



Dataiku DSS avancé : labs

DS035

Durée: 1 jour

Public :

Chefs de projet, développeurs, data-scientists, utilisateurs de Dataiku DSS

Objectifs :

Savoir installer, configurer, Dataiku DSS, l'utiliser depuis l'interface web ou des API.

Connaissances préalables nécessaires :

Connaissance des principes de la manipulation de données et du machine learning.

Programme :

Présentation, concepts DSS

Fonctionnalités :

Mise à disposition des méthodes et outils de data-sciences à partir d'une interface graphique ou de langages de requêtage ou de programmation (Python, SQL, R)

Notions de projet, data, dataset, recipes

Premiers pas avec Dataiku DSS : exemples

Connexion aux données

Connecteurs disponibles, formats de fichiers, formats spécifiques hadoop/Spark

Chargement des fichiers, emplacement des fichiers,

Système de fichiers, HDFS, Amazon S3, stockage Google Cloud, etc ...

FTP, connexion ssh, connexion aux bases de données SQL

Autres stockages : Cassandra, MongoDB, Elasticsearch, ...

Préparation des données

Scripts de préparation, échantillonnage, exécution,

les différents moteurs d'exécution: DSS, Spark, base de données SQL, , Spark

Graphiques et statistiques



Phirio

Présentation des types de graphiques disponibles et configuration
Echantillonnage et exécution,
graphiques de base, tables, nuages de points,
visualisation cartographique, etc ...
Réalisation d'exercices pratiques.
Fonctionnalités statistiques disponibles,
démonstrations

Machine learning

Fonctionnalités disponibles:
apprentissage supervisé et non supervisé,
algorithmes utilisés, scoring, deep learning, optimisation de modèles, ...

Flow/Recipes

Interface graphique de gestion des cycles de traitement, des datasets,
des traitements, paramétrage de l'exécution en parallèle,
des données géographiques,
export en format pdf ou images, etc ...
Travaux pratiques

Interfaces de programmation

Présentation des différentes interfaces : SQL, Python, R
Exemples d'interrogations SQL et Python
Interfaçage avec Spark