

Rapport Détailé du Projet

**DataCo RecEngine -
Powered by Azure AI**

Réalisé Par :

Abdelhafid AIT EL MOKHTAR

Encadré Par :

M. Hamid OUFAKIR

2025-2026

Table des Matières

| | |
|---------------------------------------|----------|
| I. Résumé du Projet :..... | 3 |
| II. Les Étapes Clés :..... | 3 |
| 1. Compte de Stockage Azure :..... | 3 |
| 1.1. Définition :..... | 3 |
| 1.2. Création et Configuration :..... | 3 |
| 2. Azure Data Factory :..... | 7 |
| 2.1. Définition :..... | 7 |
| 2.2. Création et Configuration :..... | 7 |
| 3. Azure Machine Learning :..... | 17 |
| 3.1. Définition :..... | 17 |
| 3.2. Création et Configuration :..... | 18 |
| 3.3. Pipeline ML Designer :..... | 26 |

I. Résumé du Projet :

Ce projet vise à concevoir une solution End-to-End de Data & Machine Learning sur Azure pour une Fintech, avec pour objectif la prévision de la consommation énergétique (Forecasting).

L'architecture orchestre le cycle de vie complet de la donnée :

- Ingestion automatisée via Azure Data Factory.
- Stockage structuré dans un Data Lake Gen2 (Bronze/Raw layer).
- Développement ML Low-Code via Azure ML Designer (Nettoyage, Feature Engineering Python, Entraînement).

L'infrastructure respecte des contraintes strictes : région North Europe et cluster de calcul économique (**Standard_DS11_v2**).

II. Les Étapes Clés :

1. Compte de Stockage Azure :

1.1. Définition :

Un compte de stockage Azure agit comme un conteneur universel et sécurisé dans le cloud Microsoft, regroupant sous une adresse web unique divers services de données tels que les fichiers, les objets (**Blobs**), les **tables** et les **files d'attente**. Il offre une solution hautement disponible, durable et évolutive, capable de **stocker des volumes massifs** d'informations tout en garantissant leur redondance et leur accessibilité via Internet depuis n'importe où dans le monde.

1.2. Création et Configuration :

Pour Créer un **Compte de Stockage**, on suivit les étapes suivantes :

- Accéder au groupe de ressources (**aaitelmokhtar**).
- Cliquer sur Créer une nouvelle Ressource.
- Taper compte de stockage dans la barre de recherche.
- Choisir le compte de stockage, puis cliquer sur Créer.
- Donner un **Nom** au compte de stockage.

- Choisir **Stockage Blob ou Data Lake Storage Gen2** comme Type de Stockage.
- Sélectionner **North Europe** comme Région.
- Dans la section **Avancé**, on active l'espace des noms hiérarchiques.
- Cliquer sur **Vérifier+Créer**, puis sur **Créer**.
- Accéder à la ressource de compte de stockage, et dans le menu gauche on choisit **Stockage des données/Conteneurs**.
- Ajouter un nouveau conteneur en lui attribuant un Nom.

The screenshot shows the Microsoft Azure Marketplace interface. At the top, there's a search bar with 'compte de stockage' typed in. Below the search bar are several filter options: 'Tarification : Tous', 'Système d'exploitation : Tous', 'Type de serveur de publication : Tous', 'Type de produit : Tous', 'Afficher les éléments : Tous', and 'Nom de l'éiteur : All'. A prominent message says 'Nouveau ! Obtenir des suggestions générées par l'IA pour «compte de stockage»' with a button 'Afficher les suggestions'. The results section shows four items:

| Image | Nom | Détails |
|-------|------------------------------|---|
| | Compte de stockage | Microsoft Azure Service. Utilisez des blobs, des tables, des files d'attente, des fichiers et Data. |
| | Périmètre de sécurité réseau | Microsoft Azure Service. Limite d'isolement réseau logique pour les ressources PaaS. |
| | Groupe de ressources | Microsoft Azure Service. Gérer et déployer ensemble les ressources d'une application. |
| | MapleTap Solution Template | Canadian Centre for Cyber Secur... Azure Application. Government of Canada tool for cloud network security. |

The screenshot shows the 'Compte de stockage' creation page. At the top, it says 'Compte de stockage' and 'Ajouter aux favoris'. Below that, it shows the service provider as 'Microsoft | Azure Service' and a rating of '★ 4.3 (Évaluations de 1884)'. There are two radio buttons: 'Approuvé par défaut' and 'Éligible à l'avantage Azure'. Under 'Abonnement', there are dropdown menus for 'Paiement à l'utilisation' and 'Plan', both set to 'Storage account'. A large blue 'Créer' button is at the bottom right. At the bottom of the page, there are tabs for 'Vue d'ensemble', 'Offres', 'Informations d'utilisation + support', and 'Évaluations et avis'. A note about storage solutions is present, along with a link to 'Autres produits de Microsoft' and a 'Tout voir' button. On the far right, there's a 'Envoyer des commentaires' button.

Microsoft Azure Rechercher dans les ressources, services et documents (G+)

Accueil > aaitelmokhtar > Place de marché > Compte de stockage

Créer un compte de stockage

Abonnement * Paiement à l'utilisation

Groupe de ressources * aaitelmokhtar [Créer nouveau](#)

Détails de l'instance

Nom du compte de stockage * stdatadodgen2

Région * (Europe) North Europe [Déployer sur une zone étendue Azure](#)

Type de stockage préféré Stockage Blob Azure ou Azure Data Lake Storage Gen 2

Performance * Standard: Recommandé pour la plupart des scénarios (compte universel v2)

Redondance * Stockage localement redondant (LRS)

Précédent Suivant Vérifier + créer Envoyer vos commentaires

Microsoft Azure Rechercher dans les ressources, services et documents (G+)

Accueil > aaitelmokhtar > Place de marché > Compte de stockage

Créer un compte de stockage

Informations de base Avancé Mise en réseau Protection des données Chiffrement Balises Vérifier + créer

Sécurité

Configurez les paramètres de sécurité qui impactent votre compte de stockage.

Exiger un transfert sécurisé pour les opérations d'API REST

Autoriser l'activation de l'accès anonyme sur des conteneurs individuels

Activer l'accès de clé de compte de stockage

Authorisation Microsoft Entra par défaut dans le Portail Azure

Version TLS minimale Version 1.2

Étendue autorisée pour les opérations de copie (préversion) À partir de n'importe quel compte de stockage

Espace de noms hiérarchique

L'espace de noms hiérarchique, complété par un point de terminaison Data Lake Storage Gen2, active la sémantique des fichiers et des répertoires, accélère les chargements d'analytique données Big Data et active les listes de contrôle d'accès (ACL). [Découvrir plus d'informations](#)

Activer l'espace de noms hiérarchique

Protocoles d'accès

Précédent Suivant Vérifier + créer Envoyer vos commentaires

Microsoft Azure Rechercher dans les ressources, services et documents (G+)

Accueil > stdatacodgen2_1767347059691 | Vue d'ensemble

Votre déploiement a été effectué

Le déploiement stdatacodgen2_1767347059691 a été effectué avec succès. Il a été déployé le 02/01/2026 à 10:44:26. L'ID de corrélation est 9b873260-cfc7-4aa3-9d7b-91bada39e3af.

Détails du déploiement

Etapes suivantes

Accéder à la resource

Envoyer des commentaires

Partagez votre expérience avec le déploiement

Ajouter ou supprimer des favoris en appuyant sur Ctrl+Shift+F2

Cost Management Recevez des notifications pour vous aider à respecter votre budget et à éviter des frais imprévus sur votre facture. [Configurer les alertes de coût >](#)

Microsoft Defender pour le cloud Sécuriser vos applications et votre infrastructure. [Accédez à Microsoft Defender pour le cloud >](#)

Tutoriels Microsoft gratuits Commencez l'apprentissage aujourd'hui >

Travailler avec un expert Les experts Azure sont des partenaires fournisseurs de services qui peuvent vous aider à gérer vos ressources sur Azure et constituer votre première ligne de support. [Trouver un expert Azure >](#)

The screenshot shows the Microsoft Azure Storage Containers page for the 'stdatacodgen2' account. The left sidebar navigation includes 'Conteneurs' under 'Stockage des données'. The main content area displays a table of containers with one item: 'Logs' (Last modified: 02/01/2026 10:44:53, Level: Privé, Status: Disponible). A search bar at the top right is set to 'Rechercher les conteneurs par préfixe'.

The screenshot shows the Microsoft Azure Storage Containers page for the 'stdatacodgen2' account. The left sidebar navigation includes 'Conteneurs' under 'Stockage des données'. A modal window titled 'Nouveau conteneur' is open, showing fields for 'Nom' (set to 'datalake') and 'Niveau d'accès anonyme' (set to 'Privé (aucun accès anonyme)'). A note indicates that anonymous access is disabled. The 'Créer' button is visible at the bottom right of the modal.

The screenshot shows the Microsoft Azure Storage Containers page for the 'stdatacodgen2' account. The left sidebar navigation includes 'Conteneurs' under 'Stockage des données'. The main content area now lists two containers: 'Logs' and 'datalake' (Last modified: 02/01/2026 10:47:21, Level: Privé, Status: Disponible). The search bar at the top right is set to 'Rechercher les conteneurs par préfixe'.

2. Azure Data Factory :

2.1. Définition :

Azure Data Factory est un service d'intégration de données basé sur le cloud (PaaS) qui permet de créer, planifier et orchestrer des flux de travail automatisés, appelés pipelines, pour déplacer et transformer des données à grande échelle. Agissant comme une solution **ETL** (Extract, Transform, Load) et **ELT** moderne, il fonctionne tel un chef d'orchestre capable de se connecter à plus de 90 sources disparates (sur site ou dans le cloud) pour ingérer des données brutes, les nettoyer et les organiser, afin de les rendre prêtes à être exploitées par des outils d'analyse ou d'intelligence artificielle.

2.2. Crédit et Configuration :

Afin de configurer l'environnement d'**Ingestion des Données**, on suit les étapes suivantes :

- Accéder au groupe de ressources (**aaitelmokhtar**).
- Cliquer sur Créer une Nouvelle Ressource.
- Taper **Data Factory** dans la barre de recherche.
- Choisir Data Factory, puis cliquer sur **Créer**.
- Donner un **Nom** à Data Factory.
- Sélectionner North Europe comme Région.
- Cliquer sur **Vérifier+Créer**, puis sur **Créer**.
- Accéder à la ressource du Data Factory, et lancer **ADF Studio**.
- Dans le menu gauche, sélectionne **Auteur**.
- Créer un nouveau pipeline dans **Pipelines > Nouveau Pipeline** en lui donnant un Nom.
- Dans **Activités**, cherchez **Déplacer et Transformer > Copier les Données**. Glisser-déposer dans le canvas.

The screenshot shows the Microsoft Azure Marketplace interface. The search bar at the top contains the query "data factory". Below the search bar, there are several filters: "Tarification : Tous", "Système d'exploitation : Tous", "Type de serveur de publication : Tous", "Éligible à Azure Benefit uniquement", "Type de produit : Tous", "Afficher les éléments : Tous", and "Nom de l'éditeur : All". A callout box says "Nouveau ! Obtenir des suggestions générées par l'IA pour « data factory »" with a button "Afficher les suggestions". The main area displays 128 results for "data factory", including items like "Data Factory Monitoring Dashboard" by InIntel, "Data Factory" by Microsoft, "Factory Operations Agent in Azure AI (preview)" by Microsoft, "Sight Machine Factory Namespace MCP" by Sight Machine, Inc., and "Modern Data Mart" by Ceteris AG.

This screenshot shows the product page for "Data Factory" on the Azure Marketplace. The page includes the product icon, name, average rating (3.7), number of reviews (612), and status ("Approuvé par défaut"). It also shows the subscription plan ("Plan" - "Paiement à l'utilisation") and service type ("Data Factory"). Below the product details, there's a section titled "Intégrez des silos de données avec Azure Data Factory", which describes the service as a serverless integration service. At the bottom right, there's a "Envoyer des commentaires" button.

This screenshot shows the "Créez une fabrique de données" (Create a data factory) wizard in the Azure portal. The first step, "Informations de base", is selected. It requires filling in the "Nom" (Name) field with "adf-dataco-recengine", "Région" (Region) with "North Europe", and "Version" (Version) with "V2". There is a note below the form: "Un clic pour créer une fabrique de données avec un exemple de pipeline et de jeux de données. [Essayez-le] ([https://aka.ms/adf/tryittemplate] Try Fabric instead!)"

Microsoft.DataFactory-20260102105021 | Vue d'ensemble

Votre déploiement a été effectué

Détails du déploiement

Étapes suivantes

Accéder à la ressource

Envoyer des commentaires

Partagez votre expérience avec le déploiement

Ajouter ou supprimer des favoris en appuyant sur **Ctrl+Shift+F**

Gestion des coûts

Microsoft Defender pour le cloud

Tutoriels Microsoft gratuits

Travailler avec un expert

adf-dataco-recengine

Vue d'ensemble

Bases

Type : Fabrique de données (V2)

Démarrer : Démarrage rapide

Azure Data Factory Studio

Lancer le studio

Démarrages rapides

Tutoriels

Galerie des modèles

Modules d'entraînement

Mettez à niveau votre Data Factory vers Fabric Data Factory pour profiter des dernières fonctionnalités et améliorations.

Lancer l'évaluation de la mise à niveau (préversion)

Accueil

Ressources de fabrique

Pipelines

Capture des changements de données

Jeux de données

Flux de données

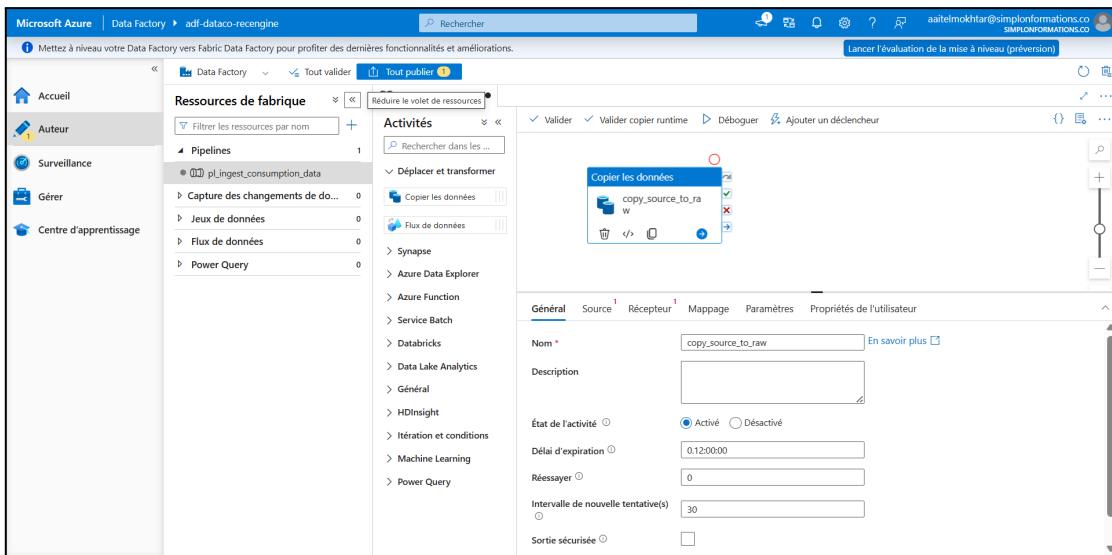
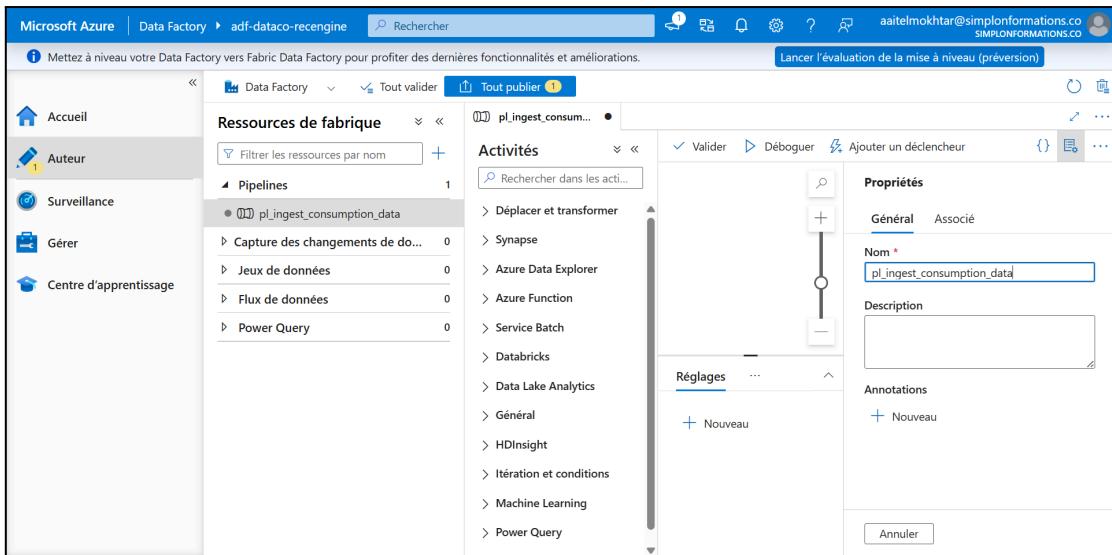
Power Query

Nouveau pipeline

Nouveau dossier

Sélectionner un élément

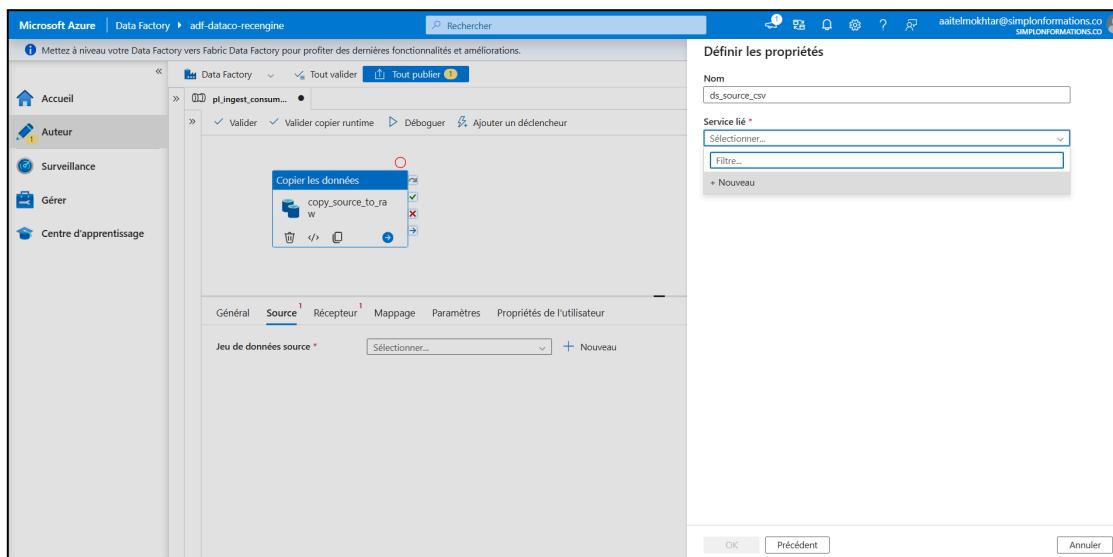
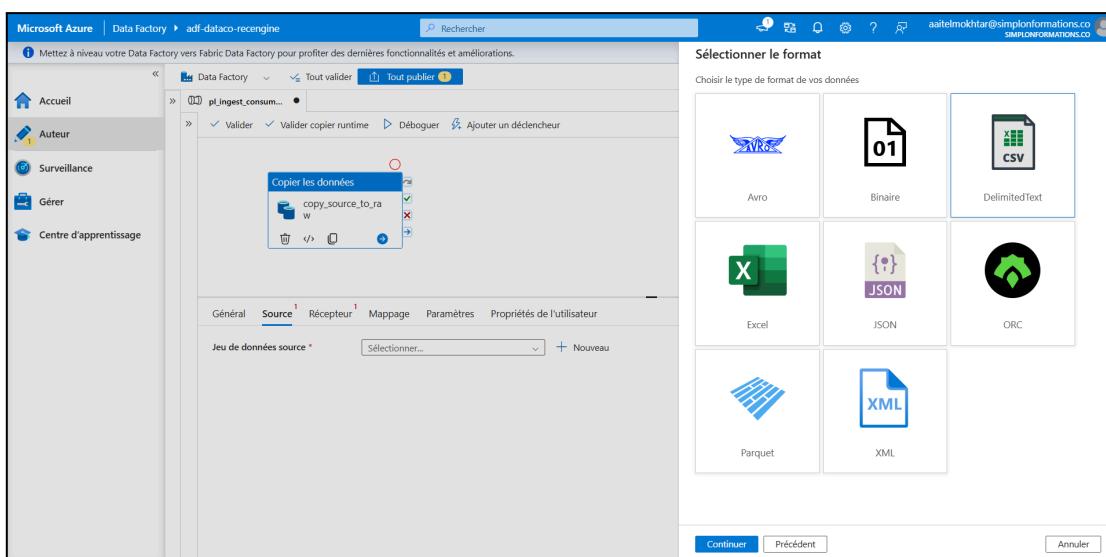
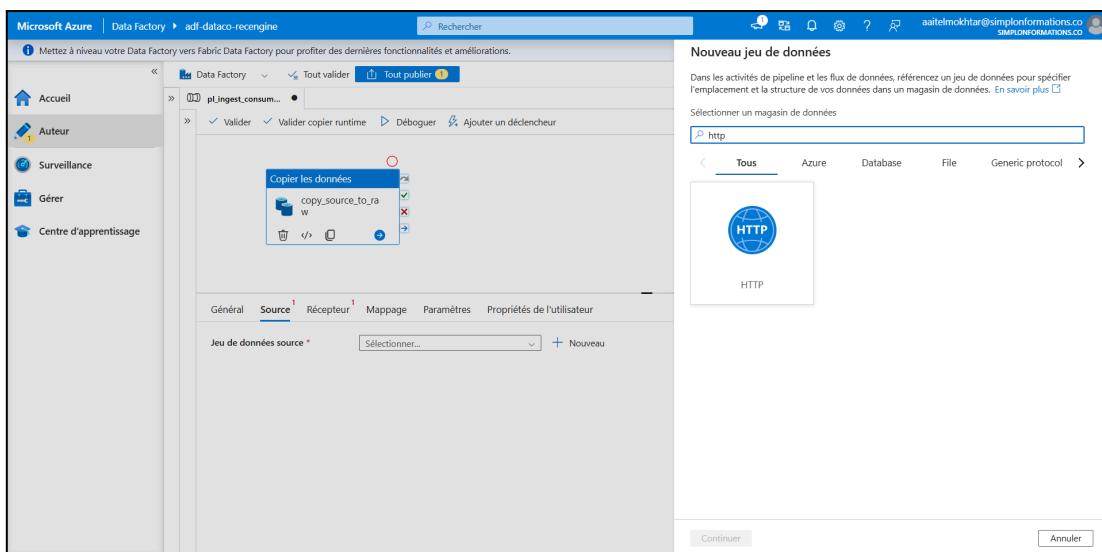
Utilisez l'explorateur de ressources pour sélectionner ou créer un élément

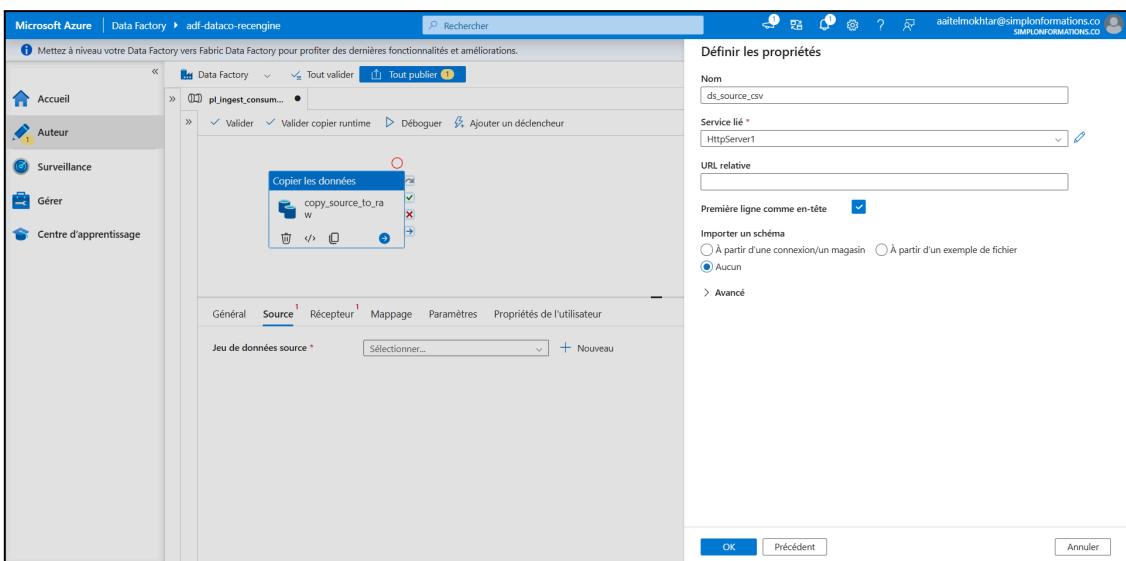
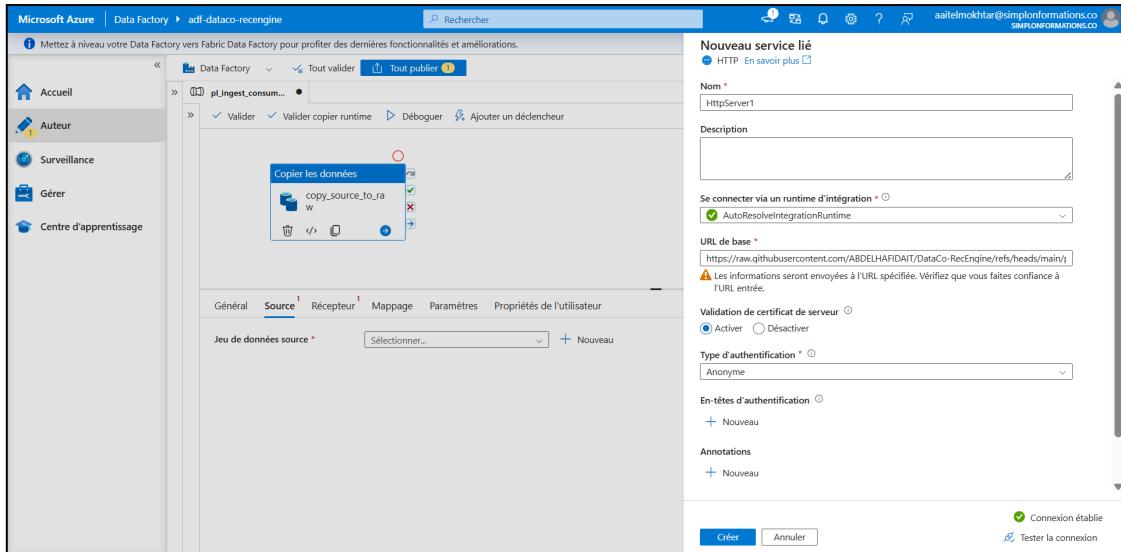


Après on passe à la configuration de l'activité **Copier des Données** :

Section 1 - Source :

- Dans la partie **Source**, on crée un **Nouveau jeu de données source**.
- Rechercher et choisir le **magasin de données HTTP** et **DelimitedText** comme format.
- Attribuer un nom à la source.
- Créer un nouveau **Service Lié** et passer **l'URL** du fichier CSV du dataset.
- Appuyer sur **OK**.

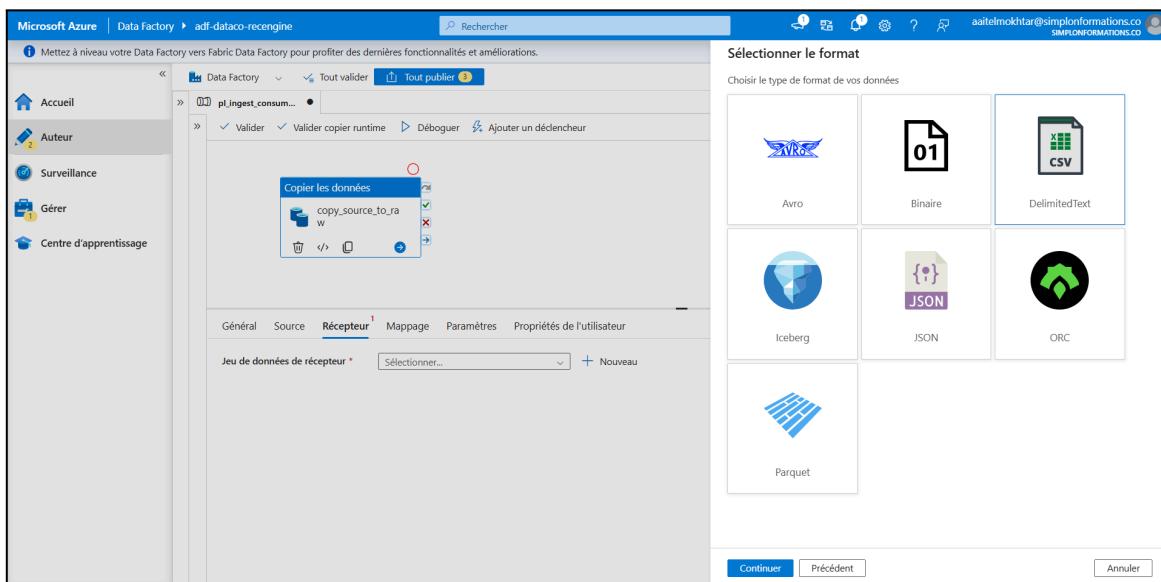
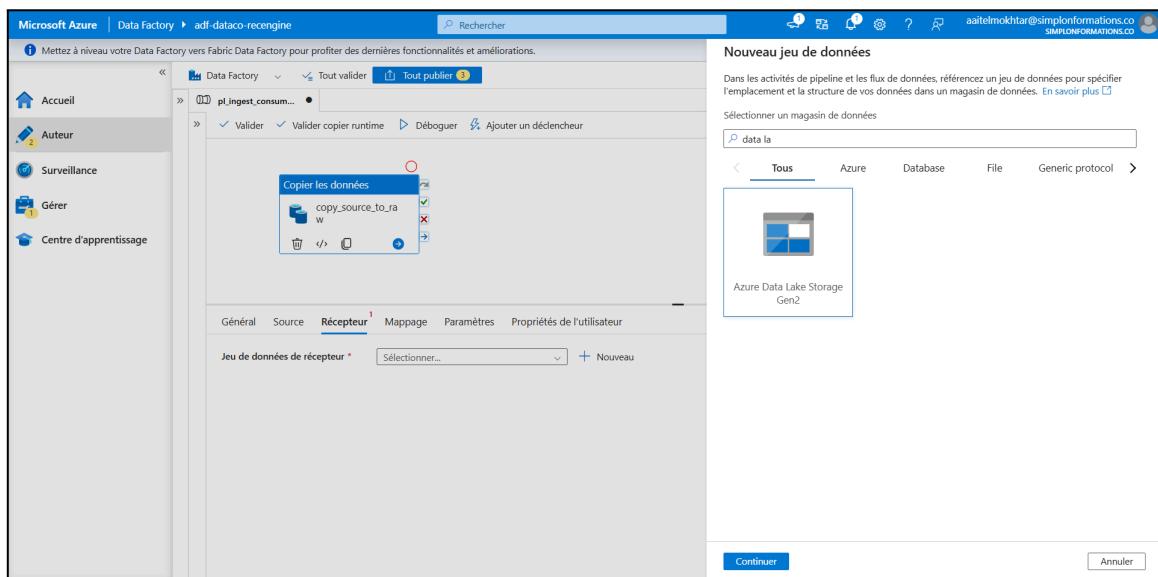


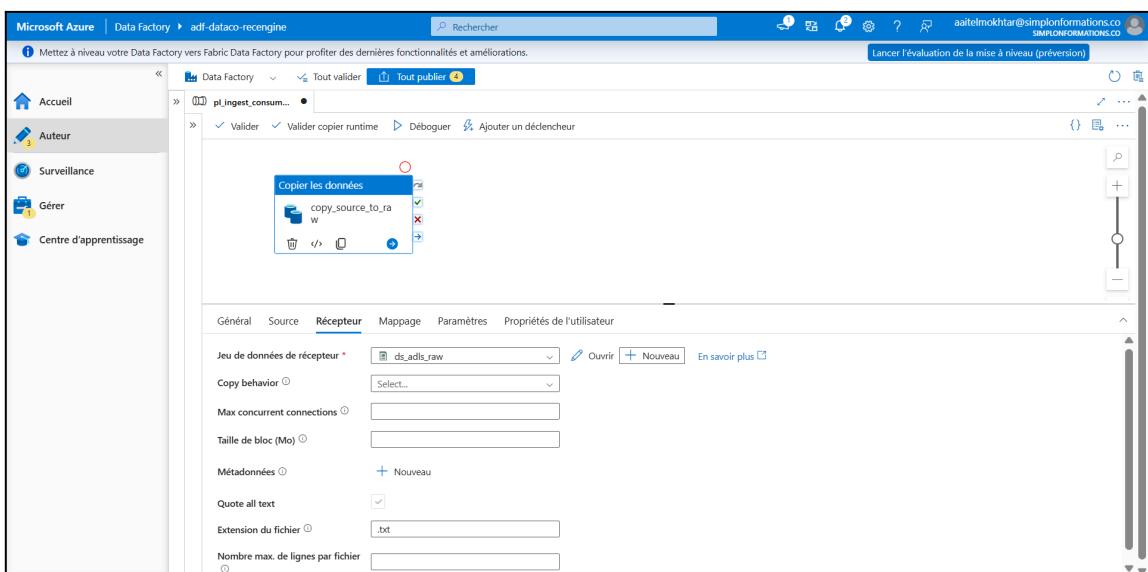
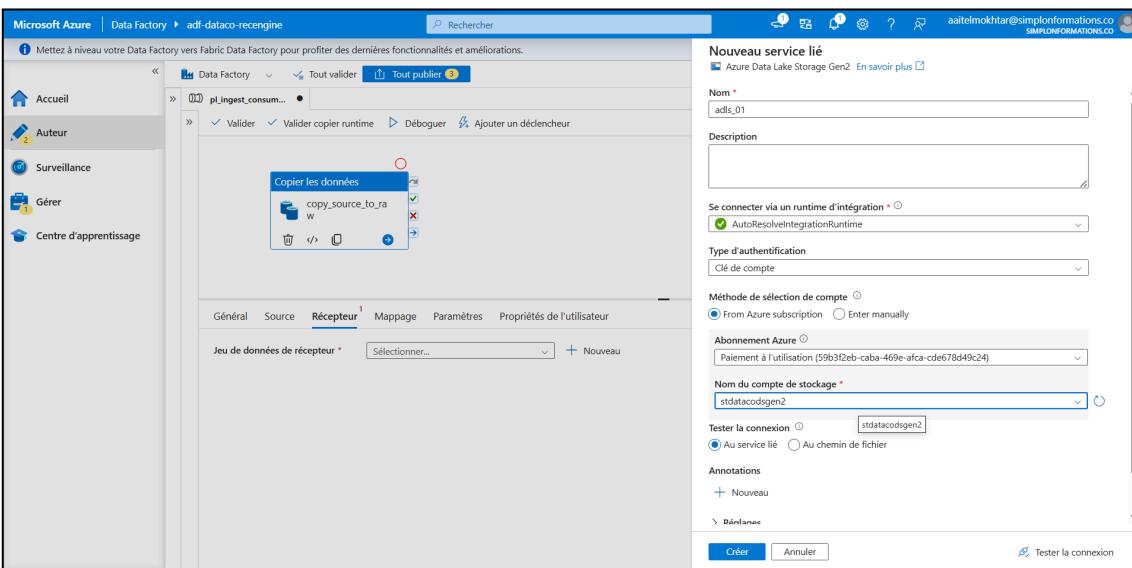
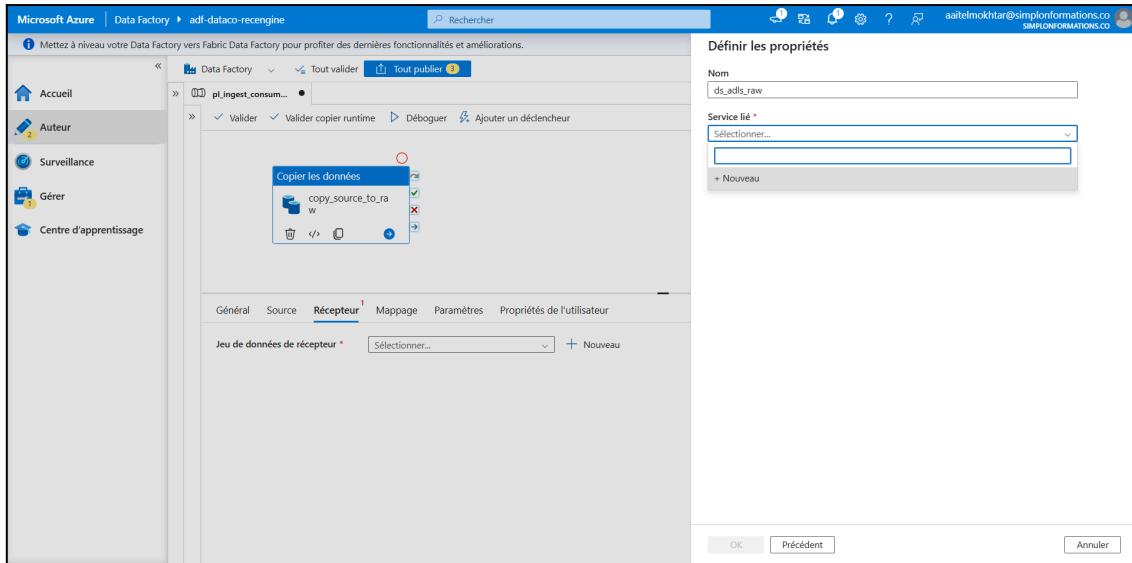


Section 2 - Récepteur :

- Dans la partie **Récepteur**, on crée un **Nouveau** jeu de données récepteur.
- Rechercher et choisir le magasin de données **Azure Data Lake Storage Gen2** et **DelimitedText** comme format.
- Attribuer un **nom** au récepteur.
- Créer un nouveau **Service Lié** et choisir l'**Abonnement** et **Nom du compte de Stockage** (celui déjà créé).
- Appuyer sur **Créer**.
- Cliquer sur **Ouvrir** le jeu de données de récepteur créé.

- Spécifier le chemin d'accès au fichier :
 - **Système de fichiers** : Taper le Nom du Conteneur déjà créé.
 - **Annuaire** : Cliquer sur Ajouter du contenu dynamique, et insérer : `@formatDateTime(utcnow(), 'raw/yyyy/MM/dd')` pour créer le chemin dynamiquement.
 - **Nom de fichier** : Donner un nom au fichier CSV.
- **Publier Tout** pour sauvegarder les changements.





The screenshot shows the Azure Data Factory interface. On the left, there's a sidebar with icons for Accueil, Auteur, Surveillance, Gérer, and Centre d'apprentissage. The main area shows a pipeline named 'pl_ingeist_consum...' with a single activity named 'ds_adls_raw'. This activity is a 'DelimitedText' type with a CSV icon. Below it is a configuration panel with tabs for Connexion, Schéma, and Réglages. Under Connexion, the service linked is 'adls_01', the file path is 'Système de fichiers / Annuaire / Nom de fichier', and the connection type is 'Aucune compression'. Other settings include separator as 'Comma (,),' line delimiter as 'Default (\r\n, or \n\r)', encoding as 'Par défaut(UTF-8)', and quote style as 'Double quote (")'. The 'Caractère d'échappement' is set to 'Backslash (\)'. There are also buttons for 'Tester la connexion', 'Modifier', 'Nouveau', and 'En savoir plus'.

This screenshot shows the 'Générateur d'expressions de pipeline' (Pipeline Expression Generator) window. It contains a text input field with the expression `@concat('raw/', formatDateTime.UtcNow(), 'yyyy/MM/dd')` and a 'OK' button at the bottom right. The background shows the same pipeline and dataset configuration as the previous screenshot.

The screenshot shows the final configuration of the 'ds_adls_raw' dataset. The 'Chemin d'accès au fichier' field now contains the generated expression `datalake / @concat('raw/', formatDateTime.UtcNow(), 'yyyy/MM/dd') / consumption_data.csv`. The rest of the configuration remains the same as the previous screenshots, including the 'Première ligne comme en-tête' (First row as header) checkbox being checked.

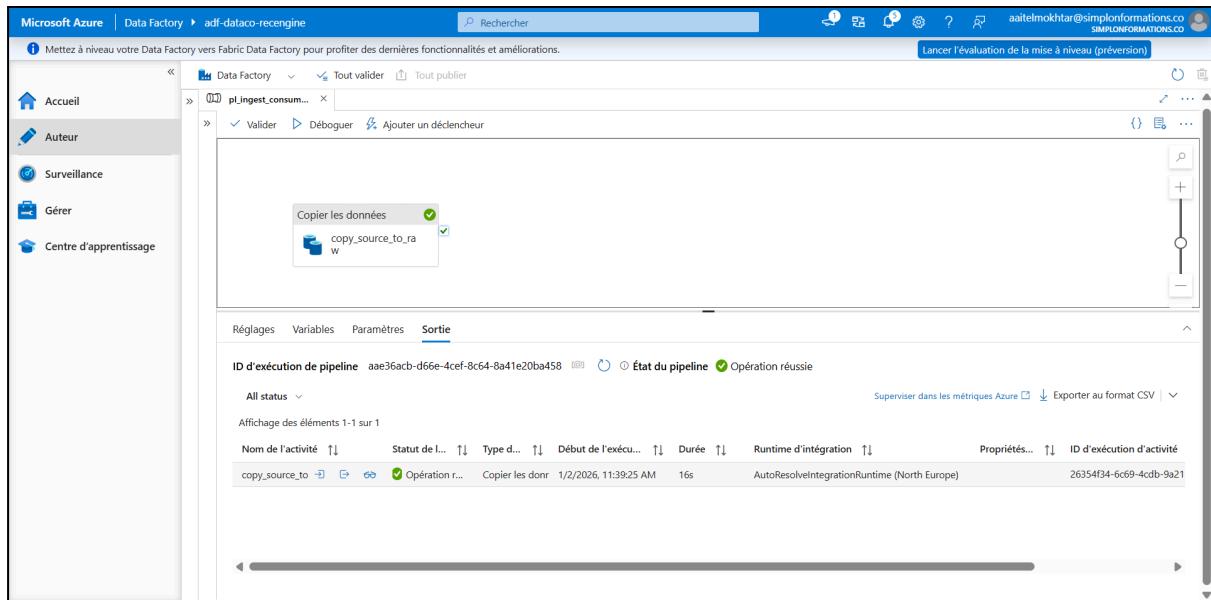
Section 3 - Mappage :

- Dans la section **Mappage**, cliquer sur **Importer les schémas** pour visualiser les colonnes du Dataset et leurs types.
- Vérifier la présence de 2 colonnes : **Datetime** et **PJME_MW**.
- Changer le type de Datetime de **String** à **TimeSpan**, et PJME_MW de **String** à **int32**.
- **Publier Tout** pour sauvegarder les changements.

The screenshot shows the Azure Data Factory interface for a pipeline named 'pl_ingest_consum...'. The 'Mappage' (Mapping) tab is active. A tooltip 'Copier les données' (Copy data) is displayed over the 'copy_source_to_rdw' mapping. The mapping table shows two columns: 'copy_source_to_rdw' and 'copy_source_to_rdw'. The 'PJME_MW' column is highlighted with a yellow border. Below the table, there are buttons for 'Import des schémas' (Import schema), 'Nouveau mappage' (New mapping), 'Effacer' (Delete), 'Réinitialiser' (Reset), and 'Supprimer' (Delete).

The screenshot shows the same pipeline 'pl_ingest_consum...' with the 'Mappage' tab still active. The 'copy_source_to_rdw' mapping table now correctly displays the column types: 'Source' (Datetime, PJME_MW) and 'Destination' (TimeSpan, Int32). The 'Type' column headers are also present. The overall interface remains consistent with the first screenshot, showing the 'Accueil' (Home) and 'Auteur' (Author) navigation items on the left.

Enfin, on exécute l'activité de **Copie de Données** en appuyant sur **Valider** puis **Déboguer**.



3. Azure Machine Learning :

3.1. Définition :

Azure Machine Learning est une plateforme de cloud computing de niveau entreprise fournie par Microsoft, conçue pour accélérer et gérer le cycle de vie complet des projets d'intelligence artificielle. Elle offre un environnement collaboratif et intégré où les data scientists et les développeurs peuvent construire, entraîner, déployer et surveiller des modèles prédictifs à grande échelle, en utilisant aussi bien des interfaces visuelles **low-code** (comme le concepteur de pipelines ou **Automated ML**) que des environnements de code complets (**Python/R SDK**). Au-delà de la simple création de modèles, ce service se distingue par sa capacité à industrialiser l'IA grâce à l'intégration de pratiques MLOps (DevOps pour le Machine Learning), facilitant ainsi l'automatisation des workflows, la gestion des versions de modèles et la gouvernance des données depuis la phase d'expérimentation jusqu'à la mise en production sécurisée.

3.2. Création et Configuration :

Premièrement, on configure l'espace de travail de Azure Machine Learning en suivant les étapes suivantes :

- Accéder au groupe de ressources (**aaitelmokhtar**).
- Cliquer sur **Créer une Nouvelle Ressource**.
- Taper **Machine Learning** dans la barre de recherche.
- Choisir **Machine Learning**, puis cliquer sur **Créer**.
- Donner un **Nom** à la ressource ML.
- Sélectionner North Europe comme Région.
- Créer un nouveau **Compte de Stockage** en lui donnant un Nom.
- Cliquer sur **Vérifier+Créer**.
- Accéder à la ressource, puis lancer **Azure Machine Learning Studio**.

The screenshot shows the Microsoft Azure Marketplace interface. The search bar at the top contains 'machine learning'. Below the search bar, there are several filters: 'Tarification : Tous', 'Système d'exploitation : Tous', 'Type de serveur de publication : Tous', 'Type de produit : Tous', 'Afficher les éléments : Tous', and 'Nom de l'éditeur : All'. A message at the top states: 'L'administrateur a activé la Place de marché privée. Seules les offres approuvées peuvent être déployées.' Below the filters, a message says: 'Nouveau ! Obtenir des suggestions générées par l'IA pour < machine learning >'. The main area displays four service cards:

- Machine Learning** (Microsoft Azure Service): Machine learning à l'échelle de l'entreprise pour créer et déployer des modèles plus rapidement. Price: À partir de 26,197 €/heure.
- VM Watira Machine Learning** (Sahara Watira for Digital Transfo...): Virtual Machine. Price: À partir de 26,197 €/heure.
- AI & Machine learning development, training & Testing** (TechLatest): Virtual Machine. Price: À partir de 0,131 €/an.
- KoçSistem Azure Machine Learning Management** (KoçSistem Bilgi ve İletişim Hizm...): Managed Services. Price: Build, train and deploy machine learning models with Azure Machine Learning Services!
- Deep Learning Studio 3.0** (Deep Cognition, Inc.): Virtual Machine. Price: A deep learning platform for you to design and train your deep learning models visually.

The screenshot shows the Azure Machine Learning service details page. At the top, it says 'Machine Learning' and 'Microsoft'. Below that, there's a summary card with the service icon, name, and a 'Ajouter aux favoris' button. It also shows a rating of ★ 4.3 (Évaluations de 155) and status indicators: 'Approuvé par défaut' and 'Éligible à l'avantage Azure'. There are dropdown menus for 'Abonnement' (Payment à l'utilisation) and 'Plan' (Azure Machine Learning), and a 'Créer' button. Below the card, there are tabs: 'Vue d'ensemble' (selected), 'Offres', 'Informations d'utilisation + support', and 'Évaluations et avis'. A note below the tabs says: 'Azure Machine Learning offre aux développeurs et aux scientifiques des données une grande variété d'expériences productives pour créer, entraîner et déployer des modèles Machine Learning. Créez un espace de travail Azure Machine Learning pour entraîner, gérer et déployer des expériences de machine learning et des services web.' At the bottom, there's a 'Média' section showing a preview of the studio interface, and a 'Envoyer des commentaires' button.

Microsoft Azure Rechercher dans les ressources, services et documents (G+/)

Accueil > aaitelmokhtar > Place de marché > Machine Learning >

Machine Learning

Créer un espace de travail Machine Learning

Groupe de ressources * aaitelmokhtar [Créer nouveau](#)

Détails de l'espace de travail

Configurez les paramètres de votre espace de travail de base, tels que sa connexion de stockage, son authentication, son conteneur, etc. [En savoir plus](#)

Nom * mlw-dataco-recengine-v2

Région * North Europe

Compte de stockage * (nouveau) mlwdatacorecen6168458906 [Créer](#)

Coffre de clés * (nouveau) mlwdatacorecen1594507864 [Créer](#)

Application Insights * (nouveau) mlwdatacorecen9229011643 [Créer](#)

Registre de conteneurs * Aucun [Créer](#)

[Vérifier + créer](#) [< Précédent](#) [Suivant : Inbound Access](#) [Enregistrer](#) [Abandonner](#)

Créer un nouveau compte de stockage

Nom * stmlwdefault

Réplication * Stockage localement redondant (LRS)

Microsoft Azure Rechercher dans les ressources, services et documents (G+/)

Accueil >

Microsoft.MachineLearningServices | Vue d'ensemble

Vue d'ensemble

Votre déploiement a été effectué

Nom du déploiement : Microsoft.MachineLearningServices Heure de début : 02/01/2026 11:09:51
Abonnement : Paiement à l'utilisation ID de corrélation : bf6e2d64-fd98-4767-bd78-145f76cd8da6

Groupe de ressources : aaitelmokhtar

Détails du déploiement

Étapes suivantes

Accéder à la ressource

Envoyer des commentaires

Partagez votre expérience avec le déploiement

Ajouter ou supprimer des favoris en appuyant sur [Ctrl+Shift+F1+F2](#)

Cost Management
Recevez des notifications pour vous aider à respecter votre budget et à éviter des frais imprévus sur votre facture. [Configurer les alertes de coût >](#)

Microsoft Defender pour le cloud
Sécuriser vos applications et votre infrastructure. [Accédez à Microsoft Defender pour le cloud >](#)

Tutoriels Microsoft gratuits
[Commencer l'apprentissage aujourd'hui >](#)

Travailler avec un expert
Les experts Azure sont des partenaires fournisseurs de services qui peuvent vous

Microsoft Azure Rechercher dans les ressources, services et documents (G+/)

Accueil > Gestionnaire des ressources | Groupes de ressources > aaitelmokhtar

mlw-dataco-recengine-v2

Azure Machine Learning workspace

Rechercher Download config.json Delete

Vue d'ensemble

Suggérer une charge de travail pour cet espace de travail ML

Bases

| | | | |
|----------------------|----------------------------|----------------------|---|
| Groupe de ressources | : aaitelmokhtar | Studio web URL | : https://ml.azure.com?tid=a2e466aa-4f86-4545-b5b8-97da... |
| Emplacement | : North Europe | Container Registry | : ... |
| Abonnement | : Paiement à l'utilisation | Key Vault | : mlwdatacorecen1594507864 |
| Storage | : stmlwdefault | Application Insights | : mlwdatacorecen9229011643 |
| Provisioning State | : Succeeded | MLflow tracking URI | : azurerm://northeurope.api.azureml.ms/mlflow/v1/subscri... |

Vue JSON

Work with your models in Azure Machine Learning Studio

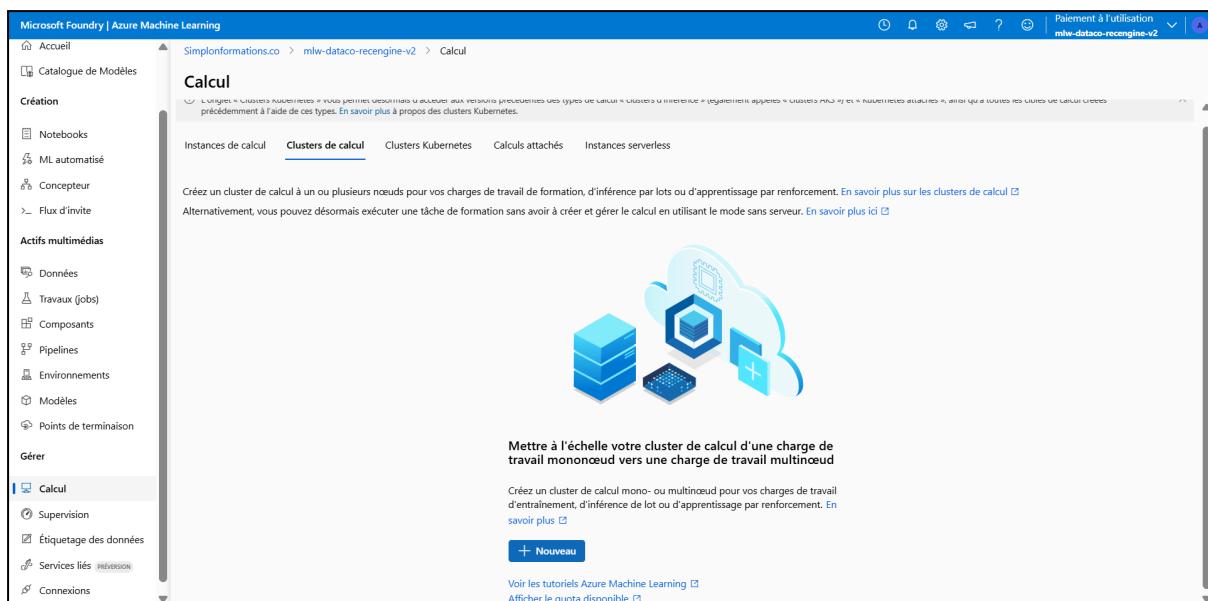
The Azure Machine Learning Studio is a web app where you can build, train, test, and deploy ML models. Launch it now to start exploring, or [learn more about the Azure Machine Learning studio](#).

[Launch studio](#)

Ajouter ou supprimer des favoris en appuyant sur [Ctrl+Shift+F1+F2](#)

Après le lancement de Azure ML Studio, on passe à la configuration de Compute Cluster en suivant les étapes suivantes :

- Dans le menu gauche, rechercher et choisir **Calcul** dans la section **Gérer**.
- Choisir et cliquer sur **Clusters de Calcul**.
- Créer un **nouveau Cluster**.
- Remplir la formulaire de création de Machine Virtuelle en assignant les champs suivants :
 - **Localisation** : North Europe.
 - **Niveau de ML** : Dédié.
 - **Type de Machine Virtuelle** : Processeur.
 - **Taille de ML** : Sélectionner “**Sélectionner parmi toutes les options**”, puis Rechercher **Standard_DS11_v2** (2 vCPUs, 14 GB RAM).
- Cliquer sur **Suivant**.
- Dans les paramètres avancés, vérifier les champs suivants :
 - **Nom du Calcul** : Donner un nom au Cluster créé.
 - **Nombre minimal de nœuds** : 0.
 - **Nombre maximal de nœuds** : 1.
 - **Secondes d'inactivité** : 120.
- Cliquer sur **Créer**.



The screenshot shows the Microsoft Foundry Azure Machine Learning interface. On the left, there's a sidebar with various menu items: Accueil, Catalogue de Modèles, Création (Notebooks, ML automatisé, Concepteur, Flux d'invite), Actifs multimédias (Données, Travaux (jobs), Composants, Pipelines, Environnements, Modèles, Points de terminaison), Gérer (Calcul, Supervision, Étiquetage des données, Services liés, Connexions). The main area is titled 'Créer un cluster de calcul'. The 'Machine virtuelle' tab is active, showing a list of virtual machine sizes. One item, 'Standard_DS11_v2', is selected and highlighted. Below the list, there are tabs for 'Précédent', 'Suivant', and 'Annuler'. The 'Paramètres avancés' tab is also visible, showing detailed configuration options for the selected VM size.

Après la configuration de Cluster de Calcul, on passe à la création du Magasin de données (**DataStore**) et Ressource de données (**Data Asset**) en suivant les étapes suivantes :

- Dans le Menu gauche, choisir **Actif multimédias > Données > Magasin des données.**
- Cliquer sur Créer un **nouveau Datastore**.
- Remplir le formulaire en renseignant les champs suivants :
 - **Nom du Datastore** : Donner un nom au datastore créé.
 - **Type** : Choisir Azure Blob Storage.

- **Compte de Stockage :** Choisir le compte de stockage déjà créé dans le groupe de ressource (aaitelmokhtar).
 - **Conteneur :** Choisir le nom du Conteneur aussi déjà créé.
 - **Type d'Authentification :** Choisir Clé de Compte.
 - **Clé de compte :** Chercher cette clé dans le compte de stockage > Sécurité + Réseau.
- Cliquer sur **Créer**.

Microsoft Foundry | Azure Machine Learning

Simplonformations.co > mlw-dataco-recengine-v2 > Données

Données

Ressources de données Magasins de données Importation des données Connexions de données

Vous pouvez créer des magasins de données OneLake utilisés pour se connecter aux données de votre point de terminaison Microsoft Fabric via CLI et SDK uniquement. En savoir plus sur le magasin de données OneLake

+ Créer Actualiser Désenregistrer Définir comme magasin de données par défaut Réinitialiser la vue

| Nom | Type | Nom du stockage |
|---------------------------------|---------------------|-----------------|
| workspaceartifactstore | Stockage Blob Azure | stmlwdefault |
| workspaceblobstore (Par défaut) | Stockage Blob Azure | stmlwdefault |

Créer un magasin de données

Nom du magasin de données * ds_raw_data

Type de magasin de données * Stockage Blob Azure

Méthode de sélection de compte

À partir d'un abonnement Azure

ID d'abonnement * Paiement à l'utilisation (59b3f12eb-caba-469e-acfa-cde678d49c24)

Compte de stockage * stdatacodgen2 (aaitelmokhtar)

Conteneur blob * datalake

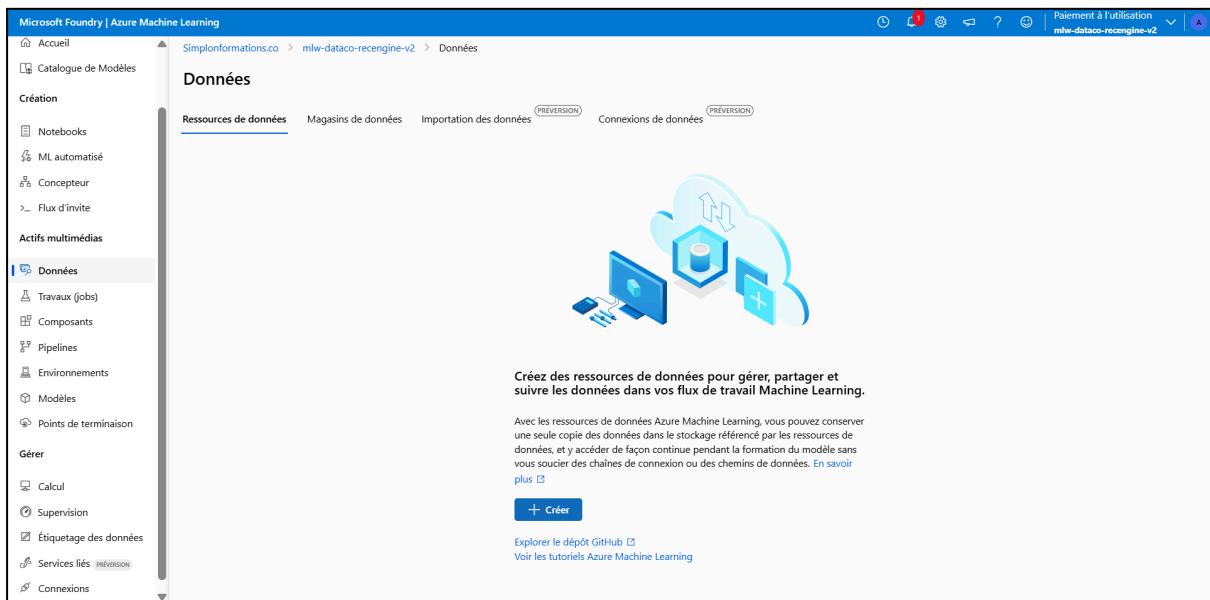
Enregistrer les informations d'identification avec le magasin de données pour l'accès aux données

Type d'authentification * Clé de compte

Clé de compte *

Créer Annuler

- Dans le Menu gauche, choisir **Actif multimédias > Données > Ressource de Données**.
- Cliquer sur Créer un **nouveau Data Asset**.
- Remplir le formulaire de **Type de Données** en renseignant les champs suivants :
 - **Nom** : Donner un nom au dataset créé.
 - **Type** : Choisir Tabulaire.
- Cliquer sur **Suivant**.
- Comme Source de Données, choisir “**À partir du Stockage Azure**”, puis cliquer sur Suivant.
- Sélectionner le **Magasin des données (Datastore)** déjà créé et cliquer sur Suivant.
- Pour le chemin de stockage, choisir le dossier “**raw/**”.
- Cliquer sur **Créer**.



Microsoft Foundry | Azure Machine Learning

Accueil Catalogue de Modèles

Création

- Notebooks
- ML automatisé
- Concepteur
- Flux d'invite
- Actifs multimédias**
- Données
- Travaux (jobs)
- Composants
- Pipelines
- Environnements
- Modèles
- Points de terminaison
- Gérer
- Calcul
- Supervision
- Étiquetage des données
- Services liés
- Connexions

Créer une ressource de données

Définir le nom et le type de votre ressource de données

Nom *: consumption_raw_dataset

Description: Description de la ressource de données

Type *: Tabulaire

Cas d'usage pour les types de données

Quand dois-je utiliser le type de fichier ?

Le type de fichier est recommandé dans la plupart des scénarios lorsque vous utilisez un seul fichier de données de l'importe quel type (y compris les données tabulaires). Ce type vous permet de spécifier un emplacement de fichier par URI dans un emplacement de stockage sur votre ordinateur local, un magasin de données attaché, un stockage blob/ADLS ou un emplacement HTTP(s) publiquement disponible(s). Il existe de nombreux types d'URI pris en charge. Dans le Azure Machine Learning CLI v2 ou le SDK Python v2, ce type de données est appelé `uri_file`. [En savoir plus sur le type `uri_file`](#)

Quand dois-je utiliser le type de dossier ?

Le type de dossier a toutes les mêmes fonctionnalités et utilise des cas que le type de fichier, mais il est utilisé lors de la spécification d'un emplacement de dossier. Dans le Azure Machine Learning CLI v2 ou le SDK Python v2, ce type de données est appelé `uri_folder`. [En savoir plus sur le type `uri_folder`](#)

Quand dois-je utiliser le type de table ?

Le type de table est le plus utile pour les scénarios avancés où vous devrez peut-être abstraire la définition de schéma pour faciliter le partage. Vous devez l'utiliser quand vous avez des

Précédent Suivant Annuler

Microsoft Foundry | Azure Machine Learning

Accueil Catalogue de Modèles

Création

- Notebooks
- ML automatisé
- Concepteur
- Flux d'invite
- Actifs multimédias**
- Données
- Travaux (jobs)
- Composants
- Pipelines
- Environnements
- Modèles
- Points de terminaison
- Gérer
- Calcul
- Supervision
- Étiquetage des données
- Services liés
- Connexions

Créer une ressource de données

Choisissez une source pour votre ressource de données

Source de données

- À partir du stockage Azure**: Créez une ressource de données à partir des services de stockage de données inscrits, notamment Azure Blob Storage, le partage de fichiers Azure et Azure Data Lake.
- À partir de fichiers locaux**: Créez une ressource de données en chargeant des fichiers à partir de votre lecteur local.
- À partir de bases de données SQL**: Créez un jeu de données à partir d'Azure SQL Database et d'Azure PostgreSQL Database.
- À partir des fichiers web**: Créez une ressource de données à partir d'un seul fichier situé sur une URL web publique.
- Azure Open Datasets**: Créez un jeu de données en un clic à partir de jeux de données prédefinis. Ces jeux de données sont créés par le grand public et publiés en tant que Azure Open Datasets.

Précédent Suivant Annuler

Microsoft Foundry | Azure Machine Learning

Accueil Catalogue de Modèles

Création

- Notebooks
- ML automatisé
- Concepteur
- Flux d'invite
- Actifs multimédias**
- Données
- Travaux (jobs)
- Composants
- Pipelines
- Environnements
- Modèles
- Points de terminaison
- Gérer
- Calcul
- Supervision
- Étiquetage des données
- Services liés
- Connexions

Créer une ressource de données

Sélectionner un magasin de données

Choisissez un type de stockage et un magasin de données qui contient vos données. Vous pouvez également créer d'abord un magasin de données pour vos données.

Type de stockage source: Stockage Blob Azure

| Nom | Nom du stockage | Créé le |
|------------------------|-----------------|----------------------|
| workspaceblobstore | stmlwdefault | Jan 2, 2026 11:10 AM |
| workspaceartifactstore | stmlwdefault | Jan 2, 2026 11:10 AM |
| ds_raw_data | stdatacodsgen2 | Jan 2, 2026 11:55 AM |

Rechercher dans le magas... Filter Colonnes

Précédent Suivant Annuler

Microsoft Foundry | Azure Machine Learning

Créer une ressource de données

Choisir un chemin de stockage
Accédez ou entrez le chemin de stockage que vous souhaitez utiliser pour cette ressource de données.

Accéder au chemin de stockage Entrer manuellement le chemin d'accès de stockage

Chemin sélectionné : raw/**

Actualiser le contenu du magasin de données Accéder au niveau supérieur

Filtre...

| Nom | Créé le | Modifié le |
|------|---------|------------|
| raw/ | ... | ... |

Paramètres avancés

Précédent Suivant Annuler

Microsoft Foundry | Azure Machine Learning

Créer une ressource de données

Paramètres
Ces paramètres déterminent la façon dont les données sont analysées. Les paramètres initiaux sont automatiquement détectés ; vous pouvez les modifier si nécessaire pour analyser les données.

Format de fichier Délimiteur Exemple Codage
Délimiteur : Virgule Exemple : Field1,Field2,Field3 Codage : UTF-8

En-têtes de colonne Ignorer les lignes
Tous les fichiers ont les mêmes en-têtes Aucun

Le jeu de données contient des données multilignes

Remarque : Le traitement des fichiers tabulaires contenant des données multilignes est plus lent car plusieurs coeurs de processeur ne peuvent pas être utilisés pour ingérer les données en parallèle. L'activation de cette option peut entraîner un ralentissement des temps de traitement.

Aperçu des données

| Datetime | PIME_MW |
|---------------------|---------|
| 2002-12-31 01:00:00 | 26,498 |
| 2002-12-31 02:00:00 | 25,147 |
| 2002-12-31 03:00:00 | 24,574 |
| 2002-12-31 04:00:00 | 24,393 |
| 2002-12-31 05:00:00 | 24,860 |
| 2002-12-31 06:00:00 | 26,222 |
| 2002-12-31 07:00:00 | 28,702 |

Précédent Suivant Vérifier Annuler

Microsoft Foundry | Azure Machine Learning

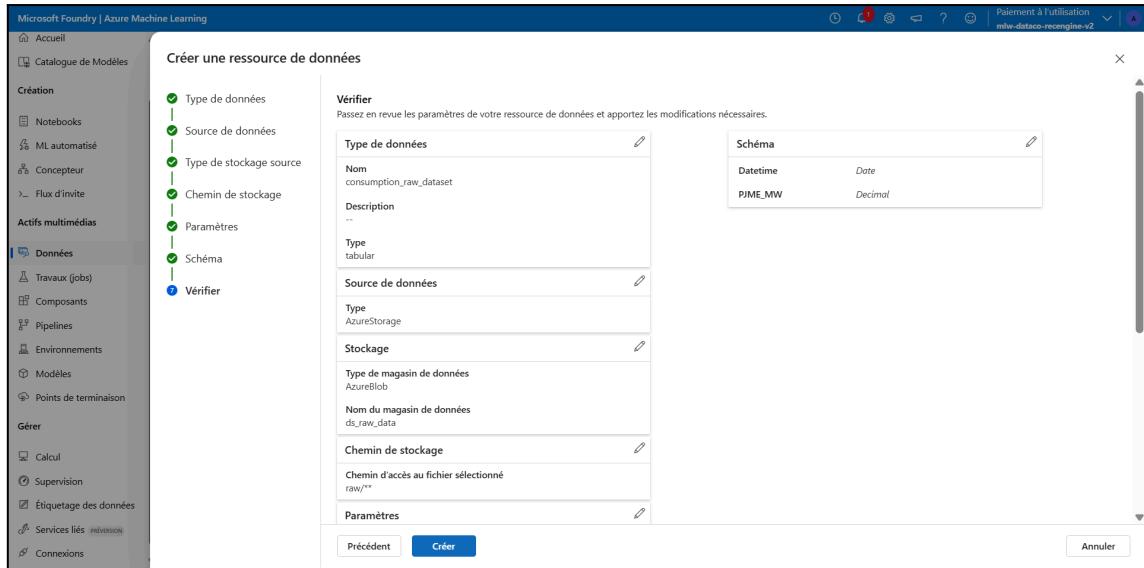
Créer une ressource de données

Schéma
Les types de colonnes sont automatiquement détectés en fonction du sous-ensemble initial des données et peuvent être mis à jour ici. Les valeurs qui ne sont pas alignées avec le type de colonne spécifié échoueront à la conversion et seront remplies de valeurs Null ou remplaçées par une valeur d'erreur. Les erreurs d'aperçu des conversions ne bloquent pas et vous pouvez continuer.

Rechercher un nom de colonne

| Inclure | Nom de la colonne | Type | Exemples de valeurs | Format de date | Propriétés |
|-------------------------------------|-------------------|---------------|--|------------------------------------|---------------------|
| <input checked="" type="checkbox"/> | Path | Chaîne | | Non applicable au type sélectionné | Non applicable a... |
| <input checked="" type="checkbox"/> | Datetime | Date | 2002-12-31 01:00:00, 2002-12-31 02:00:00 | %Y-%m-%d %H:%M:%S | Aucun |
| <input checked="" type="checkbox"/> | PIME_MW | Point décimal | 26498, 25147, 24574 | Non applicable au type sélectionné | Non applicable a... |

Précédent Suivant Annuler



3.3. Pipeline ML Designer :

On passe à la partie de création et configuration du Pipeline ML dans Designer en suivant et en respectant les étapes suivantes :

Etape 1 - Création du Pipeline et l'ajout du Dataset :

- Dans le menu gauche, choisir **Concepteur > Créeer une Pipeline**.
- Donner un **Nom** au pipeline créé.
- Configurer la Pipeline :
 - Cliquer sur **Configurer & Soumettre**.

- Créer une nouvelle expérience en lui assignant un nom, puis cliquer sur **Suivant**.
- Dans la section **Paramètres d'exécution**, sélectionner le **Compute Cluster** déjà créé.
- Cliquer sur **Vérifier+Créer**.
- Dans la partie **Données**, glisser-déposer le **dataset** créé dans le canvas.

The screenshot shows the Microsoft Foundry Azure Machine Learning Designer. The main area is titled 'Nouveau pipeline' (New Pipeline). It includes a large workspace with a blue plus sign icon, a sidebar with various navigation links such as 'Notebooks', 'Composants', and 'Pipelines', and a bottom section displaying a table of existing pipelines with columns for 'Nom' (Name), 'Type de pipeline' (Pipeline Type), 'Mis à jour le' (Last updated), and 'Créé par' (Created by).

The screenshot shows the Microsoft Foundry Azure Machine Learning Designer in 'Création' (Creation) mode. The left sidebar has 'Composants' selected. The main area shows a list of components under the 'Composeur' tab, including 'Sample data (16)', 'Data Transformation (19)', 'Computer Vision (6)', and others. There are also tabs for 'Données' and 'Script'. At the top right, there are buttons for 'Annuler' (Cancel), 'Rétablissement' (Revert), 'Valider' (Validate), 'Enregistrer' (Save), and 'Interface de pipeline' (Pipeline Interface).

Microsoft Foundry | Azure Machine Learning

Simplonformations.co > mlw-dataco-recengine-v2 > Concepteur > Création

Annuler Rétablir Valider Afficher le lignage Cloner ...

Pipeline-Created-on-01-02-2026

Rechercher par nom, étiquette et descript... Étiquettes : All Ajouter un filtre

Données Composant

95 +

- Sample data (16)
- Data Transformation (19)
- Computer Vision (6)
- Model Scoring & Evaluation (6)
- Machine Learning Algorithms (19)
- Text Analytics (7)
- Python Language (2)
- Data Input and Output (3)
- Recommendation (5)
- R Language (1)
- Feature Selection (2)
- Anomaly Detection (2)
- Statistical Functions (1)
- Model Training (4)
- Web Service (2)

Navigateur 100% 100% 100%

Modifier les détails du brouillon

Nom du brouillon * pipeline-dataco-forecast-train

Description du brouillon (facultatif) pipeline dataco forecast train

Créé le Jan 2, 2026 12:05 PM

Créé par Abd elhafid Ait el mokhtar

Heure de la dernière modification Jan 2, 2026 12:05 PM

Dernière modification par Abd elhafid Ait el mokhtar

Enregistrer Annuler

Microsoft Foundry | Azure Machine Learning

Simplonformations.co > mlw-dataco-recengine-v2 > Concepteur > Création

Annuler Rétablir ... Enregistrer Interface de pipeline

Étiquettes : All Ajouter un filtre

Données Composant

95 +

- Sample data (16)
- Data Transformation (19)
- Computer Vision (6)
- Model Scoring & Evaluation (6)
- Machine Learning Algorithms (19)
- Text Analytics (7)
- Python Language (2)
- Data Input and Output (3)
- Recommendation (5)
- R Language (1)
- Feature Selection (2)
- Anomaly Detection (2)
- Statistical Functions (1)
- Model Training (4)
- Web Service (2)

Navigateur 100% 100% 100%

Configurer la tâche de pipeline

Informations de base

Entrées & Sorties

Paramètres d'exécution

Vérifier + envoyer

Informations de base

Nom de l'expérience * Sélectionner un groupe existant Créer nouveau Nom de la nouvelle expérience * experience_02

Nom d'affichage de l'emploi pipeline-dataco-forecast-train

Description du travail pipeline dataco forecast train

Balises de travail Nom : Valeur Ajouter

Vérifier + envoyer Précédent Suivant Fermer

Microsoft Foundry | Azure Machine Learning

Simplonformations.co > mlw-dataco-recengine-v2 > Concepteur > Création

Annuler Rétablir ... Enregistrer Interface de pipeline

Étiquettes : All Ajouter un filtre

Données Composant

95 +

- Sample data (16)
- Data Transformation (19)
- Computer Vision (6)
- Model Scoring & Evaluation (6)
- Machine Learning Algorithms (19)
- Text Analytics (7)
- Python Language (2)
- Data Input and Output (3)
- Recommendation (5)
- R Language (1)
- Feature Selection (2)
- Anomaly Detection (2)
- Statistical Functions (1)
- Model Training (4)
- Web Service (2)

Navigateur 100% 100% 100%

Configurer la tâche de pipeline

Informations de base Entrées & Sorties Paramètres d'exécution Vérifier + envoyer

Paramètres d'exécution

Calcul par défaut Sélectionner le type de calcul Cluster de calcul

Sélectionner Azure ML cluster de calcul cpu-cluster-ds11, Réusite, Standard_DS11_v2, 2 processeurs virtuels (coeurs) Créer Azure ML cluster de calcul Actualiser la capacité de calcul

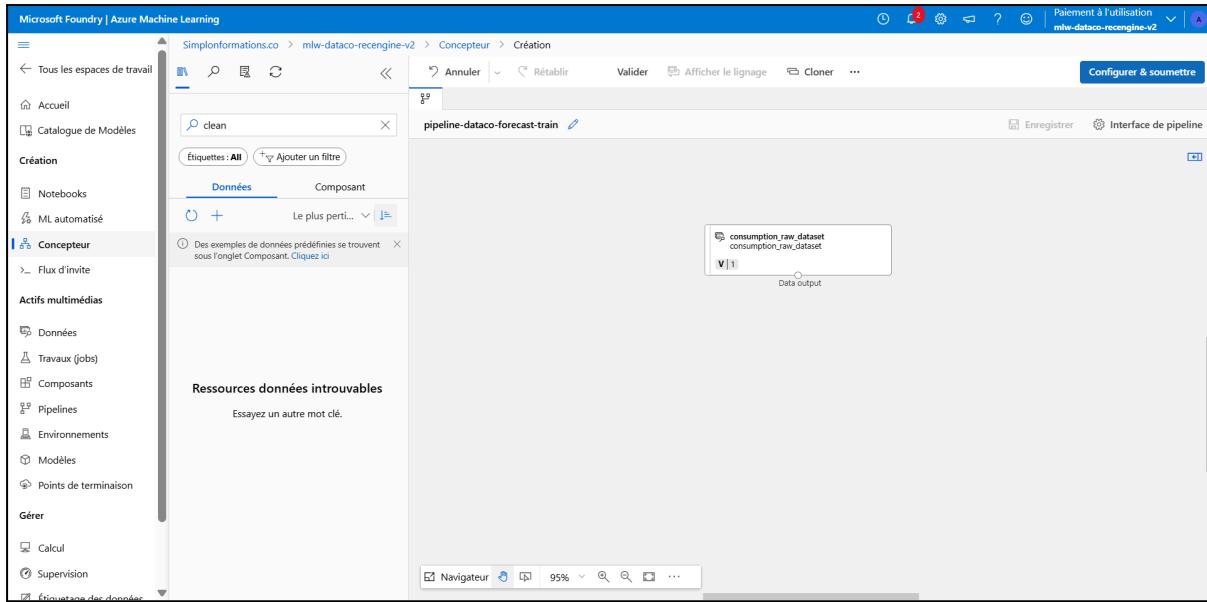
Identity

Magasin de données par défaut Sélectionner un magasin de données * workspaceblobstore

Paramètres avancés

Continuer en cas d'échec de l'étape

Vérifier + envoyer Précédent Suivant Fermer

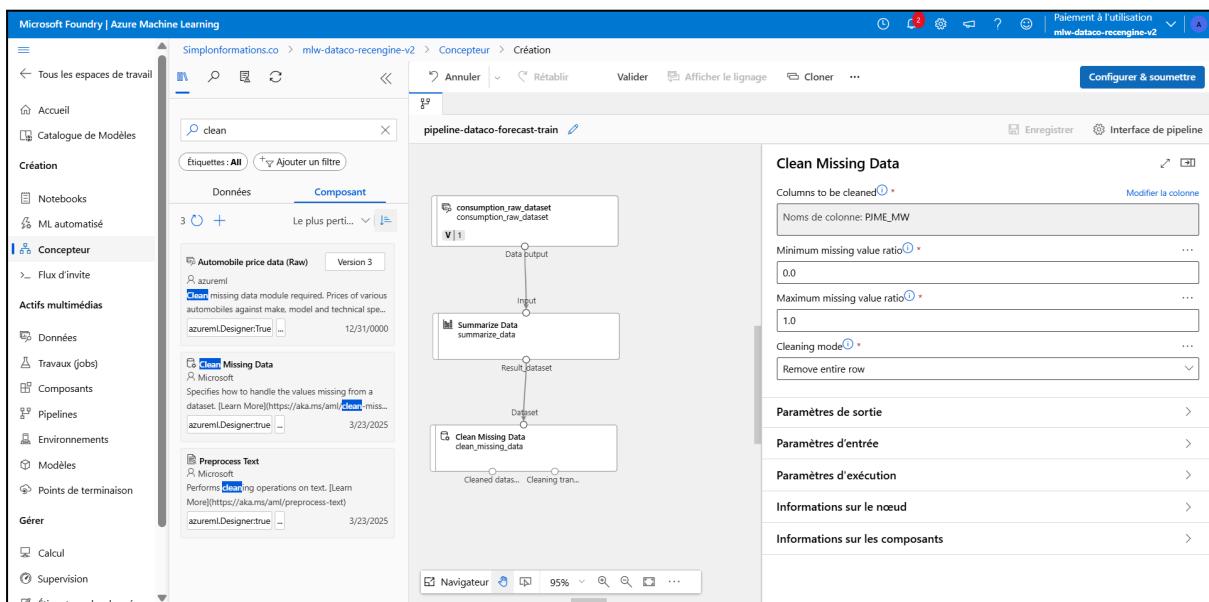
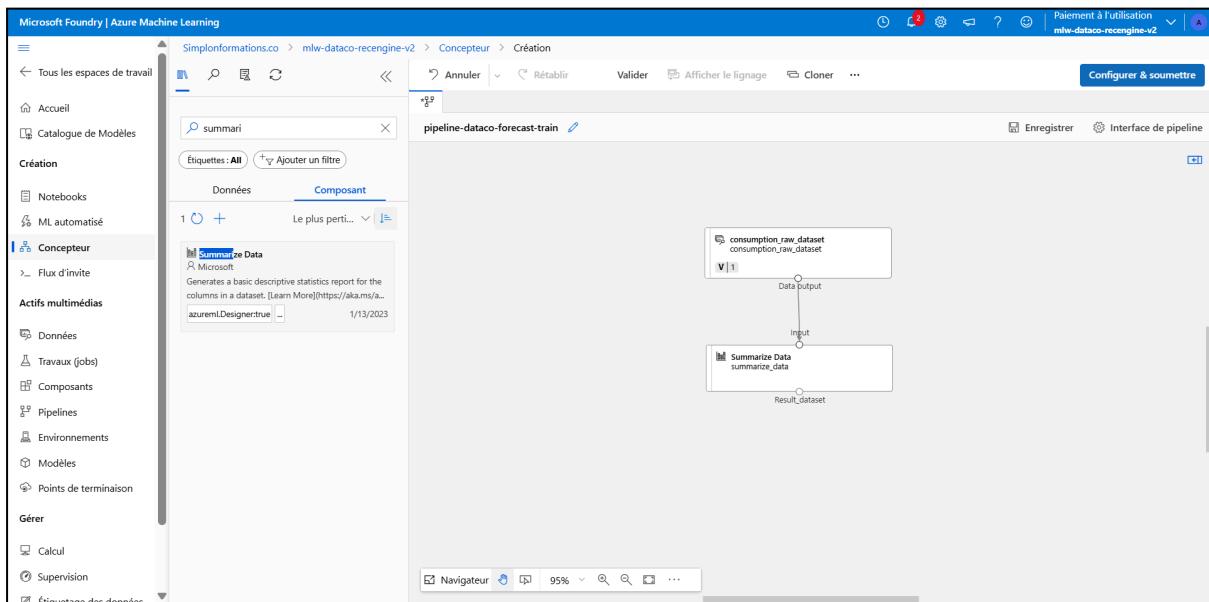


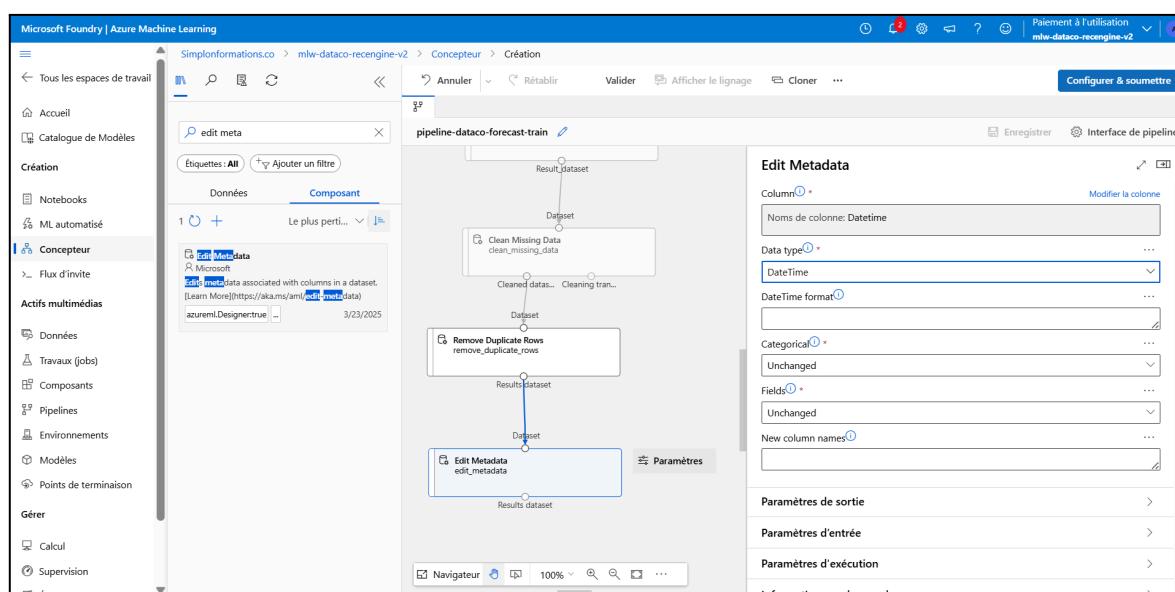
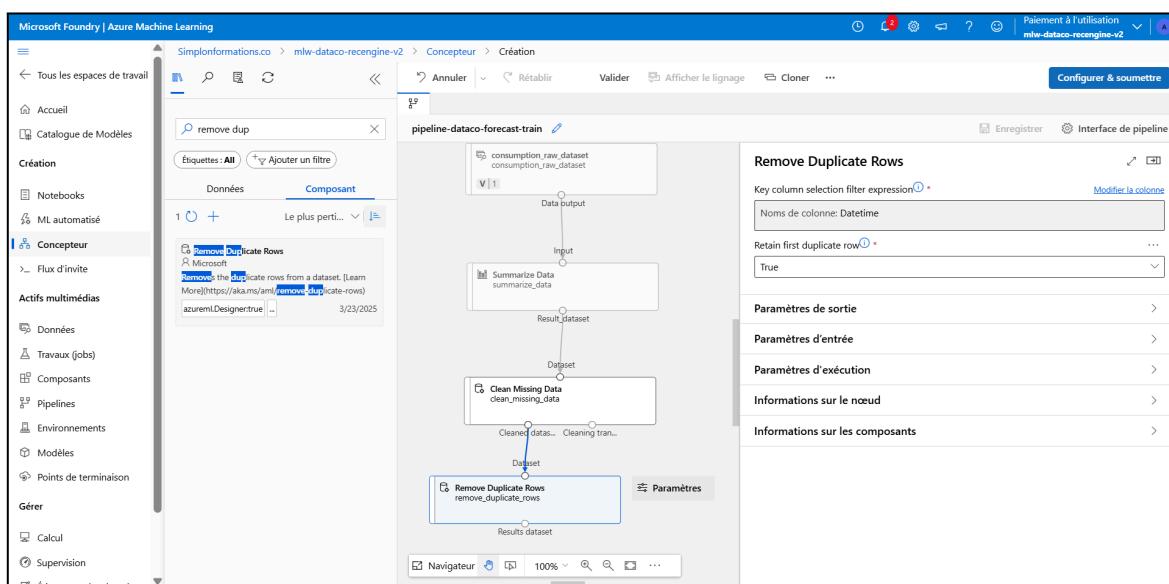
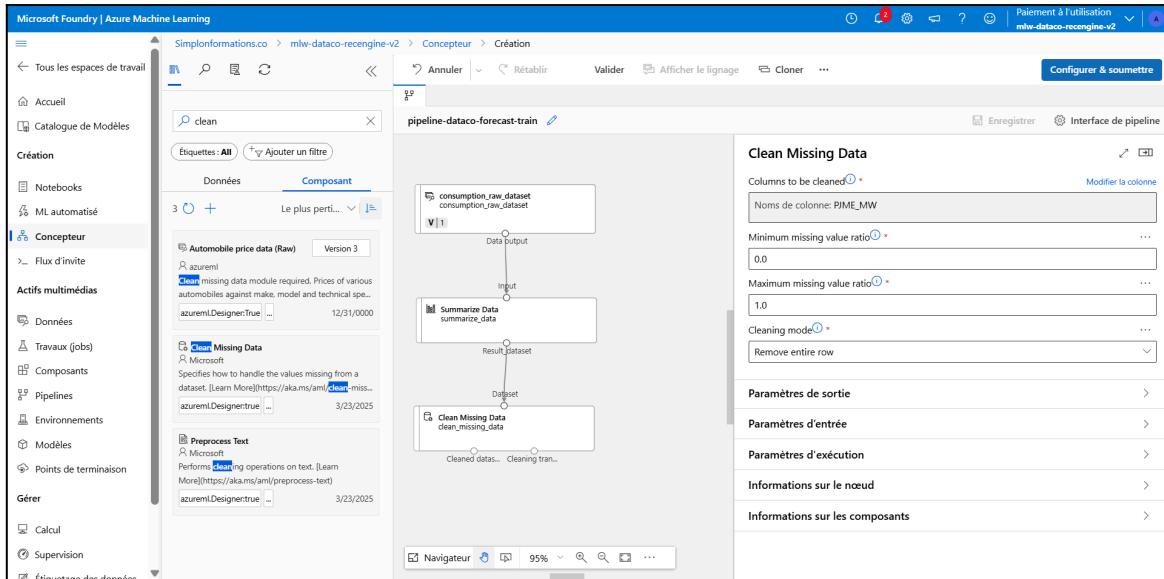
Etape 2 - Nettoyage des Données :

- **Summarize Data :**
 - Dans la section **Composants**, chercher le **Module Summarize Data**.
 - Ajouter le Module au **canvas**.
 - **Lier** le Module avec le Dataset.
- **Clean Missing Data :**
 - Dans la section **Composants**, chercher le **Module Clean Missing Data**.
 - Ajouter le Module au **canvas**.
 - Sélectionner **PJME_MW** comme **colonne à nettoyer**.
 - Sélectionner "**Remove entire row**" comme **Mode de Nettoyage**.
 - **Lier** le Module avec le Dataset.
- **Remove Duplicate Rows :**
 - Dans la section **Composants**, chercher le **Module Remove Duplicate Rows**.
 - Ajouter le Module au **canvas**.
 - Sélectionner **Datetime** comme **colonne clé** pour filtre.
 - **Lier** le Module avec le Module de **Clean Missing Data**.

- **Edit Metadata :**

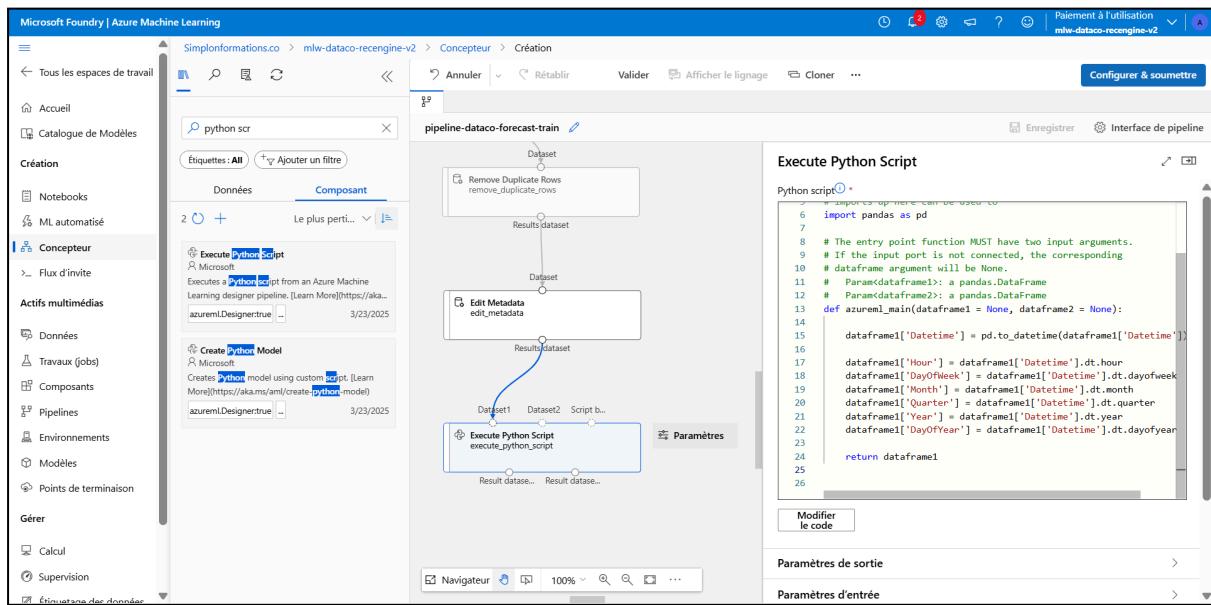
- Dans la section **Composants**, chercher le **Module Edit Metadata**.
- Ajouter le Module au **canvas**.
- Sélectionner la colonne **Datetime**.
- Sélectionner **Datetime** comme type de données (convertir).
- **Lier** le Module avec le Module de **Remove Duplicate Rows**.





Etape 3 - Feature Engineering :

- Dans la section **Composants**, chercher le **Module Execute Python Script**.
- Ajouter le Module au **canvas**.
- Modifier le Script Python par défaut avec un script pour créer les variables temporelles (**Hour, Dayofweek, Month, Quarter, Year, Dayofyear**)
- **Lier** le Module avec le Module de **Edit Metadata**.



- **Select Columns in Dataset :**
 - Dans la section **Composants**, chercher le **Module Select Columns in Dataset**.
 - Ajouter le Module au **canvas**.
 - Sélectionner les colonnes suivantes : **Hour, Dayofweek, Month, Quarter, Year, Dayofyear, PJME_MW**.
 - **Lier** le Module avec le Module de **Execute Python Script**.
- **Split Data :**
 - Dans la section **Composants**, chercher le **Module Split Data**.
 - Ajouter le Module au **canvas**.
 - Sélectionner **Split Rows** comme **Splitting Mode**.
 - Entrer **0.8** comme valeur de **Fraction**.
 - Désactiver le **Randomized Split** (valeur = **False**)
 - **Lier** le Module avec le Module de **Select Columns in Dataset**.

- **Sélection du Module à entraîner :**

- Dans la section **Composants**, chercher le **Boosted Decision Tree Regression** ou bien **Linear Regression** ou un autre modèle de Régression.
- Laisser les paramètres par défaut.

- **Train Model :**

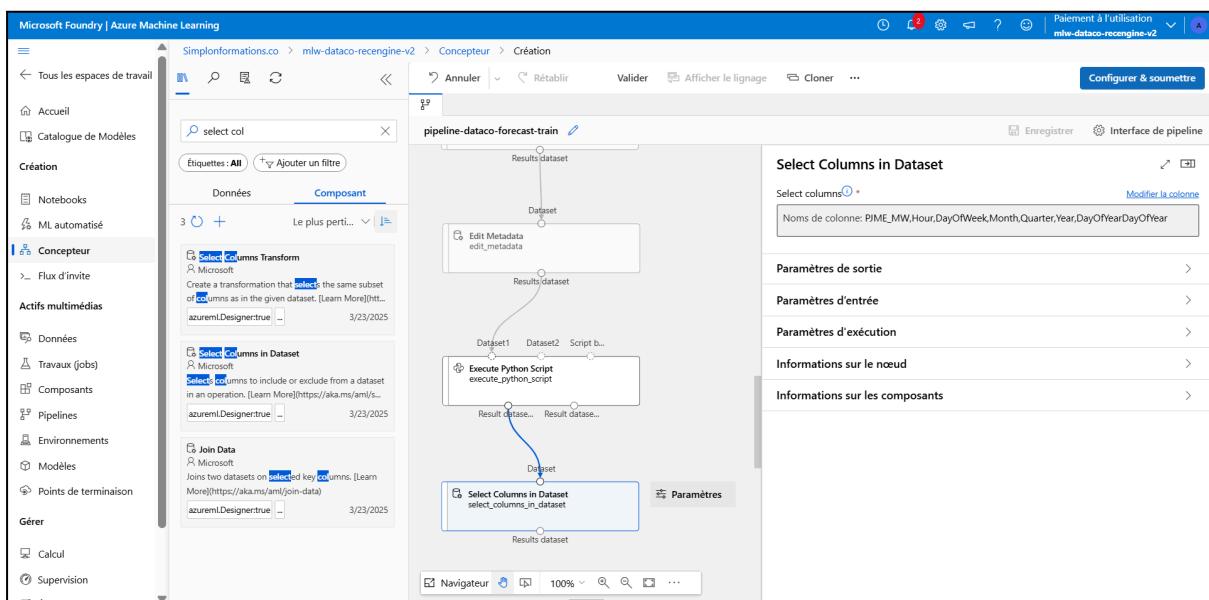
- Dans la section **Composants**, chercher le Modèle **Train Model**.
- Ajouter le Module au **canvas**.
- Sélectionner **PJME_MW** comme label colonne.
- **Lier** la 1ère entrée de **untrained model** avec Le Modèle à entraîner choisis (**Boosted Decision Tree Regression**).
- **Lier** la 2ème entrée de **Dataset** avec Le Module **Split Data**.

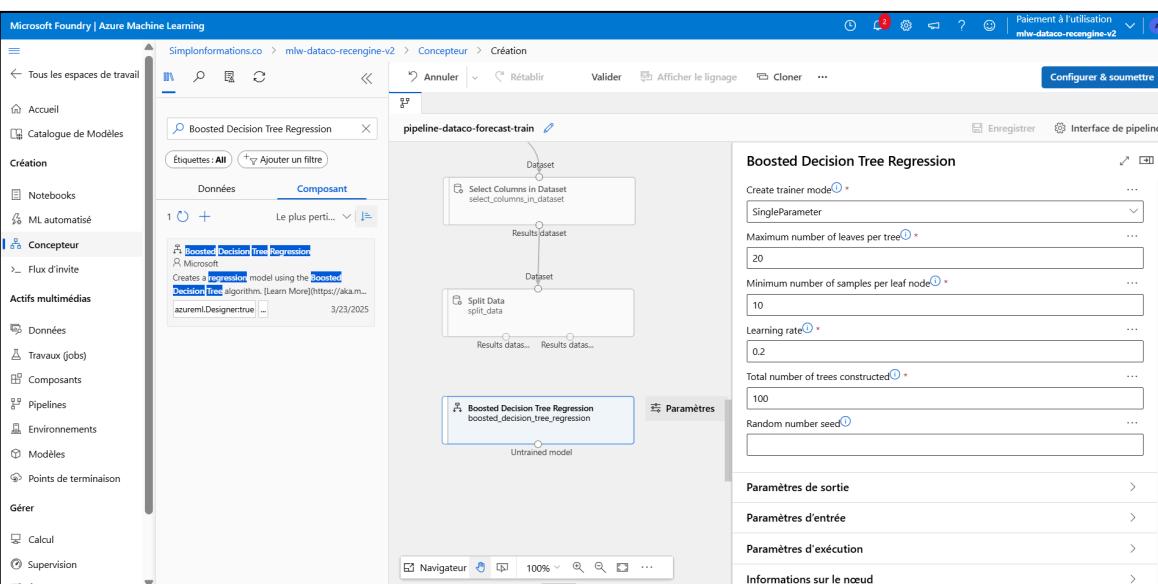
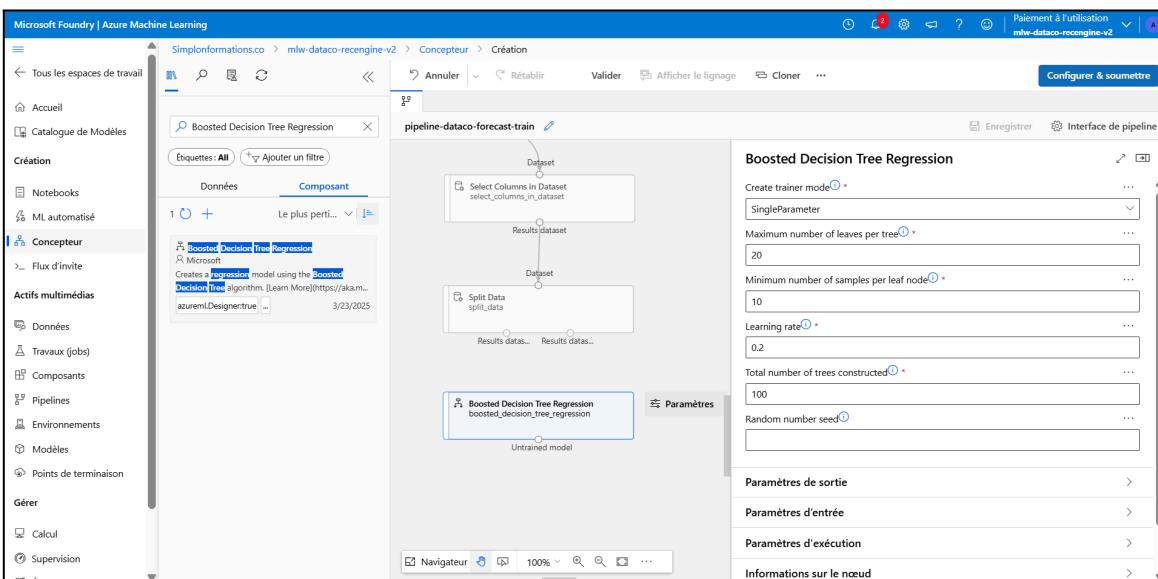
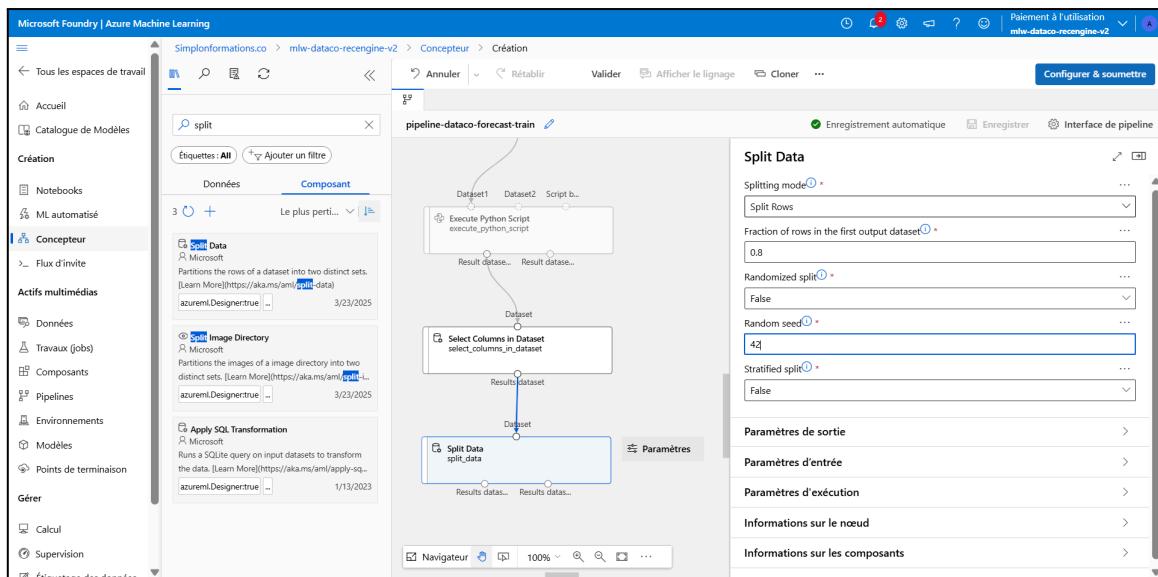
- **Score Model :**

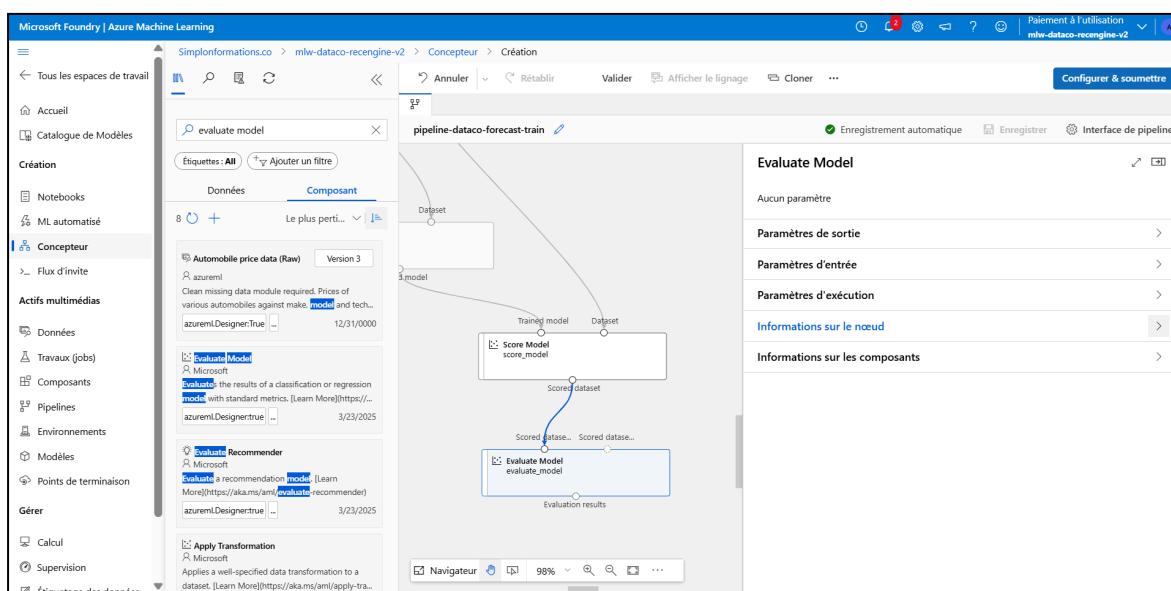
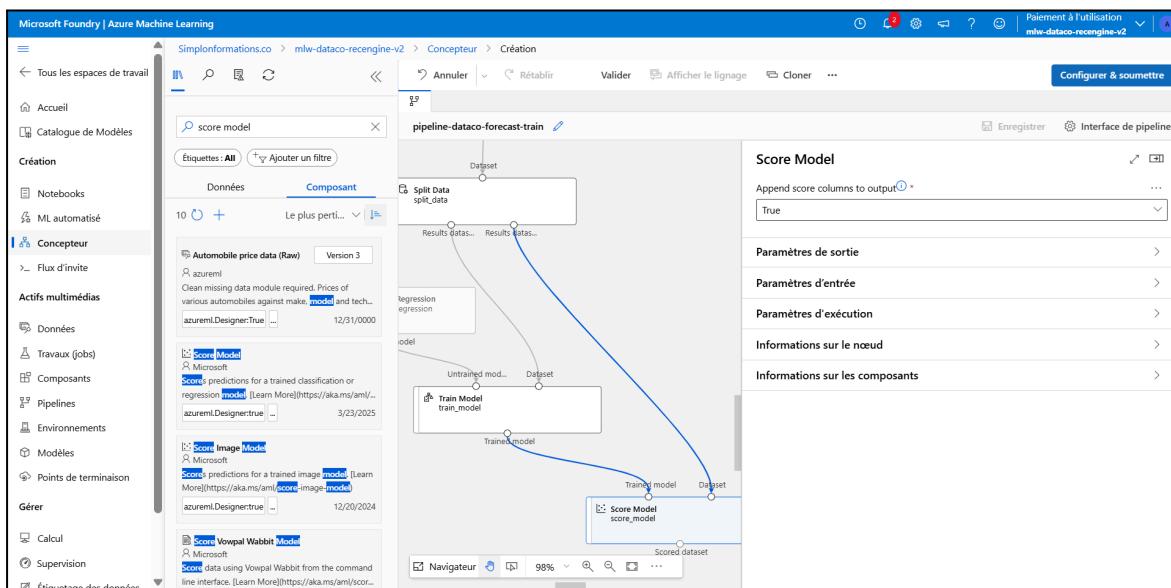
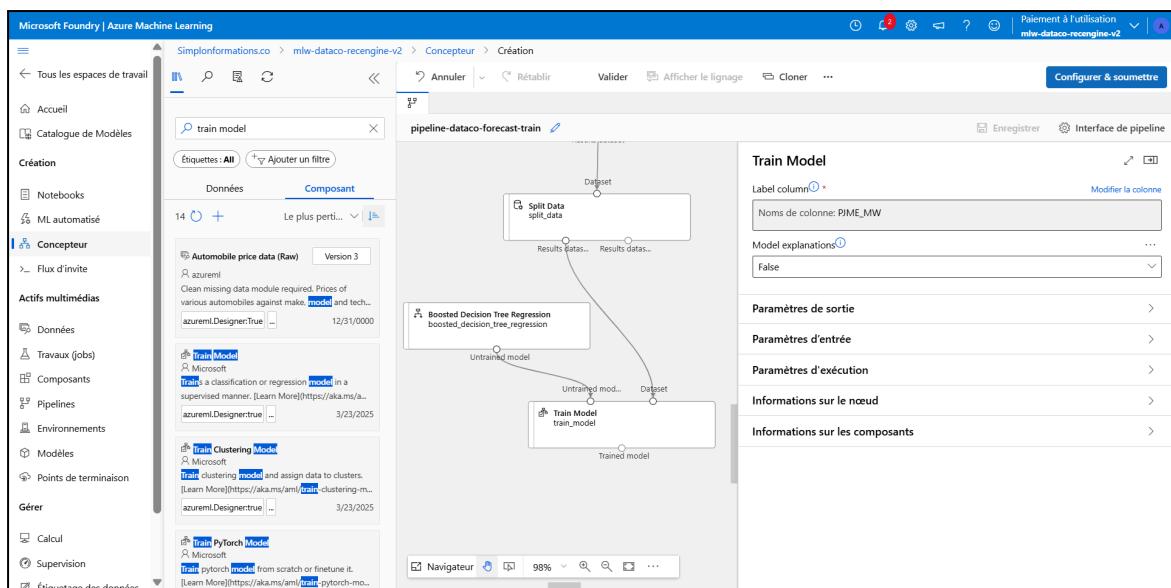
- Dans la section **Composants**, chercher le Modèle **Score Model**.
- Ajouter le Module au **canvas**.
- **Lier** le Module avec Le Module **Train Model**.

- **Evaluate Model :**

- Dans la section **Composants**, chercher le Modèle **Evaluate Model**.
- Ajouter le Module au **canvas**.
- **Lier** le Module avec Le Module **Score Model**.

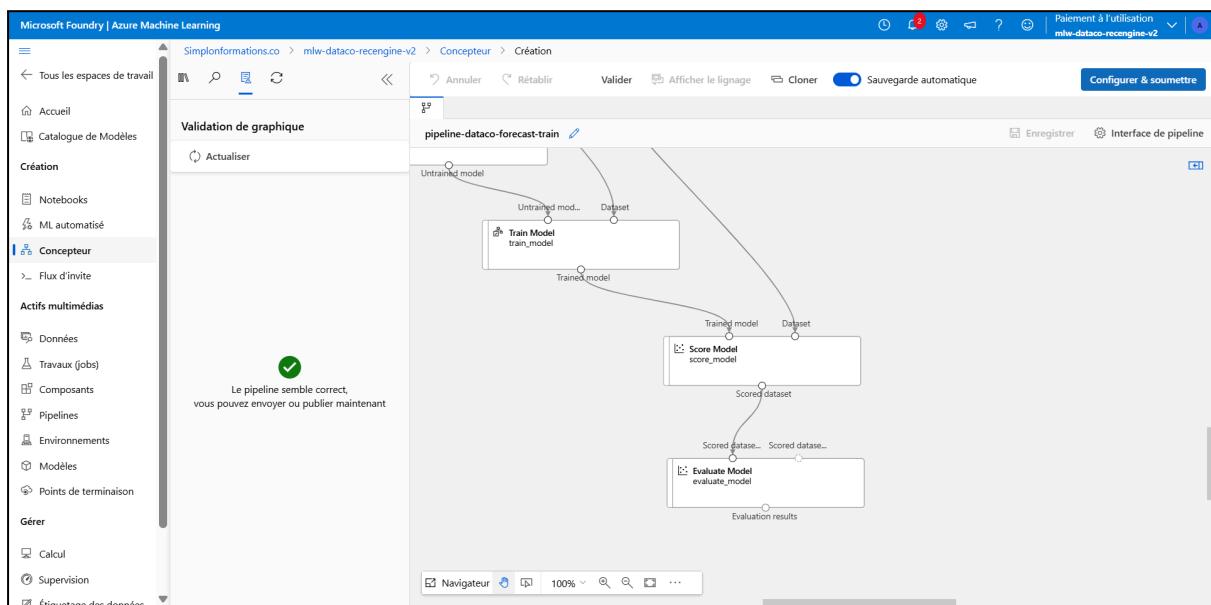


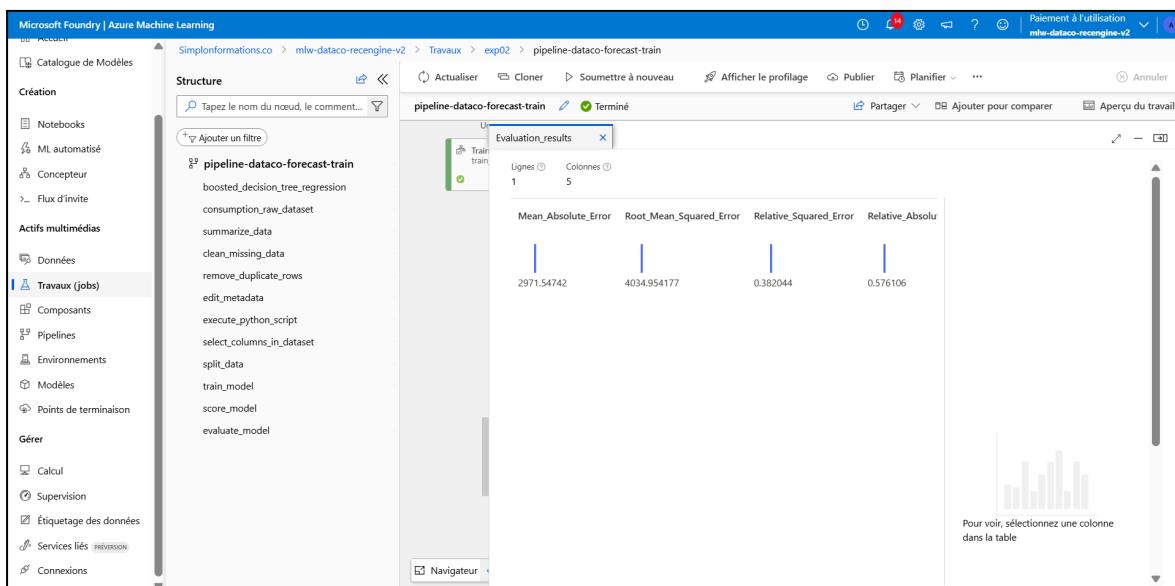
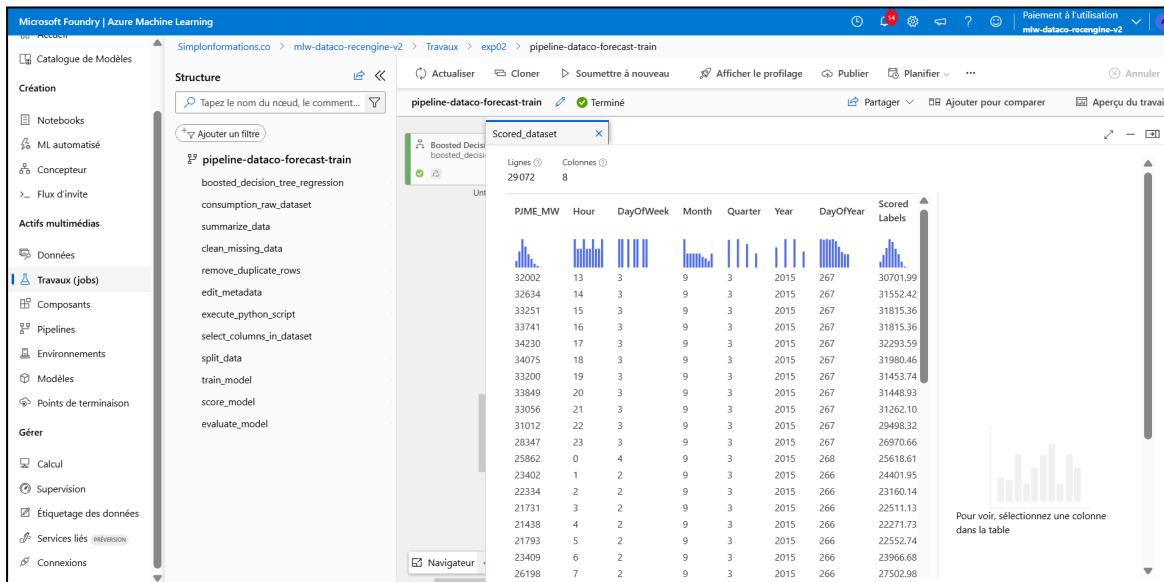
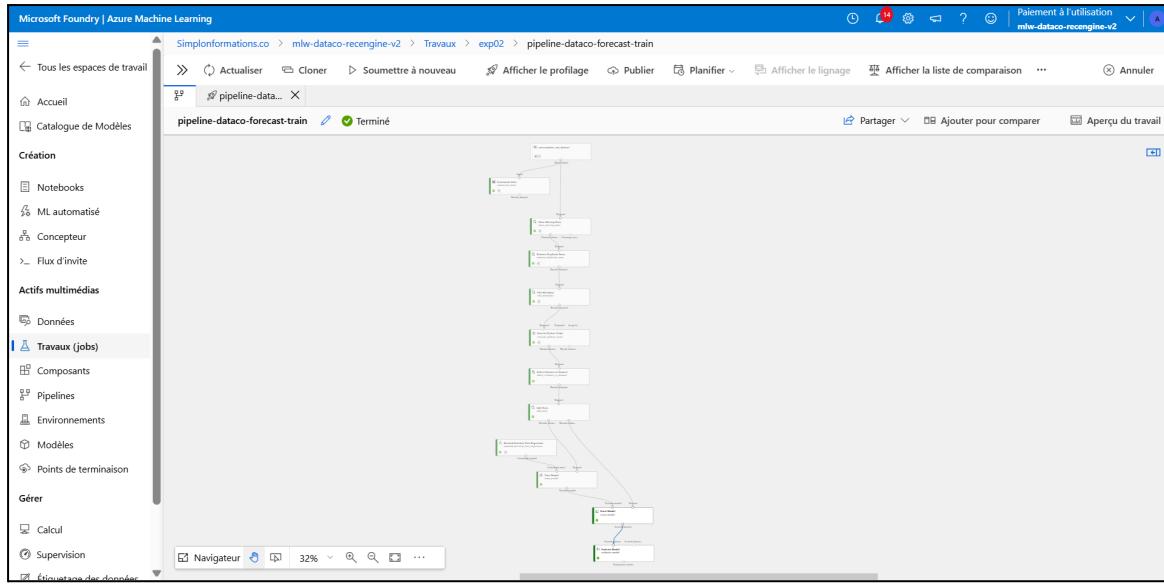




Etape 5 - Lancement du Pipeline et Sauvegarde du Modèle :

- **Valider** la Pipeline.
- Cliquer sur **Configurer & Soumettre** pour lancer la Pipeline.
- Visualiser et suivre l'avancement des étapes du pipeline via **View Profiling**.
- Après la fin de l'exécution de pipeline avec succès, on peut :
 - Visualiser les **résultats des Scores** du modèle entraîné en cliquant sur le **Module Score Model**.
 - Visualiser **les métriques (MAE, RMSE, RSE, RAE)** d'évaluation en cliquant sur **Evaluate Model**.
- **Sauvegarder** le modèle entraîné en faisant un clic droit sur le module "**Train Model**" puis "**Register model**", ensuite assigner un **nom** au modèle.
- Dans le menu gauche **Actifs multimédias > Modèles**, on trouvera le modèle Sauvegardé.





Etape 6 - Création du pipeline d'inférence et Déploiement :

- Dans le Designer, en haut à droite, cliquer sur **Create inference pipeline > Real-time inference pipeline.**
- Le système transforme le canvas :
 - Il supprime le **Train Model** et le **Split Data**.
 - Il conserve : Transformation (**Nettoyage + Python Script**) + Le **modèle entraîné + Score Model**.
 - Il ajoute : **Web Service Output**.
- Cliquer sur **Configurer & Soumettre** pour exécuter ce nouveau pipeline.
- Une fois terminé, cliquer sur **Publier**, cela donnera un **URL REST API** pour consommer les prévisions.

