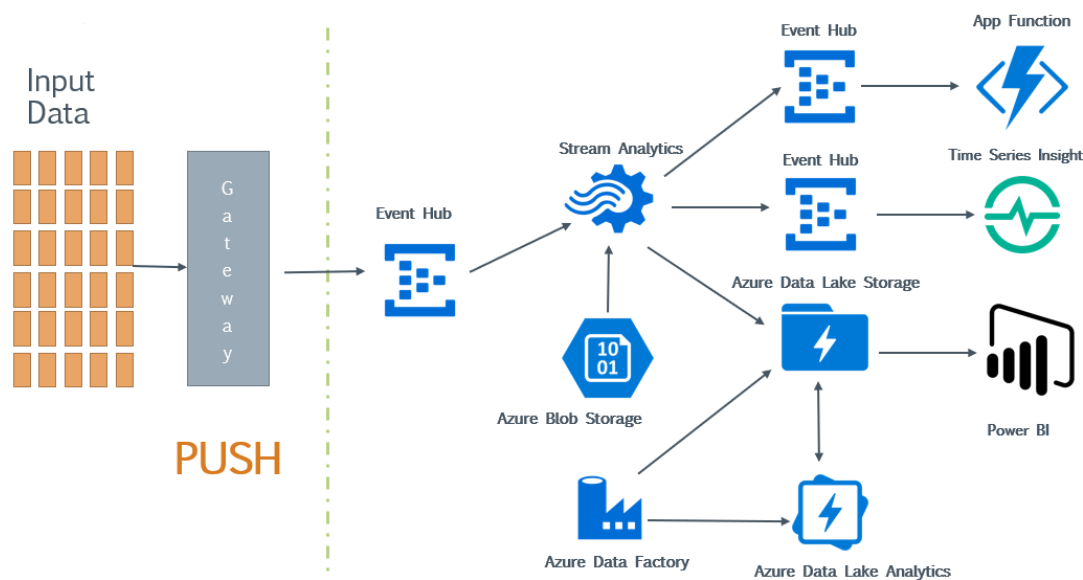


Demo tworzenie rozwiązania step-by-step

Problem:

Należy stworzyć rozwiązanie, które będzie zbierać i analizować w czasie zbliżonym do rzeczywistego dane z urządzeń (symulator) dotyczące poziomu wody w danym zbiorniku wodnym (oraz informacje dodatkowe takie jak temperatura powietrza oraz wilgotność. W przypadku kiedy poziom wody zostanie przekroczony powinien zostać wygenerowany alert. Wszystkie dane powinny być zapisywane a na ich podstawie raz dziennie powinien zostać generowany raport zawierający informacje o średnich, min oraz max wartościach poziomu wody, temperatury oraz wilgotności.

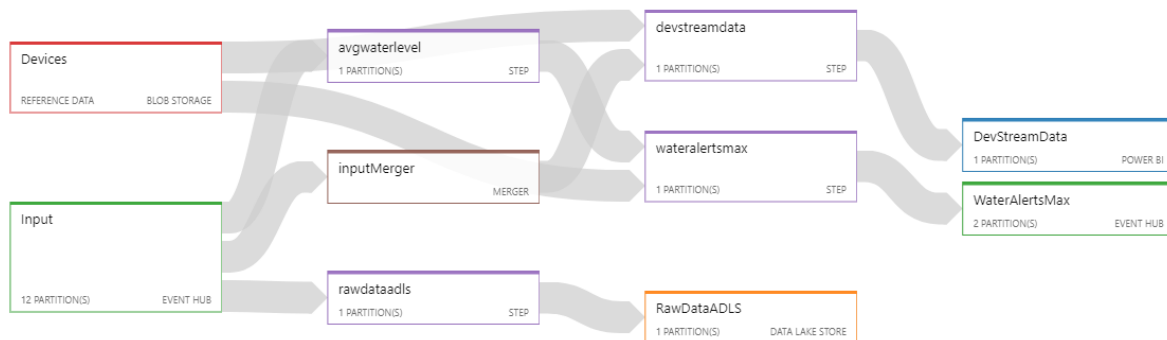
Ogólna koncepcja:



Procesowanie Hot-Path (Near to Real time)

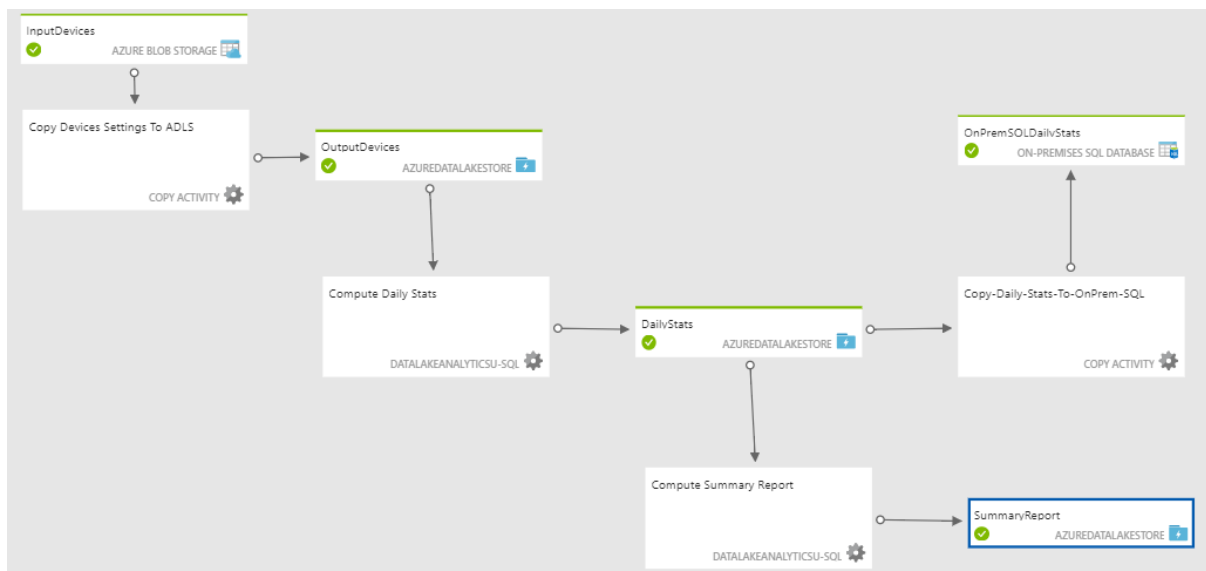
Procesowanie danych w trybie “Near to Real time” realizowane jest na platformie Azure za pomocą usługi Azure Stream Analytics. Dane pochodzące ze źródła strumieniowego (Event Hub) oraz tzw. dane referencyjne (dane znajdujące się na Azure Blob Storage w postaci pliku json) są przetwarzane przez usługę Stream Analytics, która wyniki przetwarzania zapisuje do kilku wyjść.

Informacje o alertach	Event Hub (Alias WaterAlertsMax)
Pełne przychodzące dane tzw. Raw Data	Azure Data Lake Store (Alias RawData)
Dane dla Power BI	Power Bi (Alias DevStreamData)
Dane dla Time Series Insight (opcjonalnie)	Event Hub (TimeSeries)



Procesowanie Cold-Path (Batch mode)

Procesowanie danych w trybie batchowym realizowane jest na platformie Azure za pomocą usług Azure Data Lake Analytics oraz HDInsight Spark Cluster (opcja). Dane znajdujące się na Azure Blob Storage kopiowane są do Azure Data Lake Store a następnie procesowane za pomocą usługi Azure Data Lake Analytics, wyniki są zapisywane na Azure Data Lake Store. Za proces orkiestracji przetwarzania danych odpowiada usługa Azure Data Factory (V1).



Tworzenie usług na Azure

1. Utwórz nową **Resource Group**
 1. Azure->Nowy->Resource group->Utwórz
 2. Podaj nazwę np. bigdataplatfrom
 3. Wybierz subskrypcję
 4. Wybierz region (sugerowany Europa Północna)
2. Utwórz usługę Event Hubs
 1. Azure->Nowy->Event Hubs->Utwórz

* Name
bdpmydevlab ✓
.servicebus.windows.net

* Pricing tier
Standard >

* Subskrypcja
Visual Studio Professional z subskrypcją MSDN v

* Resource group
☐ Utwórz nowy ☒ Użyj istniejącego
 bigdataplatfrom v

* Lokalizacja
Europa Północna v

Throughput Units
 1

☒ Enable auto-inflate?

Specify Upper Limit
 1

★ Proponowane | Wyświetl wszystkie

Basic	Standard
1 Consumer group	20 Consumer groups
100 Brokered connections	1000 Brokered connections
Ingress events 90.028 per million	Ingress events 90.028 per million
Message retention 1 day	Message retention 1 day
	Additional storage Up to 7 days
	Publisher policies
9.41 EUR/MONTH/TU (ESTIMATED)	18.82 EUR/MONTH/TU (ESTIMATED)

2. Podaj nazwę (bdpmydevlabs)
 3. Wybierz Pricing tier (wystarczy Basic)
 4. Wybierz subskrypcję
 5. Wybierz wcześniej stworzoną Resource Group
 6. Wybierz lokalizację (sugerowana Europa Północna)
 7. Określ Throughput Units (wystarczy 2)
3. Utwórz Event Hubs w ramach stworzonej usługi Event Hubs
 1. Otwórz stworzoną usługę Event Hubs

bdpmydevlabs Event Hub

Wyszukaj (Ctrl+/)

Overview

Kontrola dostępu (IAM)

Tagi

Diagnostowanie i rozwiązywanie...

USTAWIENIA

Shared access policies

+ Event Hub Delete

Grupa zasobów (zmień) bigdataplatfrom	Stan Active	Lokalizacja North Europe
Subskrypcja (zmień) Visual Studio Professional z subskrypcją M...	Identyfikator subskrypcji fb0d606c-7066-43ce-8b87-9112fe309305	Provisioning state Succeeded
Created wtorek, 31 października 2017	Updated wtorek, 31 października 2017	Connection Strings Connection Strings

NAMESPACE CONTENTS 2 EVENT HUBS PRICING TIER STANDARD NAMESPACE STATUS ACTIVE Show metrics data for the last:

1 hour	6 hours	12 hours	1 day
7 days	30 days		

Click for additional metrics

Requests Messages Throughput

2. Dodaj nowy Event Hub (podaj nazwę np. mydemodevices pozostałe wartości mogą zostać domyślne)

* Name
mydemodevices ✓

Partition Count
 2

Message Retention
 1

Capture

3. Dodaj kolejny Event Hub (podaj nazwę np. wateralertmax pozostałe wartości mogą zostać domyślne)

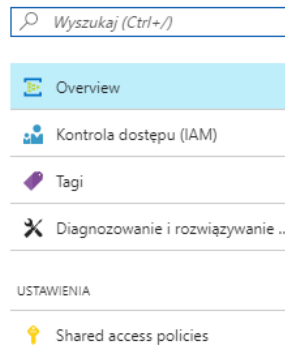
4. Skonfiguruj i uruchom Gateway (skrypt Python IoTEventGenerator-> IoTEventGenerator.py)

1. Podaj nazwę serviceNamespace = nazwa Event Hubs (patrz punkt 2) np. bdpmydevlabs

2. Podaj sharedAccessKeyName

1.Otwórz Event Hubs

2.Otwórz Shared access policy



3.Otwórz Shared access policy

4.Otwórz RootManageSharedAccessKey

5.Skopiuj Primary key

3. Podaj nazwę event huba (do którego mają trafiać eventy) np. mydemodevices

4. Uruchom skrypt Python IoTEventGenerator-> IoTEventGenerator.py

5. Utwórz usługę **Azure Blob Storage** i skopiuj dane referencyjne

1. Azure->Nowy->Blob storage->Utwórz

Utwórz konto magazynu

Koszt konta magazynu zależy od użycia i opcji wybranych poniżej.

[Dowiedz się więcej](#)

*

Nazwa

bdpbstorage

✓

.core.windows.net

Model wdrożenia

Resource Manager

Klasyczny

Rodzaj konta

Storage (ogólnego przeznaczenia w we...

Wydajność

Standardowe

Premium

Replikacja

Magazyn geograficznie nadmiarowy d...

*

Wymagany bezpieczny transfer

Wyłączone

Włączone

*

Subskrypcja

Visual Studio Professional z subskrypcj...

*

Grupa zasobów

Utwórz nowy

Użyj istniejącego

bigdataplatfrom

*

Lokalizacja

Europa Północna

☐

Przypnij do pulpitu nawigacyjnego


Utwórz

Opcje automatyzacji

2. Utwórz nowy kontener np. devices

1.Otwórz Obiekty blob

Usługi



Obiekty blob

Magazyn obiektów na potrzeby rozpoznawania danych

[Wyświetl metryki](#)
[Konfigurowanie reguł CORS](#)
[Konfiguracja domeny niestandardowej](#)

2.Dodaj kontener

+

Kontener

↺

Odśwież

Konto magazynu

karierabs

Stan

Podstawowy: Dostępne

Lokalizacja

Europa Północna

Blob Service

bdpbstorage

+

Kontener

↺

Odśwież

Konto magazynu

bdpbstorage

Stan

Podstaw.: Dostępne, dodatk.: Dostępne

Lokalizacja

Europa Północna, Europa Zachodnia

3. Skopiuj plik z danymi referencyjnymi urządzeń (Data\Devices.json)

↶

Przeładuj

↺

Odśwież

🗑️

Usuń kontener

⚙️

Właściwości kontenera

🔒

Zasady dostępu

Lokalizacja: devices

🔍

Wyszukaj obiekty blob według prefiksu (z uwzględnieniem wielkości liter)

NAZWA	ZMODYFIKOWANE
Nie odnaleziono obiektów blob.	

Piki

Wybierz plik

☐ Zastąp pliki, jeśli już istnieją

☒ Zaawansowane

Przeładuj

6. Utwórz usługę Azure Data Lake Analytics

1. Azure->Nowy-> **Data Lake Analytics** ->Utwórz

1. W ramach usługi utwórz Azure Data Lake Store

Nowe konto usługi Data Lake A...

Nazwa

bdpadla

bdpadla.azuredatalakeanalytics.net

* Subskrypcja

Visual Studio Professional z subskrypcją M...

* Grupa zasobów

☐ Utwórz nowy
 ☒ Użyj istniejącego

bigdatapatform

* Lokalizacja

Europa Północna

* Data Lake Store

Skonfiguruj wymagane ustawienia

Pakiet cenowy

☒ Płatność zgodnie z rzeczywistym użyciem
 ☐ Zobowiązanie miesięczne

Wybierz usługę Data Lake Store

Subskrypcje: Wybrano wszystko (2)

Filtruj według naz...

Wszystkie subs...

+

Utwórz nową usługę Data Lake S...

adlademosadls

Europa Północna

adlssparklab

Europa Północna

Nowa usługa Data Lake Stor

* Nazwa

bdpadls

bdpadls.azuredatalake

Pakiet cenowy

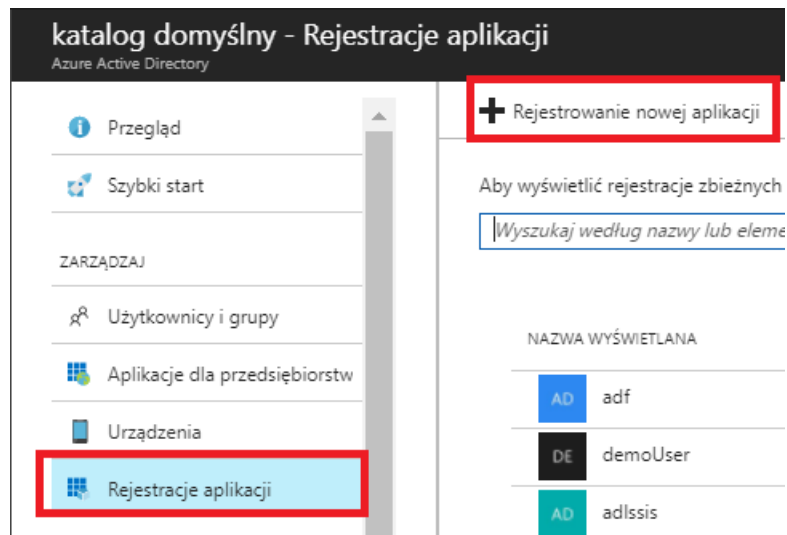
☒ Płatność zgodnie z rzeczywistym użyciem
 ☐ Zobowiązanie miesięczne

Ustawienia szyfrowania

Włączone

2. Utwórz nowy Service Principal

1. Otwórz Azure Active Directory
2. Zarejestruj nową aplikację



3. Podaj nazwę, typ aplikacji (Interface API), oraz adres url

Tworzenie
✕

* Nazwa ?

bdpdemo
✓

Typ aplikacji ?

Interfejs API/aplikacja sieci Web
▼

* Adres URL logowania ?

http://bdpdemo.org
✓

4. Utwórz nowy klucz (otwórz stworzoną aplikację Ustawienia->Klucze)

1. Podaj nazwę (np. demo)
2. Podaj datę ważności np. rok
3. Zapisz

Hasła

OPIS	WYGASA	WARTOŚĆ
<div style="border: 1px solid #ccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> demo ✓ </div> <div style="border: 1px solid #ccc; padding: 2px; margin-top: 2px;">Opis klucza</div>	<div style="border: 1px solid #ccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> Nigdy nie wygasa ▼ </div> <div style="border: 1px solid #ccc; padding: 2px; margin-top: 2px;"> Czas trwania Za 1 rok Za 2 lata Nigdy nie wygasa </div>	<div style="border: 1px solid #ccc; padding: 2px; display: flex; justify-content: space-between; align-items: center;"> Wartość zostanie wyświetlona podczas zapisywania ... </div> <div style="border: 1px solid #ccc; padding: 2px; margin-top: 2px;"> Wartość zostanie wyświetlona podczas zapisywania </div>

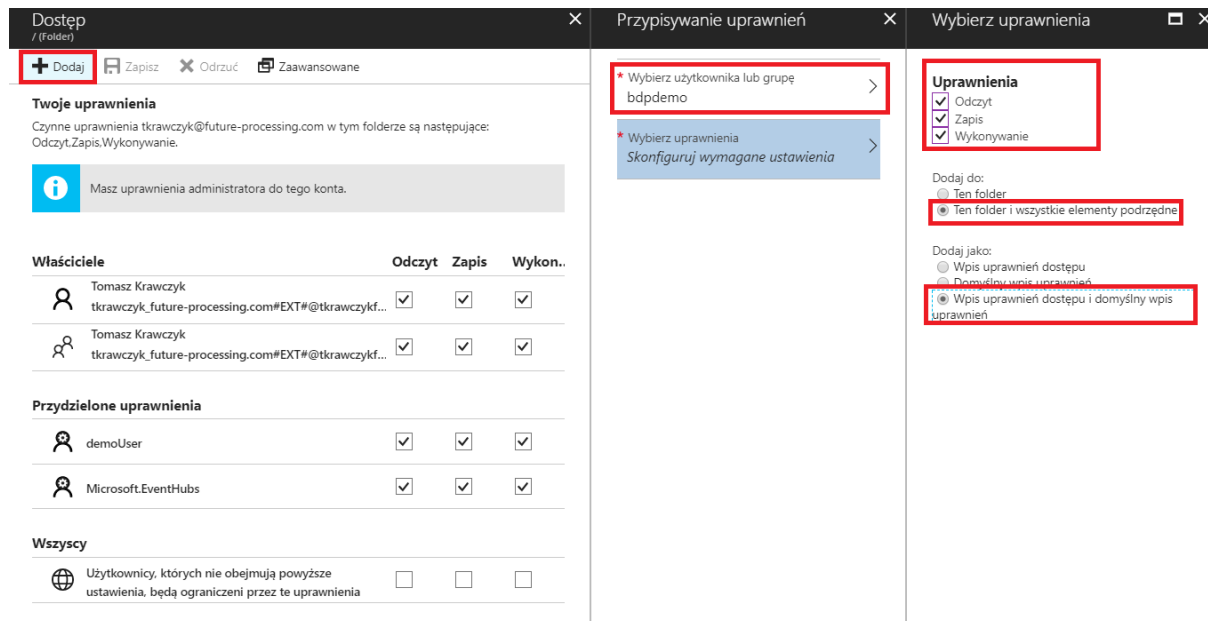
4. Po zapisaniu skopiuj wartość klucza

Zapisz
 Odrzuć
 Przekaz klucz publiczny

Skopiuj wartość klucza. Po opuszczeniu bloku nie będzie można go pobrać.

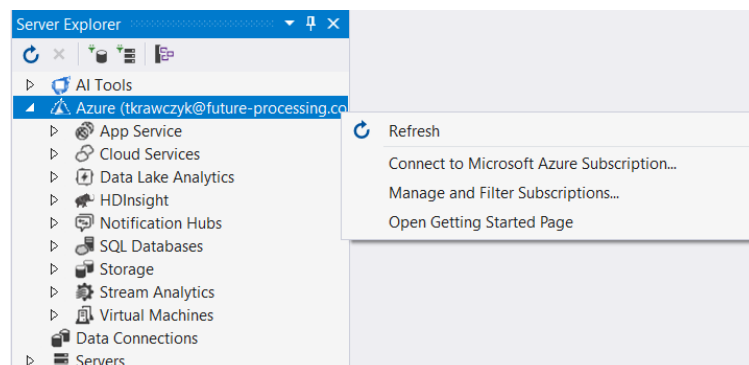
3. Nadaj uprawnienia dla nowej aplikacji do ADLS oraz ASLA

1. Otwórz ADLS Eksplorator Danych->Dostęp
2. Nadaj uprawnienia dla ADLS

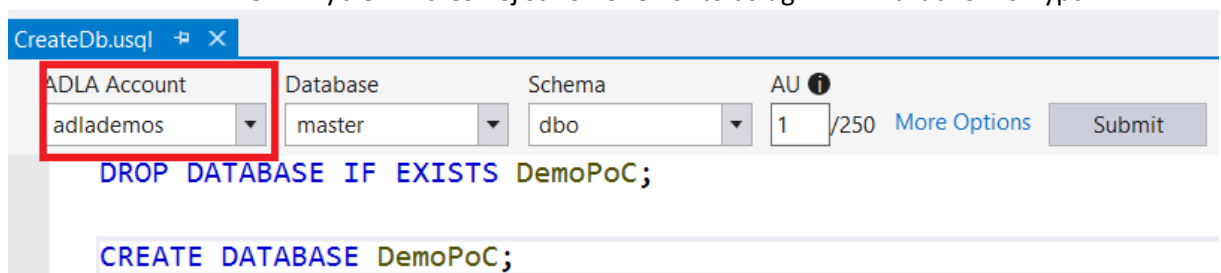


3. Utwórz bazę DemoPoc na ADLA (skrypt ColdPath->BatchProcessing->Init->CreateDB.usql)

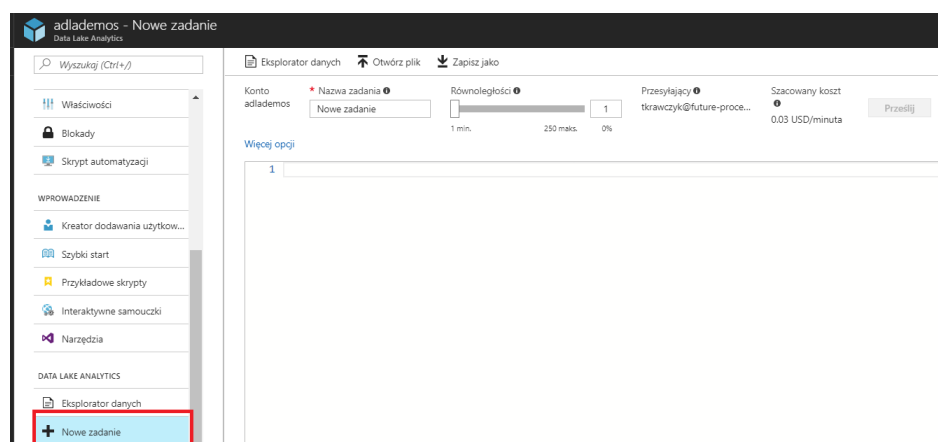
1. Otwórz rozwiązanie AzureAsBigDataPlatform.sln
2. Przejdź do projektu BatchProcessing
3. W Server Explorer połącz się do Azure



4. Otwórz skrypt CreateDb.usql
5. Wybierz wcześniej stworzone konto usługi ADLA i uruchom skrypt

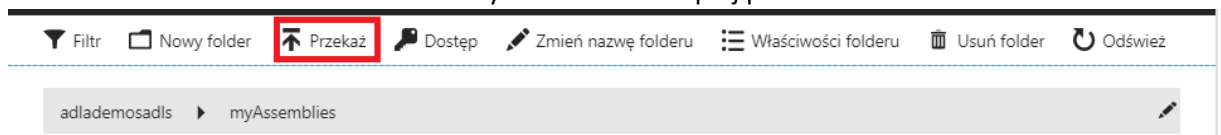


6. Skrypty można również uruchamiać bezpośrednio z portalu Azure



4. Skopiuj dodatkowe assembly na ADLS : źródło BatchProcessing\Assemblies\
cel ADLS /myAssemblies/

1. Z poziomu Azure portal przejść do ADLS->Eksplorator danych
2. Stwórz folder myAssemblies i skopiuj pliki dll



5. Zarejestruj assemblies skrypt ColdPath->BatchProcessing->Init->RegisterExtensstions.usql

6. Utwórz procedury usp_ComputeDailyStats - skrypt ColdPath->BatchProcessing->Init->SP-> ComputeDailyStats.usql

Oraz ComputeSummaryReport - skrypt ColdPath->BatchProcessing->Init->SP-> ComputeSummaryReport.usql

7. Nadaj uprawnienia dla ADLA

1. Kreator dodawania użytkowników
2. Użytkownik bdpdemo
3. Rola: Developer Usługi ADLA
4. Baza DemoPoc:Odczyt i zapis

ZAKRES	UPRAWNIENIA
adlademos (Wykaz)	Odczyt i zapis
DemoDb (Baza danych)	Brak
DemoPoc (Baza danych)	Odczyt i zapis
master (Baza danych)	Odczyt i zapis

7. Utwórz usługę Power BI (link: <https://powerbi.microsoft.com/>)

1. Sign up free
2. Try Free

POWER BI

Cloud collaboration and sharing

Use Power BI Pro to share and distribute reports with others, without any complicated setup. Get started now with a free 60-day trial of Power BI Pro.

TRY FREE >

8. Utwórz usługę Azure Stream Analytics

1. Azure->Nowy-> **Stream Analytics** ->Utwórz

* Nazwa zadania
 bdpdemo ✓

* Subskrypcja
 Visual Studio Professional z subskrypcją M... ▼

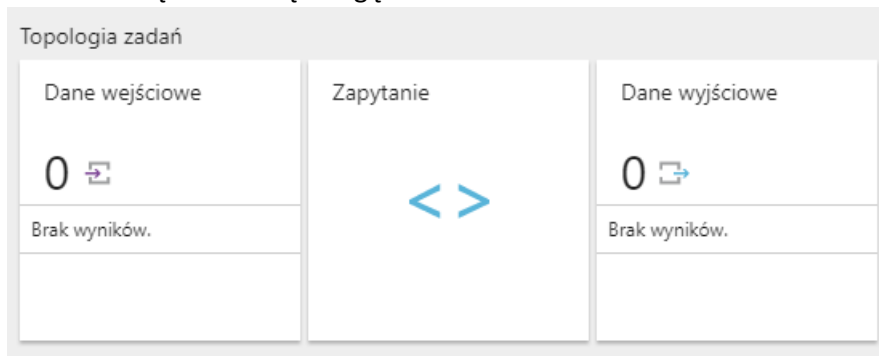
* Grupa zasobów
☐ Utwórz nowy ☒ Użyj istniejącego

bigdataplatfrom ▼

* Lokalizacja
 Europa Północna ▼

9. Utwórz nowego joba w ramach usługi Azure Stream Analytics

1. Otwórz nową stworzoną usługę



2. Skonfiguruj INPUTY oraz OUTPUTY

1. Wejdź w opcje Dane wejściowe

2. Następnie Dodaj (Input)

1. Alias wejściowy Input
2. Przestrzeń nazw magistrali usług : nazwa Azure Events np. bdpmydevlabs oraz
3. Format serializacji: JSON

* Alias wejściowy

Input_ ✓

- ☐ Podaj ręcznie ustawienia centrum zdarzeń
☒ Wybierz centrum zdarzeń z subskrypcji

Subskrypcja

Visual Studio Professional z subskrypcją MSDN ▼

* Przestrzeń nazw centrum zdarzeń ⓘ

bdpmydevlabs ▼

* Nazwa centrum zdarzeń ⓘ

- ☐ Utwórz nową ☒ Użyj istniejącego

mydemodevices ▼

* Nazwa zasad centrum zdarzeń ⓘ

RootManageSharedAccessKey ▼

Klucz zasad centrum zdarzeń

.....

Grupa konsumentów centrum zdarzeń ⓘ

* Format serializacji zdarzeń ⓘ

JSON ▼

Kodowanie ⓘ

UTF-8 ▼

Zapisz

3. Dodaj nowy input (dane referencyjne)

1. Alias wejściowy Devices
2. Wzorzec ścieżki: Devices.json
3. Format serializacji: JSON

* Alias wejściowy
Device ✓

* Typ źródła ⓘ
Dane referencyjne ▼

* Opcja importu
Wybierz magazynu obiektów blob z subsk... ▼

Konto magazynu
bdpstorage ▼

Klucz konta magazynu
.....

Kontener
devices ▼

Wzorzec ścieżki ⓘ
.....

Format daty
YYYY/MM/DD ▼

Format godziny
HH ▼

* Format serializacji zdarzeń ⓘ
JSON ▼

3. Dodaj nowy output (tu będą zapisywane eventy o przekroczeniu wartości granicznych)
 1. Nazwa (alias) WaterAlertsMax
 2. Nazwa centrum zdarzeń: wateralertsmax
 3. Format serializacji JSON
4. Dodaj nowy output (tu będą zapisywane eventy wszystkie eventy)
 1. Nazwa (alias) RawStreamData
 2. Ujęcie Data Lake Store (Nazwa konta – wcześniej stworzone konto ADLS np. bdpadls)
 3. Wzorzec prefiksu ścieżki: /mySamples/DevicesEvents/{date}/ (należy wcześniej utworzyć katalog mySamples oraz w nim DevicesEvents)
 4. Format daty YYYY-MM-DD
 5. Format serializacji zdarzeń CSV

Opcja importu
Podaj ustawienia usługi Data Lake Store rę... ▼

* Nazwa konta
adladedemosadls

* Wzorzec prefiksu ścieżki ⓘ
/mySamples/DevicesEvents/{date}/
Przykład: klaster1/dzienniki/{date}/{time}

Format daty
YYYY-MM-DD ▼

Format godziny
HH ▼


* Format serializacji zdarzeń ⓘ
CSV ▼


Ogranicznik ⓘ
przecinek (,) ▼

Kodowanie ⓘ
UTF-8 ▼

5. Dodaj nowy output (tu będą zapisywane eventy dla Power BI)
 1. Alias wejściowy : DevStreamData
 2. Ujęcie Power BI
 3. Autoryzuj

* Alias wyjściowy

* Ujście 

Power BI 

Autoryzacja połączenia
Aby skonfigurować ustawienia wyjścia, musisz przejść autoryzację w usłudze Power BI.

Autoryzuj

Nie masz jeszcze konta usługi Microsoft Power BI?
[Zarejestruj się](#)

6. Skonfiguruj query dla Joba

```

1.  --POWER BI
2.  SELECT
3.      m.humidity AS Humidity,
4.      m.temperature AS Temperature,
5.      m.waterLevel AS waterLevel,
6.      d.min AS minwaterLevel,
7.      d.max AS maxwaterLevel,
8.      m.TIMESTAMP AS time
9.      INTO [DevStreamData]
10. FROM
11.     [Input] AS m
12.     TIMESTAMP BY m.TIMESTAMP
13.     JOIN [Devices] AS d ON d.ID = m.ID
14.     WHERE m.ID = 5;
15.
16. --RAW DATA -> ADLS
17. SELECT *
18. INTO [RawData]
19. FROM [Input] AS m
20.     TIMESTAMP BY m.TIMESTAMP;
21.
22. --MAX LEVEL ALERTS
23. WITH AvgWaterLevel AS
24. (
25.     SELECT m.ID,
26.            AVG(waterLevel) AS WaterLevel
27.     FROM [Input] AS m
28.     TIMESTAMP BY m.TIMESTAMP
29.     GROUP BY
30.         TUMBLINGWINDOW(second,5),
31.         m.ID
32. )
33. SELECT d.Id,avg.WaterLevel,d.maxwaterLevel FROM
34.     AvgWaterLevel AS avg
35.     JOIN [Devices] AS d ON d.ID = avg.ID
36.     INTO [WaterAlertsMax]
37.     WHERE avg.WaterLevel > d.maxwaterLevel;
38.

```

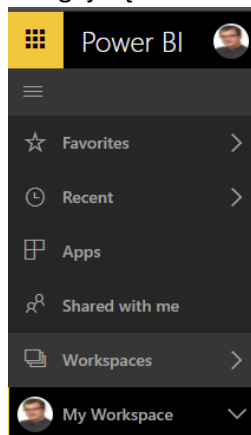
10. Uruchom Job'a

11. Uruchom Gateway (skrypt Python IoTEventGenerator-> IoTEventGenerator.py)

Stan: Dane z Gateway są przesyłane do Input (Event Hub) i procesowane przez Azure Stream Analytics Job. Wynikiem joba są dane surowe zapisywane na ADLS, dane o alertach zapisywane na Event Hub, oraz dane przesyłane do Power BI.

12. Utwórz nowy widok w Power BI

1. Zaloguj się do Power BI



2. Dodaj nowy Dashboard (po prawej stronie + Create)
3. Na dodanym dashboard'zie dodaj nowy tile (+Add tile)
4. Wybierz Real Time Data i wcześniej stworzony DataSets (DemoStreamData)
5. Następnie wybierz dowolny typ wykresu i dane które chcesz wizualizować

Po wykonaniu tej operacji uruchomieniu generatora (gateway) oraz joba Azure Stream Analytics wykres na bieżąco powinien prezentować dane zapisywane do wyjścia DemoStreamData.

13. Utwórz usługę Azure App Functions

1. Z poziomu Visual Studio 2017 zrób publisha projektu Alters-> waterdemo

1.Zrekonfiguruj plik host.json

1. AzureWebJobsStorage -connection string do Blob Storage
2. Alerts - connection string do Event Huba wateralertsmax
3. Uruchom funkcję AlertMonitor
4. Sprawdź działanie funkcji (Monitoruj)



W wyniku działania funkcji w logu powinny pojawiać się informacje o alertach o przekroczeniu poziomu wody.

Funkcja	Stan	Szczegóły Ostatnie uruchomienie (czas trwania)
AlertMonitor ([\"id\":\"waterlevel...\"])	✓	2 days ago (0 ms)
AlertMonitor ([\"id\":\"waterlevel...\"])	✓	2 days ago (0 ms)
AlertMonitor ([\"id\":\"waterlevel...\"])	✓	2 days ago (0 ms)

Dzienniki

WATER ALERT FOR OBJECT:4 MAX WATER LEVEL:60 CURRENT LEVEL:77.0

14. Stwórz na lokalnym serwerze SQL bazę danych i stwórz tabelę Measurements (projekt WaterDb)

15. Utwórz i skonfiguruj usługę Azure Data Factory (V1)

1. Azure->Nowy-> **Data Factory** ->Utwórz (wersja 1)

* Nazwa ⓘ
demoadfmy ✓

* Subskrypcja
Visual Studio Professional z subskrypcją MS ▼

* Grupa zasobów ⓘ
☐ Utwórz nowy ☒ Użyj istniejącego
demos ▼

Wersja ⓘ
Wersja 1 ▼

* Lokalizacja ⓘ
Europa Północna ▼

2. Zainstaluj dodatek do Visual Studio 2015 do tworzenia pipeline ADF (dodatek dostępny pod adresem [Microsoft Azure Data Factory Tools for Visual Studio 2015](#))

3. Skonfiguruj dane konfiguracyjne pipeline MeasurementsBatchProcessor (projekt dfdemos)

1. Skonfiguruj dane uwierzytelniające do usługi ADLA (plik ADLA-USQL.json)

1. Podaj nazwę konta ADLA (accountName)
2. Dane uwierzytelniające (servicePrincipalId, servicePrincipalKey)
patrz: Utwórz usługę Azure Data Lake Analytics -> Utwórz nowy Service Principal
3. Podaj nazwę grupy, oraz dane subskrypcji

2. Skonfiguruj dane uwierzytelniające do usługi ADLA (plik ADLDemoStore.json)

1. Podaj adres usługi ADLS(dataLakeStoreUri)
2. Dane uwierzytelniające (servicePrincipalId, servicePrincipalKey)
patrz: Utwórz usługę Azure Data Lake Analytics -> Utwórz nowy Service Principal
3. Podaj nazwę grupy, oraz dane subskrypcji

3. Skonfiguruj dane uwierzytelniające do usługi Azure Blob Storage (dane referencyjne) (plik DevicesRefs-BS.json)

1. Podaj connectionString do Azure Blob Storage

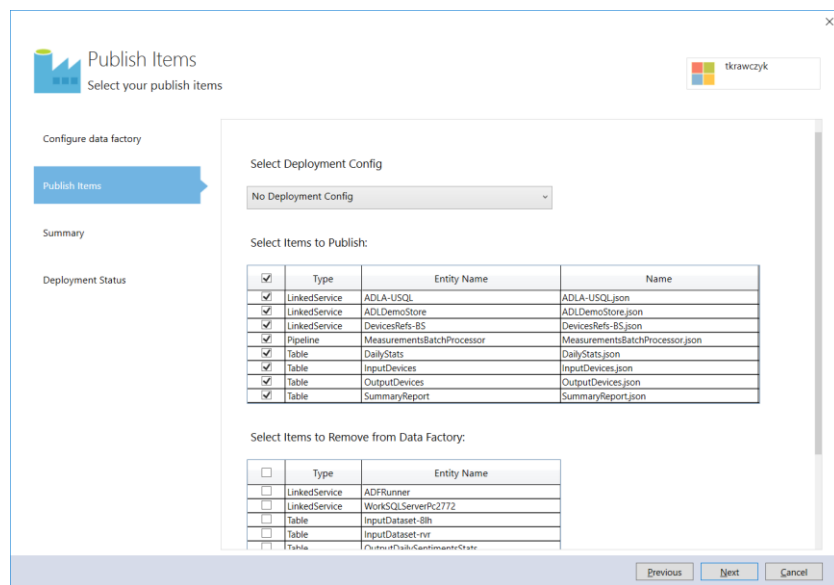
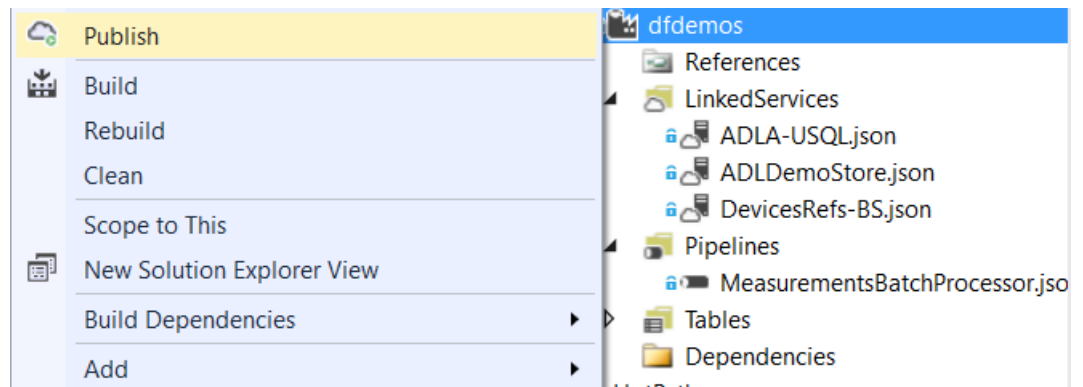
4. Skonfiguruj pipeline MeasurementsBatchProcessor (plik MeasurementsBatchProcessor.json)

1. Ustaw start oraz end (zakres kiedy pipeline będzie działać)
 - a. Godzina w przypadku daty start oznacza kiedy będzie wykonywany pipeline

5. Skonfiguruj połączenie do lokalnego serwera SQL (plik: Destination-OnPremSQL-InI.json)

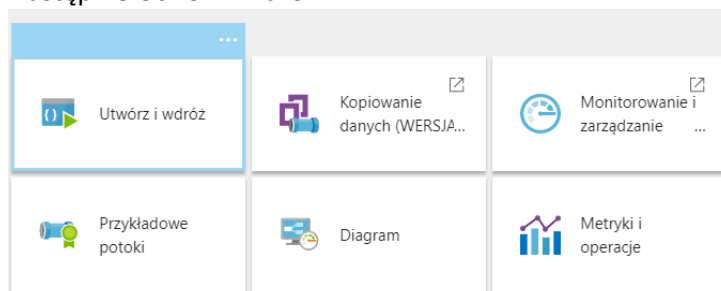
4. Zdeploy'uj rozwiązanie na platformę Azure

1. Deploy można zrobić z poziomu VS z zainstalowanym dodatkiem ADF

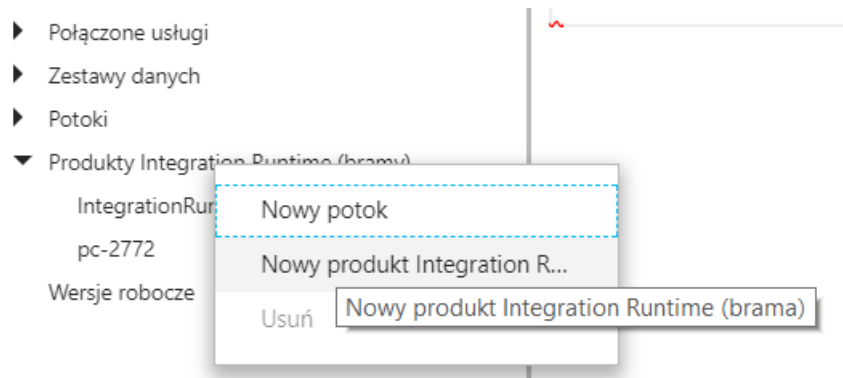


2. Deploy za pośrednictwem portalu Azure

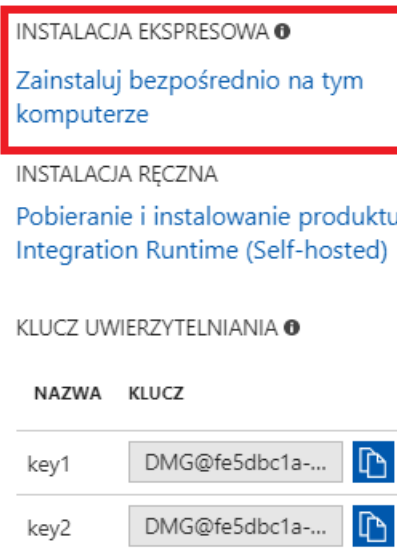
1. Przechodzimy do usługi ADF
2. Następnie Utwórz i wdróż



3. Zainstaluj na lokalnym komputerze Integration Runtime (będzie on używany do przenoszenia danych pomiędzy chmurą Azure a lokalnym serwerem SQL)
 - a. Podaj nazwę runtime np IntegrationRuntime-Lap



b. Pobierz i zainstaluj Integration-Runtime



4. Zdeploy'uj LinkedServices

- Skopiuj konfigurację ADLA-USQL -> Wdróż
- Skopiuj konfigurację ADLDemoStore -> Wdróż
- Skopiuj konfigurację DevicesRefs-BS -> Wdróż
- Skopiuj konfigurację Destination-OnPremSQL-Inl ->Wdróż

5. Zdeploy'uj Tables

- Skopiuj konfigurację DailyStats -> Wdróż
- Skopiuj konfigurację InputDevices -> Wdróż
- Skopiuj konfigurację OutputDevices -> Wdróż
- Skopiuj konfigurację SummaryReport -> Wdróż
- Skopiuj konfigurację OnPremSQLDailyStats ->Wdróż

6. Zdeploy'uj pipeline

- Skopiuj konfigurację MeasurementsBatchProcessor -> Wdróż

5. Uruchamianie pipeline (on Demand) -aktualnie pipeline ma zdefiniowany daily scheduler

- Uruchamianie całego pipeline możliwe jest poprzez uruchomienie ostatniego DataSet (zestawu danych). W przypadku tego pipeline jest to SummaryReport oraz OnPremSQLDailyStats

ZESTAW DANYCH	UTWORZONO
OutputDevices	22.11.2017, 8:20 AM UTC (22...
DailyStats	22.11.2017, 8:19 AM UTC (22...
SummaryReport	22.11.2017, 8:20 AM UTC (22...

2. Aby uruchomić dany DataSet należy:

1. Otworzyć dany DataSet
 2. Przejść do „Wycinki danych (według godziny wycinka)”
 3. Następnie wybrać dany tzw „data slice” i uruchomić go
3. Uwaga: W przypadku kiedy poprzednie „data set’y” były już uruchomione, nie activity, które generują te datasey nie zostaną uruchomione (należy wtedy po kolei uruchamiać poszczególne data sety).

W wyniku procesowania na Azure Data Lake Store w folderze /mySamples/DevicesEvents powinny się pojawić:

- W folderze Results plik z nazwą {yyyy-mm-dd}.csv zawierający informacje o średnich wartościach z danego dnia dla każdego punktu pomiarowego
- SummaryReport.csv zawierający informacje o średnich wartościach dla każdego obiektu (wyliczonych na podstawie wszystkich dziennych wartości średnich)

Oraz w tabeli Measurements powinny pojawić się dane (dziennie średnie wartości pomiarów)

Mechanizm Event Data Capturing:

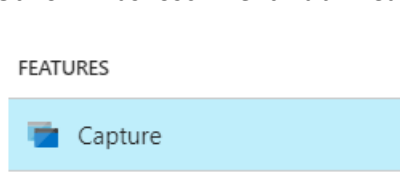
Aktualnie dane tzw. Cold Data były zapisywane na Azure Data Lake Store za pośrednictwem usługi Azure Stream Analytics (realizowane przez poniższe zapytanie)

```
--RAW DATA -> ADLS
1. SELECT *
2. INTO [RawData]
3. FROM [Input] AS m TIMESTAMP BY m.TIMESTAMP
```

Usługa Azure Event Hub zawiera mechanizm o nazwie Event Data Capture (uwaga opcja dodatkowo płatna), który pozwala na automatyczny zapis wszystkich danych trafiających do Event Hub do Azure Blob Storage lub Azure Data Lake Store (format plików Avro). (więcej informacji [Azure Event Hubs Capture](#))

1. Włączanie mechanizmu Event Data Capture

1. Otwórz właściwości Event Hub -> Capture



2. Skonfiguruj mechanizm Event Data Capture

Save changes Discard



Azure Event Hubs Capture enables you to automatically deliver the streaming data in Event Hubs to an Azure Blob storage or Azure Data Lake Store account of your choice, with the added flexibility of specifying a time or size interval. Setting up Capture is fast, there are no administrative costs to run it, and it scales automatically with Event Hubs throughput units. Event Hubs Capture is the easiest way to load streaming data into Azure, and enables you to focus on data processing rather than on data capture. Learn more about capture [here](#).

Capture

On Off

Note: Enabling Capture will result in additional charges to this account. Learn more about our pricing [here](#).

Time window (minutes)

5

Size window (MB)

10

Capture Provider

Azure Data Lake Store

Learn about setting Data Lake Store access permissions for Event Hubs [here](#).

* Data Lake Store

adlssparklab

Select Store

Selected Data Lake Store account should belong to the same subscription as current Event Hub namespace.

* Data Lake Path

/mySamples/edcAvro

Sample Capture file name formats

Capture file name format

{Namespace}_{EventHub}_{PartitionId}_{Year}-{Month}-{Day}_{Hour}{Minute}{Second}.avro

E.g. bdpmydevlabs_mydemodevices_0_2018-0-11_182659.avro

1. Określ Time Window oraz Size Window
 2. Określ Capture Provider (ADLS -store do którego kluster HDInsight Spark będzie miał dostęp)
 3. Podaj ścieżkę na ADLS gdzie będą zapisywane pliki np. /mySamples/edcAvro
 4. Określ format nazwy plików np.
{Namespace}_{EventHub}_{PartitionId}_{Year}-{Month}-{Day}_{Hour}{Minute}{Second}.avro
3. Nadaj uprawnienia do zapisu do ADLS
 1. Przejdź do folderu na ADLS, gdzie będą zapisywane pliki
 2. Przejdź do ustawień Dostęp (Access)
 3. Dodaj nowe uprawnienia Execute dla Microsoft.EventHubs

* Wybierz użytkownika lub grupę
Microsoft.EventHubs

* Wybierz uprawnienia
Skonfiguruj wymagane ustawienia

Uprawnienia

- ☐ Odczyt
☒ Zapis
☒ Wykonywanie

Dodaj do:

- ☐ Ten folder
☒ Ten folder i wszystkie elementy podrzędne

Dodaj jako:

- ☐ Wpis uprawnień dostępu
☐ Domyślny wpis uprawnień
☒ Wpis uprawnień dostępu i domyślny wpis uprawnień

2. Uruchamianie

1. Uruchom Gateway (skrypt Python IoTEventGenerator-> IoTEventGenerator.py)

W wyniku procesowania na Azure Data Lake Store w folderze mySamples/edcAvro powinny się pojawić z rozszerzeniem avro oraz nazwami zgodnymi z formatem podanym przy konfiguracji mechanizmu Event Data Capture.

Mechanizm Event Data Capturing -procesowanie danych na HDInsight Spark Cluster

Aby procesować dane znajdujące się na ADLS na klastrze HDInsight Spark należy stworzyć kluster wraz z dołączonym ADLS (szczegóły: <https://docs.microsoft.com/pl-pl/azure/data-lake-store/data-lake-store-hdinsight-hadoop-use-powershell>).

1. Uruchom HDInsight Spark Cluster
2. Uruchom Jupyter Notebook
3. Zrób upload skryptu ADLA-EDC-Avro.ipynb na klaster (Scripts \ ADLA-EDC-Avro.ipynb)
4. Uruchom poszczególne kroki skryptu

W wyniku procesowania w powinna być tabela z informacjami o średnich wartościach poziomu wody dla danych obiektów.

id	max(waterLevel)
2	18
4	57
6	98
8	59
10	113

UWAGA: Po zakończonych ćwiczenia należy usunąć wszystkie usługi (lub całą Resource Group) aby nie zużywać środków z subskrypcji Azure.