

Les données semi-structurées

Chapitre 1: Introduction

3^{ème} Année Licence Informatique
Ingénierie des Systèmes d'Information et du Logiciel (ISIL)
Préparé par M. L. FOUGHALI

Ressources utilisées :

1. XML, Ed TITTEL, SCHAUMS, 2004.
2. XML par la pratique Bases indispensables, concepts et cas pratiques (3ième édition), Thierry BOULANGER, Editions-ENI, 2015.
3. XML, Gilles CHAGNON & Florent NOLOT, PEARSON, 2007.
4. <https://waytolearnx.com/2019/03/difference-entre-donnees-structurees-semi-structurees-et-non-structurees.html>

La nature des données

- Les données est la matière première de tout algorithme. Les données peuvent être de diverses natures et plus ou moins structurées.
- Chaque type de données diffère dans la manière dont il est créé, stocké et analysé.
- Sur la base des caractéristiques, les données sont classées dans trois types:
 1. structurées,
 2. semi-structurées et
 3. non structurées.

Données structurées

- Les données structurées sont des données dont les éléments sont adressables pour une analyse efficace.
- Ils sont organisés dans un référentiel formaté qui est généralement une base de données.
- Ces données sont le plus souvent traitées lors du développement et constituent le moyen le plus simple de gérer les informations.

Id	Name	Age
1	Dida	39
2	Louey	13
3	Wael	11
4	Lamis	9

Exemple: données **relationnelles**.

Données semi-structurées

- Les données semi-structurées sont des informations qui ne résident pas dans une base de données, mais qui possèdent des propriétés organisationnelles facilitant leur analyse.
- Avec certains processus, vous pouvez les stocker dans des bases de données (cela peut être très difficile pour certaines données semi-structurées).

```
<?xml version = "1.0" encoding="UTF-8" standalone="yes"?>
<document>
  <employee><name>Dida</name><age>39</age></employee>
  <employee><name>Louey</name><age>13</age></employee>
  <employee><name>Wael</name><age>11</age></employee>
  <employee><name>Lamis</name><age>9</age></employee>
</document>
```

Exemple: données XML

Données non structurées

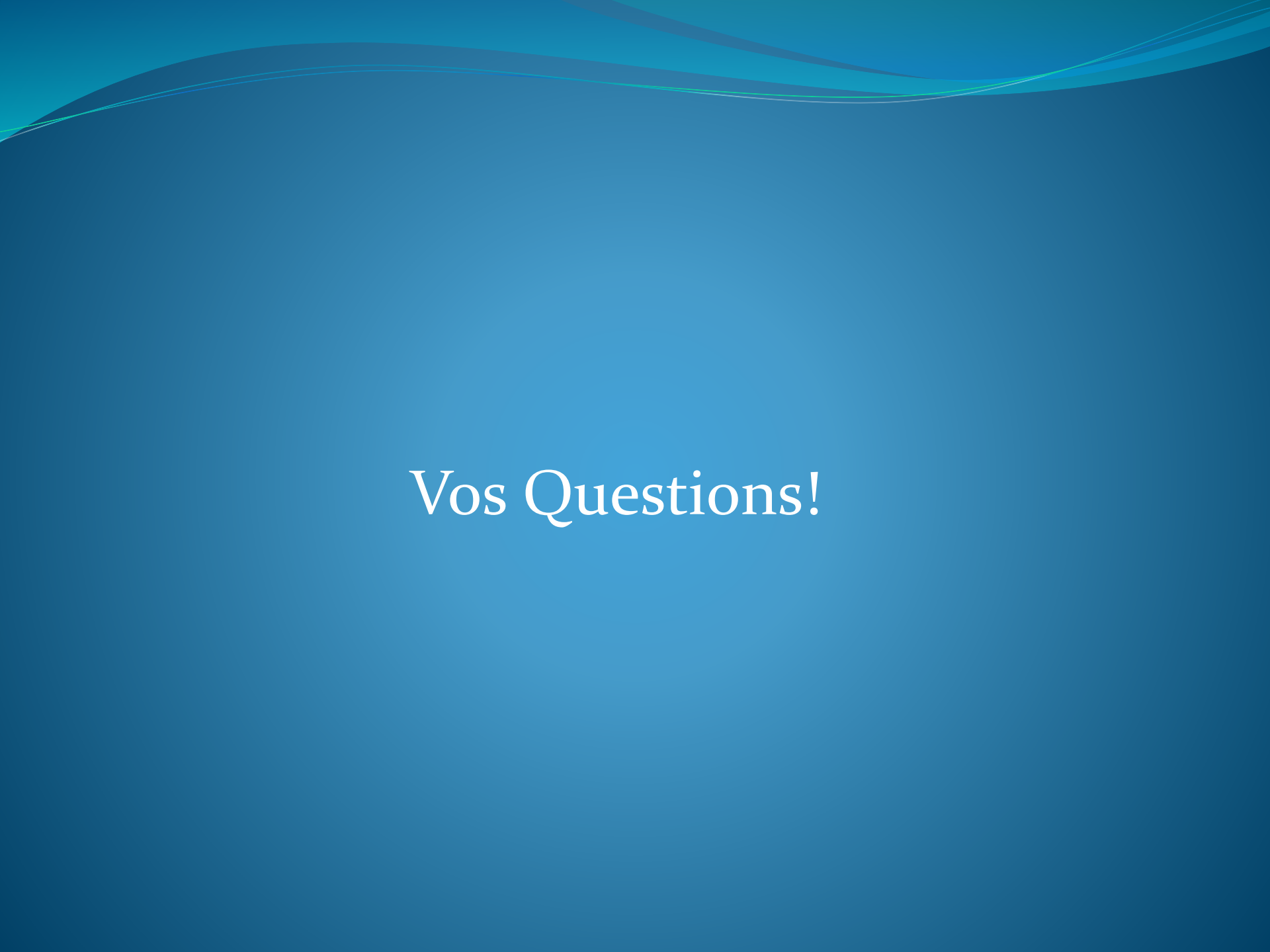
- Les données non structurées sont des données qui ne sont pas organisées de manière prédéfinie ou qui ne possèdent pas de modèle de données prédéfini.
- Elles ne conviennent donc pas à une base de données traditionnelle. Ainsi, pour les données non structurées, il existe des plates-formes alternatives pour le stockage et la gestion.
- Elles sont de plus en plus répandues dans les systèmes informatiques et sont utilisées par les entreprises dans diverses applications .

Exemple: Word, PDF, texte, logs.

Comparaison entre les données

Structurées	Semi-structurées	Non structurées
basées sur les tables de base de données	basées sur XML, JSON, ...	basées sur des caractères et des données binaires
dépendent du schéma et moins flexibles	plus flexible que les données structurées mais moins que les données non structurées	très flexibles (absence des schémas)

Conclusion: les données structurées ne représente qu'une infime partie de toutes les données informatiques.



Vos Questions!