

Azure Data Week

Presenter:
Sean Mikha

SEAN MIKHA

Open Source Technology Solutions Professional – Data Platform
Home: San Diego, California

- Specialize in Data Warehouse, Big Data, NoSQL, and Advanced Analytics workloads
 - 5+ year history with Microsoft
 - 10+ years experience with Hadoop & Big Data Projects
 - MCSA: Cloud Platform, MS Data Science Professional Degree
 - Avid fan of blockchain and crypto-currency technologies
-
- Twitter: @SeanMikha
 - LinkedIn: <https://www.linkedin.com/in/seanmikha>



Agenda

Time	Topic
830am-9am	Breakfast – Meet/Greet
9am-915am	Intros/Welcome
915am-10am	CosmosDB Overview
10am-1015am	Break
1015am-12pm	CosmosDB Labs
12pm-1pm	Lunch
1pm-145pm	Azure DB for MySQL/Postgres Overview
145pm-2pm	Break
2pm-345pm	MySQL/Postgres Labs
345pm-415pm	Wrap-up Q&A

GitHub

- Agenda
- Registration Links
- Cities
- Challenges

Let's Get Started!

1. Visit <http://bit.ly/data-hackfest> to obtain the challenge info

Details

- Access to the lab environment is using Remote Desktop Connection
- You will utilize both Windows and Linux environments

Let's Get Started!

SAVE THE INFO PROVIDED WHEN YOU REACH THIS SCREEN!

- ✓ Your On Demand Lab is ready (2 hour(s), 57 minute(s) remaining)

Here are your credentials to login to [Microsoft Azure](#) and access the On Demand Lab

Username : odl_user_5057@gbbossteamoutlook.onmicrosoft.com

Password : fkme84SHY*Qb

Name

password

Value

Microsoft

vncServerURL

<http://liftshift-vmip-7bktmalbtg3rs.eastus.cloudapp.azure.com:6080>

Service Principal Details:

Name

Application Id

Value

faea1e6f-eef4-4e5d-b808-75341aabe5e7

Application Display Name

https://odl_user_sp_5057

Application Secret Key

tvip06TWG*iD

Subscription Id

b23accae-e655-44e6-a08d-85fb5f1bb854

Tenant Id

12c5db39-b62e-4301-b848-09acda2692a5

Tenant Domain Name

gbbossteamoutlook.onmicrosoft.com

Lab Guide : <https://github.com/stuartatmicrosoft/Azure-Linux-Migration-Workshop>

[DELETE ON DEMAND LAB](#)



Azure Cosmos DB

Globally distributed, multi-model database service

Sean Mikha – SeMikha@microsoft.com

Open Source Data Solutions Lead

Intelligent Cloud – Azure Global Black Belt

Microsoft Americas



Developing planet-scale apps comes with planet-scale challenges

-  **Write accurate, globally distributed apps**
-  **Managing and versioning complex schemas**
-  **Scaling both throughput and storage based on global demand**
-  **Balancing the needs for strong and eventual consistency**
-  **Delivering highly-responsive experiences**
-  **Ensuring an always-on system**

Put your data where your users are



Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

Global distribution

- Available in all Azure regions
- Multi-homing APIs
- Comprehensive SLA
- Manual and automatic failover
- Automatic & synchronous multi-region replication

Automatically replicate all your data around the world – across more regions than Amazon and Google combined



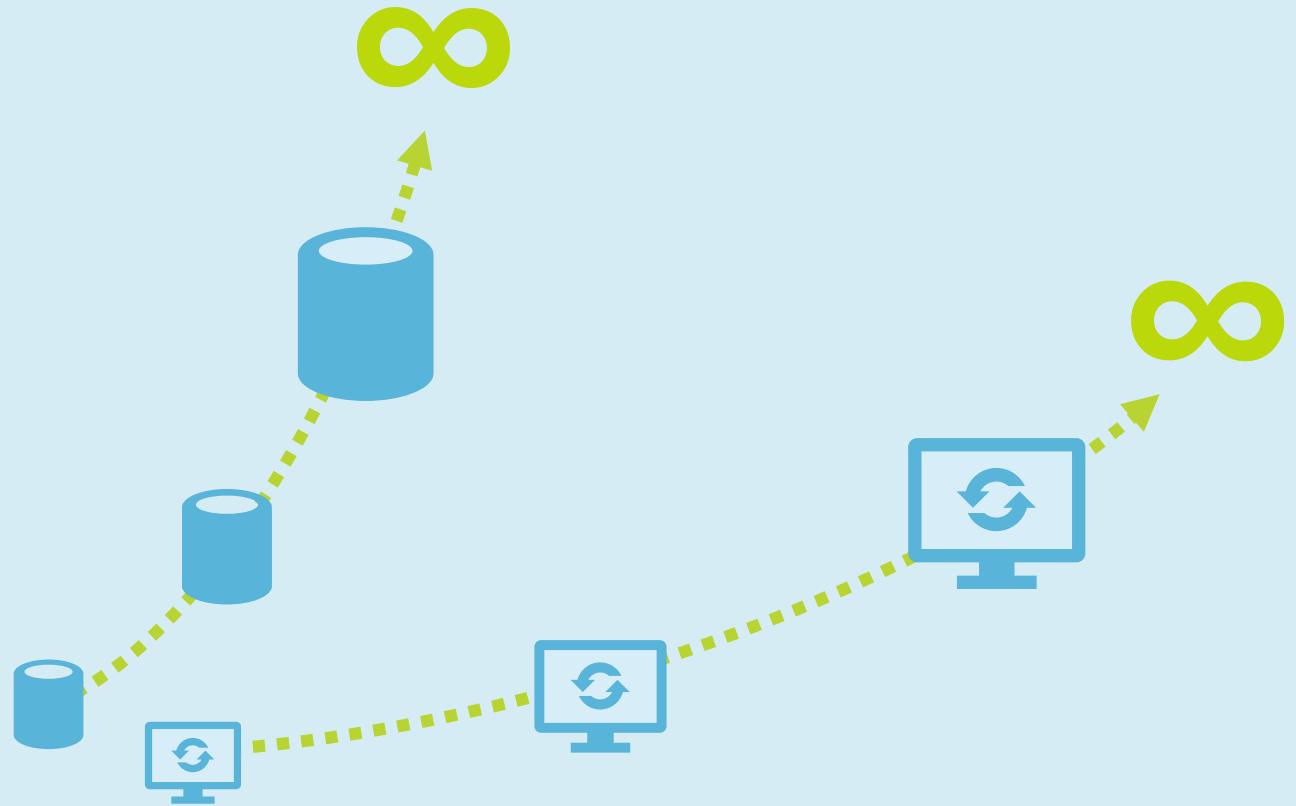
Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

Global distribution

Elastic scale-out

Independently and elastically scale storage and throughput across regions



Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

Global distribution

Elastic scale-out

Guaranteed single-digit latency

Serve <10 ms read and <15 ms write requests at the 99th percentile from the nearest region while delivering data globally



Guaranteed global millisecond latency at the 99th percentile

Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

Global distribution

Elastic scale-out

Guaranteed single-digit latency

Choice of consistency

Choose from five defined consistency levels for low latency and high availability



Strong



Bounded-stateless



Session



Consistent prefix



Eventual



Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

Global distribution

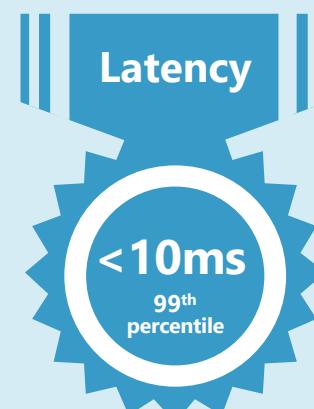
Elastic scale-out

Choice of consistency

Guaranteed single-digit latency

Enterprise-level SLAs

Only service with financially-backed SLAs for millisecond latency at the 99th percentile, 99.99% HA and guaranteed throughput and consistency



Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

Global distribution

Elastic scale-out

Guaranteed single-digit latency

Choice of consistency

Enterprise-level SLAs

Multi-model + multi API

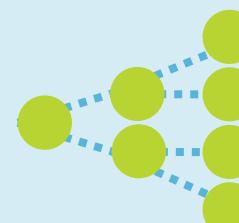
Use key-value, graph, and document with a schema-agnostic service that doesn't require any schema or secondary indexes



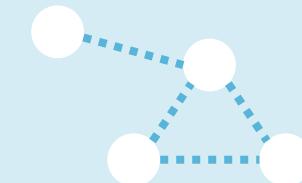
KEY-VALUE



COLUMN-FAMILY



DOCUMENT



GRAPH

Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service



Multi-model + multi API

Cosmos DB offers a multitude of APIs to access and query data including, SQL and various popular OSS APIs.



SQL

Table API



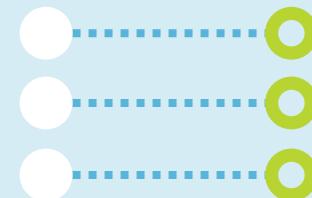
Introducing Azure Cosmos DB

A globally distributed, massively scalable, multi-model database service

Multi-model + multi API

- Database engine operates on atom-record-sequence (ARS) based type system
 - All data models are efficiently translated to ARS
- API and wire protocols are supported via extensible modules
- Instance of a given data model can be materialized as trees
- Graph, documents, key-value, column-family, ... *more to come*

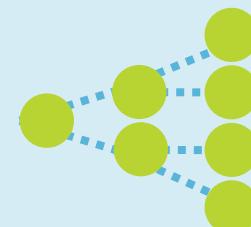
Use key-value, graph, and document with a schema-agnostic service that doesn't require any schema or secondary indexes



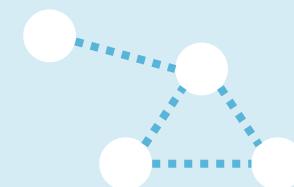
KEY-VALUE



COLUMN-FAMILY

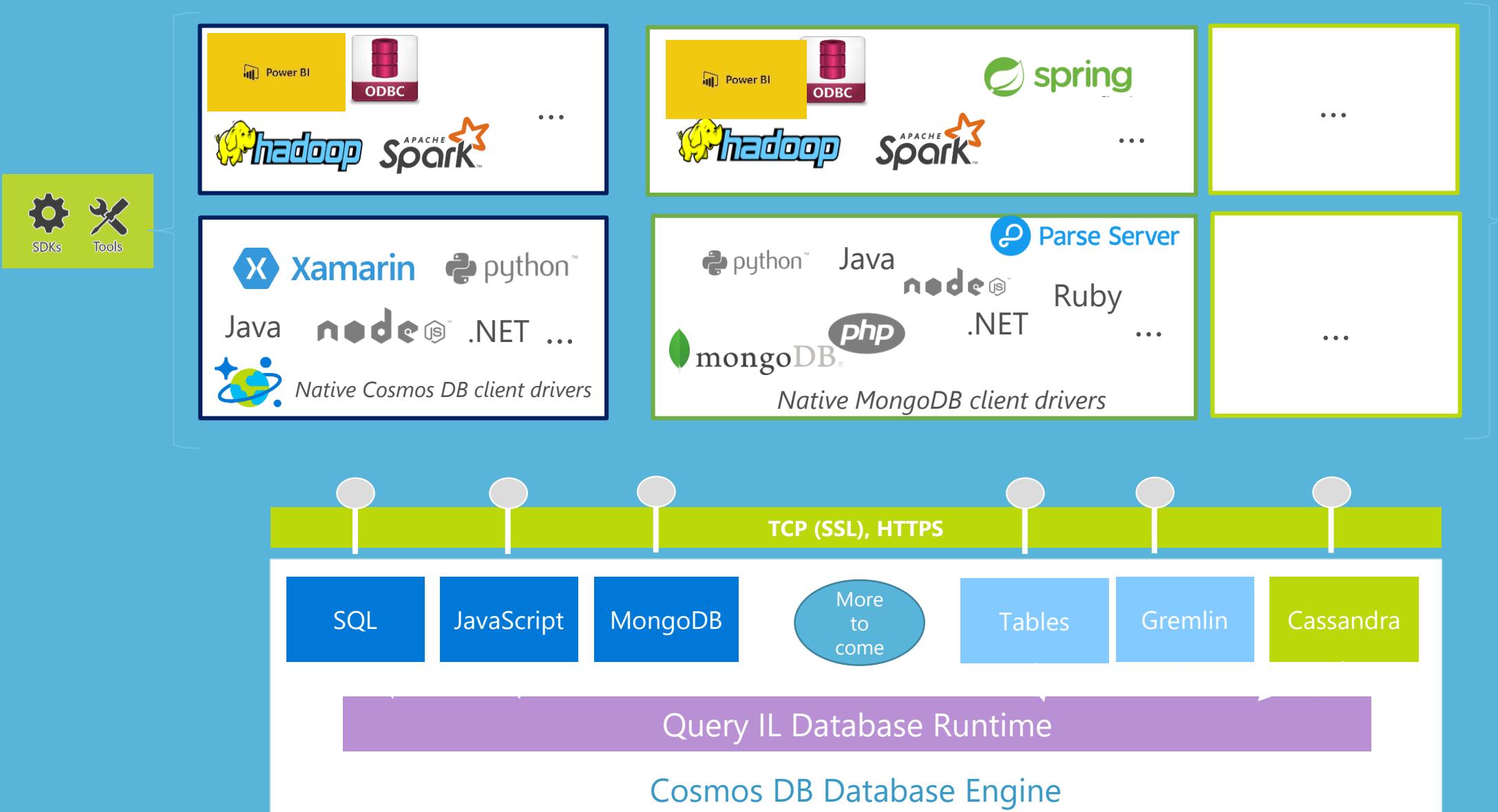


DOCUMENT



GRAPH

Native Support for Multiple APIs, formats & Wire Protocols



Security & Compliance

Enterprise grade security

Encryption at Rest

- Always encrypted at rest and in motion
- Data, index, backups, and attachments encrypted

Encryption is enabled automatically by default

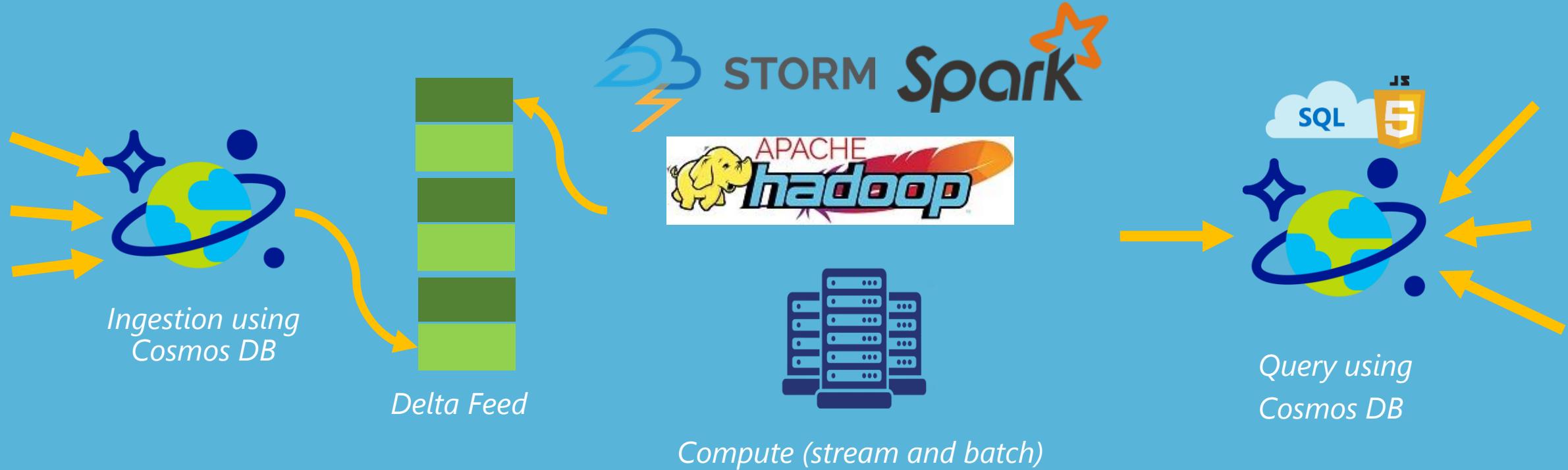
- No impact on performance, throughput or availability
- Transparent to your application

Comprehensive Azure compliance certification

- ISO 27001, ISO 27018, EUMC, HIPAA, PCI
- SOC1 and SOC2 (Audit complete, Certification in Q2 2017)
- FedRAMP, IRS 1075, UK Official (IL2) (Q2 2017)
- HITRUST (H2 2017)



Cosmos DB Change Feed



Lambda pattern with significantly lower TCO
Single scalable database solution for **both** ingestion and query

Spark Connector to Azure Cosmos DB



RDD and Dataset-based connectors available

Native integration with Spark SQL

Direct mapping to Cosmos DB partitions

Natively leverage Cosmos DB index

Predicate pushdown

Available now

Schema-agnostic, automatic indexing

d content - hash, range, geo

ine with latch free and log

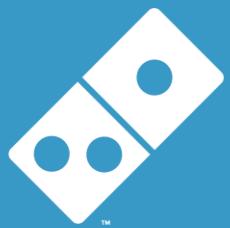
ma-extraction is built in

e

Powering global solutions

Field-tested by Microsoft's planet-scale services and industry-leading enterprises apps

Globally-distributed mission-critical apps



Domino's

Digital ordering rapidly processed across four continents

IoT



TOYOTA

Massive volumes of car sensor data drives customer service and vehicle diagnostics

Personalization



Realmadrid

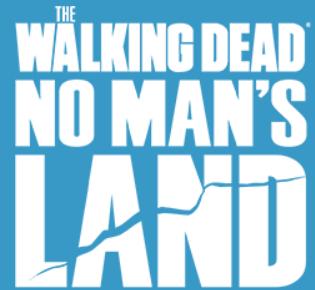
Customized, real-time experiences for fans through live and digital interactions

Retail and e-commerce



Heavy volume managed with real-time change feeds and elastic scaling

Gaming



Seamless gameplay despite traffic spikes from launch through growth

Next Steps

Getting Started

cosmosdb.com

portal.azure.com

aka.ms/cosmosdb

aka.ms/cosmosdb-Tables

aka.ms/cosmosdb-Graph

aka.ms/cosmosdb-MongoDB

aka.ms/cosmosdb-DocumentDB

cosmosdb.com/capacityplanner

Download

aka.ms/CosmosDB-emulator

Re-visit Build session recordings on [Channel 9](#).

Continue your education at [Microsoft Virtual Academy](#) online.



Azure Database for MySQL and PostgreSQL: Open Source Commitment to Choice in the Cloud

Sean Mikha – SeMikha@microsoft.com

[Open Source Data Solutions Lead](#)

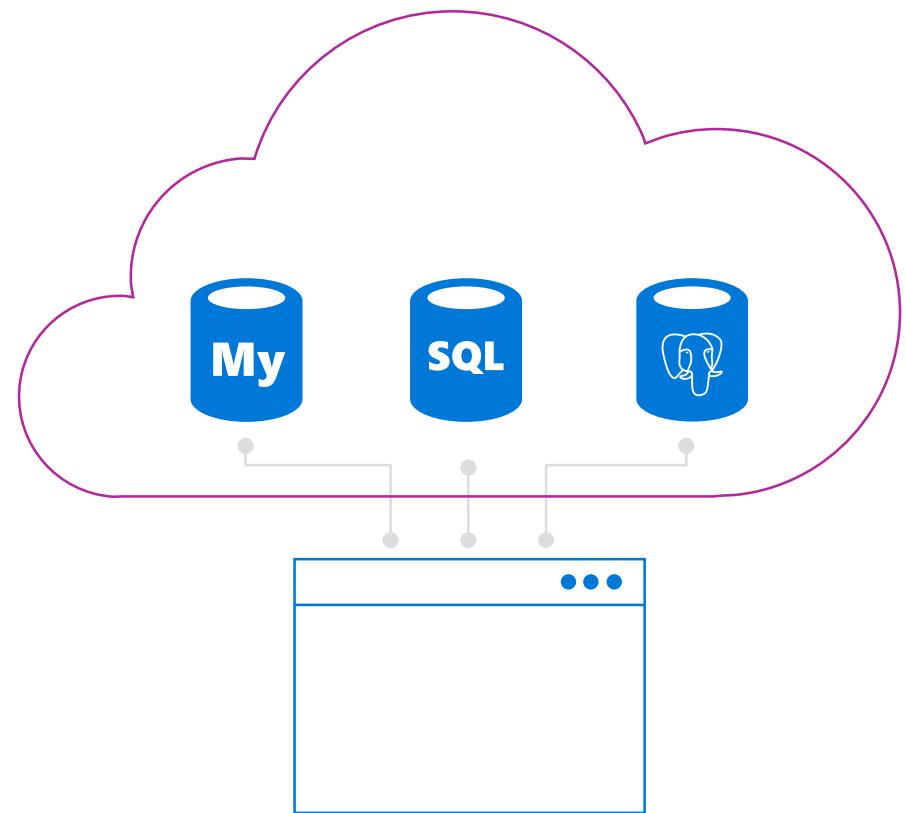
Intelligent Cloud – Azure Global Black Belt

Microsoft Americas



Azure database for Mysql and Postgresql

- Work in the DB of your choice
- Choices that enable you to focus on your app
- Create with built-in high availability
- Set up in minutes, scale on the fly
- Sustain performance with adaptive improvements
- Rest easy with unparalleled security out-of-the-box
- Automatic backups with storage management for recovery to any point up to 35 days
- Continue to use same tools, drivers and libraries



Build apps with your choice of tools and languages

Simplify and optimize with the support of all major tools, frameworks, and languages you already use

Integrated across Azure for seamless developer productivity

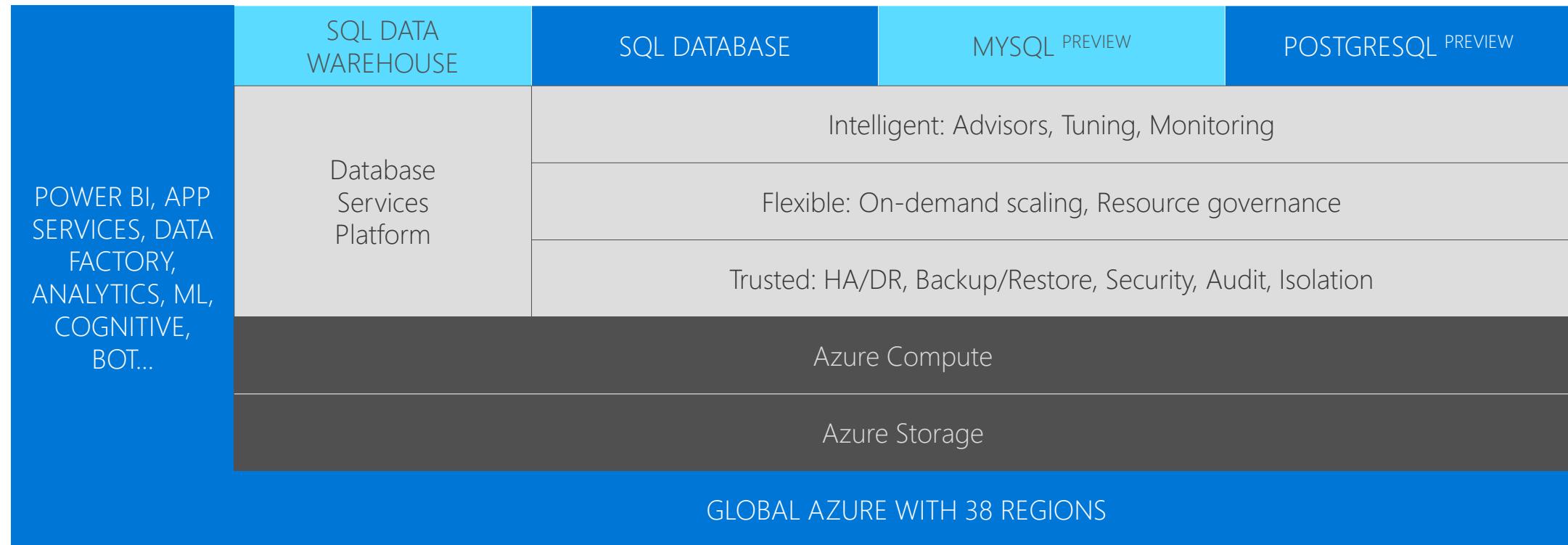
FRAMEWORKS



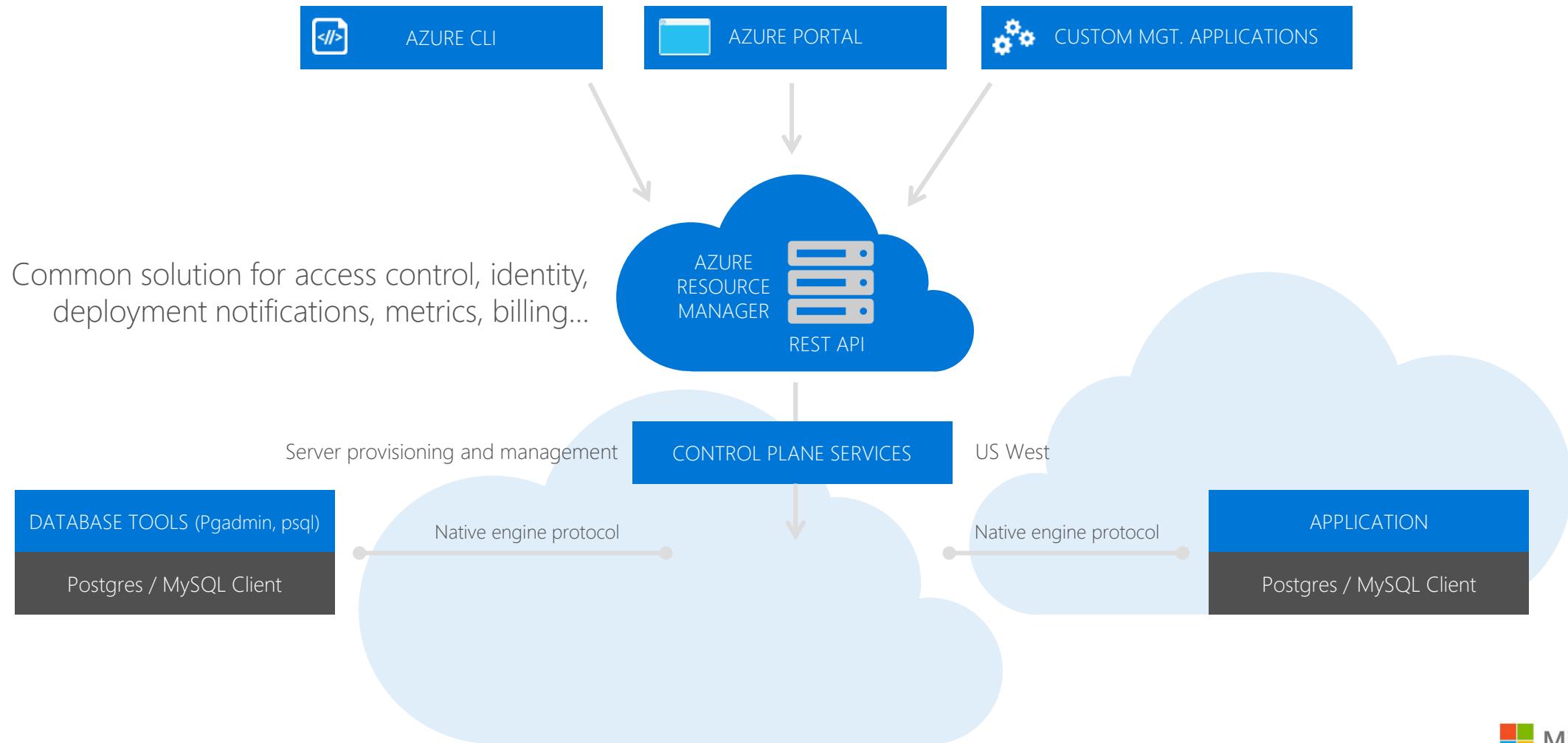
LANGUAGES



Azure Relational Database Platform



Provision, connect and manage Postgres/MySQL server



High availability built-in

- Virtually no app down time

- No need

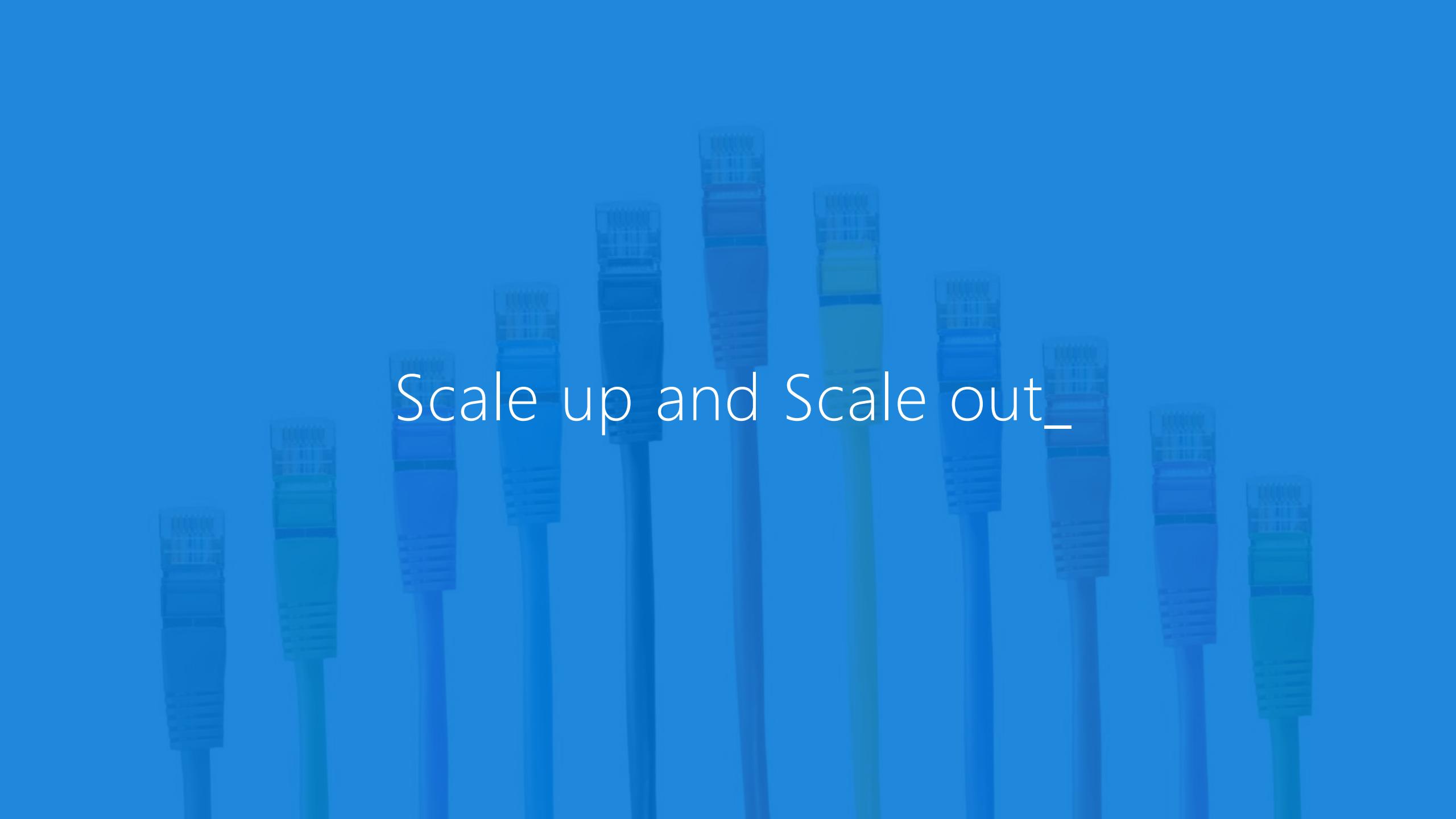
- Save cos

Scale performance on the fly

- Virtually no app down time

- Configure

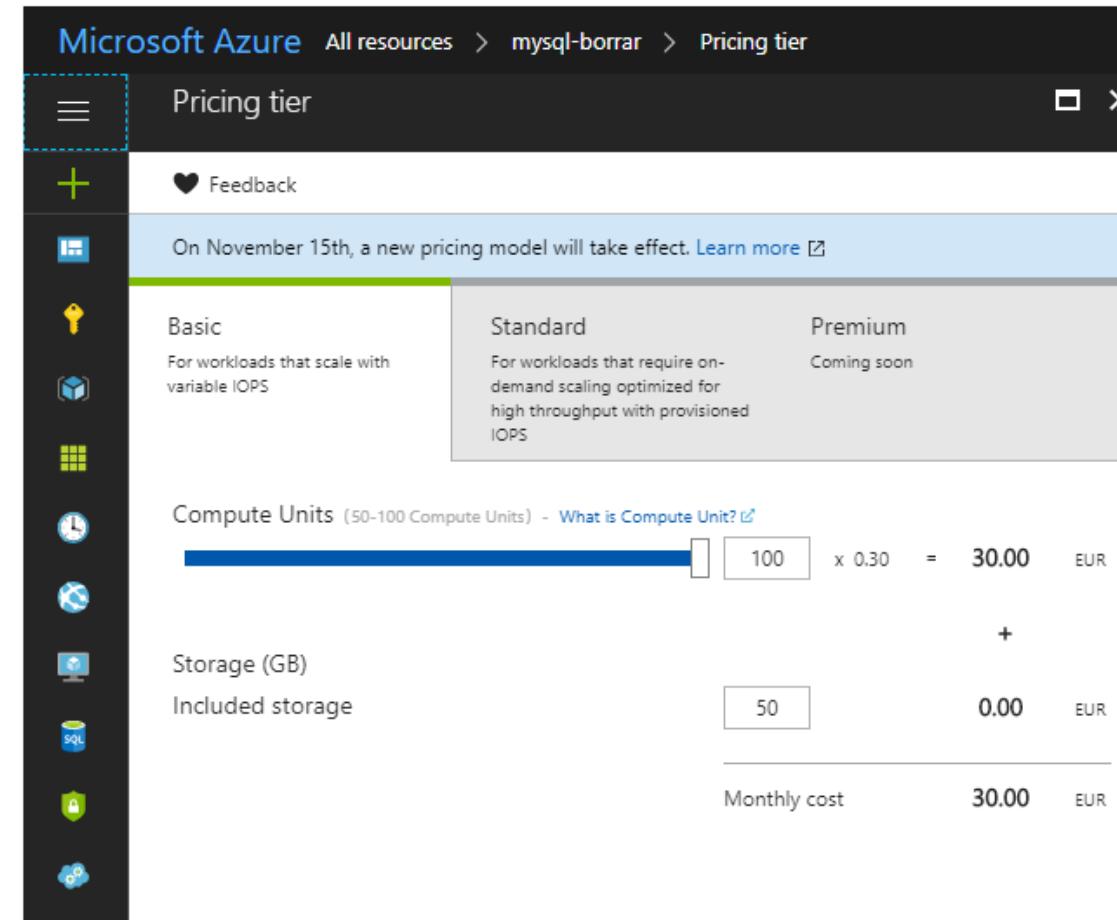
- Save costs



Scale up and Scale out_

Scale up/down using Azure Portal

1. Select Tier
2. Set Compute Units
3. Set Storage



Scale out using MySQL Replication

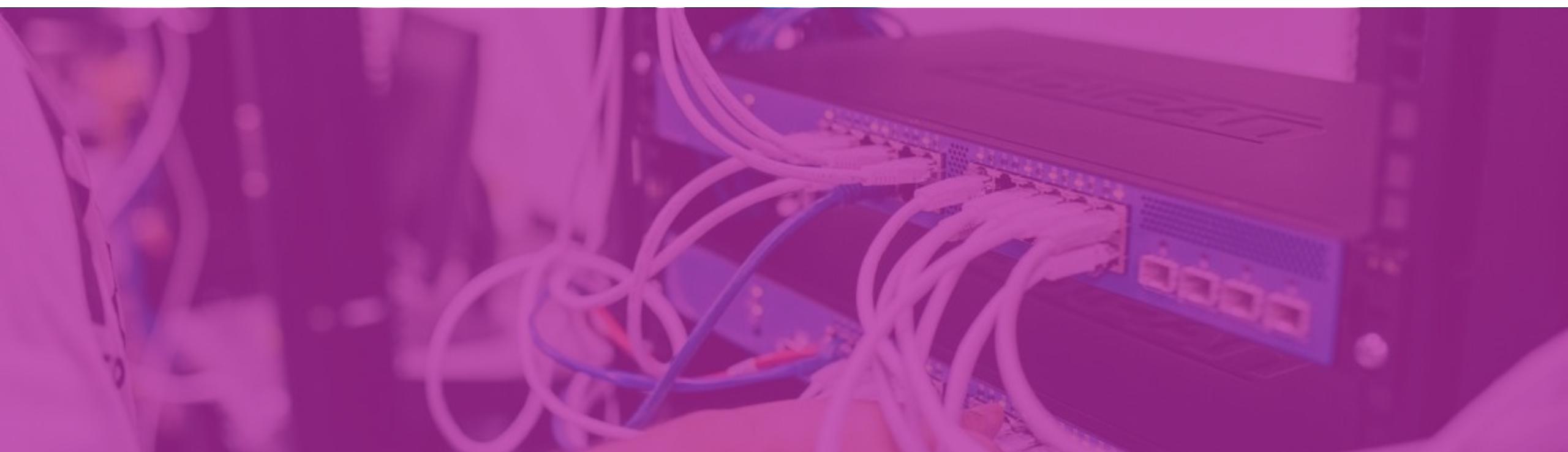
1. Get bin log position from Master
2. Connect Replica to Master using current bin log position
3. Start performance

REPLICATION BENEFITS

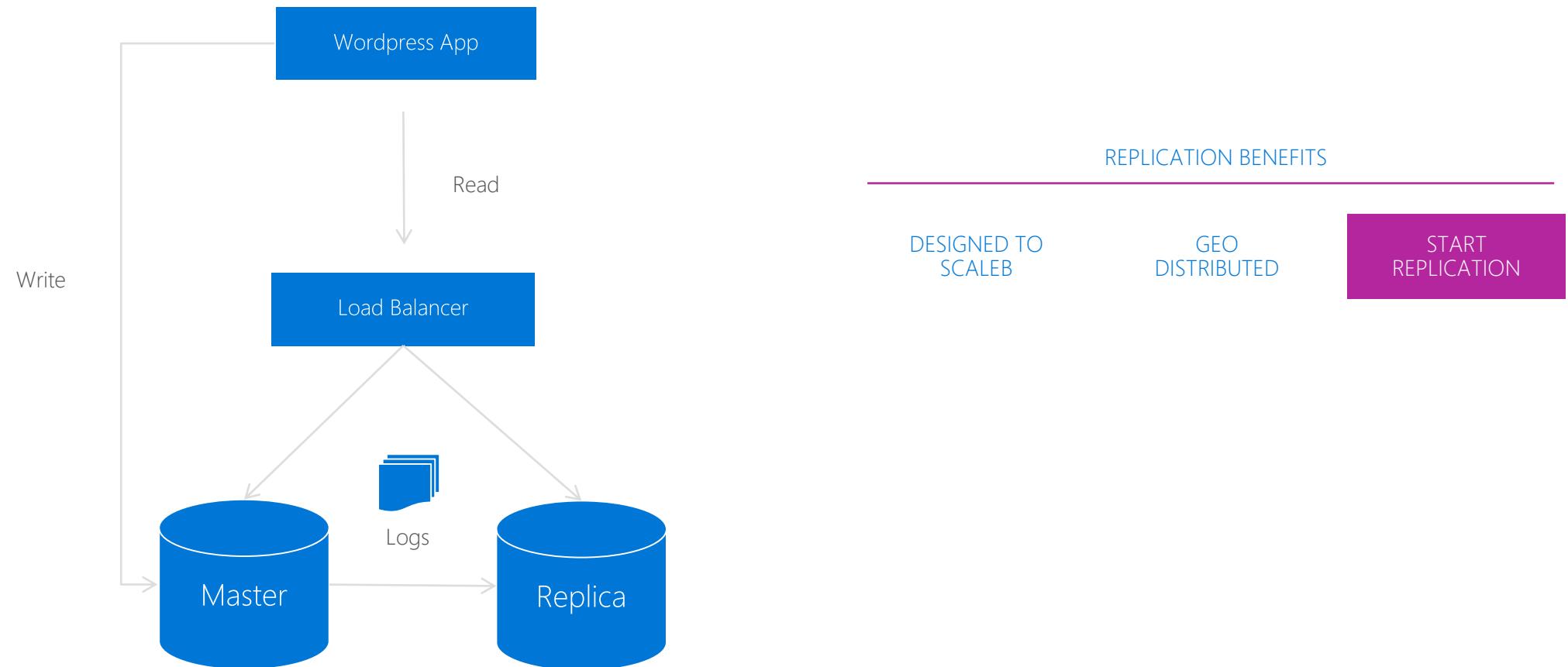
DESIGNED TO
SCALE

GEO
DISTRIBUTED

START
REPLICATION



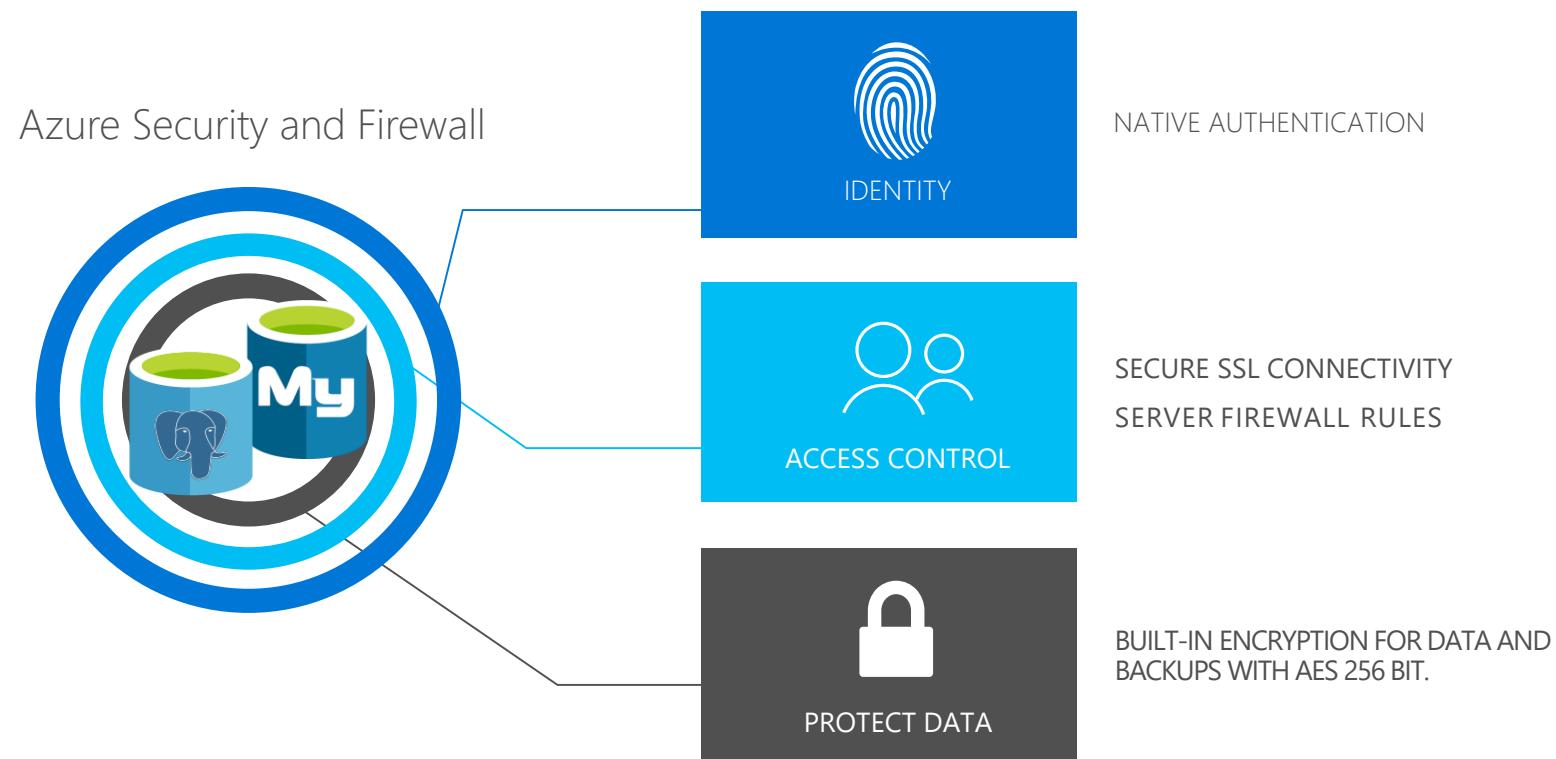
Scale out using MySQL replication



Security Options_



Security built in





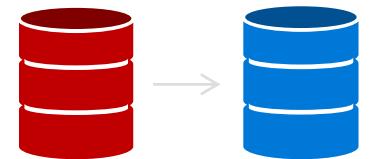
Using Attunity Replicate
to migrate database_

Accelerate migrations to the Microsoft Data Platform

Speed up database migrations now!

1. Attunity Replicate helps organizations migrate data quickly and easily with virtual no downtime
2. No additional software on source or target
3. No extra costs for Microsoft customers

Migration Source	Supported Targets
Oracle Database	Microsoft SQL Server, Azure SQL Database, Azure Database for PostgreSQL
PostgreSQL	Azure Database for PostgreSQL
MySQL	Microsoft SQL Server, Azure SQL Database, Azure Database for MySQL
Data warehouse workloads from Oracle	Azure SQL Data Warehouse
Teradata	Azure SQL Data Warehouse
Sybase ASE	Microsoft SQL Server, Azure SQL Database
IBM Netezza	Azure SQL Data Warehouse
AWS Redshift	Azure SQL Data Warehouse



For more information visit
<https://aka.ms/attunity-replicate>

Limited to database migrations only and cannot be used as a standalone product similar to Attunity Replicate.

Database migrations should be completed within a twelve-month period. The twelve months use limitation per data migration will be implemented and governed by means of the applicable end user license agreement.



The background features a dark blue gradient with two distinct patterns. On the left, there are several thin, light blue wavy lines that curve upwards from bottom-left to top-right. On the right, there is a grid-like pattern of small, semi-transparent blue circles that appear to be glowing with a light blue light.

Performance Options_

Basic Tier

Personal blogs, lightweight websites, dev and test

SERVICE TIER		BASIC
INTENDED USE CASE		Built for workloads with light compute needs and variable IO performance
COMPUTE UNITS		50, 100
STORAGE (INCLUDED)		50GB Magnetic Media
IOPS (INCLUDED)		Variable
ADDITIONAL OPTIONS		
STORAGE		Scale up to 1 TB
IOPS		NA

Standard

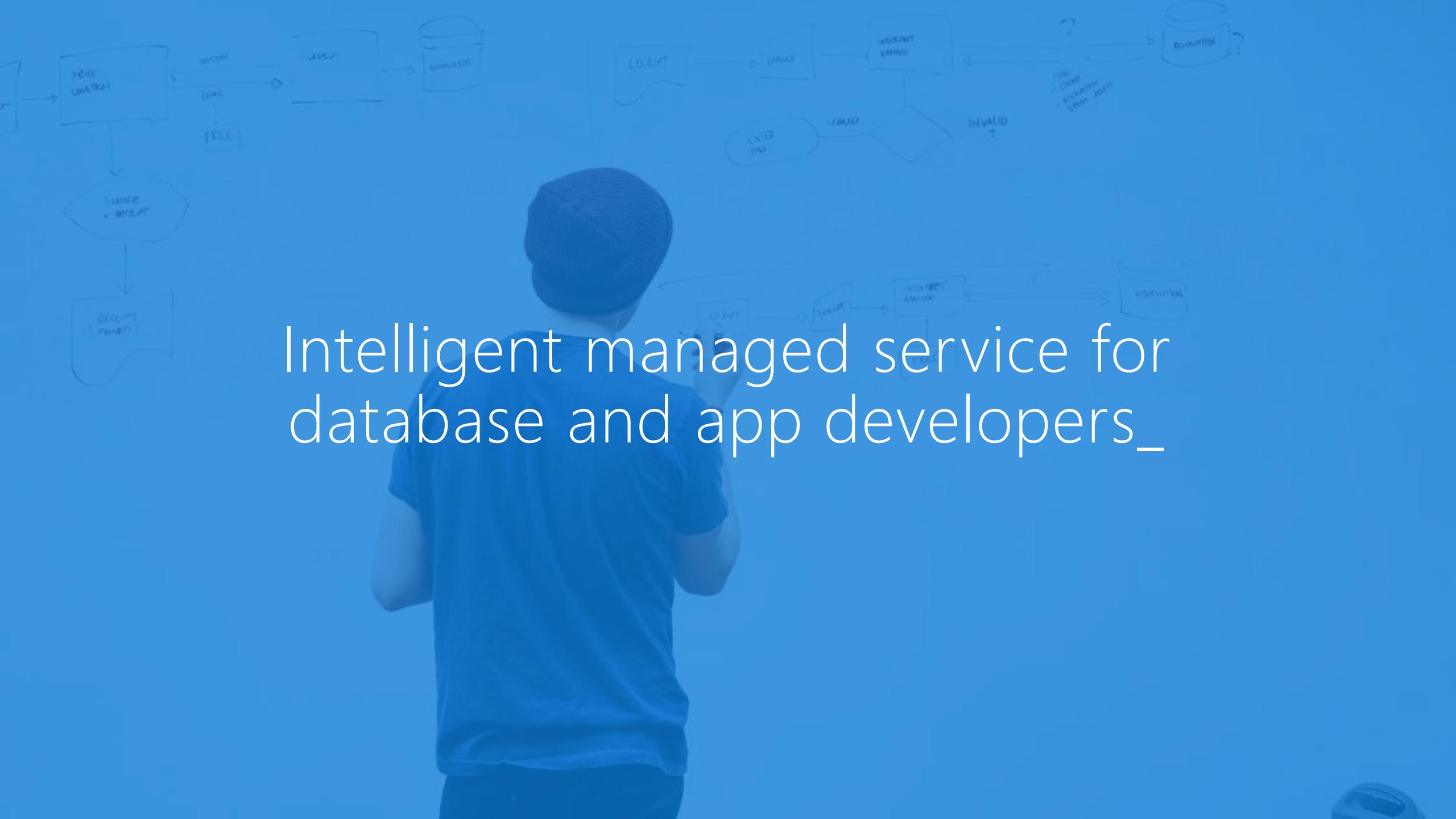
Production apps that need low IO latency

SERVICE TIER	BASIC	STANDARD BALANCED IO AND COMPUTE
INTENDED USE CASE	Built for workloads with light compute needs and variable IO performance	Ideal for most business workloads offering balanced and scalable compute and storage options
COMPUTE UNITS	50, 100	100, 200, ... 2000
STORAGE (INCLUDED)	50GB Magnetic Media	125GB Remote SSD
IOPS (INCLUDED)	Variable	375 Scales 3:1 (IOPS:GB)
<hr/>		
ADDITIONAL OPTIONS	-----	
Storage	Scale up to 1 TB	Scale up to 1 TB
IOPS	NA	Scales 3:1 (IOPS:GB)

Premium LoB & analytical apps that require very low (sub-second) latency

SERVICE TIER	BASIC	STANDARD BALANCED IO AND COMPUTE	PERFORMANCE OPTIMIZED IO AND MEMORY OPTIMIZED
INTENDED USE CASE	Built for workloads with light compute needs and variable IO performance	Ideal for most business workloads offering balanced and scalable compute and storage options	Ideal for highly transactional and analytical workloads requiring low disk latency and higher memory
COMPUTE UNITS	50, 100	100, 200, ... 2000	100, 200, ... 2000
STORAGE (INCLUDED)	50GB Magnetic Media	125GB Remote SSD	Local SSD
IOPS (INCLUDED)	Variable	375	
<hr/>			
ADDITIONAL OPTIONS			
STORAGE	Scale up to 1 TB	Scale up to 1 TB	Scale up to 1 TB
IOPS	NA	Scales 3:1 (IOPS:GB)	Scales 3:1 (IOPS:GB)

99.99% SLA | FULLY MANAGED | BUILT-IN HA | ONLINE PERFORMANCE SCALING



Intelligent managed service for
database and app developers_

Database Developer/Devops

- Infrastructure patching and upgrades
- Database availability
- Manage backups for recovery
- Data security
- Optimize database performance
- Monitoring and Alerting
- Troubleshooting

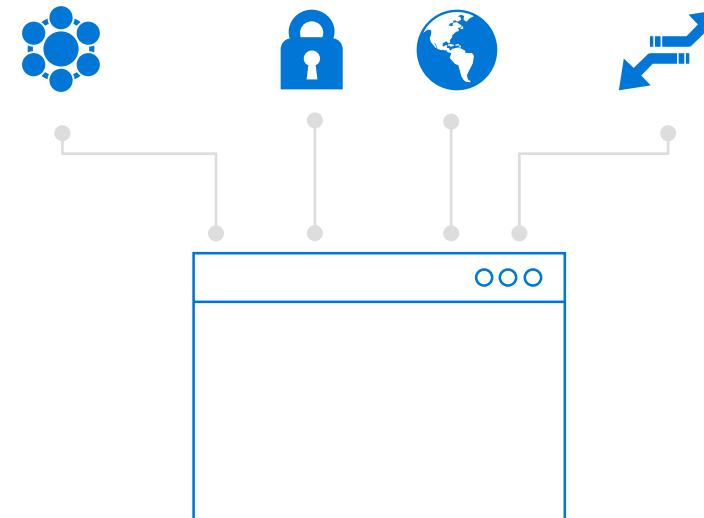
Application Developer

- Provisioning and managing databases
- No compromise availability, security and performance
- Elastic scalability on demand
- Freedom to use tools and frameworks for the task in hand
- Flexible pay-as-you-go pricing

Intelligent, managed data services for database developers

Securing your database

1. All data at-rest including backups are encrypted on disk by default with AES 256 bit encryption.
2. Connections to database are secured by default with SSL
3. No additional switch or planning required to secure your database!.

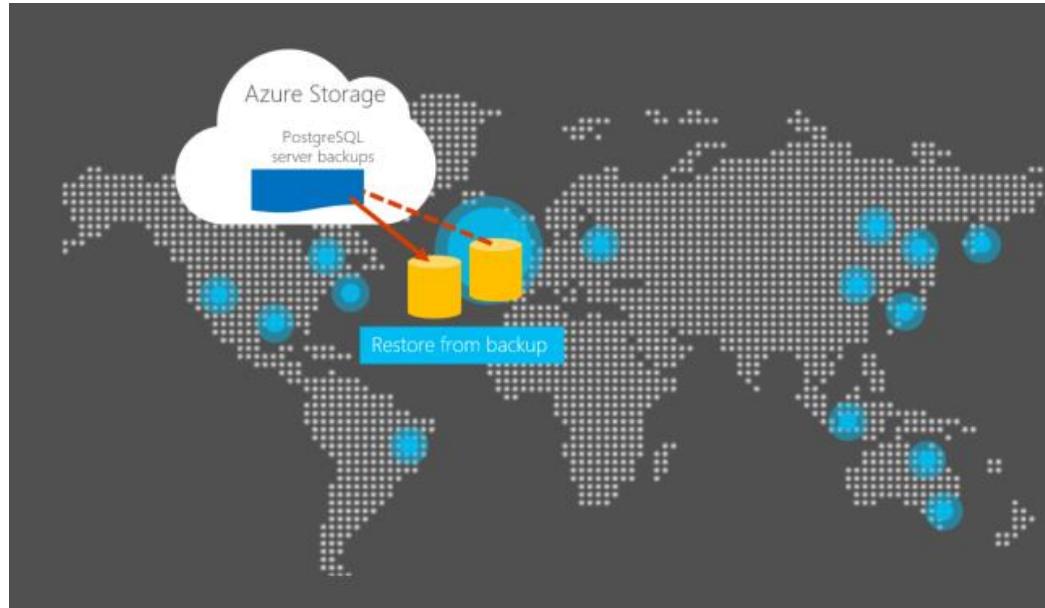


Intelligent, managed data services for database developers

Securing your database

Manage backups

1. Built-in backup with retention of backups up to 35 days for PITR for RPO < 5 minutes
2. Backups are geo-redundantly stored in another region to recover from disasters.



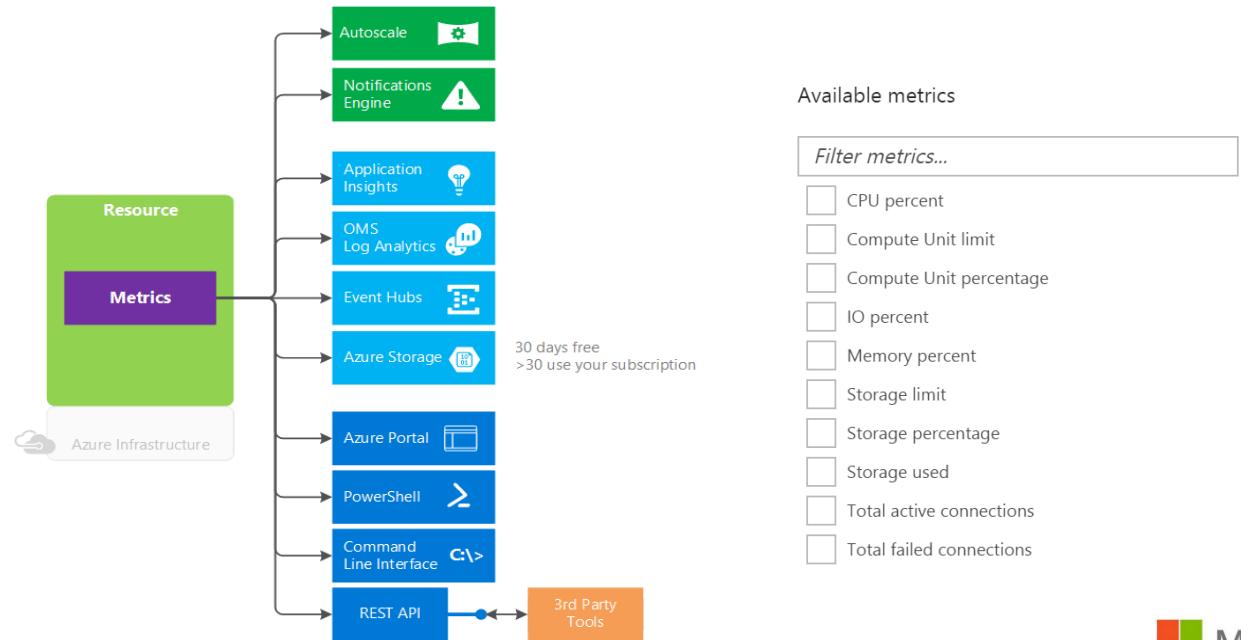
Intelligent, managed data services for database developers

Securing your database

Manage backups

Monitoring and Alerting

1. The service provides monitoring on metrics and ability to define alerts via integration with Azure Monitor service for up to 30 days.
2. Integrated with community and 3rd party monitoring tools.



Intelligent, managed data services for database developers

Securing your database

Manage backups

Monitoring and Alerting

Troubleshooting

1. Configure server log retention period for up to 7 days (consumes configured server storage)
2. Default enablement of Postgres pg_stat_statements for access to Postgres metrics and performance schema for MySQL metrics.

The screenshot shows the Azure portal interface for managing database logs. On the left, there's a sidebar with various navigation items like Overview, Activity log, Tags, Connection security, Connection strings, Server parameters, Pricing tier, Properties, Locks, Metrics, Alert rules, and the 'Server logs' item, which is highlighted with a red box. The main content area has a header with a blue bar containing the text 'Click here to enable logs and configure log parameters →'. Below this, there's a message: 'Server logs are created every 24 hours. You will be able to access each log for up to 7 days after creation.' A search bar labeled 'Search for a log file' is present. The main table lists log files with columns for NAME, LAST UPDATE TIME, and file size (1KB). A dropdown menu next to the search bar offers options: All, Last 24 hour(s), Last 2 day(s), and Last 4 day(s). The table contains approximately 15 log entries, each with a download icon.

NAME	LAST UPDATE TIME	
postgresql-2017-08-30_220000.log	Wed, 30 Aug 2017 21:59:59 GMT	1KB
postgresql-2017-08-30_210000.log	Wed, 30 Aug 2017 21:00:00 GMT	1KB
postgresql-2017-08-30_200000.log	Wed, 30 Aug 2017 20:00:00 GMT	1KB
postgresql-2017-08-30_190000.log	Wed, 30 Aug 2017 19:00:00 GMT	1KB
postgresql-2017-08-30_180000.log	Wed, 30 Aug 2017 18:00:00 GMT	1KB
postgresql-2017-08-30_170000.log	Wed, 30 Aug 2017 17:00:00 GMT	1KB
postgresql-2017-08-30_160000.log	Wed, 30 Aug 2017 16:00:00 GMT	1KB
postgresql-2017-08-30_150000.log	Wed, 30 Aug 2017 15:00:00 GMT	1KB
postgresql-2017-08-30_140000.log	Wed, 30 Aug 2017 14:00:00 GMT	1KB
postgresql-2017-08-30_130000.log	Wed, 30 Aug 2017 13:00:00 GMT	1KB
postgresql-2017-08-30_120000.log	Wed, 30 Aug 2017 12:00:00 GMT	1KB
postgresql-2017-08-30_110000.log	Wed, 30 Aug 2017 11:00:00 GMT	1KB
postgresql-2017-08-30_100000.log	Wed, 30 Aug 2017 10:00:00 GMT	1KB
postgresql-2017-08-30_090000.log	Wed, 30 Aug 2017 09:00:00 GMT	1KB
postgresql-2017-08-30_080000.log	Wed, 30 Aug 2017 08:00:00 GMT	1KB
postgresql-2017-08-30_070000.log	Wed, 30 Aug 2017 07:00:00 GMT	1KB
postgresql-2017-08-30_060000.log	Wed, 30 Aug 2017 06:00:00 GMT	1KB

Intelligent, managed data services for database developers

Securing your database

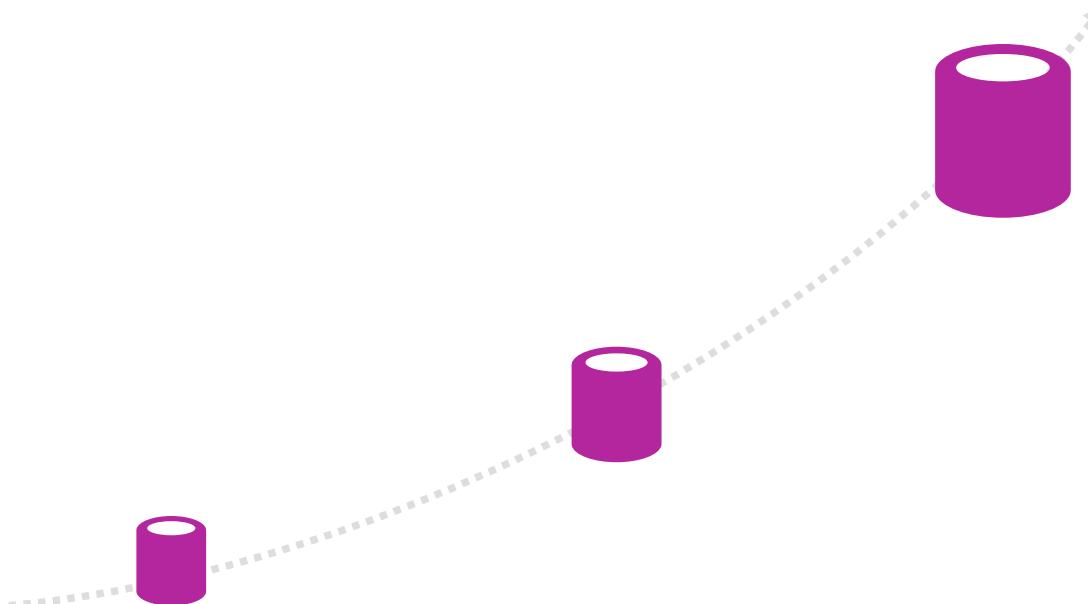
Manage backups

Monitoring and Alerting

Troubleshooting

Automated Patching

Automatic patching for Postgres and MySQL minor versions which means developers do not have to worry about managing this.



Intelligent, managed data services for database developers

Securing your database

Manage backups

Monitoring and Alerting

Troubleshooting

Automated Patching

Customize server parameters

1. Customize PostgreSQL or MySQL server parameters based on your needs
2. Ability to access server logs via portal, REST API or Azure CLI.

The screenshot shows the 'myserver4demo - Server parameters' page in the Azure portal. The left sidebar lists various management options: Overview, Activity log, Tags, SETTINGS (Connection security, Connection strings, Server parameters, Pricing tier, Properties, Locks), MONITORING (Metrics, Log Analytics), and WORKLOADS. The 'Server parameters' option is highlighted with a red box. The main content area displays a table of server parameters with their current values and descriptions. The table includes columns for Parameter Name, Value, and Description. Some parameters shown include binlog_group_commit_sync_delay, binlog_group_commit_sync_no_delay_count, character_set_server, div_precision_increment, group_concat_max_len, innodb_adaptive_hash_index, innodb_lock_wait_timeout, interactive_timeout, log_bin_trust_function_creators, log_queries_not_using_indexes, and log_slow_slave_threads.

PARAMETER NAME	VALUE	DESCRIPTION
binlog_group_commit_sync_delay	1000	Controls how many microseconds the binary log waits for all transactions in a group to commit before sending them to the slave.
binlog_group_commit_sync_no_delay_count	0	The maximum number of transactions to wait for in a group before sending them to the slave.
character_set_server	LATIN1	Use charset_name as the default server character set.
div_precision_increment	4	Number of digits by which to increase the size of the division result.
group_concat_max_len	1024	Maximum allowed result length in bytes for GROUP_CONCAT.
innodb_adaptive_hash_index	ON	Whether innodb adaptive hash indexes are enabled.
innodb_lock_wait_timeout	50	The length of time in seconds an InnoDB transaction will wait for a lock to become available.
interactive_timeout	1800	Number of seconds the server waits for active connections to send more data.
log_bin_trust_function_creators	OFF	This variable applies when binary logging is ON.
log_queries_not_using_indexes	OFF	Logs queries that are expected to retrieve a large number of rows from the table.
log_slow_slave_threads	OFF	Logs queries of statements the slave thread spends longer than the value of the long_query_time system variable.

Intelligent, managed data services for database developers

Securing your database

Manage backups

Monitoring and Alerting

Troubleshooting

Automated Patching

Customize server parameters

Add Extensions

Over 23 PostgreSQL extensions supported today with more in plan!

- address_standardizer
- address_standardizer_data_us
- btree_gin
- btree_gist
- citext
- fuzzystrmatch
- hstore
- intarray
- pgcrypto
- pgrouting
- pg_buffercache
- pg_partman
- pg_prewarm
- pg_stat_statements
- pg_trgm
- plpgsql
- postgis
- postgis_sfsgal
- postgis_tiger_geocoder
- postgis_topology
- postgres_fdw
- unaccent
- uuid-ossp

13

Azure regions
available today

West US, North Central US, East US, East US2, South Central US, West Europe, North Europe, Japan West, Japan East, East Asia, SE Asia, West India, Central India



AZURE DATABASE FOR MYSQL AND POSTGRESQL
REGION AVAILABILITY TODAY



More coming soon!

Some of our customers



GeekWire



Resources

1. Azure service page:
 - PostgreSQL: [Azure Database for PostgreSQL](#)
 - MySQL: [Azure Database for MySQL](#)
2. Documentation:
 - PostgreSQL: [Azure Database for PostgreSQL](#)
 - MySQL: [Azure Database for MySQL](#)
3. Discussion forum:
 - PostgreSQL: [MSDN](#), [StackOverflow](#)
 - MySQL: [MSDN](#), [StackOverflow](#)
4. Feedback forum:
 - PostgreSQL: [User Voice](#)
 - MySQL: [User Voice](#)
5. GitHub repo:
 - <https://github.com/Azure/azure-postgresql>
 - <https://github.com/Azure/azure-mysql>
6. Twitter: [@AzureDBPostgres](#), [@AzureDBMySQL](#)

Merci!
Thank You!

Sean Mikha – SeMikha@microsoft.
Ray Kao – Ray.Kao@microsoft.com



© Copyright Microsoft Corporation. All rights reserved.