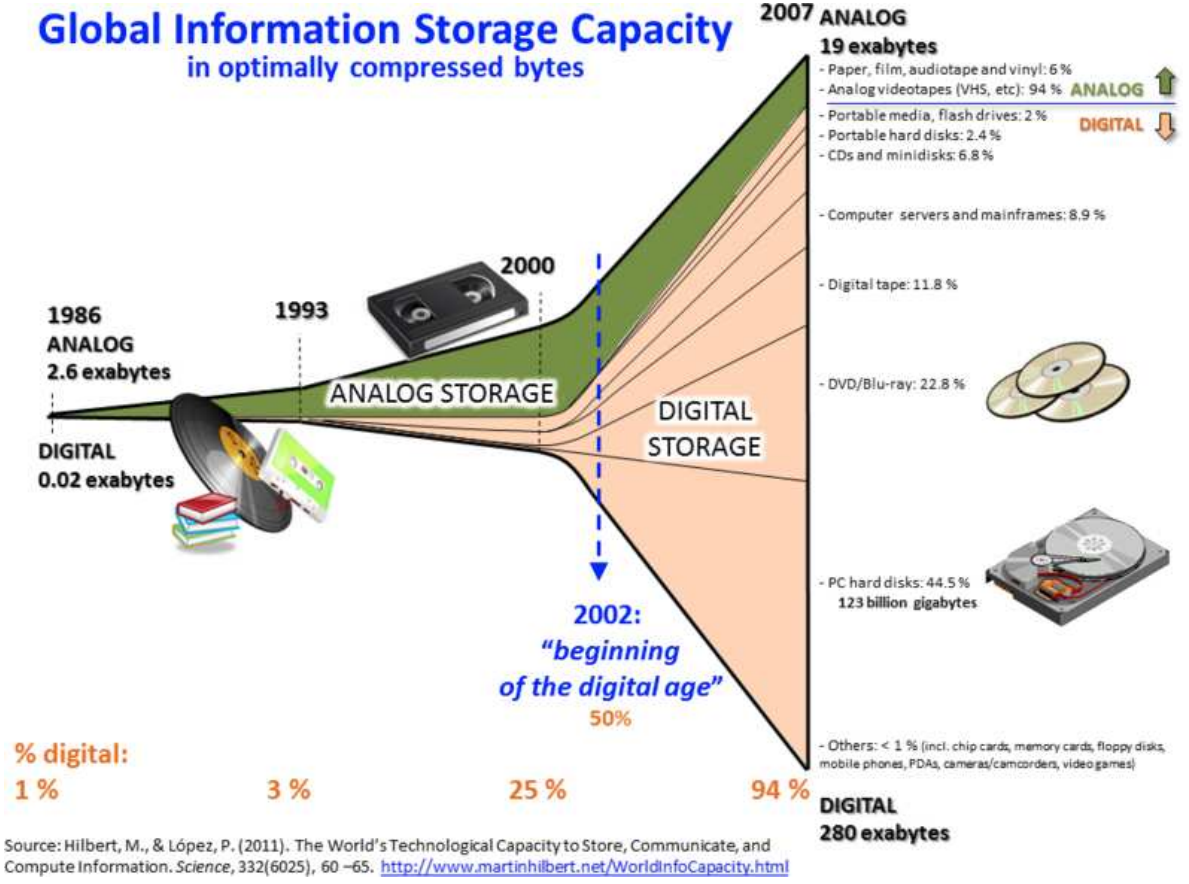
Traitement de données avec PANDAS

F. Alizard

13 mars 2018

Une introduction



Accumulation de données dans tous les secteurs d'activités (modélisation numérique, bilbiothèques numériques : classification des données, fouille de textes, annotation extraction de document)
Nature, Volume 455, N. 7209, 4 septembre 2008. Numéro spécial " Science in the petabyte era ".

Une introduction

Méthode classique

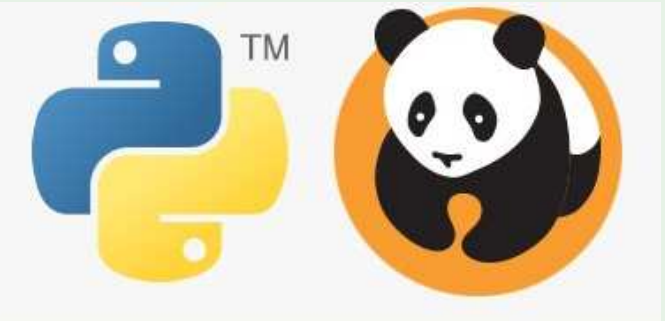
1. Utilisation d'un tableur
2. Données simples, non mixtes
3. Licences, et versions
4. Difficiles pour des données/traitements complexes (VBA, etc...)

Une alternative libre : Pandas

1. Basé sur un vrai langage de programmation (Python)
2. Traitement simple de données complexes
3. Automatisation
4. Facilité pour tracer les données grace à matplotlib

Une introduction

Python Library for Data Analysis and Statistics



Pandas : http ://pandas.pydata.org

1. Manipuler des tableaux de données avec des étiquettes pour colonnes et lignes. Structure équivalent à un tableur.
2. Tableaux : DataFrames
3. Entrée (ordonnées ou non) et sortie simple à gérer. Ecriture vers un fichier tabulé (.csv, HDF5)
4. Possibilité de nettoyage des données manquantes simple.
5. Outils de traitement statistiques intégrés

Une introduction

Plan

1. Création de DataFrames (tableaux numpy, dictionnaire, ou fichier), entrée/sortie.
2. Manipulation :
   - Indexation des données
   - Extraction
   - Opérations simples
   - Tracer
3. Modification des données
4. Regroupement des données