

# Base de données cliniques et Big Data / Clinical database and Big Data



Composante  
Polytech  
Grenoble - INP,  
UGA

- > **Langue(s) d'enseignement:** Français
- > **Ouvert aux étudiants en échange:** Non
- > **Code d'export Apogée:** KATP9M07

## Présentation

### Description

Acquérir :

- Culture générale des écosystèmes Big Data
- Notions d'architecture et d'intégration
- Méthode de conception d'un lac de données
- Méthodes et outils pour la collecte, l'organisation et l'exploration interactive de données massives et/ou complexes
- Savoir-faire l'implémentation d'une plate-forme d'exploration de données de santé centré sur une base graphe

#### PARTIE THÉORIQUE (8H)

Partie 1 : Introduction / Contexte

- Big Data : définitions
- Regroupement de données
- Approche de construction d'un lac de données : workflow associé
- Les démarches péri-informatiques (gouvernance, CNIL, ?)
- Les écosystèmes autour du Big Data
- Considérations techniques (montée en charge, haute disponibilité, redondance, ...)

Partie 2 : Mise en œuvre sur un Entrepôt de Données de Santé (EDS)

- Contexte CHU
- Approche projet

- Workflow de traitements : ETL, Traitement, Accès
- Un exemple d'architecture générale
- Quelques outils indispensables
- Exemple d'un projet de EDS au CHUGA

#### Partie 3 : Conception d'un lac de données

- Contexte, objectifs, moyens.
- Concepts des bases graphes
- Domaines courants d'application
- Interopérabilité, web sémantique
- Du modèle à la base orientée graphe
- Modèle pivot du lac, évolutions

#### Partie 4 : Implémentation d'une plate-forme interactive d'exploration de données complexes et massives

- Focus ArangoDB, Elastic, ETL, IHM
- Élaboration de données métier

#### PARTIE PRATIQUE (14H)

##### Partie 5 : Exercice de modélisation (2H)

##### Partie 6 : Mini-projet (12H)

Déployer un workflow de traitement de données : ETL, BD Graphe, Visualisation

##### Bilan des connaissances acquises (2H)

#### Acquire:

- General Culture of Big Data Ecosystems
- Architecture and integration concepts
- Method of designing a data lake
- Methods and tools for collecting, organizing and interactive exploration of massive and / or complex data
- Know-how implementation of a health data mining platform centered on a graph database.

#### THEORICAL PART (8H)

##### Part 1: Introduction / Background

- Big Data: Definitions
- Grouping of data
- Approach to build a data lake: associated workflow
- Peri-computing approaches (governance, privacy, ...)
- The eco-systems around Big Data
- Technical considerations (scalability, high availability, redundancy, ...)

##### Part 2: Implementation of a Health Data Warehouse (DHS)

- University Hospital of Grenoble Context
- Project approach
- Treatment Workflow: ETL, Treatment, Access
- An example of a general architecture

- Some indispensable tools
- Health Data Warehouse project example at the University Hospital of Grenoble

#### Part 3: Designing a data lake

- Context, objectives, means.
- Concepts of graph oriented databases
- Common areas of application
- Interoperability, semantic web
- From the model to the graph oriented database
- Pivot model of the lake, evolutions

#### Part 4: Implementation of an interactive platform for complex and massive data mining

- Focus ArangoDB, Elastic, ETL, HMI
- Business Data Development

#### PRACTICAL PART (14H)

##### Part 5: Modeling Exercise (2H)

##### Part 6: Mini-Project (12H)

Deploy a data processing workflow: ETL, BD Graph, Visualization

Assessment of acquired knowledge (2H)

## Heures d'enseignement

|   |    |     |
|---|----|-----|
| Base de données cliniques et Big Data / Clinical database and Big Data - TP | TP | 12h |
| Base de données cliniques et Big Data / Clinical database and Big Data - TD | TD | 4h  |
| Base de données cliniques et Big Data / Clinical database and Big Data - CM | CM | 8h  |

## Pré-requis recommandés

Connaissances générales en modélisation de l'information, bases de données et langage de développement.

General knowledge in information modeling, databases and development language.

**Période** : Semestre 9

## Évaluation initiale / Session principale - Épreuves

| Libellé | Nature de l'enseignement | Type d'évaluation | Nature de l'épreuve | Durée (en minutes) | Nombre d'épreuves | Coefficient de l'épreuve | Remarques |
|---------|--------------------------|-------------------|---------------------|--------------------|-------------------|--------------------------|-----------|
|         |                          |                   |                     |                    |                   | 50/100                   |           |

## Bibliographie

- Théorie des graphes, Stéphane Pelle ENSG
- [http://cours-fad-public.ensg.eu/pluginfile.php/1525/mod\\_resource/content/1/Theorie\\_des\\_graphes.pdf](http://cours-fad-public.ensg.eu/pluginfile.php/1525/mod_resource/content/1/Theorie_des_graphes.pdf)
- François Bouillé. Le modèle HBDS. ENSG 2013
- <http://cours-fad-public.ensg.eu/mod/imscp/view.php?id=254>
- Qwant et le machine learning, JRES 2017 - Sylvain Peyronnet
- <https://www.jres.org/fr/videotheque?mode=replay&id=189&resolution=360>
- Ph. GENOUD, Web des données : Les Principes-Les Standards du W3C - Journée Interopérabilité et Innovation - IGN-BRGM-OGC -7 Octobre 2014 -Paris

## Infos pratiques

### Lieu(x) ville

› Grenoble

### Campus

› Grenoble - Saint-Martin d'Hères