# Solution M3: Databases and Analytics

Main goal is to build further on what was demonstrated during the practice

## Tasks

### CosmosDB (Azure Portal)

One possible solution, using the Azure Portal, may include the following steps:

* Open a browser tab, navigate to <https://portal.azure.com> and login
* Navigate to **Resource Groups** and create a new one, for example **Homework-M3**
* Enter the resource group and click the **+ Add** button
* Search for **Azure Cosmos DB** and press **Enter**
* Click on the **Create** button
* Make sure that the **Subscription** and **Resource group** are set to the desired values
* For **Account Name** enter for example **cosdb-hw**
* Change the **API** selection to **Azure Cosmos DB for MongoDB API**
* Set the **Location** to **West Europe** for example
* Accept the default values for the rest of the parameters
* Click on the **Review + create** button
* Confirm the creation with the **Create** button
* Once the create operation is complete, click on **Go to resource** button
* You will be navigated to the **Quick start** section
* Switch to the **Overview** section to see general information about the resource
* Either click on the **Data Explorer** button or click on the **Data Explorer** option in the left menu
* Click on the **New Collection** button
* Select the **Create new** option in the **Database id** item
* Enter for example **hwdb** in the text field
* Enter **hwcollection** in the **Collection id** text field
* Switch the **Storage capacity** parameter to **Fixed (10 GB)**
* Click on the **OK** button to initiate the creation process
* Once both the database and the collection are created, expand the item tree in the data explorer and select the **Documents** node
* Click on the **New Document** button
* Enter the following:

{

    "id" : "101",

    "name" : "John Smith",

    "sex" : "M",

    "salary" : 5000

}

* Click the **Save** button
* Click the **New Document** button
* Enter the following

{

    "id" : "102",

    "name" : "Jane Hudson",

    "sex" : "F",

    "salary" : 6500

}

* Click the **Save** button
* Click the **New Document** button
* Enter the following

{

    "id" : "103",

    "name" : "Amy Watson",

    "sex" : "F",

    "salary" : 4500

}

* Click the **Save** button
* Click the **New Document** button
* Enter the following

{

    "id" : "104",

    "name" : "Jack Daniel",

    "sex" : "M",

    "salary" : 7500

}

* Click the **Save** button
* Click the **New Document** button
* Enter the following

{

    "id" : "105",

    "name" : "Drew Hewit",

    "sex" : "F",

    "salary" : 5000

}

* Click on the **Save** button
* Click on the **New Shell** button
* Enter the following

db.hwcollection.aggregate([{$group: {\_id: "$sex", avgSalary: { $avg: "$salary" } } }]);

* And press the **Enter** key
* The result will be average salary by sex

### Stream Analytics (Azure Portal)

One possible solution, using the Azure Portal, may include the following steps:

#### Resource group and storage account

* Open a browser tab, navigate to <https://portal.azure.com> and login
* Navigate to **Resource Groups** and create a new one, or reuse the one created earlier (**Homework-M3**)
* Create a storage account named for example **sachw**
* Go to the storage account and click on **Containers** option in the **Blob services** section
* Click on the **+ Container** button to create a new one
* In the **Name** field enter **blobhw** and click **Create**

#### SQL server and database

* Return to the resource group and click the **+ Add** button to create an SQL Database
* Search for **Azure SQL** and hit **Enter**
* Select **Single database** and click **Create**
* For **Database name** enter **dbhw**
* Click on the **Create new** link under **Server** to create a new database server
* Enter **srvhw** for **Server name**
* Enter **demosa** for **Server admin login**
* Enter **DemoPassword-2020** for **Password** and retype it in the **Confirm password** field
* Set the appropriate **Location.** For example, **West Europe**
* Confirm server settings with the **OK** button
* Change **Compute + storage** plan by clicking on the **Configure database**
* Select the **Basic** plan and confirm with the **Apply** button
* Click on the **Next: Networking** button
* For **Connectivity method** select **Public endpoint**
* Switch to **Yes** both **Allow Azure services** and **Add current client IP address**
* Click **Review + create** and then click **Create** to create both the database server and database itself
* Once the database is created, navigate to its **Overview** screen
* Click on the **Query editor (preview)** item in the left menu
* Enter the credentials you specified during the database creation process
* Execute the following statement to prepare the table:

CREATE TABLE tblMeasurements (

TS datetime,

Sensor varchar(25),

MeasurementCount int,

AvgTemperature decimal(10,5))

#### Stream analytics job

* Now, return to the home screen
* Next, enter in the search bar **stream analytics** text and press **Enter**
* Then, click on the **+ Add** button
* For **Job name** use **sahw**
* Make sure that the resource group is set to the one created earlier
* Change the **Streaming units** to **1**
* Other settings should be fine
* Click on the **Create** button
* Once the job is created, click on the **Go to resource** button
* In the **Overview** section of the **sahw** job click on the **Inputs** option in the **Job topology** section
* Select **+ Add stream input** to create new input stream
* In the drop-down list choose **Blob storage**
* For **Input alias** enter **StreamIn**
* Ensure that the **Subscription** is correctly set
* For **Storage account** select the one created earlier - **sachw**
* Under the **Container** select **Use existing**
* Select the **blob container** that you created earlier - **blobhw**
* In the **Path pattern** field enter **input/**
* Click on the **Save** button
* Now, it is time to configure the output stream
* Click on the **Outputs** command in the **Job topology** section
* Click on the **+ Add** button
* In the drop-down menu select **SQL Database**
* For **Output alias** enter **StreamDB**
* Make sure that the **Select SQL Database from your subscriptions** option is selected
* Pay attention to the **Subscription**
* Select the database created earlier - **dbhw**
* Make sure that the **Authentication mode** is set to **Connection string**
* In the **Table** field enter **tblMeasurements**
* For **Username** and **Password** enter the credentials specified during the database creation
* Click on **Save**
* Click on the **Query** option in the **Job topology** section
* Set the query to:

SELECT System.Timestamp AS TS,

       Sensor,

       COUNT(\*) AS MeasurementCount,

       AVG(Measurement.Value) AS AvgTemperature

INTO StreamDB

FROM StreamIn

GROUP BY Sensor, TumblingWindow(mi, 1)

* Click on **Save query** button

#### Data generation with an application

* Extract the archive provided with the practice materials – **AzureBlobJsonGenerator-bin** in a local folder of your choice
* Open the **AzureBlobJsonGenerator.exe.config** file in a text editor
* Check the **BlobContainer** setting. Set it to match the container you created earlier - **blobhw**
* Go to the storage account
* Then, navigate to the **Access Keys** option under the **Settings** section
* Copy the **Connection string** value either for **key1** or **key2**
* Paste the value in the **StorageConnectionString** setting
* Check and adjust also the **LocalTempPath** parameter, it must point to a folder on your PC, that will be used to store the files generated while the application is working
* Save and close the file

#### Start the job

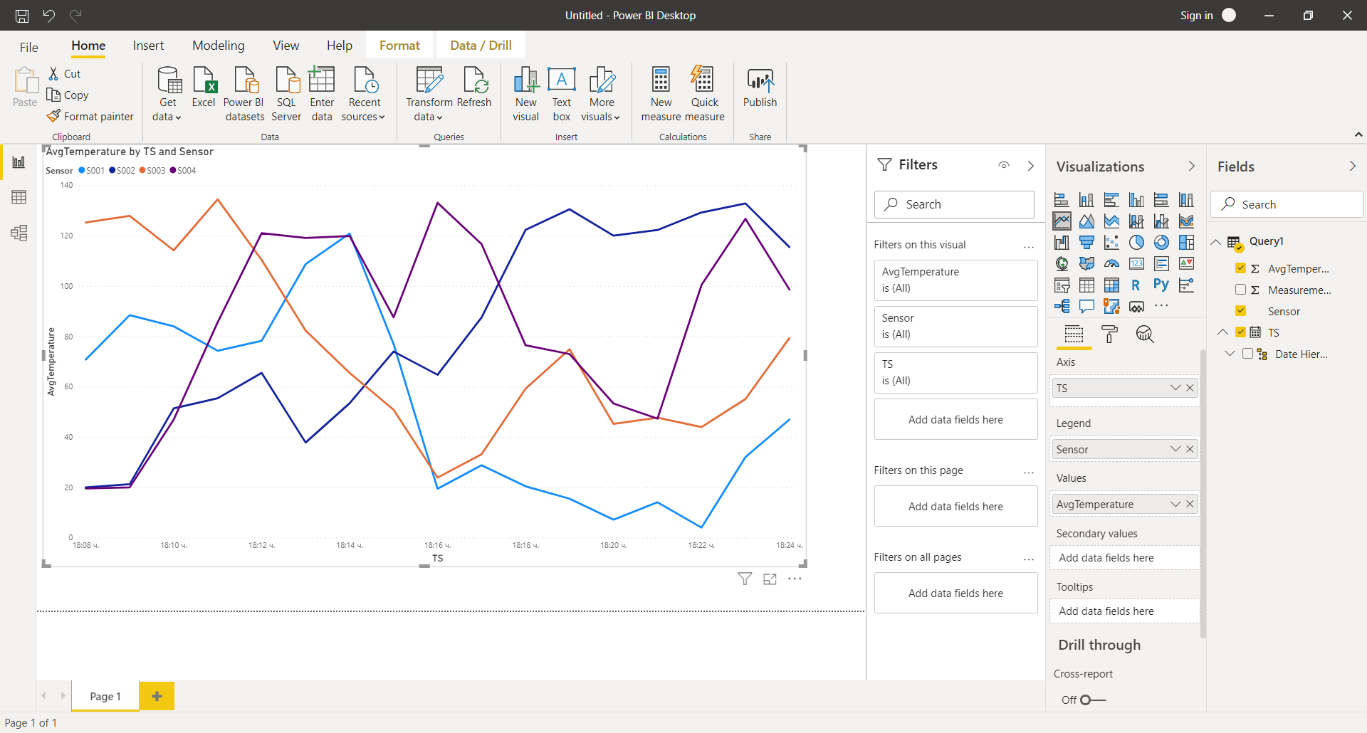
* Return to the **Overview** mode of the stream analytics job
* Click on the **Start** button
* Leave the **Job output start time** to **Now** and click on **Start**
* Wait for the job to start
* Once it is started, go to the folder where you did extract the data generation application
* Doble click on the **AzureBlobJsonGenerator.exe**
* Press the **Enter** key to initiate the data generation process
* Check both the blob container and the SQL database for data
* If you want, you can stop both the generation app and the job. It is enough to have just a small quantity of data

#### Visualization

* Download (<https://powerbi.microsoft.com/en-us/downloads/>) and install Power BI
* Start the application
* Click the **Get Data > SQL Server** or directly the button **SQL Server**
* Copy the server name from the Azure Portal (for example srvhw.database.windows.net)
* Return to the Power BI app and paste it into the **Server** text field
* Enter the database name (dbhw) into the **Database** field
* Click on the **Advanced options** link
* Paste the following statement into the **SQL statement** text field

SELECT TS, Sensor, MeasurementCount, AvgTemperature FROM [dbo].[tblMeasurements]

* Click **OK**
* On the next screen switch to **Database**
* Enter the credentials specified during the database creation
* Click **Connect**
* Click **Load**
* Select the **Line chart** visualization
* Resize it
* Select **AvgTemperature** and drop it into **Values** placeholder
* Select **Sensor** and drop it into **Legend** placeholder
* Select **TS** and drop it into **Axis** placeholder
* Click on the drop-down arrow next to the **TS** field as it is in the **Axis** placeholder
* Select **TS** instead of **Date Hierarchy**
* Now, you should see a graph showing the changes in the temperature:



## A Reminder

Do not forget to remove the resources created during the homework solution practice. The easiest option to accomplish this would be to delete the resource group