Collège de Bois de Boulogne Collecte de données  
420-A53-BB (3 heures) Nom\_\_\_Sami Bahig \_

Examen final - partie théorique

**Exercice 1 (20 pts)**

Répondre aux questions suivantes sachant qu’il peut y avoir une ou plusieurs bonnes réponses

1. **Que fait le code suivant?**

scrap = bs4.BeautifulSoup(html, 'html.parser')  
loc = scrap.find(id='price').find('h2').get\_text()

Réponse: d’abord fait le parsing de html , Objet BS avec paramètre de parseur (‘soupe’ de Html avec Beautifull soup) puis fonction pour isoler les propriétés d’une balise spécifique (‘price et ‘h2’ dans la ‘soupe’)

1. **Quels sont les types de BD NoSQL?**
   1. BD Key value\_\_\_\_\_\_\_
   2. BD document\_\_\_\_\_\_\_
   3. BD type colonnes\_\_\_\_\_\_\_
   4. BD Graphe\_\_\_\_\_\_\_
   5. BD Relationnelle\_\_\_\_\_\_\_
   6. Aucune\_\_\_\_\_\_\_

Réponse: a, b, c, et d

1. **La BD Mongo est utilisée pour le stockage de données XML**
   1. Vrai \_\_\_\_\_\_\_\_
   2. Faux \_\_\_\_\_\_\_\_

Réponse: Faux

1. **Seules les données catégorielles peuvent être représentées dans le format JSON?**
   1. Vrai \_\_\_\_\_\_\_
   2. Faux \_\_\_\_\_\_\_

Réponse: Faux

1. **Le processus de web scraping se fait selon les étapes suivantes :**
   1. Identifier les données à obtenir
   2. Identifier les conditions d’utilisation imposées par le propriétaire du site web
   3. Identifier les pages à scraper
   4. Toutes les étapes ci-dessus

Réponse: d

1. **Le format JSON est utile pour l’échange de données structurées seulement**
   1. Vrai\_\_\_\_\_\_\_\_
   2. Faux\_\_\_\_\_\_\_\_

Réponse: (b)Faux

1. **La BD Mongo est un système de gestion de bases de données orientée colonnes**
   1. Vrai\_\_\_\_\_\_\_\_
   2. Faux\_\_\_\_\_\_\_\_

Réponse: (b) Faux

1. **Un dataframe dans le module pandas représente une structure** 
   1. non relationnelle\_\_\_\_\_\_\_\_
   2. Tabulaire\_\_\_\_\_\_\_\_
   3. Aucune structure\_\_\_\_\_\_\_\_
   4. Toutes les réponses ci-dessus\_\_\_\_\_\_\_\_

Réponse: b

1. **Les éléments utilisés dans une base de données Mongo sont :**
   1. Les enregistrements\_\_\_\_\_\_\_\_
   2. Les collections\_\_\_\_\_\_\_\_
   3. Les documents\_\_\_\_\_\_\_\_
   4. Toutes les réponses ci-dessus\_\_\_\_\_\_\_\_

Réponse: b et c

1. **Apache Kafka utilise une approche de données Publisher-Observer :**
   1. Vrai
   2. Faux

Réponse: Vrai ( c’est le Publisher est appelé Producer et l’observer est appelé Consumer)