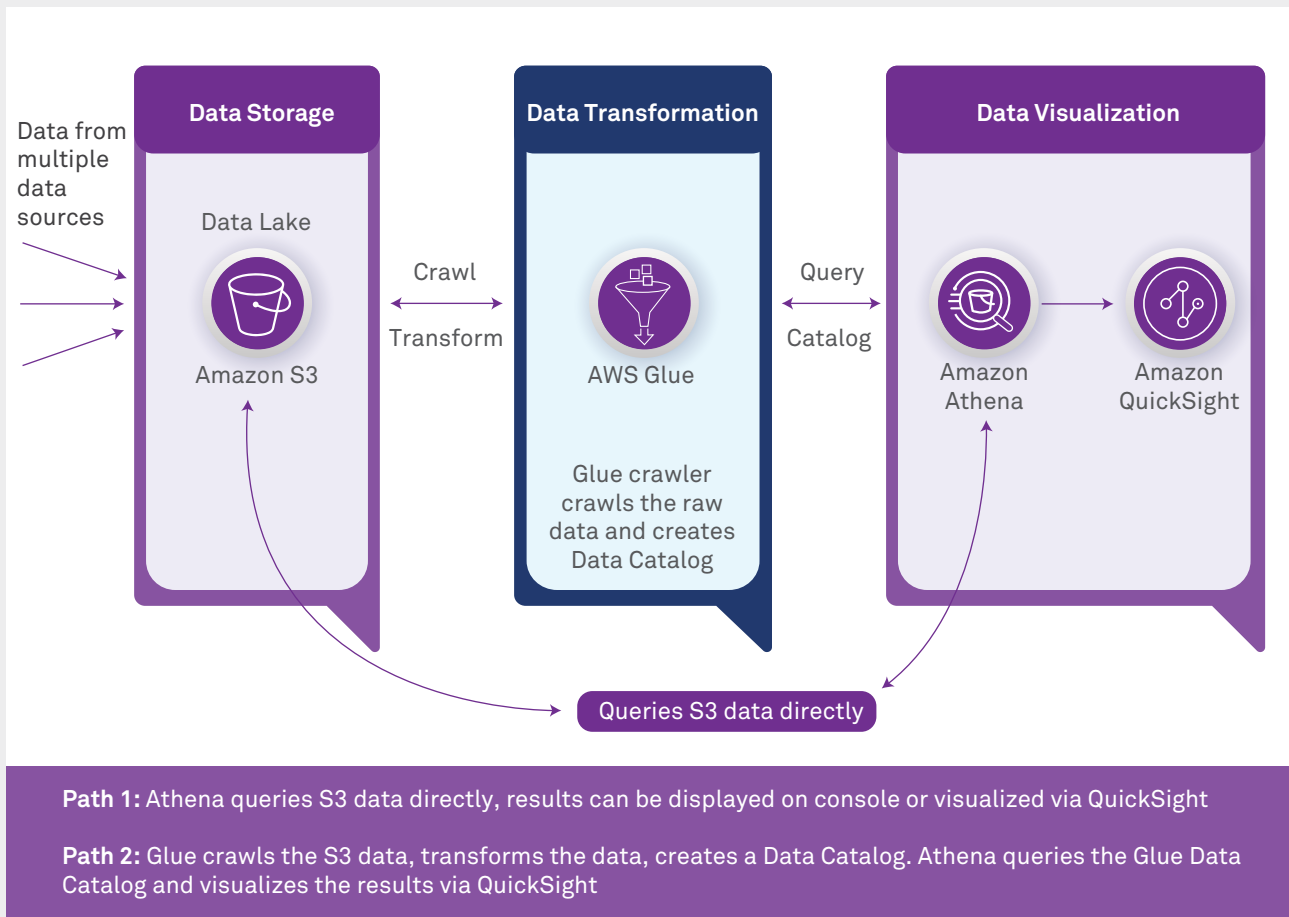
- Data in databases with flat files on AWS S3

- Any ETL tool with AWS GLUE

- Database Query feature with AWS Athena

- Reports with AWS QuickSight

Some use cases that leverage this new dynamic way of working with data:

- Where the need is to grab a quick insight of data to improve customer experience & increase market share

- Need is to analyse large data sets on a periodic basis

- Need is for a quick portable solution that requires only user access, foregoing the need to provision a database

- Need is to modify query on the fly; to explore data and get deeper insights

- Need is to integrate with various BI tools or SQL clients

- Need is to visualize logs (CloudWatch, bespoke application, database logs etc.)

- Need is for a sentiment analysis with data from blogs, review platforms, social media powered by cloud native AI/ML solutions

The very complex world of applying analytics to raw business data gets simplified by the usage of the solution depicted below. This is a paradigm shift as it helps uncover information hidden within very large datasets, including unknown correlations, trends, and preferences, etc. that enable rapid informed business decisions.

The building blocks of this solution are Data Storage, Data Transformation & Data Visualization

Path 1: Athena queries S3 data directly, results can be displayed on console or visualized via QuickSight

Path 2: Glue crawls the S3 data, transforms the data, creates a Data Catalog. Athena queries the Glue Data Catalog and visualizes the results via QuickSight

Summary:

In this rapidly accelerating and innovative world of cloud services, it is imperative to shift from the traditional way of envisioning data and embracing the rapidly evolving world of cloud services. This fosters an agile attitude to bridge the increasing gap between business expectations, IT capabilities and the need for cost-effective solutions on data analytics.

For further reading, refer to the following articles -

https://docs.aws.amazon.com/glue/latest/dg/glue-dg.pdf

https://aws.amazon.com/athena/resources/

https://aws.amazon.com/quicksight/resources/?nc=sn&loc=6&dn=1

About the author

Deepak Tyagi
Senior Architect,
MAS Cloud Services, Wipro Limited.

Deepak has 20+ years of IT industry experience. He also has Onshore-Offshore experience in Digital Transformation and Modernization & Migration solutions for cloud journey of on-premise applications to public clouds viz. AWS/Azure/GCP.

Wipro Limited (NYSE: WIT, BSE: 507685, NSE: WIPRO) is a leading global information technology, consulting and business process services company. We harness the power of cognitive computing, hyper-automation, robotics, cloud, analytics and emerging technologies to help our clients adapt to the digital world and make them successful. A company recognized globally for its comprehensive portfolio of services, strong commitment to sustainability and good corporate citizenship, we have over 175,000 dedicated employees serving clients across six continents. Together, we discover ideas and connect the dots to build a better and a bold new future.

For more information, please write to us at **info@wipro.com**