



Projet PROCOM : Migration d'une solution de Business Intelligence existante vers le cloud Azure

Guide de déploiement de l'architecture

Groupe 22 : Solène LEMERCIER, Mohamed Amine KROUT, Nour ABBOUD, Uyen NGUYEN, Menglin XIONG
Année : 2023-2024

Sommaire :

Introduction :	3
I/ Création d'un compte Azure	4
II/ Importation/Déploiement du projet sur ce nouveau compte	5
III/ Et après : quelles améliorations ?	14

Introduction :

Dans ce rapport, nous présenterons comment redéployer l'entièreté du projet sur un nouveau compte Azure. Dans une première partie, nous présenterons la création d'un compte Azure étudiant au sein d'IMT Atlantique. Dans une seconde partie, nous verrons l'importation du projet. Enfin, nous présenterons les améliorations envisageables de notre architecture.

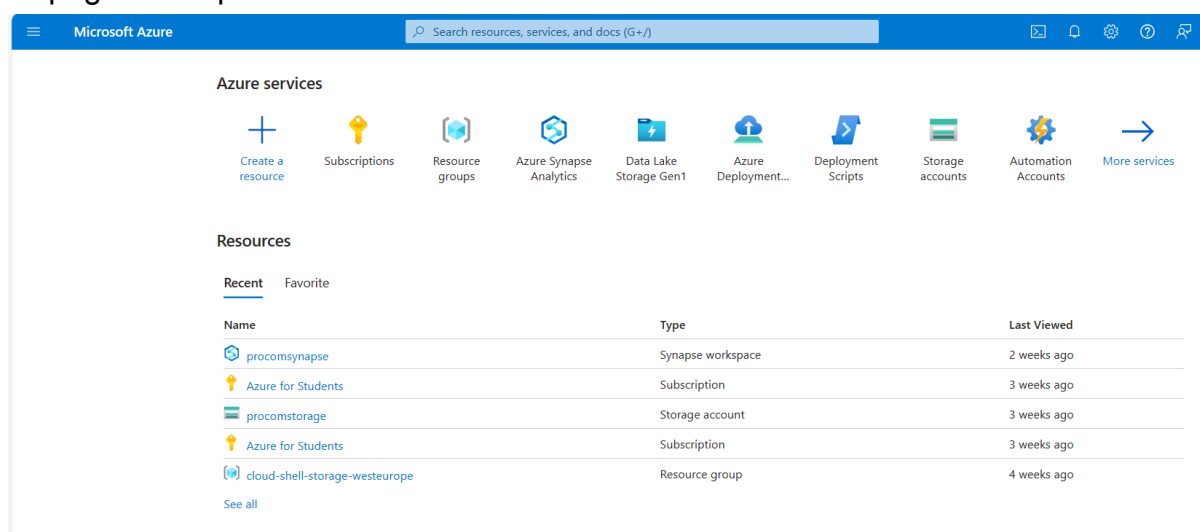
I/ Création d'un compte Azure

Dans un premier temps, il est nécessaire d'avoir un compte Azure pour réaliser le déploiement du projet.

Pour se créer un compte Azure via IMT Atlantique il faut faire les étapes suivantes :

1. Se connecter à ce site [Connexion \(imt-atlantique.fr\)](https://connexion.imt-atlantique.fr) avec les identifiants et mots de passe de votre compte IMT Atlantique (pas l'adresse mail mais l'identifiant type x24xxxxx)
2. Une fois connecté, allez dans la rubrique **Profil**
3. Cocher la case **Communiquer mes données pour l'utilisation de Microsoft Azure**
4. Puis **Enregistrer**. Les informations mises à jour seront prises en compte et un compte sur Azure sera automatiquement créé entre 12h35 et 23h.
5. Le compte créé, connectez vous sur [Se connecter à Microsoft Azure \(microsoftonline.com\)](https://microsoftonline.com) et les 100 dollars gratuits pour les étudiants sont normalement actifs et vous disposez donc d'un crédit à utiliser tout le temps nécessaire de votre scolarité à IMT Atlantique. Si il n'y a pas les crédits alloués aux étudiants, consultez ce lien : [Azure for Students – Crédit de compte gratuit | Microsoft Azure](#)

La page sur laquelle vous vous trouvez est semblable à celle-ci :



II/ Déploiement/Importation du projet sur Azure

Dans cette partie, nous décrirons comment déployer le projet. Nous avons au préalable écrit un fichier setup.ps1 pour pouvoir créer l'ensemble des services nécessaires automatiquement

Le fichier est un script shell configuré de manière à être exécuté directement par l'utilisateur via le Powershell fourni dans l'environnement Azure.

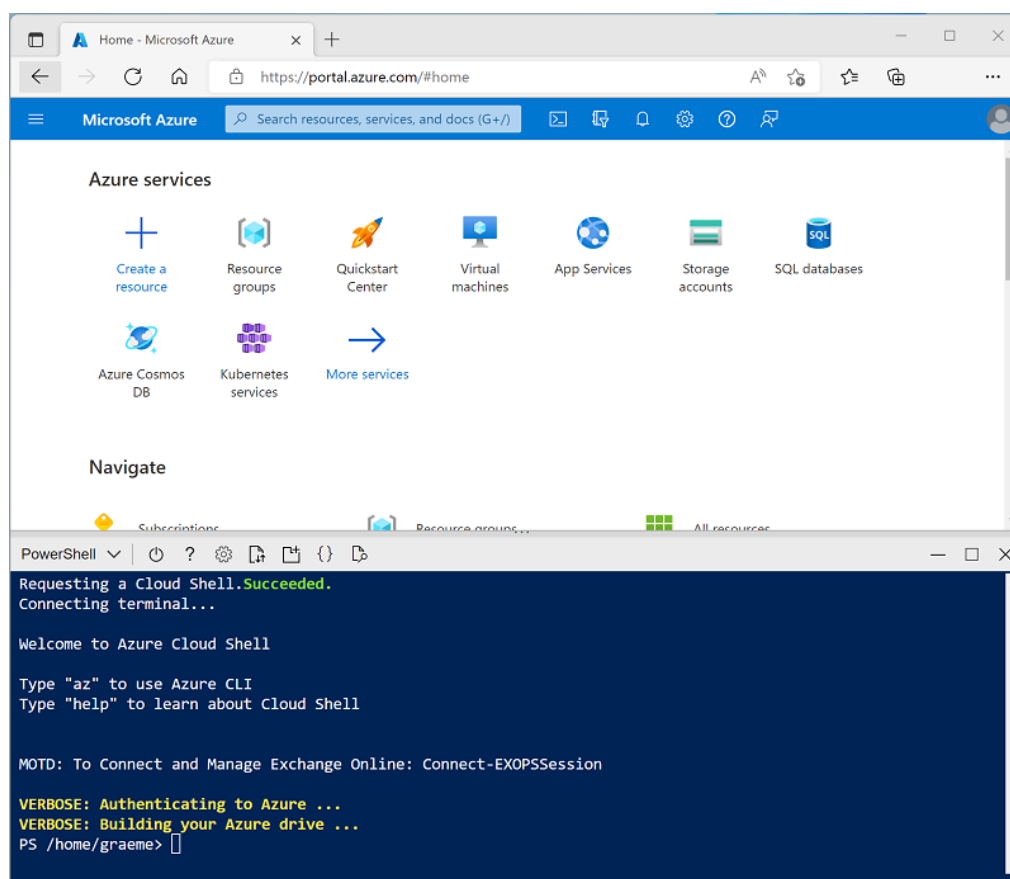
L'idée a été inspirée par les travaux pratiques fournis par Microsoft Learn.

L'exécution du fichier configurera l'environnement pour qu'il puisse être utilisé directement par les étudiants sans autre connaissance de la configuration. Il créera le compte de stockage, créera un service synapse et importera des fichiers via github dans l'espace de travail synapse. De plus, il créera le serveur SQL et la base de données serverless qui hébergeront le puits du pipeline configuré plus tard. En outre, il faudrait après créer un Linked Service entre la base de données et l'espace Synapse.

Remarque : Dans le script, on a spécifié une variable qui s'appelle **\$preferred_list** cette variable définit les régions parmi lesquelles on choisit au hasard celle disponible pour héberger les services dont on aura besoin durant le projet. Nous avons limité cette liste à deux régions : **northeurope** et **westeurope**. It may be the case that these regions are not available, one of the two

Pour effectuer l'importation du projet :

1. Utilisez le bouton [**>**_] à droite de la barre de recherche en haut de la page pour créer un nouveau Cloud Shell dans le portail Azure, en sélectionnant un environnement PowerShell et en créant un espace de stockage si vous y êtes invité. Le Cloud Shell fournit une interface de ligne de commande dans un volet situé en bas du portail Azure, comme illustré ici :



2. Dans le powershell, il faut écrire la commande qui permettra de cloner le github afin de configurer les environnements de stockage des données sources (Azure Data Lake Gen 2), de traitement (Azure Synapse Analytics) et de stockage des données transformées (Azure Sql server et Azure Sql Database) :


```
rm -r exercise -f
```

```
git clone
```

```
https://github.com/amine-krouit/procom-buisness-intelligence.git exercise
```
3. Une fois le clonage du github effectué, il faut se placer dans le bon dossier afin d'exécuter le script de configuration :


```
cd exercise
```

```
./setup.ps1
```
4. Il faut ensuite se laisser guider par les instructions apparaissant dans le powershell (choisir l'abonnement Azure qu'on souhaite utiliser et créer un mot de passe)
5. Le chargement et la création des différents services peut prendre du temps.
6. Une fois que tout a été exécuté, cliquez sur le groupe de ressource qui a été créé avec un nom commençant par procom-buisness-intelligence-XXXXXX (nom lié au nom du github qui a été cloné) puis cliquez sur le serveur qui porte le nom 'sqlserver-procom'. Au sein de l'environnement du serveur, il faut sélectionner le bouton 'Networking' puis cocher le paramètre : " et sauvegarder le serveur mis à jour.

Azure services

Create a resource Subscriptions Resource groups Azure Synapse Analytics Data Lake Storage Gen1 Azure Deployment... Deployment Scripts Storage accounts Automation Accounts More services

Resources

Recent Favorite

Name	Type	Last Viewed
sqlserver-procom	SQL server	a minute ago
procom-business-intelligence-n0recq2	Resource group	2 minutes ago
Azure for Students	Subscription	3 hours ago
procomsynapse	Synapse workspace	5 hours ago
procomstorage	Storage account	5 hours ago
Azure for Students	Subscription	21 hours ago
cloud-shell-storage-west europe	Resource group	7 days ago

See all

procom-business-intelligence-n0recq2 Resource group

Search Create Manage view Delete resource group Refresh Export to CSV Open query Assign tags Move Delete Export template JSON View

Overview Activity log Access control (IAM) Tags Resource visualizer Events Settings Deployments Security Deployment stacks Policies Properties

Essentials

Resources Recommendations

Filter for any field... Type equals all Location equals all Add filter

Showing 1 to 4 of 4 records. Show hidden types No grouping List view

Name	Type	Location
datalaken0recq2	Storage account	West Europe
sqldatabase-procom (sqlserver-procom/sqldatabase-procom)	SQL database	West Europe
sqlserver-procom	SQL server	West Europe
synapsen0recq2	Synapse workspace	West Europe

sqlserver-procom SQL server

Search

DTU quota Properties Locks

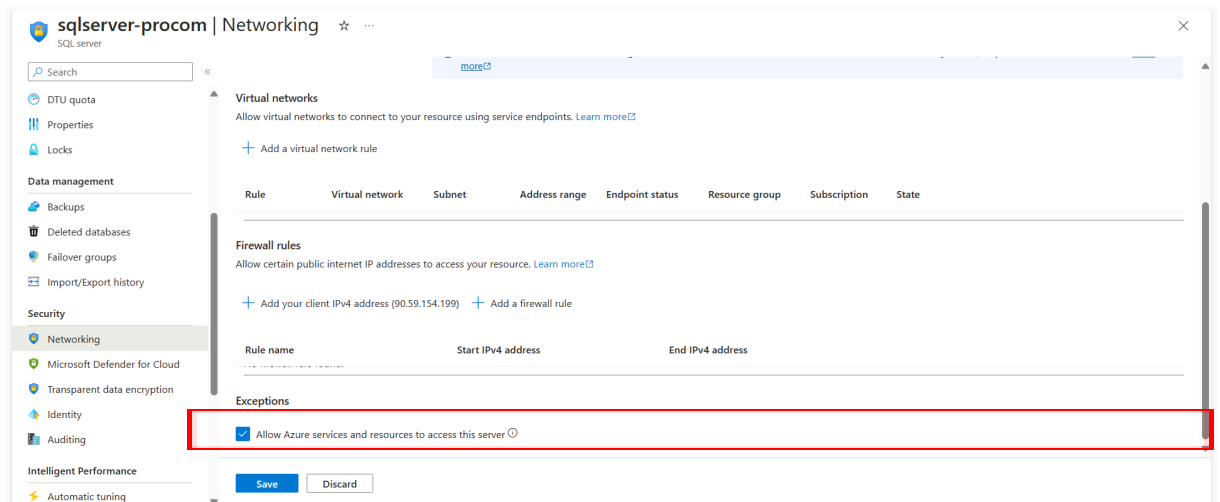
Data management

Backups Deleted databases Failover groups Import/Export history

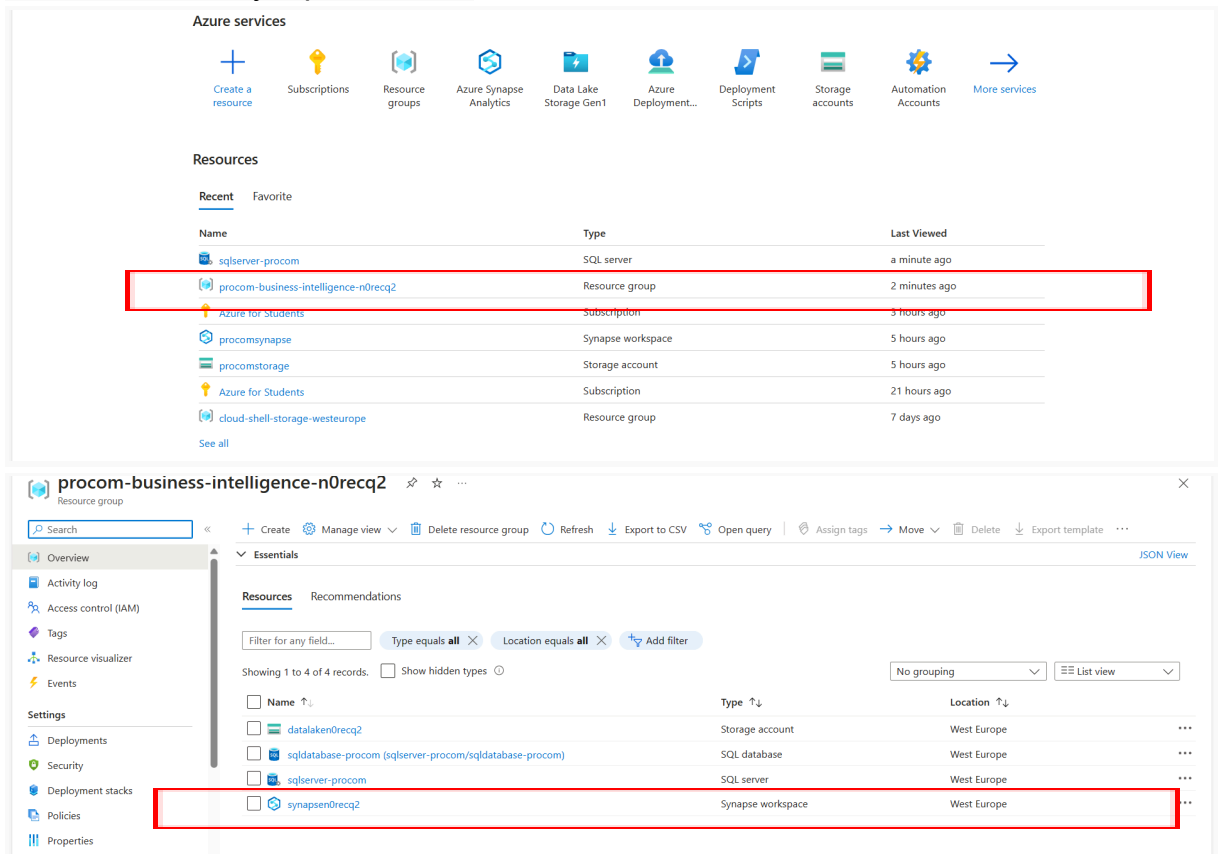
Security

Networking Microsoft Defender for Cloud Transparent data encryption Identity Auditing

Intelligent Performance



7. En retournant dans le groupe de ressources, cliquez sur synapse-XXXXXX et ensuite d'ouvrir Synapse Studio :



[+ New dedicated SQL pool](#)
[+ New Apache Spark pool](#)
[+ New Data Explorer pool \(preview\)](#)
[Refresh](#)
[Reset SQL admin password](#)
[Delete](#)

[^ Essentials](#)
[JSON View](#)

Resource group (move) : [procom-business-intelligence-n0recq2](#)
 Status : Succeeded
 Location : West Europe
 Subscription (move) : [Azure for Students](#)
 Subscription ID : 2c5a5adb-fa1b-4f41-b99f-2aefa8e0ff0
 Managed virtual network : No
 Managed Identity object ... : ac00d12e-7929-4584-8fcc-79c2dab90081
 Workspace web URL : <https://web.azuresynapse.net?workspace=%2fsubscriptions%2f2...>
 Tags (edit) : [Add tags](#)

Networking : [Show firewall settings](#)
 Primary ADLS Gen2 acco... : <https://datalaken0recq2.dfs.core.windows.net>
 Primary ADLS Gen2 file s... : files
 SQL admin username : SQLUser
 SQL Microsoft Entra admin : solene.lemercier@imt-atlantique.net
 Dedicated SQL endpoint : synapsen0recq2.sql.azuresynapse.net
 Serverless SQL endpoint : synapsen0recq2-ondemand.sql.azuresynapse.net
 Development endpoint : <https://synapsen0recq2.dev.azuresynapse.net>

Getting started

Open Synapse Studio
Start building your fully-integrated analytics solution and unlock new insights.
[Open](#)

Read documentation
Learn how to be productive quickly. Explore concepts, tutorials, and samples.
[Learn more](#)

8. Une fois dans Synapse Studio, ouvrir le volet Integrate et cliquer sur le (+) au-dessus de Pipelines afin d'importer le template de la pipeline que l'on a téléchargé en I/1. :

[Synapse live](#)
[Validate all](#)
[Publish all](#)

[+ Integrate](#)

Filter resources by name

Pipelines

Pipeline
 Link connection
 Copy Data tool
 Browse gallery
 Import from pipeline template

9. Il suffit de sélectionner le bon nom correspondant au stockage de données sources

[synapsrhe74by](#)
[Activity log](#)
[Access control \(IAM\)](#)
[Tags](#)
[Diagnose and solve problems](#)

[Settings](#)
[Microsoft Entra ID](#)
[Properties](#)
[Locks](#)

[Analytics pools](#)
[SQL pools](#)
[Apache Spark pools](#)
[Data Explorer pools \(preview\)](#)

[Security](#)

[+ New dedicated SQL pool](#)
[+ New Apache Spark pool](#)
[+ New Data Explorer pool \(preview\)](#)
[Refresh](#)
[Reset SQL admin password](#)
[Delete](#)

[^ Essentials](#)
[JSON View](#)

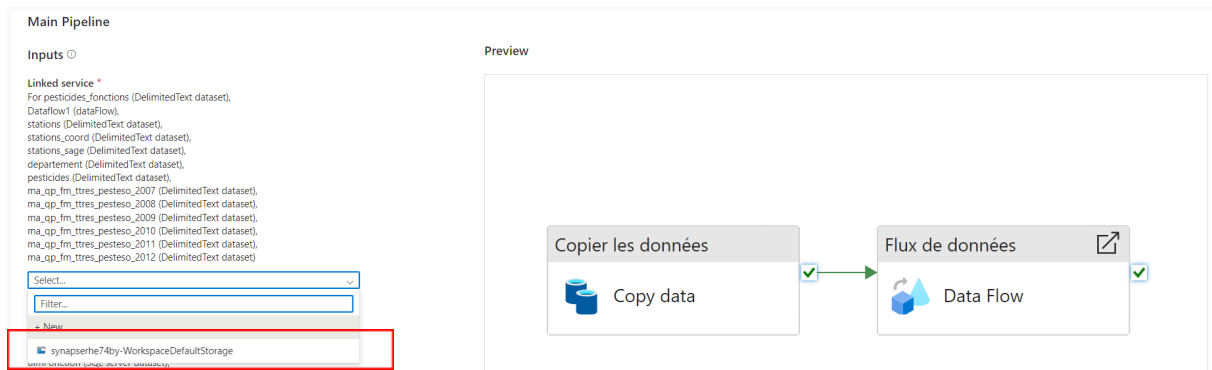
Resource group (move) : [procom-business-intelligence-the74by](#)
 Status : Succeeded
 Location : East US 2
 Subscription (move) : [Azure for Students](#)
 Subscription ID : 2c5a5adb-fa1b-4f41-b99f-2aefa8e0ff0
 Managed virtual network : No
 Managed Identity object ... : 33f5f5c-ec3-4a8b-afc9-d6a76956dbe0
 Workspace web URL : <https://web.azuresynapse.net?workspace=%2fsubscriptions%2f2...>
 Tags (edit) : [Add tags](#)

Networking : [Show firewall settings](#)
 Primary ADLS Gen2 acco... : <https://datalakerhe74by.dfs.core.windows.net>
 Primary ADLS Gen2 file s... : files
 SQL admin username : SQLUser
 SQL Microsoft Entra admin : solene.lemercier@imt-atlantique.net
 Dedicated SQL endpoint : synapsrhe74by.sql.azuresynapse.net
 Serverless SQL endpoint : synapsrhe74by-ondemand.sql.azuresynapse.net
 Development endpoint : <https://synapsrhe74by.dev.azuresynapse.net>

Getting started

Open Synapse Studio
Start building your fully-integrated analytics solution and unlock new insights.
[Open](#)


Read documentation
Learn how to be productive quickly. Explore concepts, tutorials, and samples.
[Learn more](#)



10. En ce qui concerne le service qui va stocker les données transformées (tables de dimension et de faits), il faut en créer un nouveau, mais celui-ci doit être lié au serveur SQL que nous avons créé via le script. Cliquez sur (+ New) et un onglet va s'ouvrir.

This screenshot shows the 'Linked service' configuration window. It lists several SQL server datasets: 'dimStationTable', 'dimPesticideTable', 'factPrelevementsTable', 'dimFonction', and 'dimGroupeFonction'. Below the list, there are 'Select...' and 'Filter...' dropdowns. A red rectangular box highlights the '+ New' button at the bottom of the window.

11. Dans cet onglet, il faut remplir les paramètres avec le bon nom du serveur qui a été créé dans le groupe de ressources (sqlserver-procom.database.windows.net), le bon nom de la base de données qui a été créé au sein du serveur dans le groupe de ressource (sqldatabase-procom), le bon nom d'administrateur (SQLUser) et le mot de passe que vous avez entré lors de l'exécution du script. L'ensemble des informations précédentes se trouvent dans la page d'accueil d'Azure dans le groupe de ressource créée. Le nom du linkedServer doit correspondre à SqlServer1. Il faut ensuite valider.

New linked service
 SQL server [Learn more](#)

i Choose a name for your linked service. This name cannot be updated later.

Name *

Description

Connect via integration runtime *
☒ AutoResolveIntegrationRuntime

Connection string **Azure Key Vault**

Server name *

Database name *

Authentication type

User name *

Password **Azure Key Vault**

Password *

12. La configuration est alors terminée. Il suffit de cliquer sur 'publish all' pour pouvoir sauvegarder et cliquer encore sur 'publish'.

Synapse live

Integrate

Pipelines 1

- Main Pipeline

Activities

- Synapse
- Move and transform
- Azure Data Explorer

☐ Data flow debug

Copy data ☒ **Data flow**

Publish all

You are about to publish all pending changes to the live environment. [Learn more](#)

Pending changes (19)

NAME	CHANGE	EXISTING
✓ Pipelines		
Main Pipeline	(New)	-
✓ Datasets		
pesticides_fonctions	(New)	-
stations	(New)	-
stations_coord	(New)	-
stations_sage	(New)	-
departement	(New)	-
pesticides	(New)	-
ma_qp_fm_ttres_pesteso_...	(New)	-
ma_qp_fm_ttres_pesteso_...	(New)	-
ma_qp_fm_ttres_pesteso_...	(New)	-
ma_qp_fm_ttres_pesteso_...	(New)	-
ma_qp_fm_ttres_pesteso_...	(New)	-
ma_qp_fm_ttres_pesteso_...	(New)	-
ma_qp_fm_ttres_pesteso_...	(New)	-
dimStationTable	(New)	-
dimPesticideTable	(New)	-

Publish **Cancel**

Quelques actions utiles :

- Pour obtenir une prévisualisation des tables : sélectionner 'Debug' puis valider avec **Ok**

✓ Validate **▶ Debug** ▼ ⚡ Add trigger ☐ Data flow debug

Copy data

Copy data

Data flow

Data Flow

Turn on data flow debug

Integration runtime
AutoResolveIntegrationRuntime

AutoResolveIntegrationRuntime

Region	AutoResolve
Compute size	Small

Debug time to live ⓘ
1 hour

OK Cancel

- Pour run la pipeline: sélectionner 'add trigger' puis 'trigger now'

✓ Validate ▶ Debug ▼ ⚙ Add trigger ⏻ Data flow debug

Trigger now

New/Edit

Copy data

Data flow

Data Flow

Pipeline run

⚠ Trigger pipeline now using last published configuration.

Parameters

Name	Type	Value
No records found		

OK Cancel

III/ Et après : quelles améliorations ?

Il y a toujours des améliorations possibles pour ce projet. Actuellement, l'automatisation du processus que nous avons mis en place présente des limites, car certaines configurations doivent toujours être effectuées manuellement (création du linked service SQL). Ce serait plus pratique si tout pouvait être fait en un seul clic. De plus, en ce qui concerne la création des rapports PowerBI, nous avons suivi une méthode consistant à connecter notre base de données à PowerBI Desktop, puis à déployer les rapports sur Azure. Dans l'idéal, nous pourrions trouver une solution pour créer directement les rapports sur le service PowerBI sur Azure.

En outre, des améliorations peuvent également être apportées au niveau de la pipeline elle-même. Cela pourrait inclure la séparation des processus et l'amélioration des opérateurs utilisés afin de minimiser les erreurs dans les résultats présentés. (notamment lors d'une jointure, la colonne utilisée pour la jointure se dédouble il faut donc la supprimer après, peut-être qu'il existerait une solution pour éviter que la colonne se dédouble et donc de la supprimer après)

En résumé, ces améliorations restent des détails, mais elles ne compromettent en rien la qualité du travail réalisé. Ce travail nous permet avec succès de migrer d'une solution de business intelligence traditionnelle à une solution plus innovante et autonome utilisant le cloud.