



E

Consultant Big Data, Ph.D

SYNTHESE DE COMPETENCES

Compétences techniques

- **Langages** : Python, Scala, Matlab, SQL, Bash
- **SGBD** : Access, MySQL,
- **NoSQL** : MongoDB (document), Neo4J (graph)
- **Stockage et management des données** : Hortonworks suite, MapR (Hadoop), Kafka, Elasticsearch
- **Data Integration** : Talend
- **Data Analysis** : Spark, Pandas, Numpy, Scikit-Learn (Machine learning)
- **Data Visualization** : Tableau, Matplotlib, Seaborn, Plotly

Compétences fonctionnelles :

- Compétences dans l'analyse d'images et de données à l'aide de Python, Matlab et Spark
- Connaissance de bases de données SQL et NoSQL.
- Data integration avec Talend et data visualization avec Tableau.
- Bonnes capacités d'apprentissage et d'auto-formation.
- Expérience du travail en autonomie et en équipe.

Langues:

- Français (écrit et parlé)
- Anglais (écrit et parlé)





EXPERIENCES PROFESSIONNELLES

2019 – Consultant Big Data (Consultant Jems) - RCI Bank and Service (projet Formation), Paris

Environnement technique : MapR, Spark, Talend, MapR-DB, Scala

Environnement fonctionnel :

Création d'une architecture fonctionnelle et technique de type Big Data

Réalisation d'un algorithme d'étude de la similitude des données client -> Création d'une base de données client unique

Analyse des données clients dans le cadre de la réalisation d'une campagne marketing

2012 - 2017 Doctorat de Biophysique Laboratoire d'instabilité génétique et d'organisation du noyau - CEA Fontenay-aux-Roses

Environnement technique: Matlab, Génétique des levures, Imagerie cellulaire

Environnement fonctionnel: Mise au point et réalisation d'un projet de recherche concernant l'étude par microscopie de du comportement des chromosomes chez *Saccharomyces Cerevisiae*

Réalisation de simulations informatiques basées sur des équations de physique de polymère afin de simuler le comportement des chromosomes.

Présentation des résultats lors de conférences internationales (Gif-sur-Yvette, Trieste)

Rédaction d'articles scientifiques dans des revues internationales à comité de lecture

2013 - Team Advisor - Compétition iGEM (MIT) - Equipe Paris Bettencourt, Paris

Environnement technique : MatLab, Biologie Moléculaire, Génétique des Mycobactéries, Imagerie Cellulaire

Environnement fonctionnel :

Management d'une équipe de jeune diplômés dans le secteur des biotechnologies/bio-informatique dans le cadre de la mise au point d'un projet de recherche dans le cadre d'un projet portant sur la mise au point de nouvelles approches thérapeutiques concernant le traitement des infections à *Mycobacterium Tuberculosis*

Rédaction d'un article scientifique dans une revue internationale à comité de lecture

2012 - Stage de Recherche - Laboratoire de modélisation de physiologie cellulaire - ENS Paris / Laboratoire d'instabilité génétique et d'organisation du noyau - CEA Fontenay-aux-Roses





Environnement technique: Matlab, Génétique des levures, Imagerie cellulaire

Environnement fonctionnel: Mise au point et réalisation d'un projet de recherche concernant l'étude par microscopie de l'impact de l'apparition des cassures double-brin sur la mobilité des chromosomes cassés chez *Saccharomyces Cerevisiae*

Réalisation de simulations informatiques basées sur des équations de physique de polymère afin de simuler le comportement des chromosomes.

2011 - Stage de recherche - Laboratoire de Chimie Théorique - UPMC, Paris

Environnement technique : Gaussian - Molekel

Environnement fonctionnel : Réalisation d'un projet de recherche visant à étudier grâce à des outils de chimie théorique la stabilité des acides nucléiques arseniés

Rédaction d'un article scientifique dans une revue internationale à comité de lecture

2011 - Team - Compétition iGEM (MIT) - Equipe Paris Bettencourt, Paris

Environnement technique : MatLab, Techniques de Biologie Moléculaire

Environnement fonctionnel :

Élaboration et réalisation d'un projet de recherche fondamentale concernant les mécanismes de communication inter-cellulaire au sein d'une population de *Bacillus Subtilis*

FORMATION

2019 FORMATION BIG DATA Jems Datafactory

OBJECTIFS :

- Définition du périmètre du Big Data.
- Enjeux stratégiques des projets Big Data en entreprise
- Maîtriser les différents types de données (structurées, non-structurées et semi-structurées)
- Comprendre et maîtriser Les architectures Big Data à base du projet Hadoop et son écosystème (spark, hive, yarn,...)
- Comprendre et maîtriser Les architectures Big Data à base des bases de données Nosql



JEMS - 6 bis rue des Graviers 92200 Neuilly-sur-Seine - www.jems-group.com

Votre contact JEMS : Etienne D'OLEON – 07 83 87 10 39



- Maîtriser et être capable de proposer des architectures lambda

Contenue de la formation :

- Introduction à hadoop et son écosystème (yarn, hive, pig,...)
- Travaux pratiques sur l'écosystème hadoop (hive, yarn , pig)
- Introduction à Spark et son écosystème (sparkSQL, sparkR, sparkML,...)
- Cas d'utilisation de Spark.
- La notion de RDD et de Data Frames
- Introduction aux lambda architectures (streaming, kafka, cassandra, elasticsearch)
- Mise en pratique avec Spark scala & python (eclipse, intellij, Zeppelin & jupyter)
- Spark – Cas d'usages (Spark avec hdfs, MongoDB, cassandra, elasticsearch)
- Introduction à l'écosystème elasticsearch (plugins, index, mapping, shard, replications, ...) & les bases de données NoSQL (MongoDB, cassandra, Hbase,...)
- Travaux pratiques sur elasticsearch : logstash (collecte et parsing de données), elasticsearch (stockage, replicas, sharding, mapping) et kibana (Visualisation)
- Travaux pratiques sur cassandra et mongodb
- Visualisation avec Tableau

2013 - 2017 Licence Economie / Gestion, Université Paris Saclay (4 ans)

2012 - 2017 Thèse de Doctorat Biophysique, Université Paris Diderot (5 ans)

2010 - 2012 Master de Bio-informatique, Centre de Recherche Interdisciplinaire (2 ans)

2005 - 2008 Licence de Génétique, Biologie Moléculaire, Université Pierre et Marie Curie (3 ans)

Awards:

MIT iGEM 2011, Best presentation

MIT iGEM 2013 Team Advisor - World Grand Prize, Best Medical Project



JEMS - 6 bis rue des Gravières 92200 Neuilly-sur-Seine - www.jems-group.com

Votre contact JEMS : Etienne D'OLEON – 07 83 87 10 39