

Formation de Pyspark : (deux semaine, 10 jours)

Cours de Big Data Analytics Using Spark

Accessible par: <https://learning.edx.org/course/course-v1:UCSanDiegoX+DSE230x+3T2020/home>

Il faut ouvrir un compte sur le site EDX. Le cours est accessible gratuitement. Il y a une possibilité de prendre la certification en passant l'examen. Dans ce cas un frais de 220€ doit être payé.

Jour	Cours	Etat advancement
Jour 1	<ul style="list-style-type: none">• Introduction and course information• Installation de Spark• Map-reduce and Spark	<input type="checkbox"/>
Jour 2	<ul style="list-style-type: none">• The Memory Hierarchy• Spark basics• Quiz 1• Commencer le devoir (programming assignment 1)	<input type="checkbox"/>
Jour 3	<ul style="list-style-type: none">• Finir avec programming assignment 1• Pyspark and RDDs• Spark SQL and DataFrames• Quiz 2• Commencer le Programming assignment 2	<input type="checkbox"/>
Jour 4	<ul style="list-style-type: none">• Finir programming assignment 2• PCA and Weather Analysis• Finir Programming assignment 3	<input type="checkbox"/>
Jour 5	<ul style="list-style-type: none">• Visualizing PCA residuals 1 et 2• Programming assignment 4	<input type="checkbox"/>
Jour 6	<ul style="list-style-type: none">• K-mean and Intrinsuc Dimensions• Programming assignment 5	<input type="checkbox"/>
Jour 7	<ul style="list-style-type: none">• Finir le programming assignment 5• Decision Trees, random forests and boosting	<input type="checkbox"/>

	<ul style="list-style-type: none"> • Programming assignment 6 	
Jour 8	<ul style="list-style-type: none"> • Finir le programming assignment 6 • Boosting et ensembles • Programming assignment 7 	<input type="checkbox"/>
Jour 9	<ul style="list-style-type: none"> • Finir le programming assignment 7 • Neural networks and tensorflow • NN, Tensorflow, Estimator API • Programming assignment 8 	<input type="checkbox"/>
Jour 10	<ul style="list-style-type: none"> • Finir le programming assignment 8 	<input type="checkbox"/>

Les ressources utile:

1. Learn Pyspark
 - a. Explication de tous les fonctions et méthodes nécessaire de pyspark avec exemple y compris les algorithmes machine learning.
 - b.
2. Machine Learning with Pyspark
 - a. Explication plus approfondi sur les algorithmes ML en pyspark avec beaucoup d'exemple. Implémentation et des exemples sont bien expliqué.
3. Spark, the definitive guide
 - a. Livre le plus complet pour apprendre Spark. Les exemples sont présenté en scala et pyspark.