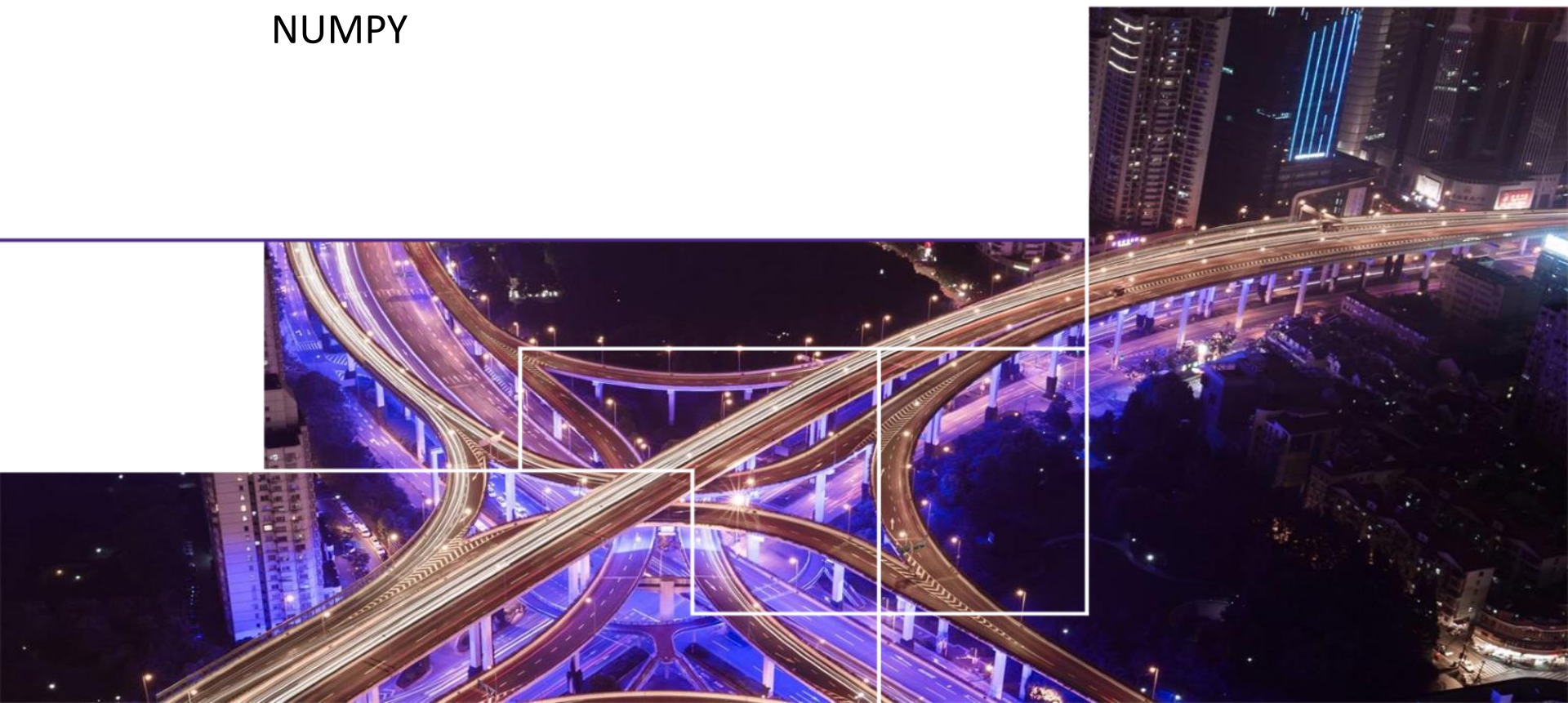




NUMPY





- Découvrir Numpy, la bibliothèque de référence de manipulation de tableau et matrices sous python
- Manipuler l'ensemble des méthodes de Numpy couramment utilisées en datascience

QU'EST CE QUE NUMPY ?

Numpy : "Numerical Python"

- Package python écrit en **C** pour être plus **efficace**.
 - Un seul pc avec Numpy peut sur certains problèmes être plus efficace qu'un cluster de machine sur le cloud.
- Orienté pour gérer de manière efficace des **tableaux (matrices) de grandes dimensions**.
 - En datascience, (presque) tout est matrice
 - En datascience, (presque) tous les jeux de données ont beaucoup de dimension
- Certaines des opérations de Pandas utilisent en fait Numpy derrière.



Numpy vs Python

- Python n'a pas de type tableau, juste des listes de listes
- Les tableaux peuvent-être plus efficaces que les listes, mais nécessitent des **éléments de même taille**
 - Les algorithmes de datascience/Machine Learning aussi, et manipulent le plus souvent des matrices
- Numpy a **des fonctionnalités plus adaptées** pour le calcul scientifique et sur des données de grandes tailles

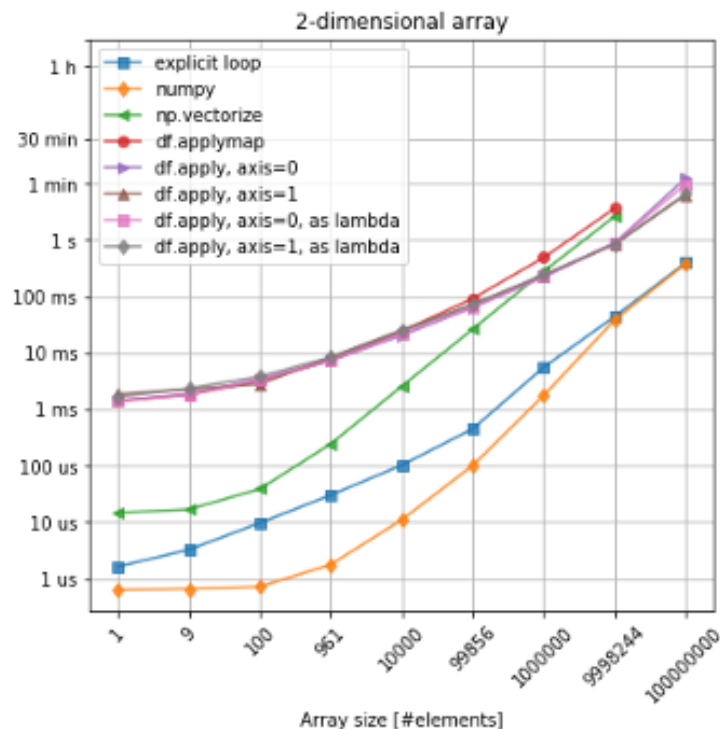
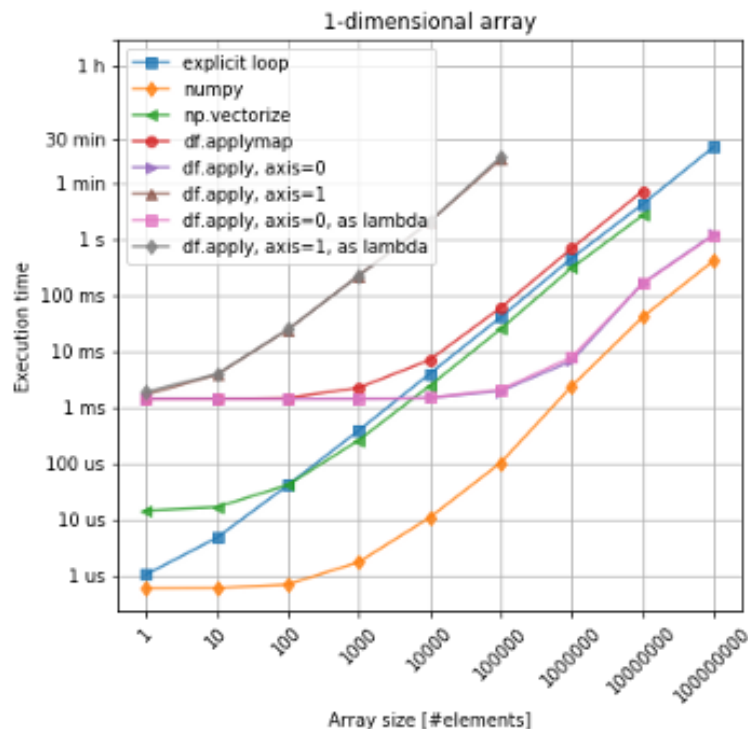
	Numpy	Python (list)
Code	<pre>a = numpy.array([1, 2, 5, 7, 8]) a[2:5] = 3</pre>	<pre>a=[1,2,5,7,8] a[1:5] = 3</pre>
Résultat	<pre>array([1, 2, 3, 3, 3])</pre>	<pre>TypeError: can only assign an iterable</pre>
Conclusion	a est bien modifié !	Ce n'est pas possible !

Numpy vs Pandas

- Pandas permet de gérer plus facilement des types de données multiples
 - Lorsque que l'on charge et explore un jeu de donnée, on préfère souvent pandas.
- Numpy est adapté lorsque les données ont été rendus homogènes (à minima chaque échantillon de même taille, souvent de même type)
 - Traitement de toute façon nécessaire avant d'appliquer la majorité des traitements de datascience et machine learning
- Numpy est **plus efficace** que pandas dans de nombreux cas
 - En gestion mémoire. Le jeu de donnée Iris chargé 100000 fois :
 - Pandas: 1506 MB
 - Numpy: 686 MB

POURQUOI NUMPY ?

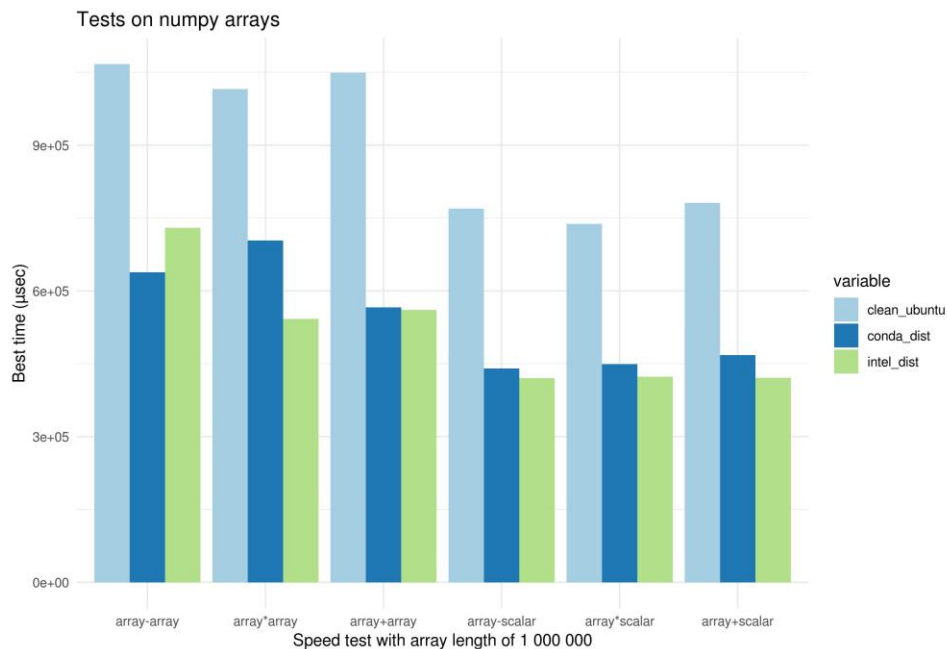
- Numpy est **plus efficace** que pandas dans de nombreux cas



Appliquer une méthode (ici $x \rightarrow x^{0.5}$) sur chaque élément de tableaux à 1 et deux dimensions

POURQUOI NUMPY ?

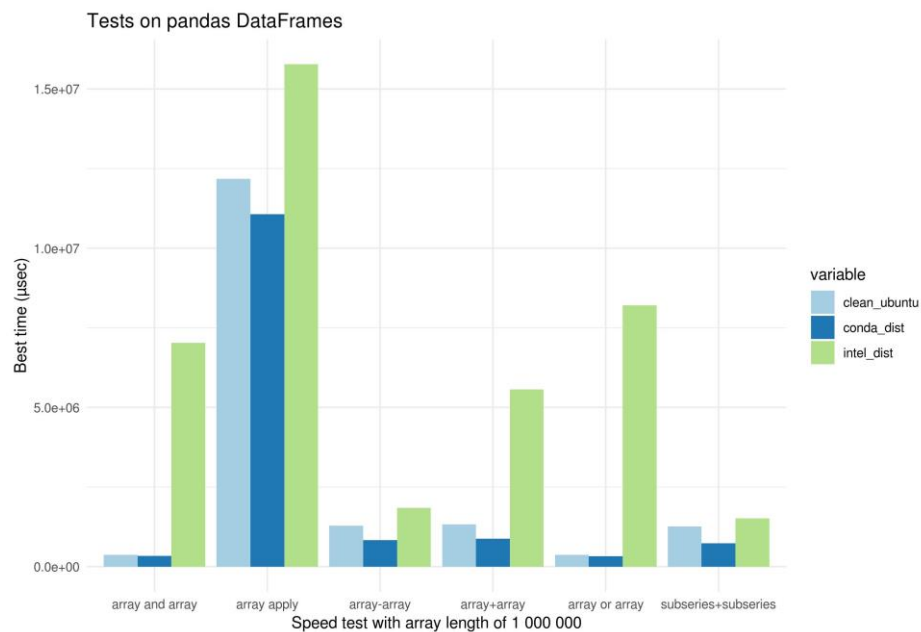
- Numpy est **plus efficace** que pandas dans de nombreux cas
- Et peut bénéficier de backends optimisés pouvant encore accélérer ses calculs



Utilisations des backends conda et intel par Numpy

POURQUOI NUMPY ?

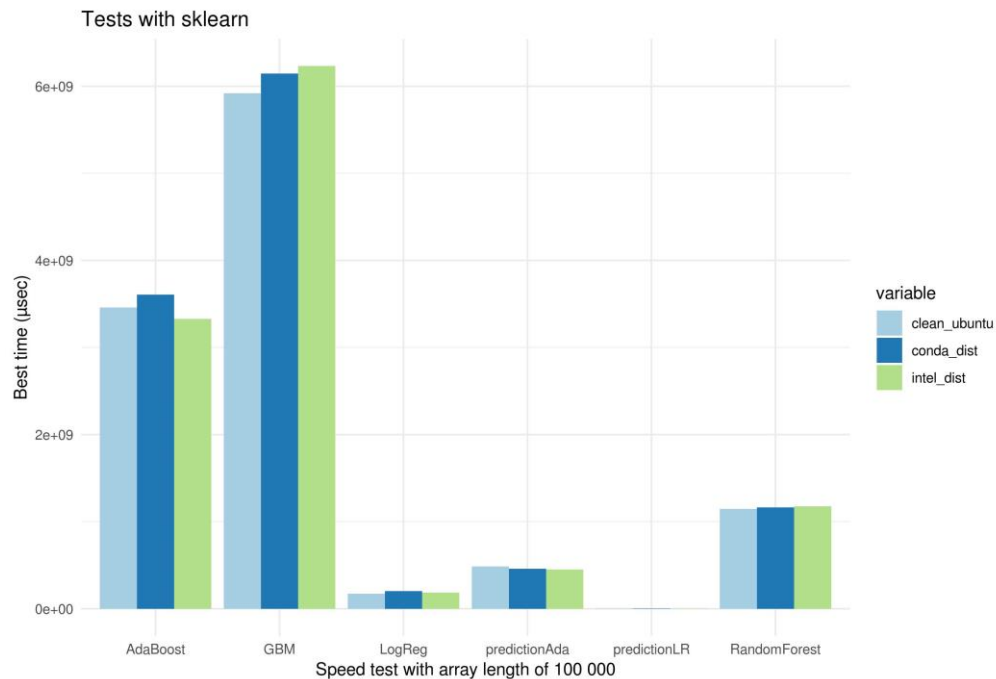
- Numpy est **plus efficace** que pandas dans de nombreux cas
- Et peut bénéficier de backends optimisés accélérant les calculs.



Utilisations des backends par pandas

POURQUOI NUMPY ?

- Numpy est **plus efficace** que pandas dans de nombreux cas
- Et peut bénéficier de backends optimisés accélérant les calculs.



Utilisations des backends par sklearn

LA SUITE EN LIVECODING