

Lenovo Solution Brief for

# Data Science and AI with Azure SQL on Lenovo ThinkAgile MX Series

Lenovo

## Introducing Azure SQL for AI & ML

Data is an important asset for every business, and the hunt for business insights from that data has never been as important as it is today. To gather actionable insights from raw data quickly and efficiently empowers business decision makers to gain a deeper understanding of each aspect of their business, helping them to react to new developments and take the right course of action as soon as possible.

For example, consider a retail scenario. The business analyst notices that sales are dropping for specific retail stores. The business analyst wants to drill down to understand the details on what's causing the drop in sales. By being able to run the analysis (aggregating, joining of data from multiple data sources, filtering and so on) on a large amount of data, it enables deep analysis of customer behavior and trends in the retail stores. Microsoft SQL Server powers these mission-critical applications.

Many companies begin a digital transformation to modernize their data platform to keep pace with the ever-growing requirements on the type of data that needs to be stored. The volume in which the data is being acquired is continuously increasing. A high percentage of these are looking to the Microsoft Azure SQL family of offerings to obtain a unified experience across their entire SQL portfolio adding a full range of deployment options from edge to cloud.

As a reminder, an Azure service like SQL runs on Azure Stack HCI, Microsoft's hybrid cloud solution for customers that wish to run workloads on-premises and extend easily to Azure for hybrid capabilities such as back-up, site recovery, storage, cloud-based monitoring and more. Microsoft Azure Stack HCI with Azure Arc on Lenovo servers is the perfect hybrid cloud solution at the edge to help our joint customers on their modernization journey.

Azure Stack HCI is a new HCI host operating system from Microsoft, delivered as an Azure service, providing the latest and up-to-date security, performance, and feature updates. Azure Stack HCI builds on the foundation of the Microsoft Windows Server Software Defined program and provides a certification path for Storage Spaces Direct solutions.

Lenovo Azure Stack HCI offerings use the Microsoft Azure Stack HCI operating system on the host nodes, and optionally include Windows Server 2022 Datacenter in case you require unlimited guest OS virtual machine licenses.

The Azure Stack HCI OS license provides the following benefits:

- **Storage Spaces Direct:** State of the art software-defined storage from Microsoft with multiple high-performance resiliency options, deduplication, compression and more.
- **Hyper-V:** Hypervisor is included in the license
- **Software-Defined Networking:** Features such as virtual network encryption, firewall auditing, virtual network peering allows you to get the benefits of a more secure software defined network with Azure Stack HCI.
- **Windows Admin Center (WAC):** A web-based management portal software is recommended for managing an Azure Stack HCI cluster. Deployment and update features in WAC make deployment extremely simple and easy to perform. Additionally, Lenovo's XClarity plugin allows you to deploy their hardware as well as software from the same interface, enabling single pane of management. Cluster-aware updating features makes it easy to streamline software and firmware updates in a single maintenance window.
- **Azure Stack HCI** is hybrid by design, and you can benefit from native integration with Azure ARC and Azure Monitor and connect to Azure for a variety of Azure hybrid services seamlessly. Fleet management for hosts and VMs allows you to monitor and manage clusters at scale.

Deploying hyperconverged infrastructure has become the de-facto standard for organizations looking to modernize their aging infrastructure. Large storage deployments are increasingly being replaced by HCI based solutions for most general-purpose workloads. HCI has proven to deliver better efficiency and price performance in the datacenter. Additionally, customers have been choosing a hybrid approach, migrating certain workloads to the cloud, while keeping other workloads on-premises.

Lenovo ThinkAgile MX Series (Appliances and Certified Nodes) combine the Storage Spaces Direct technology included in this new host OS, with industry leading Lenovo servers to deliver HCI building blocks that build your infrastructure solutions. Lenovo ThinkAgile MX appliances map to Microsoft Azure Stack HCI Integrated Systems and ThinkAgile MX Certified Nodes map to Microsoft Azure Stack HCI Validated Nodes.

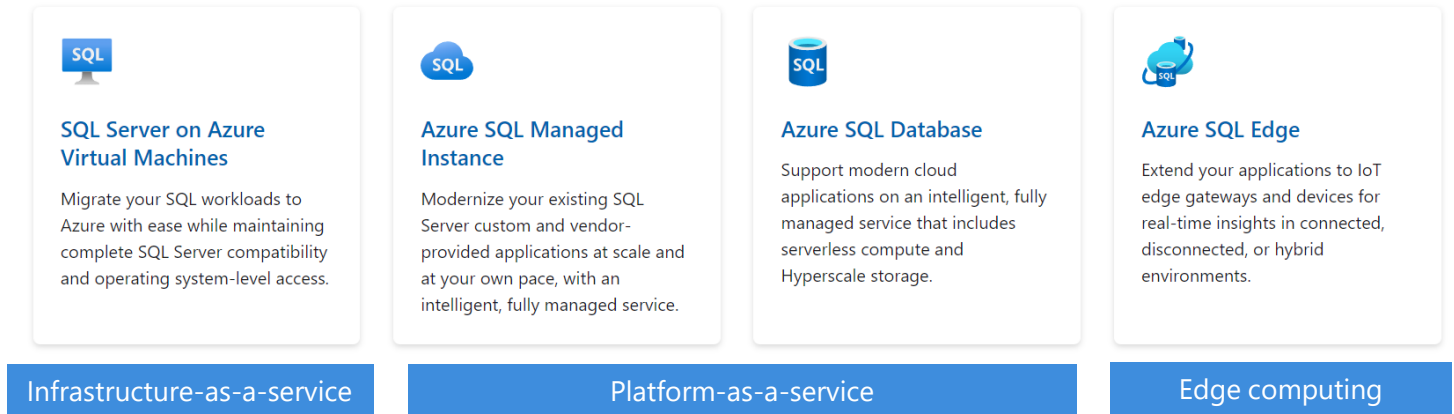
ThinkAgile MX Appliances include Azure Stack HCI, and optionally include Windows Server 2022 Datacenter in case you require unlimited guest OS virtual machine licenses. ThinkAgile MX Certified nodes optionally include the Azure Stack HCI software or Windows Server 2022 Datacenter depending on their needs.

Lenovo has developed ready to deploy solution configurations based on ThinkAgile MX 2U Appliances that are optimized for both OLTP and DW workloads. These configurations represent starting points for 1TB and 5TB databases, respectively. They are highly available 2 node 2U clusters featuring 3rd Gen Intel Xeon Scalable processors and can be found on [Lenovo's Data Center Solutions Configurator](#). They can be customized to fit the exact size of a customer database.

Lenovo's complete Azure SQL for AI & ML solution delivers a well configured ThinkAgile MX All Flash Appliance equipped with Azure Stack HCI which is delivered as an Azure service when the system is deployed. Customers are given a choice as to how to deploy their SQL database to support AI & ML.

The complete hardware and software solution can easily be ordered as is or customized by visiting Lenovo's Data Center Solution Configurator and browsing the Microsoft Azure Deployment Ready Solutions category at the top of the page. Our Azure customers who are not in need of fresh infrastructure can order the Azure SQL family service of choice by contacting their local sales rep or business partner and inquiring about Azure services from Lenovo via CSP, courtesy of Lenovo Cloud Marketplace.

## Azure SQL family of database services



### Choose Azure SQL Databases and...

Do more with fully managed and always up to date services



Use your existing SQL experience in the cloud and at the edge



Save with the lowest total cost of ownership



Protect your data with built-in, real-time intelligent security



Figure 1. Flexibility of Azure SQL

Azure SQL permits customers to spend more time innovating and less time patching, updating, and backing up their databases. Azure is the only cloud that automatically applies the latest updates and patches and provides the latest SQL Server features and functionality. Your databases are always up to date—eliminating end-of-support hassles. Even complex tasks like performance tuning, configuring high availability, disaster recovery, and backups are automated, freeing you to focus on applications.

Azure SQL is built on the same familiar SQL Server technology, so existing SQL customers don't need to relearn their SQL skills when they make the move to Azure cloud. These customers can develop their application once using their existing SQL skills and deploy it on any Azure SQL cloud database on Azure. Azure is the only cloud with a consistent SQL code base that stretches from edge to cloud.

Azure SQL Database is continuously innovating by building [artificial intelligence](#) into the database engine to improve performance, reduce resource usage, and simplify management. The most prominent use of artificial intelligence is the automatic tuning feature that has been globally available since January 2016. Automatic tuning uses artificial intelligence to continuously monitor database workload patterns and recognize opportunities to improve the database performance. Once confidence is built that a certain tuning action would improve the performance of a database, the Azure SQL Database service automatically does the tuning action in a safe and managed fashion. The service monitors each tuning action and the benefits to performance are reported to the database owners. In the infrequent case of a performance regression, the service quickly reverts the tuning action. [Click here to read more about automatic tuning.](#)

## Enter AI & ML

As part of this digital transformation, advanced analytics plays an important role. Specifically, companies have been either building up data science teams within their companies or leveraging external resources to do data science. They use data science to distill data assets into valuable nuggets of insights that can help them proactively deliver personalized customer experiences (personalized Web sites, product recommendations, customer lifetime value and so on), reduce downtime for equipment (predicting remaining useful lifetime) and more. The potential use of data science and how it can literally change businesses is exciting and evolving at a rapid pace to keep up with [data evolution](#).

One example of data science application to real world use case is retail environments attempting to manage customer churn. Customer churn models have been used extensively (in retail and by telecommunication providers). For example, customers leveraging mobile services offered by telecommunication providers have a rich variety of choices and can easily switch between service providers. Managing customer churn is important to reduce customer acquisition costs and maintain a high-quality service. In addition, retail companies are using churn models to predict customers that are most likely to churn and to understand the key factors that lead to those churns.

This presents an exciting opportunity for database professionals and developers to either work with data scientists or put on a data scientist hat to build predictive models that can help with challenges like customer churn, credit loan risk assessment, identification and prediction of disease, optimizing shipping and logistics routes in real-time, detection of frauds, healthcare recommendations, automating digital ads and countless other challenges commonly facing business and IT today. The possibilities for developers to turn all these raw data assets sitting in the database to golden, strategic insights is exciting.

## Why Doing Data Science with SQL Server Matters

What does doing data science mean and why does it matter to the database administrator? Today, most data scientists first figure out how to connect to many data sources (databases included), in order to bring the data out from the database and then use the historical data to train and subsequently test the machine learning models that they've built.

A data scientists typical process is to read the data from the database into the client platform that they're using for building the ML model. After reading the data, they combine that data with other data sources to begin the work of transforming data into the right shape to suit the data modeling process. Once all the transformation steps are completed, the data scientist develops the models and validates it using test data before figuring out an operational plan for the model to be deployed to production so that applications, can consume them for use cases of the like that was previously discussed.

At this point, if you're a person with database experience, you might ask, "Why should I have to move the data out from the database to do data science? Should we push the processing of joins and aggregations into the database?" A join represents the collection of data from different sources for your machine learning model. An aggregate function is a mathematical computation involving a range of values that results in just a single value expressing the significance of the accumulated data it is derived from.

To finish the previous point, when collecting data for your machine learning model, you must 'join together' different sources to create your data set.

Wouldn't it make more sense to do this in the database where the historical data resides?

Its important to note that data movement is expensive. If the machine learning models can run where the data is stored, this removes the risky need of moving data between the database and the client application. Second, a new working copy of the data is extracted from the database and stored external to the database. The implication is that many of the security policies and audits that apply to data stored in the database can no longer be enforced. Lastly,, if the computation of joins and aggregations can be done where the data is located, you can leverage decades of database innovations (leveraging indexes—clustered and non-clustered, in-memory tables, column stores, high-availability and so on). If training the model can be done where the data is stored, it can and will lead to performance improvements.

Database administrators and developers have tremendous knowledge and value they inherently bring to a data science project. By enabling them to do data science and AI right where the data resides, there are numerous benefits to be realized. These include being able to take advantage of the enterprise-grade performance, scale, security and reliability that you've come to expect from SQL Server over the years. Most importantly, you eliminate the need for expensive, time consuming and risky data movement!

Figure 2 illustrates the difference between doing data science and AI outside of the database, as it has traditionally been done, vs inside the database where the data resides. From the application perspective, a developer doesn't need to learn new methods to tap the power of AI teeming in SQL Server. They connect to the data the same way they connect to a database today, and invoke SQL Server-stored procedures, which encapsulates the R or Python code. The stored procedure has just become an intelligent AI stored procedure.

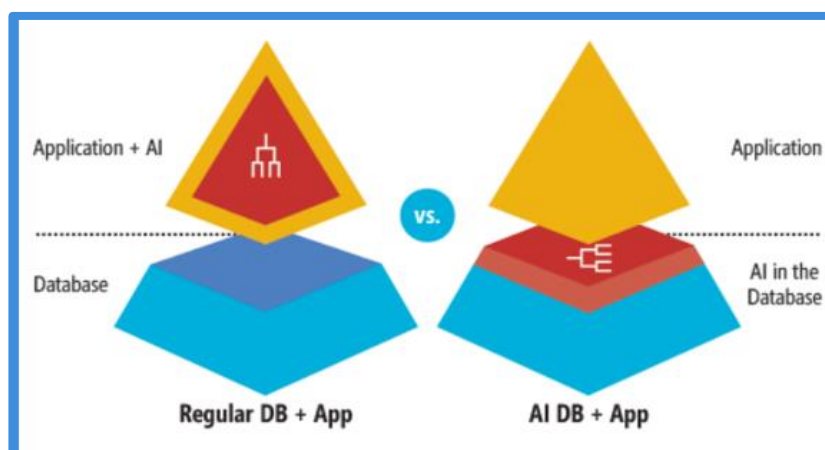


Figure 2. Doing Data Science and AI where the Data is stored

A stored procedure is a prepared SQL code that you can save, so the code can be reused repeatedly. So, if you have a SQL query that you write repeatedly, save it as a stored procedure, and then just call it when needed to execute it. [Click here to learn more](#)

## Using Stored Procedures for Machine Learning and AI

There are many benefits of using a SQL stored procedure:

**It can be easily modified:** We can easily modify the code inside the stored procedure without the need to restart or deploying the application. For example, If the T-SQL queries are written in the application and if we need to change the logic, we must change the code in the application and re-deploy it. SQL Server Stored procedures eliminate such challenges by storing the code in the database. so, when we want to change the logic inside the procedure, we can just do it by simple ALTER PROCEDURE statement.

**Reduced network traffic:** When we use stored procedures instead of writing T-SQL queries at the application level, only the procedure name is passed over the network instead of the whole T-SQL code.

**Reusable:** Stored procedures can be executed by multiple users or multiple client applications without the need of writing the code again.

**Security:** Stored procedures reduce the threat by eliminating direct access to the tables. we can also encrypt the stored procedures while creating them so that source code inside the stored procedure is not visible. Use third-party tools like ApexSQL Decrypt to decrypt the encrypted stored procedures.

**Performance:** The SQL Server stored procedure when executed for the first time creates a plan and stores it in the buffer pool so that the plan can be reused when it executes next time.

By encapsulating the machine learning and AI models as part of the SQL Server stored procedure, it lets SQL Server serve AI with the data. There are other advantages for using stored procedures for operationalizing machine learning and AI (ML/AI) but a key advantage is the following:

Applications can leverage existing database drivers to connect to SQL Server: Many programming languages have database drivers that enable them to connect to SQL Server. These database drivers (OLEDB, ODBC, JDBC, MSPHPQL and Node.js Driver for SQL Server) are used by application developers to develop cutting-edge applications that “talk” to SQL Server.

Being backed by the full spectrum of SQL Server enterprise-ready capabilities: Some considerations include ...

- Where can I host the model for doing inference?
- Which users can access the ML/AI model?
- When the model is used for prediction (aka “inference”), it might need to access specific data in the database. In what security context should the model and associated R/Python code execute?
- How can I ensure the R/Python code doesn't use up all the SQL Server resources?

SQL Server provides enterprise-ready capabilities from using row-level security to limit the data that can be accessed, to providing database admins with both server and database scoped database audits, to enabling ownership-chaining for SQL Server securable, to being able to sign stored procedures with a certificate or asymmetric key, resource governance and more. These enterprise-ready SQL Server capabilities can be used by the ML/AI stored procedures as is, without requiring the data scientist to reinvent the wheel for serving data at scale. Most important, the DBAs today can leverage their existing skills to secure and manage the ML/AI stored procedures.

SQL Server supports Python and R which will allow developers to implement ML/AI models that natively process data stored in SQL Server databases. Those ML/AI models can be directly persisted in the underlying database servers and scaled as part of SQL Server clusters. More importantly, developers will have access to these capabilities using the familiar SQL Server tool set.

- Eliminate data movement and leverage database security
- Operationalize scripts and models by taking advantage of T-SQL stored procedures and model management in SQL Server
- Achieve Enterprise grade performance and scale - Python / R Analytics
- Take advantage of SQL Server security, compliance, query performance and resource governance

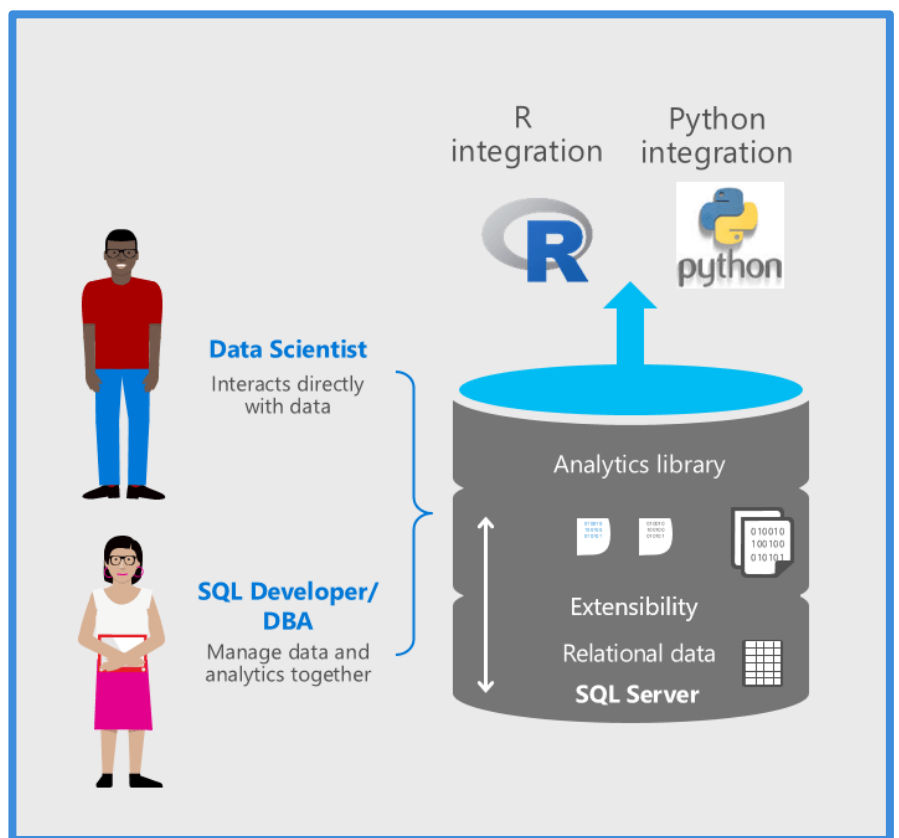
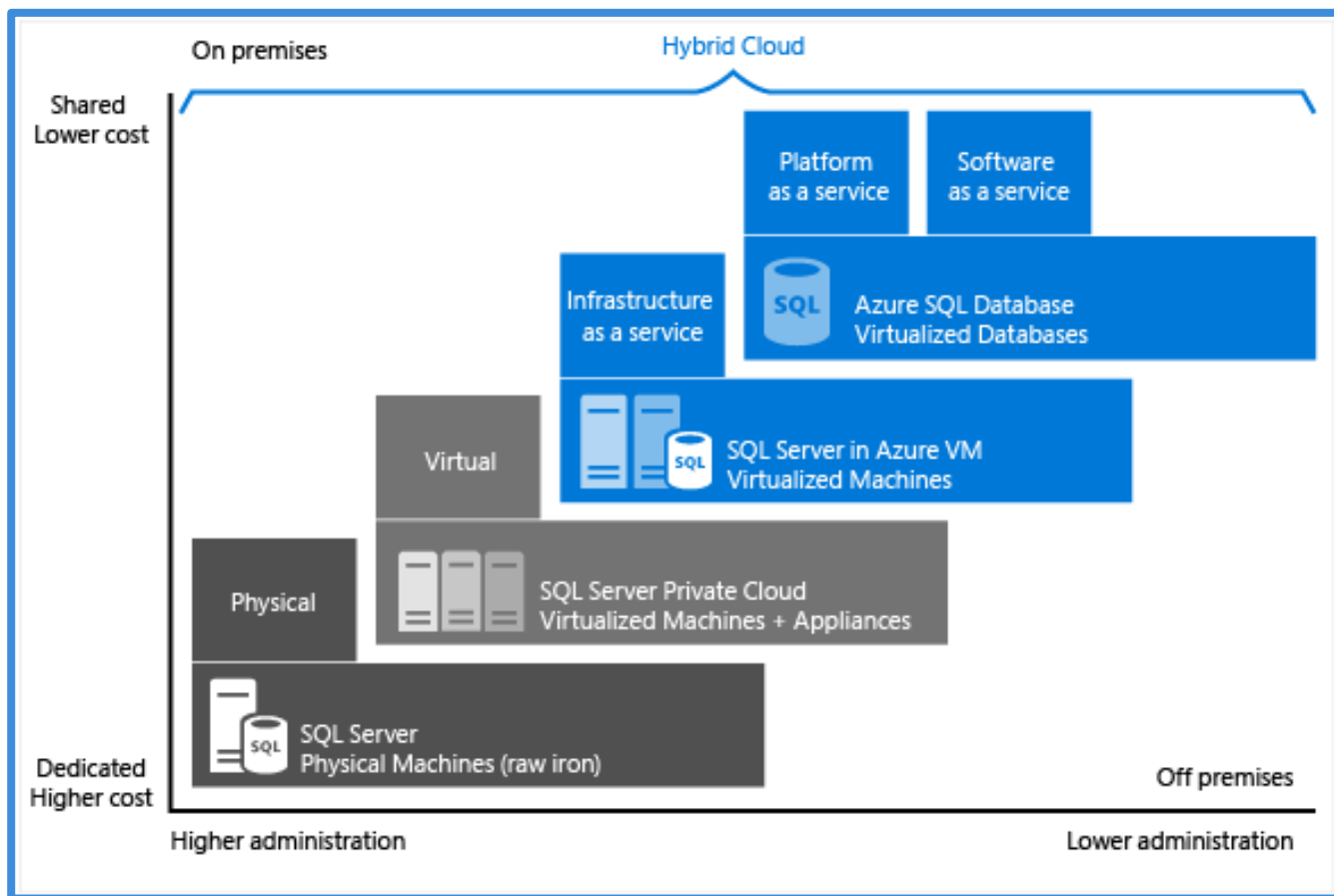


Figure 3. Leverage the skills of a DBA

There are three primary considerations for selecting the right Azure SQL service for your requirements: Your overall objective, the extent to which you require operating system (OS) access or control, and the version of SQL Server you prefer. If you require control of the underlying OS due to compliance or software dependencies and prefer to remain on your current version of SQL Server, then rehosting to SQL Server on Azure Virtual Machines would be your best option. If, however, OS control is not a requirement and you prefer to be on the latest "version less" SQL Server, then modernizing your existing applications or supporting modern cloud applications on fully managed solutions like Azure SQL Database or Azure SQL Managed Instance would be optimal.



## Usage Scenarios

	SQL Server on Azure VMs IaaS	Azure SQL Managed Instance PaaS	Azure SQL Database PaaS	Azure SQL Edge Edge Compute
<b>Scenario</b>	<ul style="list-style-type: none"> <li>Rehost rich SQL apps to the current SQL Server version</li> <li>Migration of single/fewer apps to the cloud</li> <li>Rehost sunset applications</li> </ul>	<ul style="list-style-type: none"> <li>Modernization and migration of existing SQL applications to the newest SQL Server version with minimal code changes.</li> </ul>	<ul style="list-style-type: none"> <li>Build modern cloud applications on the newest SQL Server version</li> </ul>	<ul style="list-style-type: none"> <li>Securely store, analyze, and stream data at the edge</li> <li>Extend SQL applications to IoT edge devices</li> </ul>
<b>Benefits</b>	<ul style="list-style-type: none"> <li>100% SQL Server compatibility</li> <li>Full control on the OS and/or SQL Server level</li> <li>Hybrid high availability/disaster recovery scenarios</li> <li>SSRS, SSAS, and SSIS support</li> </ul>	<ul style="list-style-type: none"> <li>Rich, instance-centric programming model</li> <li>Fully-managed, no patching or maintenance required</li> <li>VNet integration</li> <li>AI-driven performance and security</li> </ul>	<ul style="list-style-type: none"> <li>Simplicity and flexibility of SLA-backed deployments and scale</li> <li>AI-driven performance and security</li> <li>Available serverless compute and Hyperscale storage capabilities</li> <li>Fully-managed, no maintenance required</li> </ul>	<ul style="list-style-type: none"> <li>Flexible connected, disconnected, and hybrid deployments</li> <li>Built-in data streaming, time series, and AI capabilities</li> <li>Consistent SQL code and native data movement</li> </ul>

Figure 4. Azure SQL Usage Scenarios



## Data Science and AI with Azure SQL on Lenovo ThinkAgile MX Series



### Why Lenovo?

Lenovo is a US\$70 billion revenue Fortune Global 500 company serving customers in 180 markets around the world. Focused on a bold vision to deliver smarter technology for all, we are developing world-changing technologies that power (through devices and infrastructure) and empower (through solutions, services and software) millions of customers every day.

### For More Information

To learn more about this Lenovo solution contact your Lenovo Business Partner or visit:

<https://www.lenovo.com/us/en/servers-storage/solutions/>

Lenovo Cloud Marketplace - [link](#)

### References:

Lenovo ThinkAgile MX Series – [link](#)

Azure SQL – [link](#)

Azure Migration Center – [link](#)

© 2022 Lenovo. All rights reserved.

**Availability:** Offers, prices, specifications and availability may change without notice. Lenovo is not responsible for photographic or typographical errors. **Warranty:** For a copy of applicable warranties, write to: Lenovo Warranty Information, 1009 Think Place, Morrisville, NC, 27560, Lenovo makes no representation or warranty regarding third party products or services. **Trademarks:** Lenovo, the Lenovo logo, System x, ThinkServer are trademarks or registered trademarks of Lenovo. The Intel logo, Xeon and Xeon Inside are registered trademarks of Intel Corporation in the U.S. and other countries. Other company, product, and service names may be trademarks or service marks of others.

CRN: CLDMS03XX22

09/2022