



Korcomptenz



Serverless Analytics

Guide your digitization journey with the velocity of a proven cloud approach

Contents

1.	Abstract	2
2.	Understanding what is serverless computing	2
3.	Managed Services or Serverless Technologies—Which Is Right for Your Business?	3
4.	Traditional serverless versus containerized serverless	3
5.	Why build a serverless data platform?	3
6.	Why Data Analytics?	4
7.	Serverless Data Analytics with Azure	5
8.	Serverless Data Analytics with AWS	5
9.	Benefits of serverless	6
10.	Options when choosing a data analytics technology	7
11.	AWS to Azure services comparison	7
12.	Architecture	9
13.	Security Layer	9
14.	Serverless Best Practices to secure the Data	10
15.	Serverless SQL pool with Power BI	10
16.	Why Korcomptenz	11
17.	Acknowledgments	12

Abstract

C-Level executives, do you often feel that you spend a significant amount of time and resources managing costs? Or do you often wonder about the impending costs on IT assets and how to optimize them? Gartner, in a recent survey, has expected that IT services are going to raise this year, with the projected IT spending growth rate being 5.5.% for 2023. With soaring expenses, you need to find alternative tactics to reduce the burden of IT expenses.

Most enterprises need a significant amount of resources and time to manage servers, which adds cost. If you could do away with servers, you could save significant capital expenditure. If enterprises can use fewer resource to manage servers, they can focus on building great Analytics that can help their business leaders take right decision for growth of their company. Serverless computing helps you do just that because the infrastructure that you need to run and scale your apps is managed for you. Serverless computing is the abstraction of servers, infrastructure, and operating systems. Serverless computing is driven by the reaction to events and triggers which are all taking place in near real-time—in the cloud. As a fully managed service, server management and capacity planning are invisible to the developer.

There are various challenges to process and compute this data, like storage, server management, clustering, algorithm deployment, etc. As most of things are done manually, it is difficult to design the architecture for data analysis in Azure. Serverless computing is a mechanism to provide pay-per-use backend services to clients. A serverless provider lets users create and deploy code without worrying about operating and managing servers. Serverless computing simplifies operations associated with data analysis in cloud environments. Companies collect and store more data than ever. This increased volume of data makes serverless analytics applications more appealing. Serverless analytics platforms are scalable, cheaper, and allow great flexibility for ingesting new datasets. In this paper, we present serverless architecture for data analytics, also we show how to implement, maintain, and governance of a serverless big data application on Azure.

Understanding what is serverless computing

Serverless is an implementation model where the cloud provider is accountable to execute code by distributing the resources dynamically. Enable business analysts, IT professionals, and data scientists to collaborate seamlessly, providing a single version of data truth that delivers insights across your organization.

Serverless computing represents a natural next step of the “as a service” and resource sharing trends in cloud computing. Specifically, “function as a service” offerings such as Azure Functions allow developers to write blocks of code with well-defined entry and exit points, delegating all aspects of execution to the cloud provider. Typically, these blocks of code are stateless, reading from and writing to various “state as a service” offerings (databases, message queues, persistent stores, etc.).

Standard serverless deployments are characterized by asynchronous, loosely coupled and event-driven processes that touch relatively small amounts of data. Most serverless applications are user facing, even if users are not directly involved in the processing pipeline. This paper explores serverless architectures for a completely different use case: large-scale analytical data processing by data scientists. One key feature is that we realize a pure pay-as-you-go cost model, in that there are zero costs for idle capacity. The primary contribution of our work is a demonstration that it is indeed possible to build an analytical data process

Managed Services or Serverless Technologies –Which Is Right for Your Business?

In both Managed services and Serverless computing, you are not managing servers; they have their own significant and unique role in optimizing resources and reducing overhead. Below are some useful tips while choosing one of them which will best fit for your business.

Managed services: If you are new or planning on moving to the cloud, managed services is a great choice for your organization. A managed services provider can offer the expertise you need, when you need it, at a fixed monthly rate. Allowing a third party that has the resources, training, and capacity to manage troubleshooting certain aspects of your business will allow you to shift your focus to running your business and meeting core objectives along the way.

Serverless technologies: If you are already in the cloud and are ready to take the next step to being cloud native, serverless technologies are a great choice for your organization. Not only does serverless computing reduce your reliance on servers and the costs associated with them it also allows enables elasticity, scalability, faster-time-to-market, and increased productivity within your organization.



Traditional serverless versus containerized serverless

The FaaS concept is based on the paradigm of running a function code without dealing with problems of underlying infrastructure such as servers or VMs. Functions are designed to act in accordance with both the stateless model and the fire-and-forget model. No cloud resources are involved while no functions are activated. Functions are not designed for handling long-running processes, but they can be run as a large set of isolated instances that can be created quickly. FaaS solutions come with limitations on, for example, memory usage or execution-time of individual functions, which causes correspondingly high granularity of the division of a whole task.

The basic concept of FaaS is derived from serverless computing. A similar concept, Container as a Service (CaaS), supports containerization and holds several advantages resulting from serverless idea. Within CaaS, containers can be instantiated with an autoscaling option without worrying about a runtime infrastructure

The Azure Functions service is made up of two key components: a runtime and a scale controller. The Functions runtime runs and executes your code. The runtime includes logic on how to trigger, log, and manage function executions. The Azure Functions runtime can run anywhere. The other component is a scale controller. The scale controller monitors the rate of events that are targeting your function, and proactively scales the number of instances running your app

Kubernetes-based Functions provides the Functions runtime in a Docker container with event-driven scaling through KEDA. KEDA can scale in to 0 instances (when no events are occurring) and out to n instances. It does this by exposing custom metrics for the Kubernetes auto scaler (Horizontal Pod Auto scaler). Using Functions containers with KEDA makes it possible to replicate serverless function capabilities in any Kubernetes cluster. These functions can also be deployed using Azure Kubernetes Services (AKS) virtual nodes feature for serverless infrastructure

Why build a serverless data platform?

Serverless data analytics provides many benefits and the agility required of a modern data platform solution.

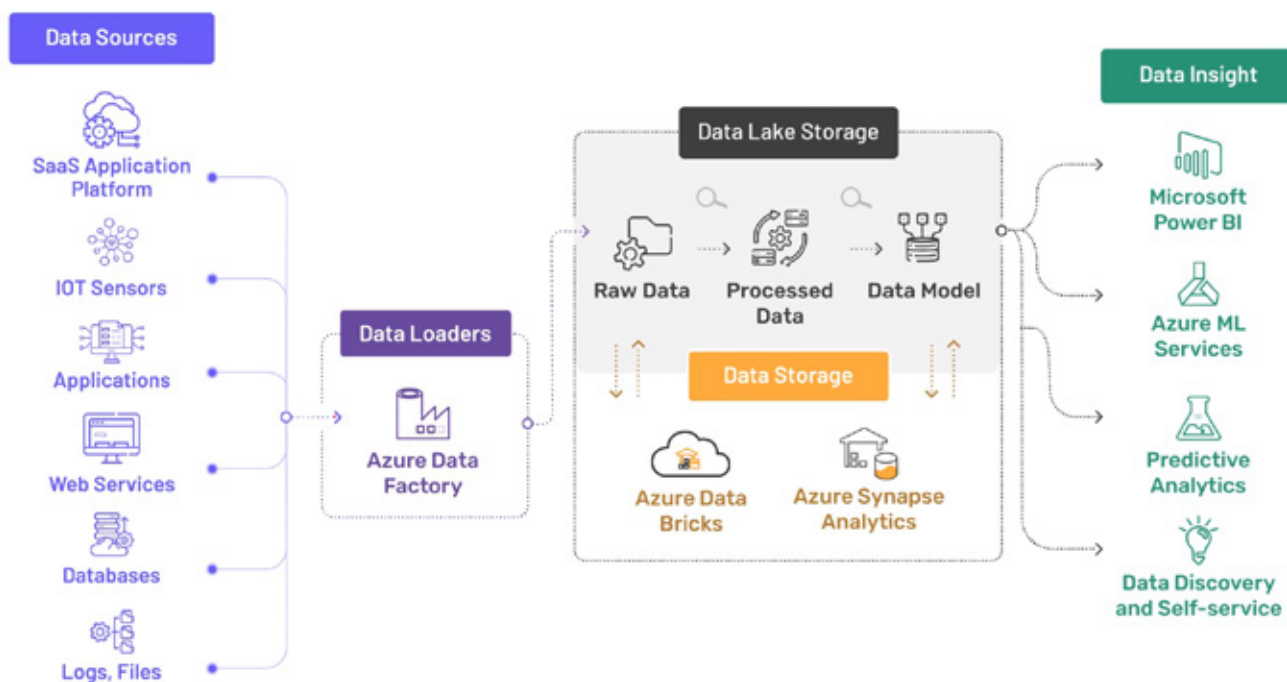
- No server management implies that users do not need to provision or maintain any servers
- Flexible scaling entails that the application can be scaled automatically via units of consumption (throughput, memory) rather than units of individual servers
- Pay for value is to pay for the use of consumption units rather than server units
- Automated high availability ensures that the system must provide built-in availability and fault tolerance.



Why Data Analytics?

Data Analytics enables each one at every level of your organization to make confident decisions. It creates a data-driven culture with business intelligence for all. Every organization should get self-service analytics at enterprise scale which will reduce the added cost, complexity, and security risks of multiple solutions with an analytics platform that scales from individuals to the organization. By using smart tools for strong results which will find and share meaningful insights with hundreds of data visualizations, built-in AI capabilities, tight Excel integration, and pre-built and custom data connectors. Also keep your data secure with industry-leading data security capabilities including sensitivity labelling, end-to-end encryption, and real-time access monitoring.

Sample of our modern data platform for real-time Analytics with Azure



Serverless Data Analytics with Azure

Azure provides a package of products and services designed to capture, organize, analyze, and visualize large amounts of data consisting of the following services:

HDInsight:

managed Apache distribution that includes Hadoop, Spark, Storm, or HBase

Machine Learning:

used to build and apply predictive analytics on data.

Data Factory:

provides data orchestration and data pipeline functionality.

Stream Analytics:

real-time data analysis.

Azure Synapse Analytics:

an enterprise analytics service that accelerates time to insight, across data warehouses and big data systems.

Data Lake Analytics:

large-scale analytics service optimized to work with Data Lake Store

Azure Databricks:

a unified analytics platform for data analysts, data engineers, data scientists, and machine learning engineers.

Power BI:

a business analytics service that provides the capabilities to create rich interactive data visualizations.

Data Lake Store:

analytics service that brings together enterprise data warehousing and big data analytics. Query data on your terms, using either serverless or dedicated resources—at scale.

Azure Logic Apps:

Serverless technology for connecting apps, data and devices anywhere, whether on-premises or in the cloud for large ecosystems of SaaS and cloud-based connectors.

Use Case Scenario

- ♦ A client using Azure Synapse successfully improved their just-in-time supply management, by reducing approximately 30%-40% of the time that was involved in managing issues caused by legacy apps. To solve the issue of time lag of arrival of data and its availability for data analytics, they switched over to a serverless analytics platform, Azure Synapse, was adopted—with the availability of data on a real-time basis, to take actionable insights.
- ♦ A manufacturing company used Artificial Intelligence through drones to capture cracks in the machineries that were not possible to be investigated humanly. With the captured information, they created a monthly report for related information, which assisted in finding areas that needed immediate attention and avoiding breakdowns.

Serverless Data Analytics with AWS

AWS also provides an assortment of products that aid in developing and managing large amount of data:

AWS Lambda:

→ This serverless compute service that allows you to code without the worry of managing services. It is available a wide range of SaaS applications. With Lambda, you can scale as per your requirement and enables you to run applications even during peak hours without any downtime.

Amazon Kinesis:

→ It makes data management and processing possible so that you get real-time insights for taking quick actions.

AWS Serverless Application Model (SAM):

→ It is an open-source software that is used to develop serverless application on AWS. taking quick actions.

AWS CloudFormation:

→ It helps you model your AWS resources so that you can build applications sans the need of creating the AWS infrastructure.

AWS IoT Core:

→ It enables you to connect across multiple IoT devices with various AWS services, enabling them for interact while being offline too.



Benefits of serverless

There are several benefits of having serverless approach. Serverless Analytics provides the following advantages to CFOs in their leadership of internal financial operations

No infrastructure management: Using fully managed services enables developers to avoid administrative tasks and focus on core business logic. With a serverless platform, you simply deploy your code, and it runs with high availability.

Boost team performance: Improve team agility and performance by using a fully managed platform to build, deploy and operate applications. Build for any application pattern and environment—hybrid, cloud, and edge. Proactively manage applications with intelligent monitoring and analysis tools.

Dynamic scalability: With serverless computing, the infrastructure dynamically scales up and down within seconds to match the demands of any workload.

Increase developer velocity: Reduce the time spent on tasks that are non-core to the business by freeing developers from infrastructure provisioning and management. Build and deploy faster using developer-friendly APIs, low-code/no-code services, and ready-to-use machine learning and cognitive models.

Faster time to market: Serverless applications reduce the operations dependencies on each development cycle, increasing development teams' agility to deliver more functionality in less time.

Cognitive computing: Enable your serverless apps to see, hear, speak, understand, and interpret your user needs through natural methods of communication using Azure Cognitive Services via an API or deployed as containers on Kubernetes.

More efficient use of resources: Shifting to serverless technologies helps organisations reduce TCO and reallocate resources to accelerate the pace of innovation.

Conversation bots: Use Azure Bot Services to build intelligent bots that interact naturally with your users through channels such as text/SMS, Skype, Microsoft Teams, Slack, Office 365, and Twitter.

Serverless application environments: With a serverless application environment, both the back end and front end are hosted on fully managed services that handle scaling, security and compliance requirements. innovation.

Machine learning models: Build, train, and deploy models on Azure Machine Learning, from the cloud to the edge.

Success story of client

A food industry giant was looking for solutions related to increasing its customer base by increasing its number of outlets. They reached us for an appropriate strategy—we deployed serverless analytics to analyze real-time data to observe the pattern and identify the gap between the nearest store and customer demand, which helped the client find the right location for opening new outlets.



Options when choosing a data analytics technology

There are several options for analysis, visualizations, and reporting in Azure, depending on your needs:



Capability	Power BI	Jupyter Notebooks	Zeppelin Notebooks	Microsoft Azure Notebooks
Connect to big data cluster for advanced processing	Yes	Yes	Yes	No
Managed service	Yes	Yes 1	Yes 1	Yes
Connect to 100s of data sources	Yes	No	No	No
Offline capabilities	Yes 2	No	No	No
Embedding capabilities	Yes	No	No	No
Automatic data refresh	Yes	No	No	No
Access to numerous open-source packages	No	Yes 3	Yes 3	Yes 4
Data transformation/cleansing options	Power Query, R	40 languages, including Python, R, Julia, and Scala	20+ interpreters, including Python, JDBC, and R	Python, F#, R
Pricing	Free for Power BI Desktop	Free	Free	Free
Multiuser collaboration	Yes	Yes, (through sharing or with a multiuser server like JupyterHub)	Yes	Yes (through sharing)

AWS to Azure services comparison

The below comparison helps you understand how Microsoft Azure services and Amazon Web Services (AWS).

Type	AWS service	Azure service	Description
Serverless computing	AWS Lambda	Azure Functions, WebJobs in Azure App Service	Azure Functions is the primary equivalent of AWS Lambda in providing serverless, on-demand code. AWS Lambda functionality also overlaps with Azure WebJobs, which let you schedule or continuously run background tasks.
Serverless relational database	Amazon Aurora Serverless	Azure SQL Database serverless	Database offerings that automatically scales compute based on the workload demand. You're billed per second for the actual compute used (Azure SQL)/data that's processed by your queries (Azure Synapse Analytics Serverless).



Type	AWS service	Azure service	Description
NoSQL/Document	DynamoDB, SimpleDB, Amazon DocumentDB	Cosmos DB	Cosmos DB is a globally distributed, multi-model database that natively supports multiple data models including key-value pairs, documents, graphs and columnar.
Caching	ElastiCache	Cache for Redis	An in-memory-based, distributed caching service that provides a high-performance store typically used to offload nontransactional work from a database.
Big data processing	EMR	Azure Data Explorer	Fully managed, low latency, distributed big data analytics platform to run complex queries across petabytes of data.
Big data processing	EMR	Databricks	Apache Spark-based analytics platform.
Big data processing	EMR	HDInsight	Managed Hadoop service. Deploy and manage Hadoop clusters in Azure.
Big data processing	EMR	Data Lake Storage	Massively scalable, secure data lake functionality built on Azure Blob Storage.
Data orchestration / ETL	Data Pipeline, Glue	Data Factory	Processes and moves data between different compute and storage services, as well as on-premises data sources at specified intervals. Create, schedule, orchestrate, and manage data pipelines.
Data orchestration / ETL	Glue	Azure Purview	A unified data governance service that helps you manage and govern your on-premises, multicloud, and software as a service (SaaS) data.
Data orchestration / ETL	Dynamo DB	Table Storage, Cosmos DB	NoSQL key-value store for rapid development using massive semi-structured datasets.
Analytics and visualization	Kinesis Analytics	Stream Analytics, Azure Data Explorer, Data Lake Analytics, Data Lake Store	Storage and analysis platforms that create insights from large quantities of data, or data that originates from many sources.
Analytics and visualization	QuickSight	Power BI	Business intelligence tools that build visualizations, perform ad hoc analysis, and develop business insights from data.
Analytics and visualization	CloudSearch	Cognitive Search	Delivers full-text search and related search analytics and capabilities.
Analytics and visualization	Athena	Data Lake Analytics	Provides a serverless interactive query service that uses standard SQL for analyzing databases.
Analytics and visualization	Athena	Azure Synapse Analytics	Azure Synapse Analytics is a limitless analytics service that brings together data integration, enterprise data warehousing, and big data analytics. It gives you the freedom to query data on your terms, using either serverless or dedicated resources at scale.
Analytics and visualization	Elasticsearch Service	Elastic on Azure	Use the Elastic Stack (Elastic, Logstash, and Kibana) to search, analyze, and visualize in real time.
Containers and container orchestrators	Amazon Elastic Container Service (Amazon ECS), AWS Fargate	Azure Container Instances	Azure Container Instances is the fastest and simplest way to run a container in Azure, without having to provision any VMs or adopt a higher-level orchestration service.
Containers and container orchestrators	Amazon Elastic Container Registry (Amazon ECR)	Azure Container Registry	Container registries store Docker formatted images and creates all types of container deployments in the cloud.
Containers and container orchestrators	Amazon Elastic Kubernetes Service (EKS)	Azure Kubernetes Service (AKS)	EKS and AKS let you orchestrate Docker containerized application deployments with Kubernetes. AKS simplifies monitoring and cluster management through auto upgrades and a built-in operations console. See Container runtime configuration for specifics on the hosting environment.
Containers and container orchestrators	AWS App Mesh	Azure Service Fabric	Distributed systems platforms help you develop, deploy, and host scalable microservices-based solutions without managing VMs, storage, or networking.

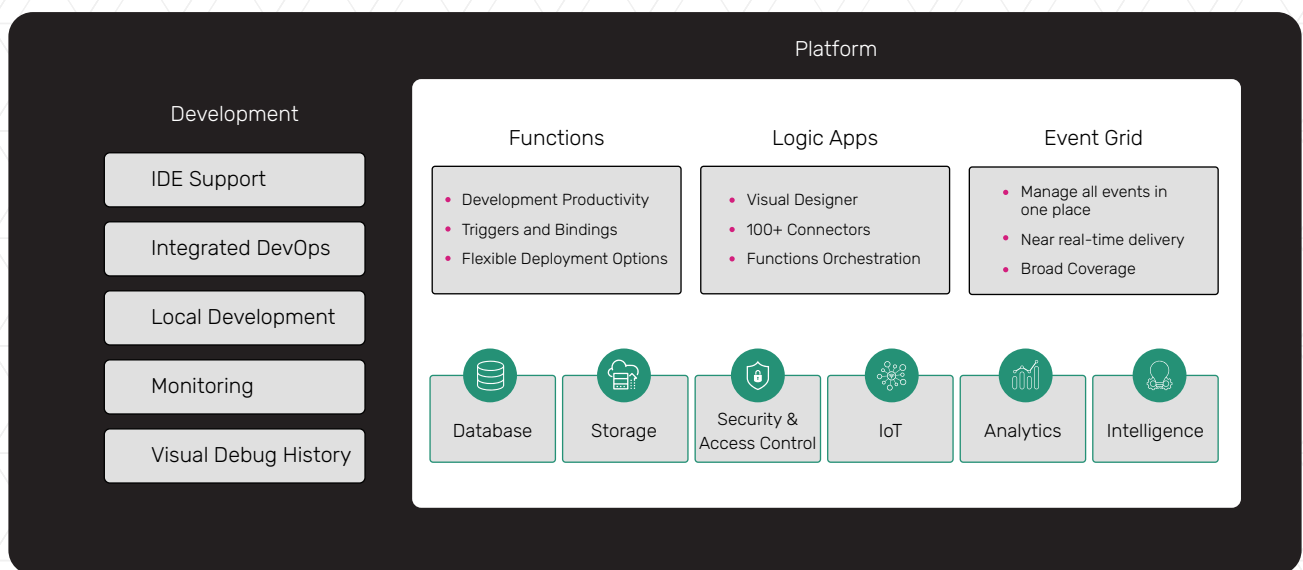


Architecture

We can think of the concepts that support a serverless architecture as three layers that sit atop existing compute, network, and storage resources. Data Analytics will be on top of below 3 mentioned layers.

- The serverless fabric removes infrastructure and operations concerns from the developer's responsibilities. The developer can write code and build a working application without managing servers.
- The event-driven programming model provides a framework for the creation of that code.
- Functions provide the packages, patterns, and reference architectures that are needed to assemble the application.

A serverless architecture includes the platform, related services, and development tools:



Security Layer

Azure is built on trust. The Azure approach to trust is based on the following foundational principles:

- 1 **Security:** Azure leverages leading security technologies to help organizations manage and control user identity and access, which are central elements in securing your environment.
- 2 **Compliance:** Through rigorous and widely recognized formal standards that are certified by independent third parties, Microsoft helps organizations comply with constantly shifting requirements and regulations.
- 3 **Privacy:** We will have control over where our data is located, who can access it, and on what terms. One can access customer data at any time and for any reason.
- 4 **Resiliency:** Microsoft Azure helps to avoid many potential disasters and quickly recover if organization gets hit by disaster.
- 5 **Intellectual Property (IP) protection:** Trust in the cloud also encompasses clarity and confidence that our intellectual property will be protected against frivolous infringement claims. Microsoft Azure IP Advantage and the Shared Innovation Initiative can offer that assurance.



Serverless Best Practices to secure the Data

Below are some recommendations for using serverless Data Analytics to protect the data all the time.

- To avoid sensitive data leakages from the cloud storage infrastructure, vendors can provide a set of security capabilities, like cloud storage configurations, multi-factor authentication, and encryption of data in transit and at rest.
- Organizations that use cloud storage should get familiar with the available storage security controls that their cloud vendor provides.
- Privileges should be assigned to a serverless application that requires access to a resource and should be constrained to a limited time.
- Granting permissions to a user beyond the scope required for an action can allow that user to obtain or change information in unwanted ways. A careful delegation of access rights can limit attackers from damaging a system.
- To avert successful attacks, organizations must have real-time monitoring and logging and eventing to gain insights into how well a system is functioning. It's also a crucial part of maintaining quality-of-service targets.
- Maintaining an inventory list of software packages and other dependencies and their versions.
- Scanning software for known vulnerable dependencies—especially when we are adding new packages or upgrading package versions. Vulnerability scanning should be part of ongoing continuous integration and delivery process.
- Removing unnecessary dependencies, especially when serverless functions no longer need them.
- Consuming third-party packages only from trustworthy resources and making sure that the packages have not been compromised.
- Upgrading deprecated package versions to the latest versions and applying all relevant software patches.

Serverless SQL pool with Power BI

Use serverless SQL pool with Power BI Desktop & create a report

◆ Prerequisites

Power BI Desktop - needed to visualize the data and create a report.

Azure Synapse workspace - needed to create database, external data source, and view.

Optional: A SQL query tool, such as Azure Data Studio, or SQL Server Management Studio (SSMS).

Values for the following parameters:

Parameter	Description
Serverless SQL pool service endpoint address	Used as server name
Serverless SQL pool service endpoint region	Used to determine the storage used in the samples
Username and password for endpoint access	Used to access endpoint
Database you'll use to create views	The database used as starting point in the samples



◆ **Create database:**

Create database. This database to view metadata, not to store actual data.

◆ **Create data source:**

A data source is necessary for the serverless SQL pool service to access files in storage. Create the data source for a storage account that is in the same region as endpoint. Although serverless SQL pool can access storage accounts from different regions, having the storage and endpoint in the same region provides better performance.

◆ **Prepare view:**

Create the view based on the external demo data for Power BI to consume by running the Transact-SQL (T-SQL) scripts

◆ **Create Power BI report:**

Create the report for Power BI Desktop using the following steps:

- Open the Power BI Desktop application and select Get data.
- Select Azure > Azure SQL Database.
- Type the name of the server where the database is in the Server field, and then type Name of Database in the database name. Select the Import option and then select OK.
- Select preferred authentication method like AAD or SQL Login
- Select the view which was created and then select Load.
- Wait for the operation to complete, and then a pop-up will appear stating There are pending changes in your queries that haven't been applied. Select Apply changes.
- Click apply changes.
- Wait for the Apply query changes dialog box to disappear, which may take a few minutes.
- Once the load completes, select the following columns in this order to create the report:

Why Korcomptenz?

Korcomptenz comes with expertise that helps you in using your data into actionable insights that could deliver real-time solutions. You can deploy predictive analytics to accelerate your business results by taking informed decisions with the right data literacy. By going serverless, we enable you to reduce your CAPEX as well as OPEX, while you pay only for the services used without worrying about infrastructure management or sourcing. Our data analytics strategy will aid in making matured decisions with a multi-faceted and dynamic approach. You and your team can stay focused on creating better business opportunities and delivering enhanced value to your clients, while we #FocusOnYou.



Acknowledgements

Although Korcomptenz does not share any code with Microsoft, we'd like to acknowledge Microsoft Power BI team and Azure Serverless team for their contributions to developments of the Serverless Data Analytics concept

References

- azure.microsoft.com
- [Data Visualization | Microsoft Power BI](#)
- [Tutorial: Analyze Azure Open Datasets in Synapse Studio - Azure Synapse Analytics | Microsoft Docs](#)
- [Tutorial: Use serverless SQL pool to build a Logical Data Warehouse - Azure Synapse Analytics | Microsoft Docs](#)
- [Connect to Synapse SQL with SQL Server Management Studio \(SSMS\) - Azure Synapse Analytics | Microsoft Docs](#)
- [Serverless Data Processing \(amazon.com\)](#)
- [Serverless \(amazon.com\)](#)



Korcomptenz

Discover how we #FocusOnYou at www.korcomptenz.com

DISCLAIMER:

The content provided in this document is intended solely for general information purposes. The content was compiled with reasonable care and attention at the time of its release. However, it is possible that some information in this document is incomplete, incorrect, outdated, or inapplicable to particular circumstances or conditions. Korcomptenz does not accept liability for direct or indirect losses resulting from using, relying or acting upon information in the document.

This document may contain logos, trademarks, service marks or other insignia owned by third party organizations. The use of any such items does not constitute an endorsement, sponsorship, or any formal association with the respective owner. The respective owner retains all rights and title to their intellectual property.

Get in touch with us:

+1 (973) 601 8770 | sales@korcomptenz.com

