



NEW STARS  
OF DATA

# Introduction to Azure SQL Database Serverless

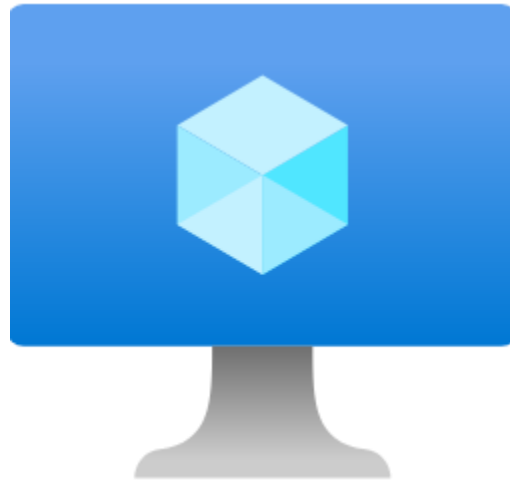
# Shaun Atkinson

- 10 years experience working with SQL Server
- Co-Organiser of the Manchester data platform user group in the uk and co-organiser of SQL Saturday Manchester
- Volunteer at SQL Bits



# SQL Server Options in Azure

Virtual Machine



Azure SQL Managed Instance



Azure SQL DB



# Basic, Standard, Premium Service Tiers

- DTU based purchase model
- 5 DTU to 4000 DTU's
  - Basic – Dev, low demand workload
  - Standard – Typical production workloads
  - Premium – High I/O intensive workloads
- Billed on DTU's and storage provisioned regardless of it being used

# Azure SQL Database Provisioned

- General Purpose service tier
- vCore based purchase model
- Regular usage databases with high compute utilisation or multiple databases with elastic pools
- Manual or scripted scaling / downtime
- Billed Per Hour
- Billed for vCores and database storage provisioned regardless of it being used

# Azure SQL Database Serverless

- vCore based purchase model
- Optimises price-performance for workloads
- Automatic Scaling / No downtime
- Automatic Pause & Resume
- Billed per second
- Billed for vCores used and storage provisioned
- Same Azure SQL Database features

# Memory

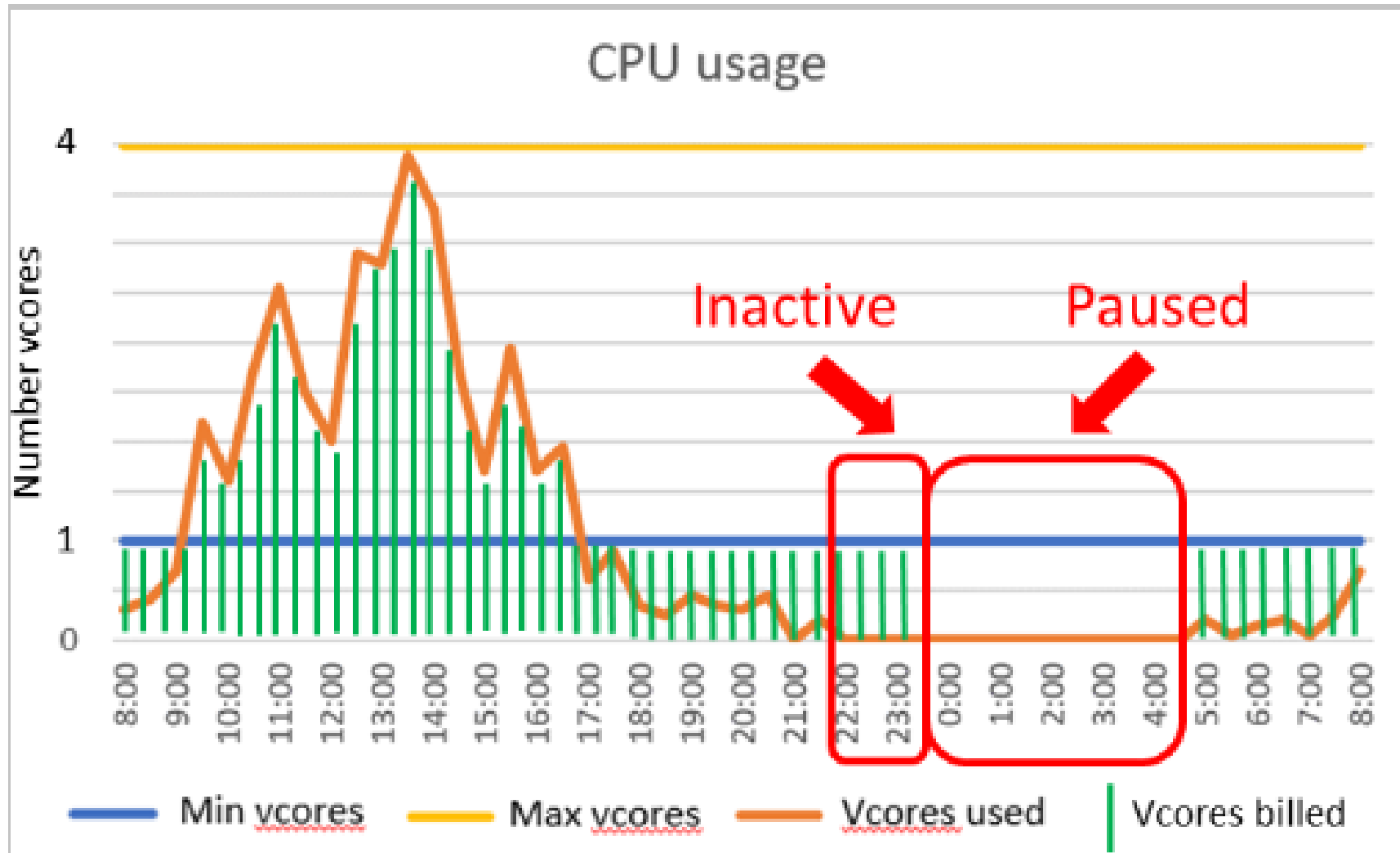
- Per Vcore
- Allocated based on min and max vCores configured
- Memory is reclaimed from SQL cache when CPU or cache use is low and falls below a threshold for a period of time
- Reduces costs but can impact performance

# Serverless Use Cases

- Intermittent unpredictable usage with low compute or long periods of inactivity
- New databases where usage pattern maybe unknown
- Provisioned databases that are scaled up and down regularly
- Demo databases



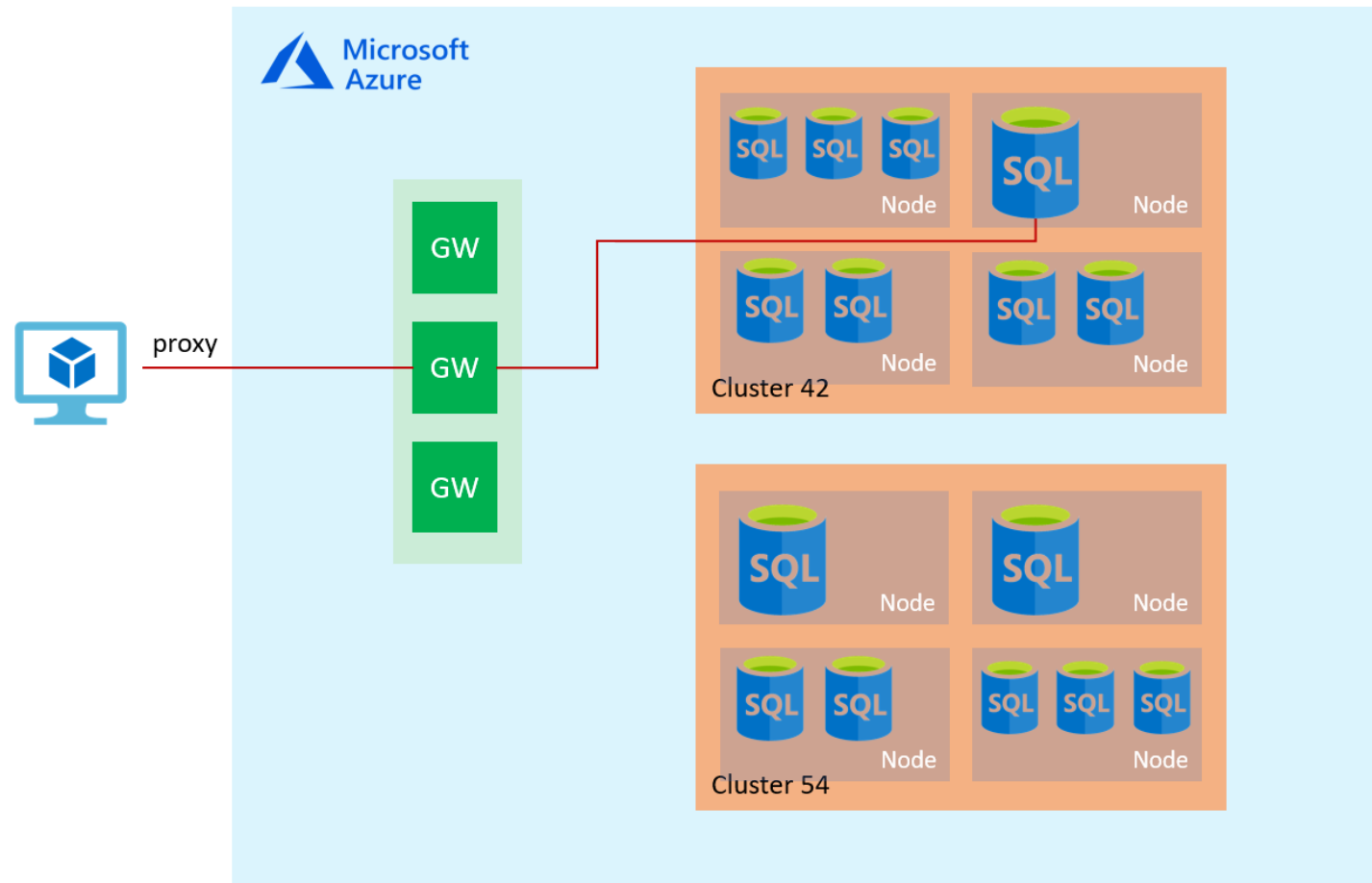
# Serverless Workload Example



# Auto-Scaling

- Min and Max vCores provisioned provide the range for auto-scaling
- Min vCores is 0.5, max vCores 40
- Max Memory 120GB
- Max Storage 4TB
- Memory and I/O range available for autoscaling depends on workload and number of vCores provisioned
- Scaling and Compute disconnected so auto-scaling unaffected by database size
- Potential delay if internal load balancing is required

# Azure SQL Database Architecture



<https://docs.microsoft.com/en-us/azure/azure-sql/database/connectivity-architecture>

# Auto-Resume

- Auto-Resume is triggered by an event e.g. a login
- Auto-Resume takes 30 to 45 seconds on average
- If database is paused the first connection will Auto-Resume the database is likely to fail
- Set application timeout between 30 and 60 seconds
- Running a query against a paused database in SSMS 18.1 or earlier will Auto-Resume all Serverless databases on that server

# Auto-Pause Delay

- Auto-Pause Delay auto-pauses a database after a period of inactivity
- The Auto-Pause Delay period of inactivity begins:
  - Sessions = 0
  - CPU Usage = 0
- Auto-Pause Delay can be disabled

# Features That Don't Support Auto-Pause

- Geo-replication (active geo-replication and auto-failover groups).
- Long-term backup retention (LTR).
- The sync database used in SQL data sync.
- The job database used in elastic jobs.
- Service updates

Demo

# Limitations

- If database is paused first connection attempt is likely to fail
- Supported on General Purpose only, Not supported on Hyperscale or Business Critical
- Not available for Managed Instance
- Only supported on Gen 5 hardware
- Doesn't work in national clouds
- Azure Hybrid Benefit (AHB) and reserved capacity discounts do not apply



# The Future

- Hyperscale support
- Pause / Resume Improvements
- Increased Processor support beyond Gen 5 (M-Series & FSv2 Series)
- Build-in retry logic
- Availability zone support

# Summary

- Serverless is best suited to databases that have intermittent workloads or test / dev / demo databases that are rarely used
- Only available on General Purpose tier and Gen5 hardware at present
- Connection failures will happen so implement retry logic
- The best thing about serverless is that it can save you money!

# Resources

- Serverless compute tier resource limits - <https://docs.microsoft.com/en-gb/azure/azure-sql/database/resource-limits-vcore-single-databases#general-purpose---serverless-compute---gen5>
- Azure SQL Database pricing - <https://azure.microsoft.com/pricing/details/sql-database/single/>
- ADO Retry Logic - <https://docs.microsoft.com/en-us/sql/connect/ado-net/step-4-connect-resiliently-sql-ado-net?view=sql-server-ver15>
- PHP Retry Logic - <https://docs.microsoft.com/en-us/sql/connect/php/step-4-connect-resiliently-to-sql-with-php?view=sql-server-ver15>

# Resources

- Stack Overflow database download - <https://www.brentozar.com/archive/2015/10/how-to-download-the-stack-overflow-database-via-bittorrent/>
- Azure SQL Database connectivity error messages - <https://docs.microsoft.com/en-us/azure/azure-sql/database/troubleshoot-common-errors-issues>
- Serverless Overview - <https://docs.microsoft.com/en-us/azure/azure-sql/database/serverless-tier-overview>