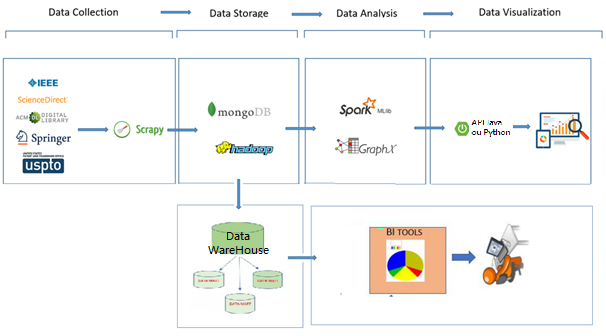
Mini Projet: Big Data & BI

L’objectif du projet est la création d’un petit Framework BI pour l’analyse du BIG Data. L’architecture proposée est illustré dans la figure suivante



Elle fait intervenir les es technologies et les compétences suivantes:

* La collection de données (Scraping) avec Scrapy.
* Le Stockage de données volumineuses avec Hadoop et MongoDB.
* L’analyse des Données avec Apache Spark.
* La visualisation des résultats dans une interface Web en se basant sur l’utilisation de Python et l’api Flask (ou sur les bibliothèques JavaScript).
* Mise en place d’un datawarehouse et analyse et reporting par une plateforme BI (Tableau, Pentaho, …)

**Etapes du Projet:**

1. Travailler avec le Framework Scrapy pour collecter ou crawler les sites web qui contiennent les données qui nous intéressent ;
2. Travailler avec la Base de données MongoDB et Hadoop pour stocker les données collectées, qui sont de nature différentes (structurés, semi structurés et non structurés) ;
3. Analyse et visualisation :

**3. 1 : Analyse Big Data :**

* + - * Analyse par le Framework Apache Spark.
      * Visualisation des résultats de l’analyse par le développement Python et l’api Flask (On pourra le faire avec avec Angular et JavaScript).

**3. 2 : Analyse BI :**

* + - * Analyse SID et Mise en place d’un DWH dans un SGBDR.
      * Analyse et reporting par une plateforme BI au choix (Tableau, Pentaho, …)

**Etude de cas :**

* **Data Collection:**

En utilisant les capacités de scrapping, on collectera un corpus de papiers scientifiques à partir des sites librairies IEEE, ScienceDirect et ACM. Des analyses seront ensuite faites et des génération de rapports.

On s’interessera tout particulièrement aux productions scientifiques sur le “Blockchain”

* **Data sources:**

- [IEEE](https://www.ieee.org/)

- [Science Direct](https://www.sciencedirect.com/)

- [ACM libraries](https://dl.acm.org/)

* **Results:**

Lancer de multiples analyses. Par exemple, on visualisera comme résultat, l’évolution des “Blockchain“ . Vous pouvez ensuite généraliser à plusieurs domaines de recherche en informatique