



Phirio

Dataiku DSS

DS034

Durée: 2 jours

1 610 €

25 au 26 mars
27 au 28 mai

4 au 5 septembre
12 au 13 novembre

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), principe de "Self Service Analytics"

Cluster DSS : Design, Deployer, Api, Automation et Govern Nodes

Notions de projet, data, dataset, recipes

Outils collaboratif : Wiki, tags, discussions...

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

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

Le laboratoire
Fonctionnalités disponibles:
apprentissage supervisé et non supervisé,
algorithmes utilisés, scoring, deep learning...
Evaluation, optimisation, compréhension du modèle
Déploiement et cycle de vie du modèle

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