

DataTH X Microsoft Azure

AZURE DATA FUNDAMENTALS TRAINING

EP 2: Modern Data Analytics



ตารางเรียน Azure Data Fundamentals Training

↓ We are here

CLASS 1

วันจันทร์ 17 Jan

เวลา 7:00 - 9:00 PM

เราจะเรียนอะไร?

- Core Data Concepts
- Explore relational data in Azure



CLASS 2

วันพฤหัสบดี 20 Jan

เวลา 7:00 - 9:00 PM

เราจะเรียนอะไร?

- Explore non-relational data in Azure
- Explore modern data warehouse analytics



CLASS 3

วันจันทร์ 24 Jan

เวลา 7:00 - 9:00 PM

เราจะเรียนอะไร?

- Azure Data Architecture with Microsoft Thailand



DATA TH



Microsoft
Azure

หากมีคำถาม ถามได้ที่ไหน



ถามคำถามไว้ใน chat ด้านขวา หัวข้อ Q&A

โดยจะมาตอบช่วงหลังเบรก หรือช่วงท้าย CLASS

หากพิมพ์เข้ามาในกล่องแชททั่วไป คำถามอาจจะโดนดันจนไม่ได้ตอบ



DATATH

Microsoft
Azure

FAQ: คอร์สนี้ต้องสมัครสมาชิก Azure เพื่อเรียนมั้ย ?

สมัครหรือไม่สมัครก็ได้

ในคอร์สนี้จะเป็นคอนเซปต์ความรู้ของ Azure ซึ่งเราจะมีกดใช้บางบริการให้ดู แต่จะดูเฉย ๆ หรือทำตามเพื่อเรียนรู้ไปด้วยกันก็ได้ (ขออนุญาตไม่หยุดรอ เพราะไม่ใช่ Workshop)

ถ้าทำตาม อย่าลืมกดลบบริการที่สร้างมาทุกครั้งหลังจากจบคลาส ไมอย่างนั้นถ้าลืมลบบริการไหน อาจจะโดนหักเงินไปเรื่อย ๆ



DATATH



Microsoft
Azure

WARNING: THAI & ENGLISH CONTENT



เนื้อหาจะมีทั้งภาษาไทย เพื่อความเข้าใจง่าย

**และบางศัพท์จะขอใช้ภาษาอังกฤษ เพื่อให้ทุกท่านคุ้น
ชินกับตอนสอบ Certificate ที่จะเจอศัพท์ภาษาอังกฤษ**



CLASS 2

CH 3: EXPLORE

NON-RELATIONAL DATA

1 HR

- BREAK 5 min -

5 MIN

CH 4: MODERN DATA WAREHOUSE
ANALYTICS

1 HR

Q & A

15 MIN



AZURE DATA FUNDAMENTALS TRAINING

CH3: NON-RELATIONAL DATA



Chapter 3



เรียนรู้เกี่ยวกับบริการ non-relational data บน Azure



เรียนรู้เกี่ยวกับการ provision บริการ non-relational data บน Azure



ทดลองเล่นกับข้อมูลแบบ non-relational data stores บน Azure

Lesson 1: เรียนรู้เกี่ยวกับบริการ non-relational data บน Azure



Lesson 1 objectives



use-cases และการใช้งาน Azure Table storage



use-cases และการใช้งาน Azure Blob storage



use-cases และการใช้งาน Azure File storage



use-cases และการใช้งาน Azure Cosmos DB

มาทำความรู้จักกับ Azure Table storage

Key (Customer ID)	Value (Customer Data)					
C1	AAAAA	BBB	101 Block Street		YY	999 888
C2	MM	NN	21 A Street	5 B Avenue		
C3	DDD	EEE	FFF	111	222	66 C Road

ดูเผิน ๆ Azure table storage คล้ายกับ SQL DB มาก ความต่าง คือ แต่ละแถวจะมีคอลัมน์อะไรก็ได้ ไม่ต้องเหมือนกัน และไม่มีคอนเซปต์ relational data เช่น relationship, foreign key, index เลย

มาทำความรู้จักกับ Azure Blob storage

Block blobs

- ❑ ขนาดการเก็บข้อมูลสูงสุด 4.7TB
- ❑ เหมาะกับการเก็บไฟล์ขนาดใหญ่ที่มีการเปลี่ยนแปลงนาน ๆ ครั้ง
- ❑ แบ่งข้อมูลออกเป็นกล่อง ๆ ที่จุได้ 100 mb เรียกว่า block
- ❑ 1 Block blob จะจัดเก็บ block ได้สูงสุด 50,000 blocks

Page blobs

- ❑ ขนาดการเก็บข้อมูลสูงสุด 8TB
- ❑ แบ่งข้อมูลออกเป็นกล่อง ๆ ที่จุได้ 512 byte เรียกว่า page
- ❑ ใช้สำหรับ virtual disk storage สำหรับ virtual machines

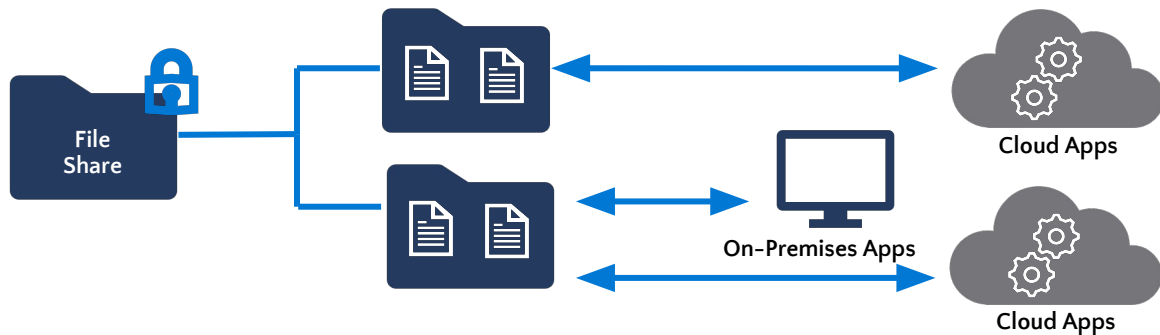
Append blobs

- ❑ ขนาดการเก็บข้อมูลสูงสุด 195GB
- ❑ เป็น block blobs ที่ใช้เพื่อให้ append ข้อมูลได้รวดเร็ว
- ❑ แบ่งข้อมูลออกเป็นกล่อง ๆ ที่จุได้ 4 mb

มาทำความรู้จักกับ Azure File storage

Azure File Storage จะสร้าง File Share บน Cloud ช่วยให้เราสามารถแชร์ไฟล์หากันได้จากทุกที่ทั่วโลก ที่มีอินเทอร์เน็ต

- ใช้ระบบชื่อ Server Message Block 3.0 (SMB) เพื่อแชร์ไฟล์
- สามารถแชร์ไฟล์ได้สูงสุดถึง 100 TB ใน 1 storage account
- บริการแบบ Fully managed - ข้อมูลของเราจะถูกทำสำเนา (replicate) และมีการเข้ารหัสไฟล์ทั้งหมดที่ถูกจัดเก็บ (encryption at rest)



Azure Cosmos DB คืออะไร?

Azure Cosmos DB คือบริการฐานข้อมูลแบบ NoSQL ที่รวดเร็วสำหรับผู้ใช้จากทั่วโลก

- Cosmos DB จัดการข้อมูลแบบเป็น Document ที่ Partition (ทำงานเหมือนการ Index ใน RDBMS)
- เข้าถึงข้อมูลได้แบบ Real-time ด้วยการอ่านและเขียนข้อมูลที่รวดเร็วมากกกกกก
- Document แต่ละอัน เก็บข้อมูลได้สูงสุด 2 MB
- ใช้ร่วมกับ Azure Blob Storage เพื่อเก็บข้อมูลที่มีขนาดใหญ่ มาก ๆ เป็น Blob ได้
- ใช้ข้อดีของ Azure ในด้านการ Scale และการเก็บข้อมูล เพื่อให้ได้ประสิทธิภาพสูง



ตัวอย่างการใช้งานจริง Azure Cosmos DB

Web and retail

ด้วยฟีเจอร์ multi-master replication ของ Azure Cosmos DB (เหมือน Read Replica แต่ใช้กับการเขียนได้ด้วย)

ทำให้ Data Engineer จะสามารถสร้างเว็บและมือถือที่ทำงานได้ภายใน 10ms ทุกที่บนโลก

Gaming

การจะพัฒนาเกมให้สนุกได้ ต้องมีการทำ personalized content เช่น สถิติในเกม, การเชื่อมต่อกับ social media, ตารางคะแนน, ตารางอันดับผู้เล่น

ซึ่งจำเป็นต้องมีฐานข้อมูลที่ทำงานได้รวดเร็ว และรองรับผู้เล่นที่มาจากทั่วโลกได้

IoT scenarios

ปัจจุบันในตลาด IOT (Internet of things) หรืออุปกรณ์ที่ช่วยสร้างข้อมูลเซนเซอร์ หลายแสนเครื่องได้รับการออกแบบและถูกจัดจำหน่ายไปมากมายทั่วโลก

เทคโนโลยีเช่น Azure IoT Hub ช่วยให้ Data Engineer สามารถสร้างระบบจัดเก็บข้อมูลจาก IoT ได้ และถูกส่งมาเก็บแบบ Real-time โดยการใช้ Azure Cosmos DB

Lesson 1: Knowledge check (continued on next slide)



What are the elements of an Azure Table storage key?

- ☐ Table name and column name
 - ☐ Partition key and row key
 - ☐ Row number
-



When should you use a block blob, and when should you use a page blob?

- ☐ Use a block blob for unstructured data that requires random access to perform reads and writes. Use a page blob for discrete objects that rarely change.
 - ☐ Use a block blob for active data stored using the Hot data access tier, and a page blob for data stored using the Cool or Archive data access tiers
 - ☐ Use a page block for blobs that require random read and write access. Use a block blob for discrete objects that change infrequently
-



Why might you use Azure File storage?

- ☐ To share files that are stored on-premises with users located at other sites
- ☐ To enable users at different sites to share files
- ☐ To store large binary data files containing images or other unstructured data

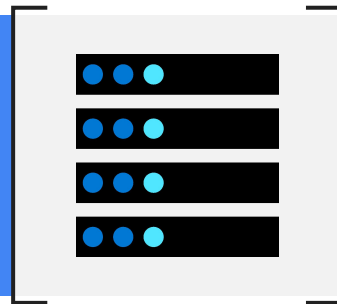
Lesson 1: Knowledge check (continued)



You are building a system that monitors the temperature throughout a set of office blocks, and sets the air conditioning in each room in each block to maintain a pleasant ambient temperature. Your system has to manage the air conditioning in several thousand buildings spread across the country/region, and each building typically contains at least 100 air-conditioned rooms. What type of NoSQL data store is most appropriate for capturing the temperature data to enable it to be processed quickly?

- ☐ Send the data to an Azure Cosmos DB database and use Azure Functions to process the data
- ☐ Store the data in a file stored in a share created using Azure File Storage
- ☐ Write the temperatures to a blob in Azure Blob storage

Lesson 2: เรียนรู้เกี่ยวกับการ provision บริการ non-relational data บน Azure



Lesson 2 objectives



Provision non-relational data services



Configure non-relational data services



รู้จักกับปัญหาการเชื่อมต่อ (connectivity issues) เบื้องต้น



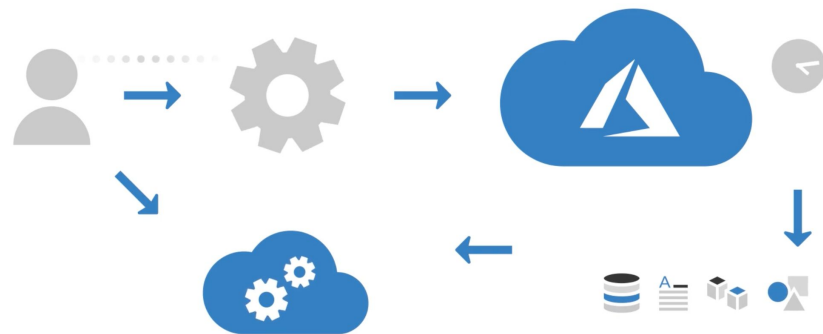
เรียนรู้เกี่ยวกับความปลอดภัยของข้อมูล (data security)

การ Provision คืออะไร

Provision คือ การแจ้ง Azure ว่าต้องการสร้างบริการไหน เช่น ต้องการสร้าง Azure SQL DB ใหม่

ในการ Provision เราต้องตั้งค่าบริการด้วย เช่น จะใช้เซิร์ฟเวอร์ขนาดไหน, username - password อะไร ฯลฯ

หลังจากนั้น Azure ก็จะสร้างบริการนั้นเพื่อให้เราใช้งาน ซึ่งบางบริการ หลังจากสั่ง Provision อาจจะต้องรอนานาที่



Cosmos DB APIs

ขณะสร้าง Cosmos DB เราสามารถเลือก API ได้ โดยตัวเลือกนี้จะกำหนดการรับส่งข้อมูลเข้า - ออกจาก Cosmos DB

Core (SQL) API

ทำให้ Cosmos DB
รองรับการใช้ภาษา
SQL ในการงานดึง
ข้อมูล

Gremlin API

ทำให้ Cosmos DB
รองรับการใช้งาน
แบบ Graph
Database

MongoDB API

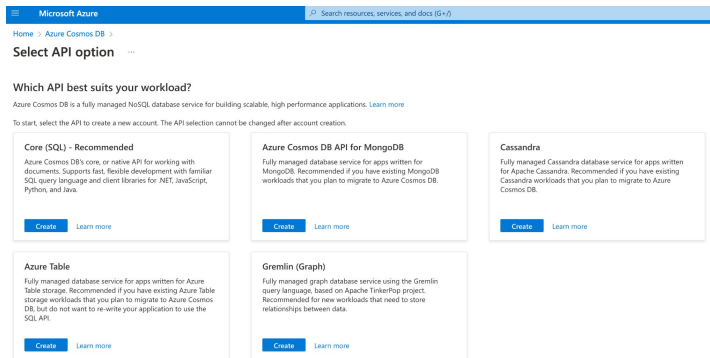
ทำให้ Cosmos DB
รองรับโค้ดที่เขียน
ขึ้นมาเพื่อใช้กับ
MongoDB

Table API

ทำให้ Cosmos DB
รองรับโค้ดที่เขียน
ขึ้นมาเพื่อใช้กับ
Azure Table
Storage

Cassandra API

ทำให้ Cosmos DB
รองรับโค้ดที่เขียน
ขึ้นมาเพื่อใช้กับ
Cassandra



Consistency ใน Azure Cosmos DB (1)

กรณีที่มีฐานข้อมูลในเซิร์ฟเวอร์หลายตัว การ Consistent คือ การที่ฐานข้อมูลในทุกเซิร์ฟเวอร์มีค่าล่าสุด

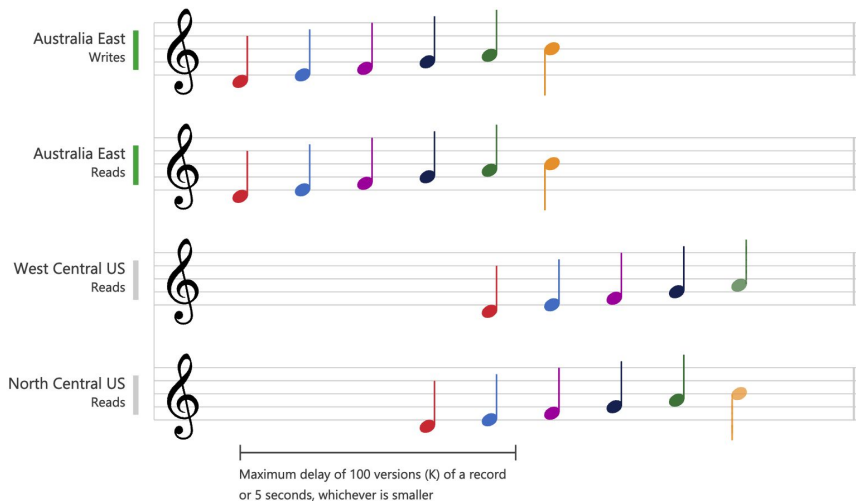
สิ่งที่ต้องแลกเปลี่ยน คือ ยิ่งข้อมูล Consistent มาก ยิ่งทำให้รับข้อมูลได้ช้า (throughput น้อย)

Option 1) Strong

ข้อมูลจะ Consistent ตลอดเวลา

Option 2) Bounded Staleness

ข้อมูลจะดีเลย์เล็กน้อย (ตั้งค่าเป็นเวลา หรือเป็นจำนวนครั้งการเขียนข้อมูลได้)



Consistency ใน Azure Cosmos DB (2)

Option 3) Session

ข้อมูลจะ Consistent ทันที ถ้าอยู่ใน Session (Application) เดียวกัน

ถ้าอยู่ Region เดียวกันจะค่อย ๆ Consistent แบบ Option 4 (Order ตรงซัวร์)

ถ้าอยู่คนละ Region จะค่อย ๆ Consistent แบบ Option 5 (Order อาจไม่ตรง)

Option 4) Consistent Prefix

ข้อมูลจะค่อย ๆ Consistent อาจจะเร็วบ้าง ช้าบ้าง แต่เดี๋ยวก็นั่นแหละ การันตีว่าลำดับข้อมูลจะตรงกัน



Consistency ใน Azure Cosmos DB (3)

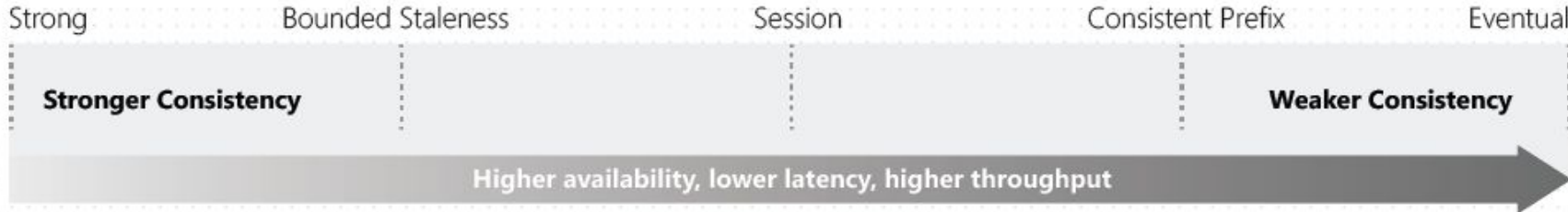
Option 5) Eventual

ข้อมูลจะค่อย ๆ Consistent อาจจะมีบ้าง ช้าบ้าง
แต่เดี๋ยวก็ตรงกัน ลำดับข้อมูลอาจจะไม่ตรงกัน



Consistency ใน Azure Cosmos DB (4)

กรณีที่มีฐานข้อมูลในเซิร์ฟเวอร์หลายตัว การ Consistent คือ การที่ฐานข้อมูลในทุกเซิร์ฟเวอร์มีค่าล่าสุด
สิ่งที่ต้องแลกเปลี่ยน คือ ยิ่งข้อมูล Consistent มาก ยิ่งทำให้รับข้อมูลได้ช้า (throughput น้อย)

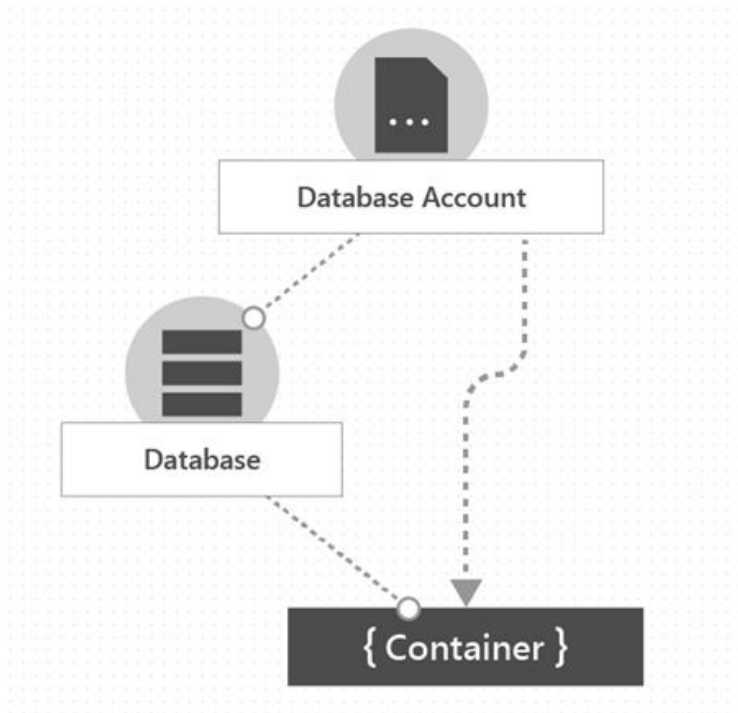


Demo: มาลอง provision บริการ Cosmos DB

Cosmos DB สามารถ provision ผ่าน Azure Portal หรือผ่านเครื่องมืออื่น ๆ เช่น Azure CLI, Azure Powershell, Azure Resource Manager Template ได้



Demo: สร้าง Database กับ Container เพื่อเก็บข้อมูลใน Cosmos DB



Database Account: Cosmos DB

Database: ที่เก็บ Container(s)

Container: ที่เก็บ items (ข้อมูล)

Demo: ลองปรับค่า Consistency ใน Cosmos DB

Home > jpwscosmosdb | Default consistency

jpwscosmosdb | Default consistency
Azure Cosmos DB account

Search (Ctrl+/)

Tags

Diagnose and solve problems

Quick start

Notifications

Data Explorer

Settings

Features

Replicate data globally

Default consistency

Firewall and virtual networks

Private Endpoint Connections

CORS

Keys

Add Azure Cognitive Search

Add Azure Function

Advanced security (preview)

Locks


Save Discard

STRONG BOUNDED STALENESS SESSION CONSISTENT PREFIX **EVENTUAL**

Eventual consistency is the weakest form of consistency wherein a client may get the values which are older than the ones it had seen before, over time.

Understand Eventual consistency

In the absence of any further writes, the replicas within the group will eventually converge. Eventual consistency is ideal where the application does not require any ordering guarantees. Examples include count of Retweets, Likes or non-threaded comments.

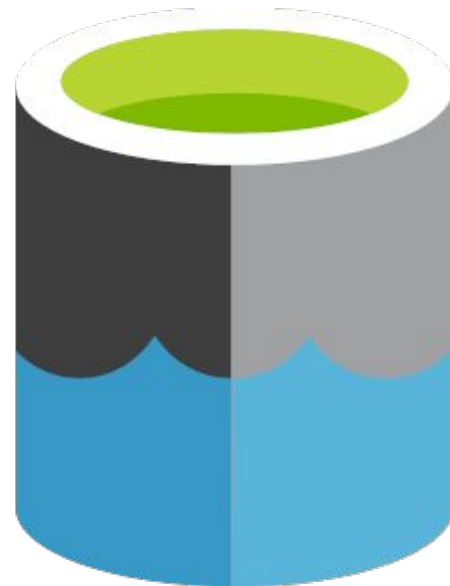


Azure Data Lake Storage (ADLS) Gen 2

ADLS Gen 2 คือ พีเจอาร์เสริมของบริการ Azure Blob Storage ที่เหมาะกับการเก็บข้อมูล Big Data

จุดเด่นที่น่าสนใจของ ADLS Gen 2:

- **รองรับ Hadoop API (HDFS ระบบเก็บข้อมูลของ Hadoop)** - ใช้งานได้เหมือนเป็น HDFS ตัวหนึ่ง
- **เก็บข้อมูลแบบแบ่งโฟลเดอร์ได้จริง (hierarchical namespace)** - โฟลเดอร์ใน Azure Blob Storage ปกติ เป็นโฟลเดอร์ปลอม ทำให้การทำงานกับไฟล์ เช่น การลบโฟลเดอร์ หากเป็น Azure Blob จะทำได้ช้า



Demo: มาลองสร้าง Azure Data Lake Storage

Create storage account

Basics Advanced Tags Review + create

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more](#)

PROJECT DETAILS

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

* Subscription

* Resource group [Create new](#)

INSTANCE DETAILS

The default deployment model is Resource Manager, which supports the latest Azure features. You may choose to deploy using the classic deployment model instead. [Choose classic deployment model](#)

* Storage account name

* Location

Performance ☒ Standard ☐ Premium

Account kind

Replication

Access tier (default) ☐ Cool ☒ Hot

Home > New > Create storage account

Create storage account

Basics Advanced Tags Review + create

SECURITY

Secure transfer required ☐ Disabled ☒ Enabled

VIRTUAL NETWORKS

Allow access from ☒ All networks ☐ Selected network

☒ All networks will be able to access this storage account. [Learn more](#)

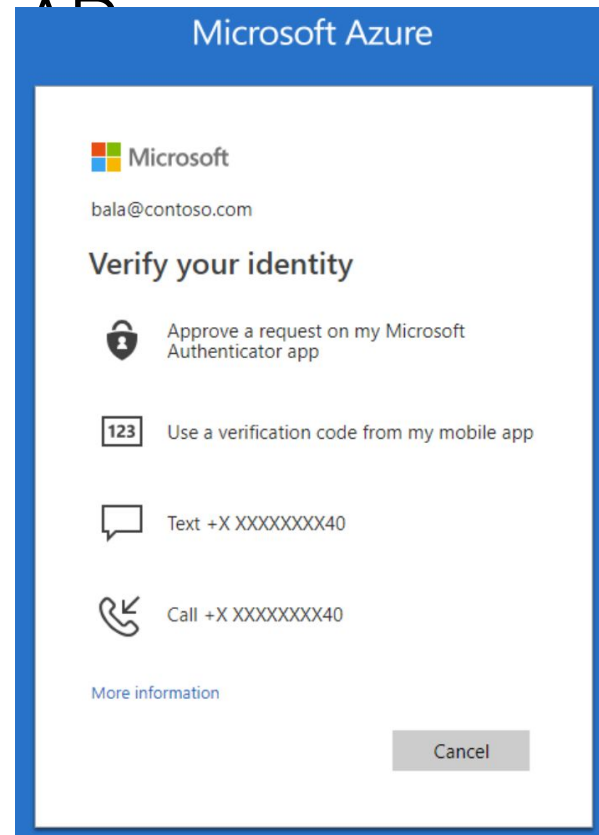
DATA LAKE STORAGE GEN2 (PREVIEW)

Hierarchical namespace ☐ Disabled ☒ Enabled

การทำ authentication ด้วย Azure AD

Azure AD (Azure Active Directory) เป็นบริการที่ช่วยให้ผู้ใช้ หรือ Application จากภายนอก สามารถล็อกอินเข้าใช้งานบริการต่าง ๆ บน Azure ได้

Azure AD รองรับการทำ **Multi-Factor Authentication (MFA)** นั่นคือ การให้ผู้ใช้ยืนยันตัวตนผ่านช่องทางอื่น เช่น sms, รับโทรศัพท์



Lesson 2: Knowledge check



What is provisioning?

- ☐ The act of running series of tasks that a service provider performs to create and configure a service
 - ☐ Providing other users access to an existing service
 - ☐ Tuning a service to improve performance
-



What is a security principal?

- ☐ A named collection of permissions that can be granted to a service, such as the ability to use the service to read, write, and delete data. In Azure, examples include **Owner** and **Contributor**.
 - ☐ A set of resources managed by a service to which you can grant access
 - ☐ An object that represents a user, group, service, or managed identity that is requesting access to Azure resources
-



Which of the following is an advantage of using multi-region replication with Cosmos DB?

- ☐ Data will always be consistent in every region
- ☐ Availability is increased
- ☐ Increased security for your data

Lesson 3: ลองใช้งานบริการ non-relational data บน Azure



Lesson 3 objectives



Upload data เข้าสู่บริการ Cosmos DB และลองดึงข้อมูล



Upload และ download ข้อมูลใน Azure Storage account

วิธีการ Query จาก Azure Cosmos DB

Aggregation Function Basics

`COUNT(<fields_to_count>)`

`SUM(<numeric_fields>)`

`AVG(<numeric_fields>)`

`MAX(<numeric_fields>)`

`MIN(<numeric_fields>)`

SQL API examples

```
SELECT COUNT(*) FROM Products p
```

```
SELECT SUM(p.quantity) FROM Products p
```

```
WHERE p.expired = 0
```

```
SELECT AVG(p.price) AS 'Average Price'  
FROM Products p
```

```
SELECT p1.ID, p.Name, p1.Description,  
p1.Price FROM Products p1  
WHERE p1.Price = (SELECT MIN(p2.Price) FROM  
Product p2)
```

Upload และ download ข้อมูลใน Azure Storage account

Azure Storage Account หรือ ADLS Gen 2 สามารถใช้งานได้ เหมือนกับที่เก็บไฟล์ทั่วไป อัปโหลด - ดาวน์โหลดไฟล์ได้

นอกจากนั้น ยังสามารถเปิดใช้ บริการอื่น ๆ ได้แก่

- File Share
- Queue
- Table Storage

Microsoft Azure

Home > Storage accounts > perthadlsfs >

perthadlsfs Container

Search (Cmd+/)

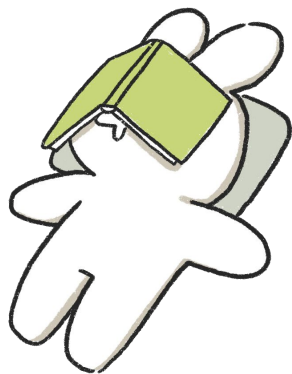
Upload Add Directory Refresh Rename Delete Change tier Acquire lease Break lease

Authentication method: Access key (Switch to Azure AD User Account)
Location: perthadlsfs

Search blobs by prefix (case-sensitive)

Name	Modified
<input type="checkbox"/> data-2021-08-30-03-41-22-utc.jpg	1/20/2022, 6:11:53 PM

สำหรับคนที่อยากเล่น Lab แบบเต็ม ๆ บน Microsoft Learn สามารถเล่นได้ที่นี่ <https://go.datath.com/dp900-lab1>



BREAK TIME



DATATH



Microsoft
Azure

หากมีคำถาม ถามได้ที่ไหน



ถามคำถามไว้ใน chat ด้านขวา หัวข้อ Q&A

โดยจะมาตอบช่วงหลังเบรก หรือช่วงท้าย CLASS

หากพิมพ์เข้ามาในกล่องแชททั่วไป คำถามอาจจะโดนดันจนไม่ได้ตอบ



DATATH



Microsoft
Azure