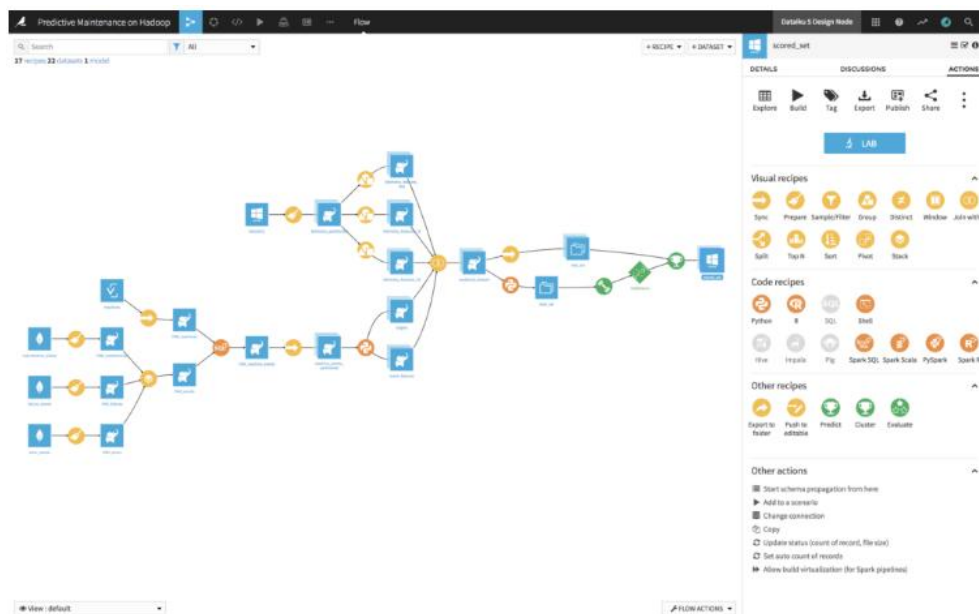


Introduction à Dataiku

Dataiku, C'est quoi?

<https://www.dataiku.com/product/>

La plateforme tout-en-un de science des données et d'apprentissage machine qui rassemble tout le monde pour avoir un impact transformateur sur les entreprises



- All-in-One Platform

- Collaborative Data Science

Il existe d'autres solutions tout-en-un sur le marché :

Figure 1: Magic Quadrant for Data Science and Machine Learning Platforms



Source: Gartner (March 2021)

- Talend : <https://www.talend.com/products/data-fabric/>

- Alteryx: <https://www.alteryx.com/fr/products/apa-platform>

Présentation de DSS

Dataiku DSS (Data Science Studio) est une plateforme collaborative de science des données qui permet aux équipes de travailler ensemble sur des projets de data science. Elle offre une gamme complète d'outils pour la préparation des données, l'analyse, la modélisation et la mise en production des modèles de machine learning. Voici quelques fonctionnalités clés de Dataiku DSS :

Préparation des données : DSS offre des outils pour nettoyer, transformer et enrichir les données. Il permet également d'automatiser les flux de travail de préparation des données.

Visualisation des données : Il fournit des outils pour visualiser les données sous forme de graphiques et de tableaux de bord interactifs.

Machine Learning : DSS intègre des outils pour créer, entraîner et évaluer des modèles de machine learning. Il prend en charge les bibliothèques courantes telles que Scikit-Learn, TensorFlow, Keras, etc.

Collaboration : DSS facilite la collaboration entre les data scientists, les analystes et les ingénieurs en fournissant des fonctionnalités de partage de projets, de versionnage et de commentaires.

Mise en production : Il permet de déployer les modèles en production et de gérer leur cycle de vie. DSS offre également des capacités de surveillance et de gestion des modèles déployés.

Intégration : DSS s'intègre avec une variété de sources de données, y compris les bases de données SQL, les entrepôts de données, les services de stockage cloud, les API, etc.

1- Prise en main de DSS:

Valider le module Core designer : <https://academy.dataiku.com/path/core-designer>

- intro : <https://academy.dataiku.com/path/core-designer/basics-101/>
- explore Data : <https://academy.dataiku.com/path/core-designer/basics-102>
- Dashboard: <https://academy.dataiku.com/path/core-designer/basics-103>

2- Bonus:

Avancer sur le maximum des quickstar suivants:

<https://academy.dataiku.com/page/quick-start-programs>

3- Livrables:

- Un document pdf qui démontre avec des captures écran de chaque étape la bonne validation des ressources validées.

=> à enregistrer dans votre dossier apprenant sur le drive et notifier votre formatrice par MP sur discord.

Documentation Utile:

* *Data preparation*

- Recipes in DSS:

<https://knowledge.dataiku.com/latest/courses/basics/prepare-data/concept-recipe.html>

- Join Recipe: <https://knowledge.dataiku.com/latest/courses/lab-to-flow/join/join-summary.html>

* *Exploring Datasets*

- Connecting to and Exploring Data:

*Datasets in DSS:

<https://knowledge.dataiku.com/latest/courses/basics/create-dataset/concept-dataset.html>

- Schema: <https://knowledge.dataiku.com/latest/courses/basics/explore-data/concept-schema.html>

- Storage Type and Meaning:

<https://knowledge.dataiku.com/latest/courses/basics/explore-data/concept->

[storage-meaning.html](https://knowledge.dataiku.com/latest/courses/basics/explore-data/concept-storage-meaning.html)

- Sampling: <https://knowledge.dataiku.com/latest/courses/basics/explore-data/concept-sampling.html>)
- Analyze: <https://knowledge.dataiku.com/latest/courses/basics/explore-data/concept-analyze.html>
- Connections: <https://knowledge.dataiku.com/latest/courses/basics/create-dataset/concept-connection.html>

*** Charts**

- Charts: <https://knowledge.dataiku.com/latest/courses/basics/explore-data/concept-charts.html>

*** Reporting & Insights**

- Dashboards in DSS: <https://knowledge.dataiku.com/latest/courses/lab-to-flow/reporting/concept-dashboards.html>

*** Ressources documentation**

- <https://academy.dataiku.com/page/learning-paths>
- <https://academy.dataiku.com/page/certifications>