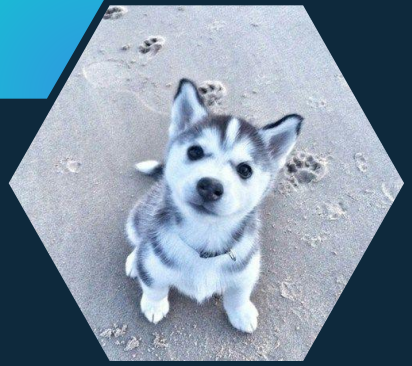
Data Technologies Overview (with Azure!)



Hello!

Charmaine Lee

Twitter: @charmaine_klee

Anthony Chu

Twitter: @nthonychu

Gora Kong

Twitter: @gorakong

Types of Data Stores

Relational

SQL Database

SQL Data Warehouse

MySQL

PostgreSQL

Non-Relational

Table Storage

Redis

Cosmos DB

Search

Big Data

Blob Storage

Data Lake Storage

HDInsight

No-DB?

Service Fabric

Why different data stores?

- Shift towards microservices
- Domain Driven Design
- “Best tool for the job”

Relational Databases

SQL Database

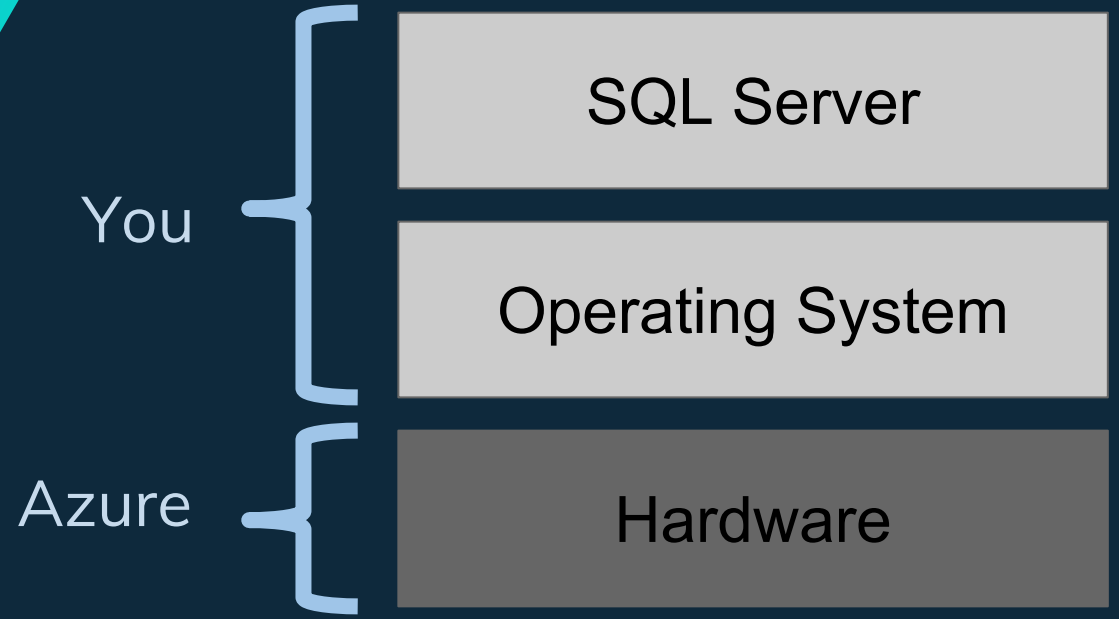
SQL Server
on premises

You

- SQL Server
- Operating System
- Hardware

SQL Database

SQL Server on VM



SQL Database

Azure

- SQL Server
- Operating System
- Hardware

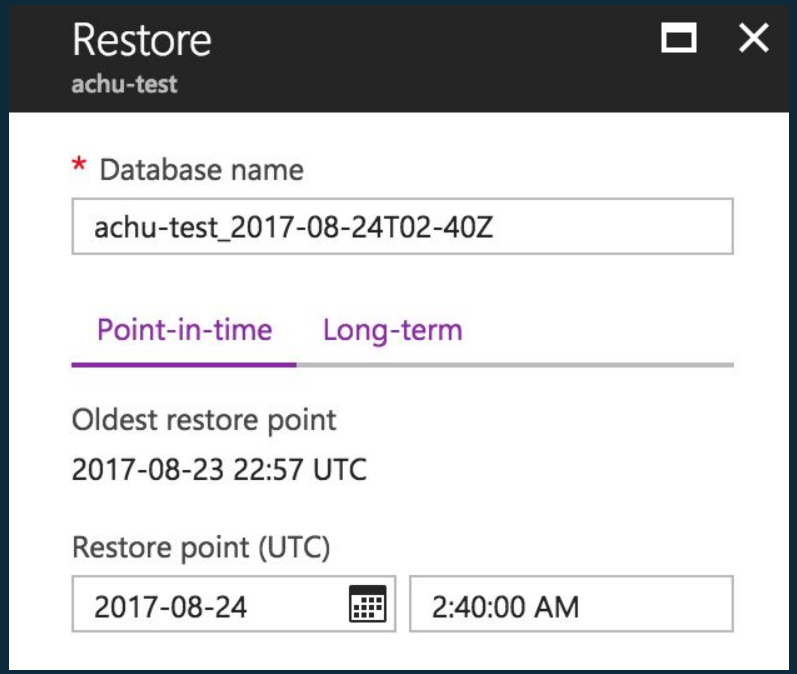
SQL Database

SQL Database

- Starts at $6/month
- Scale on the fly
- 99.99% uptime SLA

SQL Database

Point-in-time restore



SQL Database

Geo-replication

- Readable replicas
- Manual failover*

SQL Database

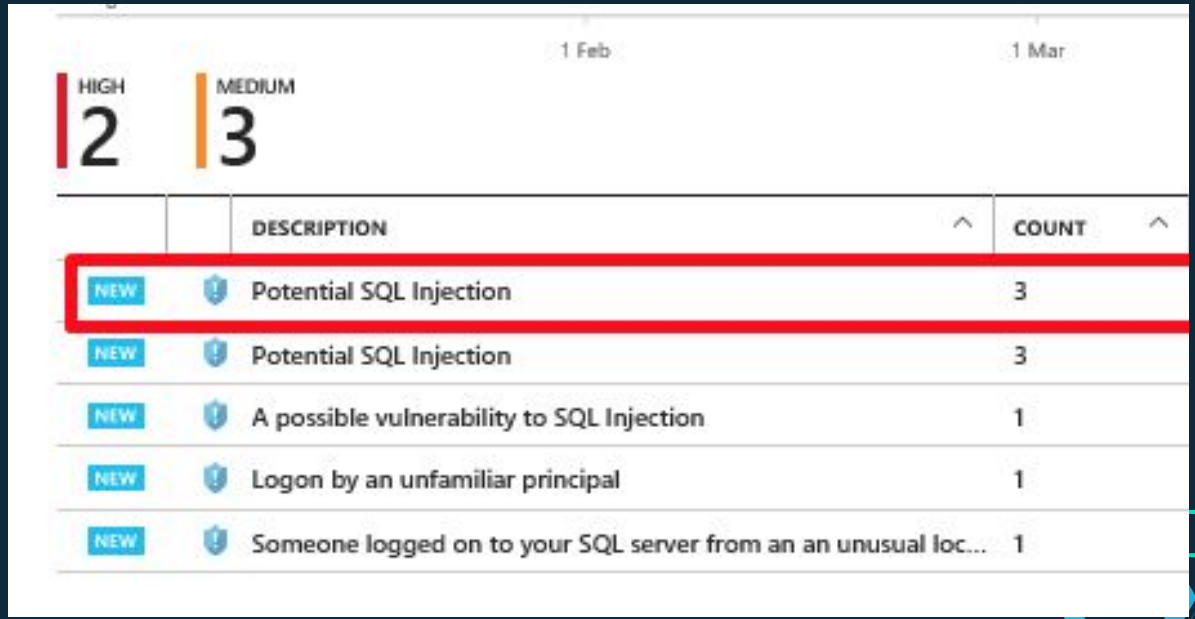
- Encryption at-rest and in-flight
- Row-level security
- Auditing

SQL Database

Dynamic data masking

SQL Database

Automatic
threat detection



SQL Database

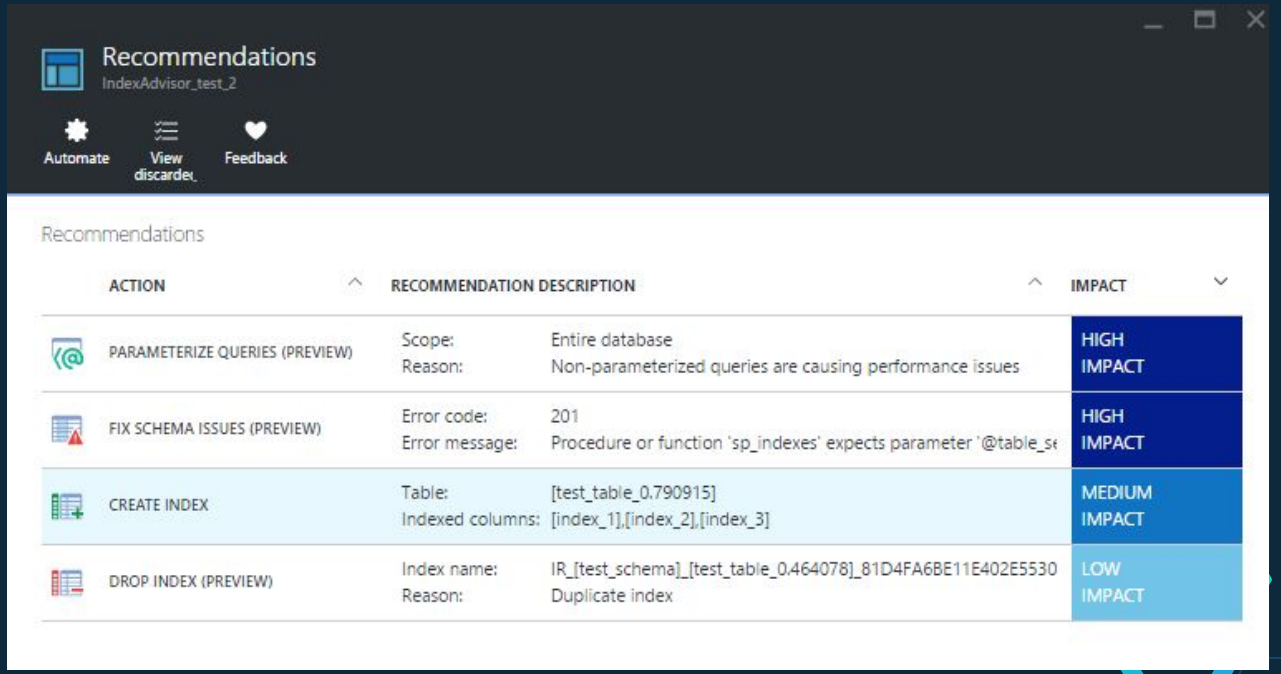
- Familiar tools
  - SQL Server Mgmt Studio
  - Visual Studio & VSCode
  - In-portal query

SQL Database

Query Performance Insight

SQL Database

Database Advisor



SQL Database

SQLDB Managed Instance

- Addresses gap between SQLDB and SQL Server
- Migration service
- Private preview

SQL Database

Stretch Database

- SQL Server 2016
- Transparently migrate cold data to the cloud
- Query “stretched” data

SQL Data Warehouse

A SQL-based, fully-managed, petabyte-scale cloud data warehouse.

Key Features:

- Scalability
- Security & Reliability
- Integrated

SQL Data Warehouse ARCHITECTURE

- Massively Parallel Processing (MPP) engine

- PolyBase

  Query Azure Blob Storage & Data Lake Store

  Import/Export data directly

SQL Data Warehouse

SCALABILITY

- Decoupled storage & compute

  Scale up/down, pay for only what you use

- Scale based on Data Warehouse Units (DWUs)

  Used computational resources

  higher DWUs → higher performance → higher costs

- Pause on demand

SQL Data Warehouse

SECURITY

- Automatic threat detection
- Authentication

RELIABILITY

- 99.9% uptime
- Disaster recovery
    - Data stored using geo-redundant blobs

SQL Data Warehouse INTEGRATED

- Use in conjunction with SQL Server tools
  - SSDT, SSIS, SSAS, SSRS
- Azure Data Factory, Machine Learning, Stream Analytics
- Power BI
- Cortana Analytics Suite
  - For building analytic solutions

MySQL / PostgreSQL

- High availability
- High security
- Elastically scales

MySQL

- Relational database service
- Create DB in minutes
- Predictable & easy scaling

MySQL

- Easy DB admin & management
  - DB patching
  - Backups
  - Monitoring
  - Security

PostgreSQL

- Object-RDBMS
- Versions 9.5 & 9.6
- Can export current data to Azure

Relational Databases

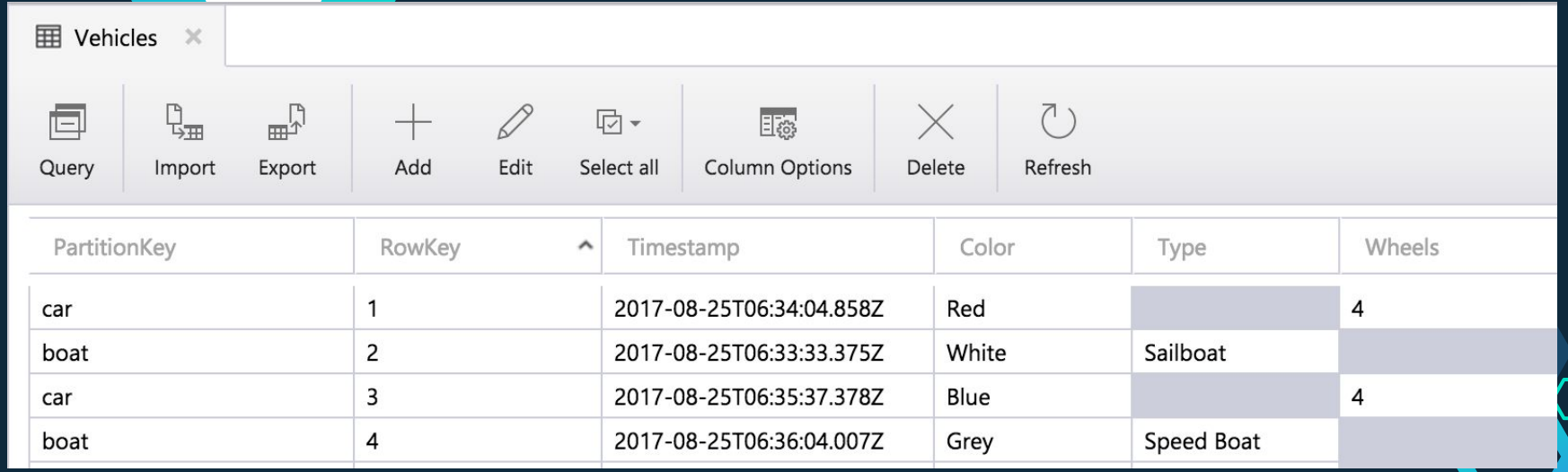
- SQL DB
- SQL Data Warehouse
- MySQL
- PostgreSQL

Non-relational Databases (NoSQL)

Table Storage

- Cost-effective key-value store
- Semi-structured data
- Flexible schema
- Single index
- Fast
- 7 cents per GB
- 0.36 cents per 100k transactions

Table Storage



Redis Cache

- Fully-managed Redis as a service
- Highly secure: Uses Redis authentication and supports SSL connections
- Scales up: Cache is available in sizes up to 530 GB (Premium)

Redis Cache

- 3 tiers:

  Basic: Single node, multiple sizes

  High-availability SLA

  Standard: Two-node primary/secondary

  Premium: All of the Standard features + more

- Data persistence (Premium only)

  → data persists even with hardware failure

Why Redis Cache?

- ## High Performance

  Low latency, high throughput Redis engine

- ## Great features

  Support for high performance data types, transactions, Lua scripting, keys with limited TTL, etc.

- ## Easy to use and manage

  Monitor health & performance of cache through Azure portal

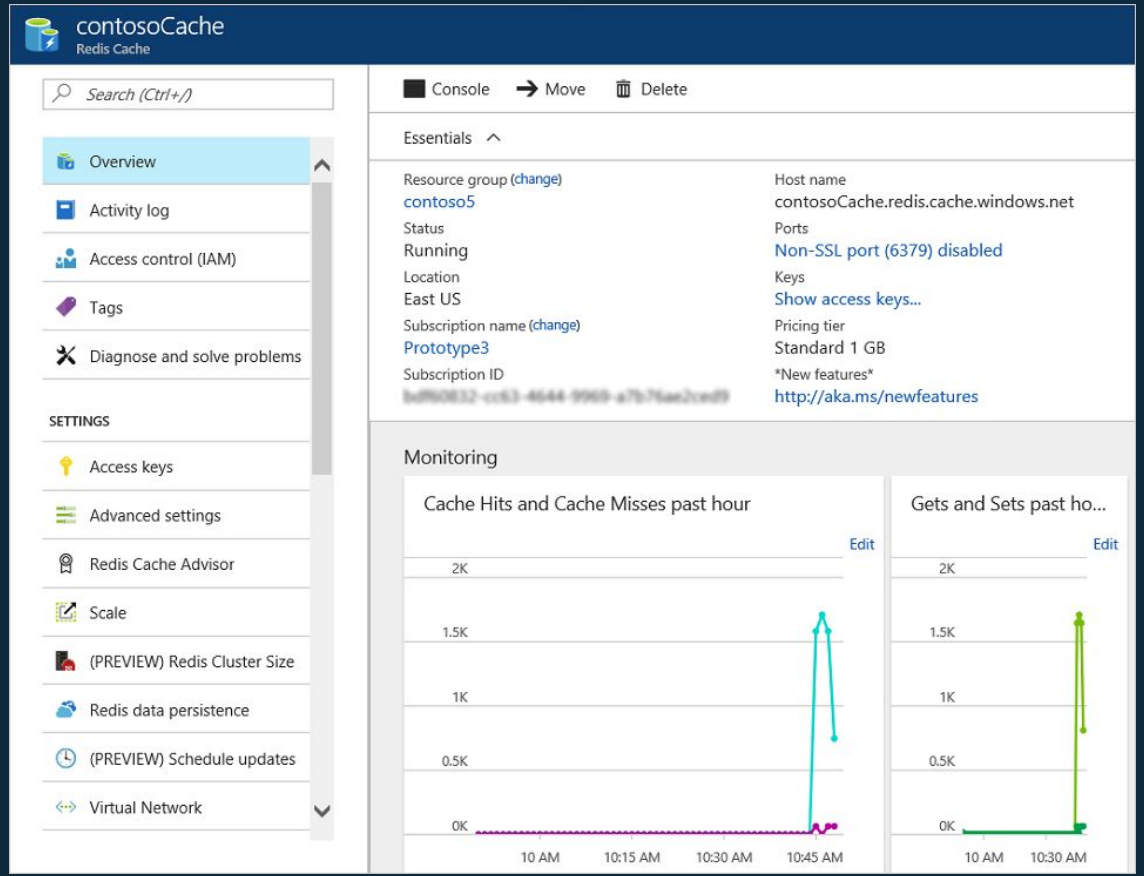
Monitoring a Redis Cache

Performance metrics that can be monitored:

- Uptime and Availability
- Cache Hits & Misses
- Get/Set Commands
- Total Operations
- Evicted & Expired Keys
- Memory, Bandwidth & CPU utilization

Full support for customization of monitored parameters, & alerts

Monitoring a Redis Cache



Cosmos DB

- One step from DocumentDB, one giant leap for databases
- Globally distributed, multi-model database service
- Low latency, high availability & consistency

Turnkey Global Distribution

- Distribute data with just one click!

1

Azure Regions

Elastic Scaling & High Availability

- Scale throughput and storage at anytime and anywhere in the world
- App is always “on”

3

Consistency

- Can choose between 5 well-defined consistency models

Consistency

Consistency Levels and guarantees

Consistency Level	Guarantees
Strong	Linearizability
Bounded Staleness	Consistent Prefix. Reads lag behind writes by k prefixes or t interval
Session	Consistent Prefix. Monotonic reads, monotonic writes, read-your-writes, write-follows-reads
Consistent Prefix	Updates returned are some prefix of all the updates, with no gaps
Eventual	Out of order reads

4 No database schema/index management

Data models

- Graph
- Key-value
- Table

APIs

- DocumentDB API
- MongoDB API
- Table API
- Graph (Gremlin) API

Azure Search

- Cloud search service for web and mobile app development

Azure Search

Features

- Full text search
  - Searching in a full-text db instead of metadata
- Data integration
  - Azure Search Indexes accept data from any source
- 56 languages
  - Handles language-specific linguistics

Azure Search

Features

- Monitoring and reporting
  - Search traffic analytics
- Geospatial support
  - Proximity-based results
- UX features
  - Search suggestions, filters, sorting, etc.

Non-Relational Databases

- Table Storage
- Redis
- Cosmos DB
- Search

Big Data

- Store everything
- Unstructured data
- Batch query and process

Blob Storage

- Cost-effective object storage
- Different access tiers
- Massively-scalable
- Geo-redundant

Data Lake Storage

- “No limits”
  - Petabyte-size files
  - Trillions of object
  - Scalable throughput
- Always encrypted, RBAC
- HDFS-compatible

Data Lake Analytics

- On-demand analytics jobs
- Query from
  - Data Lake Store
  - Blob Storage
  - SQLDB and SQLDW
- U-SQL
  - C#
  - Python / R

HDInsight

- Managed Hadoop
- Query Data Lake Store
- Spark, Storm, Kafka, HBase

Service Fabric

- Store data with the application
- Stateful Services
- Stateful Actors

Summary

Relational

SQL Database

SQL Data Warehouse

MySQL

PostgreSQL

Non-Relational

Table Storage

Redis

Cosmos DB

Search

Big Data

Blob Storage

Data Lake Storage

HDInsight

No-DB?

Service Fabric

SQLSaturday Sponsors!

Titanium
& Global Partner

Gold

Silver

Bronze

Without the generosity of these sponsors, this event would not be possible! Please, stop by the vendor booths and thank them.

Thanks!

Charmaine Lee

Twitter: @charmaine_klee

Anthony Chu

Twitter: @nthonychu

Gora Kong

Twitter: @gorakong