

centrale_201606

November 26, 2021

1 Big Data, Azure, Machine Learning, Python

Présentation à Centrale Paris - Juin 2016.

[Xavier Dupré](#)

xavier.dupre AT gmail.com

Senior Engineer at **Microsoft France** on [Azure ML](#), **Teacher in Computer Science** at the [ENSAE](#)



```
[1]: from jyquickhelper import add_notebook_menu
      add_notebook_menu()
```

```
[1]: <IPython.core.display.HTML object>
```

1.1 Introduction

Expérience

- Microsoft, Bing, Azure
- Enseignements, ENSAE, Formation professionnelle (Institut des actuaires) (+ 200 élèves, + 60h de cours)

Expertise

- Machine Learning (PhD)
- Python
- Map Reduce
- Azure

Microsoft in universities

- [Microsoft, partenaire de la filière Data Science de l'ENSAE ParisTech avec Microsoft Azure Machine Learning \(2014\)](#)
- [Developing the Next Wave of Data Scientists \(2015-2016\)](#)
- [Microsoft is one of the sponsors of the DataScienceGame \(2016\)](#)



1.1.1 Que feriez-vous si ?

- Statistiques descriptives sur un fichier de 1Go, 10Go, 100Go ?
- Apprendre une régression logistique sur ... ?
- Apprendre une forêt d'arbre sur ... ?
- Et si vous deviez le faire toutes les semaines ?
- Comment représenter 10M de points sur une carte ?

Retour des étudiants à propos du hackathon :
C'était bien de se confronter à des données pas très propres.

1.1.2 Démarrage d'un cluster Hadoop sur Azure

1.2 Cas concret 1 : système de recommandation, moteur de recherches

- Système de recommandation, application chez Bing aux recherches associées
- Quelques approches théoriques
- La mise en pratique avec aux gros volumes de données
- Optimisation sur Internet, apprentissage par renforcement

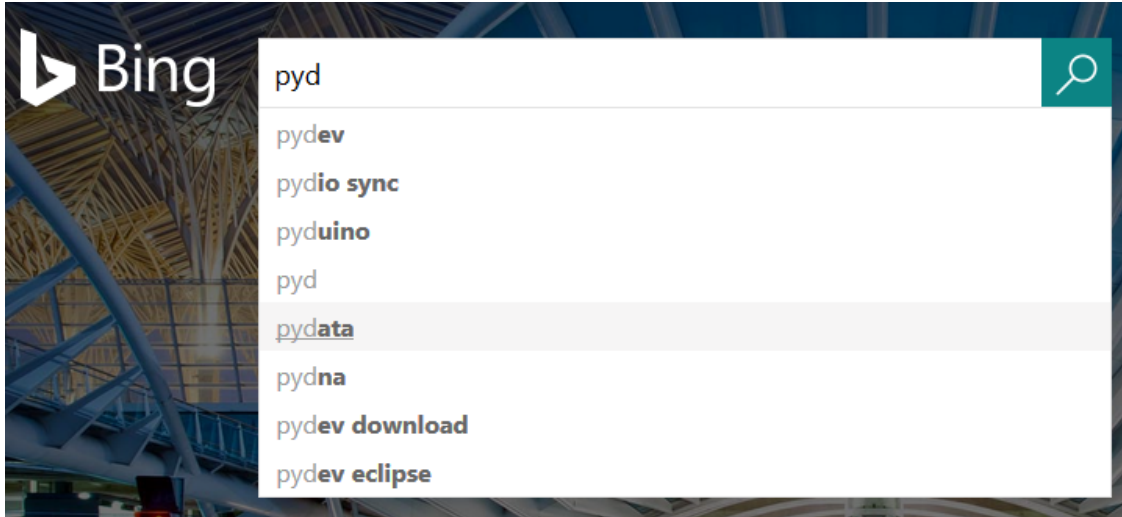
1.3 Outils Big Data chez Microsoft / Azure

- Cluster Map/Reduce, Azure ML, Machine Virtuelle, Power BI
- Trois exemples d'utilisation (Machine Virtual, Azure ML, Cluster Hadoop) sur des jeux de données académiques
- Quelques exercices

1.4 Cas concret 2 : suggestions sur Internet

- Réflexions autour de l'impact d'un système de suggestions

- Sur les utilisateurs
 - Sur les logs
- Métriques : comment mesurer l'impact ?



[2] :