



## Hamza MOUNAH

Big Data

+8 années d'expérience

Tél : 0652599728

Email : hamzamounah@gmail.com

# Data Architect

## Compétences

### Principales Compétences

BIG Data, Data Mining, analyses des données, Architectures, Développement, Réseau, Cloud et virtualisation, Web sémantique et RDF.

### Distribution

Hortonworks , Cloudera, Mapr, Snowflake

### Cloud

Aws, Microsoft Azure, SnowFlake

### Data Store

DynamoDB, Mysq/ MariaDB, HBASE, HDFS, HIVE, MongoDB, OpenLink Virtuoso, Cassandra  
Partitionning & Bucketing – Data compression (Gzip, BZip2, Snappy,..)  
Formats de fichiers (Sequence File, RCFile, Avro, Orc, Parquet,...)

### Sécurité

Rsa, SSL/TLS, VPN, Ldap, Knox, Ranger, Kerberos

### Architecture Pattern

Data Lake, Lambda, Kappa, Data Ops, Data Mesh.

### Système d'exploitation

Linux, Windows, Mac Os, Centos

### Langages de développement

PYTHON - JAVA - SCALA - SQL – HQL - SparQL - C- SHELL / BASH - GROOVY - RUBY - R

### FrameWorks et Outils

Spark, Talend, Dataiku, Presto, Livy, Nifi, ELK, Kafka, Oozie, Sqoop, Zookeeper, Flume, Hive/Pig, Eclipse, IntelliJ

### Devops

Ansible, Chef, SonarQube, CheckMarx, Nagios, ControlM, Jenkins, Jmeter, Ycsb, Maven, Git, Svn, Nexus, Artifactory

### Méthodologie

AGILE, SCRUM , UML, CONCEPTION ORIENTEE OBJET, Merise

## Formation et certifications

2022 Dataiku DSS certification x 4

2017 Certification Hortonworks HDP Developer <http://bcert.me/sjudfogn>

2015 Études d'ingénieur en Systèmes d'informations, ESIEA (75) Option Big Data

2011 Licence Mathématiques & Physiques Université Joseph Fourier - GRENOBLE

2008 Bac Scientifique option Mathématique (Mention Bien)

# INTERVENANT : H.M.

## Expériences

### Crédit Agricole- Big data (depuis 10/2021)

#### *Data Architect & Infrastructure As a Code:*

- Membre de l'équipe socle big data dans le pole cluster grandes clientèles.
- Mise en place d'une plateforme CDP.
- Migration des uses cases de HDP 2.7 vers CDP 7.3.
- Mise en place d'une offre de service ELK premium / Opendistro.
- Mise en place d'une offre de service IA (Dataiku).
- Mise en place d'une offre de service KubeFlow.
- Contribution à la mise en place d'une offre Spark kubernetes.
- Mise en place d'une offre confluent-kafka.

Environnement : Ansible, Python, Kvm, CDP, HDP, Hive, Elastic Search, Elastic Open distro, Dataiku, KubeFlow, Spark, kubernetes, kafka, zookeeper.

### Société Générale - Big data (02/2018 – 09/2021)

#### *Tech lead Data :*

- Membre du comité des techleads Big Data à la SG.
- Responsable technique d'une équipe de 7 personnes (3 à paris et 4 en Inde).
- Migration use case Refinitiv on AWS.
- Comité d'architecture move to Cloud.
- Mise en place du workflow Livy pour exécuter les jobs Spark via API.
- Migration des traitements de HDP 2.6.4 to CDP 7.1.5 .
- Adaptation des traitements à la version Hive 3.
- POC benchmarking Apache Presto and Apache Iceberg.
- Mise en place de l'architecture du data market.
- POC d'une architecture de Cloud hybride entre le cluster HDP on-premise et le cloud Microsoft Azure datalake.

### Société Générale - Big data (02/2018 – 09/2021)

#### *Tech lead Data :*

- Mise en place d'un service d'abonnement à la donnée « Datalake As A Service ».
- Mise en place d'une solution de Data Quality.
- Administration du cluster Hadoop (HDP, Ldap, edge, kerberos, Ranger...)
- Mise en place d'un manuel de bonnes pratiques pour les développeurs Talend.
- Mise en place d'une solution de centralisation des logs « data logger » avec la stack kafka, nifi et ELK.
- Conception et développement de jobs Nifi à des fins de monitoring.
- Mise en place d'une application de vérification des flux collectés dans le datalake et chargement des données dans les tables Hive (Météo des flux)
- Mise en place d'un projet " Instant Payment " qui consiste à croiser plusieurs sources de données pour contrôler les virements frauduleux.
- Traitement massive des données de virements pour générer des alertes sur les top conformités (projet réglementaire).

#### *Devops :*

- Mise en place d'une architecture CI/CD pour les projets Talend.
- Mise en place du workflow CI/CD pour les jobs spark.

Environnement : Aws DynamoDB, EMR, Glue, Lambda, API Gateway, VPC, Talend Big Data, Spark Scala, Hive, Hue, Sqoop, Splunk, Nifi, ELK, Hbase, MongoDB, Presto, Livy, Jenkins, Nexus, Ansible, Kafka, Iceberg, Microsoft Azure.

# INTERVENANT : H.M.

## Expériences

---

### LVMH – Data Engineer (11/2017 – 02/2018)

*Projet : Mise en place d'un DataLake Azure (Cloud)*

- Mise en place de deux datalake (Europe & Asie) pour migrer l'ensemble des données du groupe LVMH vers ces nouveaux Azure Data Lake (Cloud)
- Développement de mécanisme de supervision et d'ingestion des données, en contrôlant l'intégrité des données à la fin du processus

Environnement: Azure Data Lake Storage, Azure Data Factory, Azure SQL Data Warehouse, Spark, Sqoop

### INA (Institut National de l'Audiovisuel) – Data Engineer (06/2017 – 11/2017)

*Projet : catégorisation et de recommandation (analyse prédictive)*

- Mise en place d'un Cluster HDP Hortonworks ( 6 nœuds )
- Développement des traitements Talend Spark pour migrer et préparer des données
- Catégorisation d'un corpus de vidéo (Machine Learning)
- Mise en place d'un algorithme de recommandation personnalisée en fonction de l'historique de navigation sur le site de INA MEDIAPRO (Machine Learning)
- Classification des tours de paroles dans les émissions (Machine learning)

Environnement: Cluster HDP, Talend for Big Data, Spark, Machine Learning, Dataiku, Machine learning ( Kmeans, logisticRegression, RandomForest, DecisionTree) .

### INA (Institut National de l'Audiovisuel) – Data Engineer (12/2016 – 06/2017)

*Projet : Migration des données vers un DataLake*

- Mise en place des modèles logique transverse des données
- Analyse des données avec l'outil dataiku
- Stockage des données dans le datalake

Environnement: MariaDB, Sqoop, ElasticSearch, MongoDB, talend DI, dataiku

### Enedis– Data Engineer (11/2016 – 12/2016)

*Projet : POC migration des données de Teradata vers un DataLake*

- Mise en place d'un cluster Hortonworks POC ( 3 nœuds )
- Etude de la faisabilité d'une migration des données de Teradata vers Hdfs
- Orchestration de la Migration
- Etude de la sécurité des données dans un datalake

Environnement: Hortonworks, Teradata, Sqoop, Oozie, Knox, Ranger, ranger-audits(Solr)

### AMF (Autorité des Marchés Financiers – Data Engineer (09/2016 – 11/2016)

*Projet : Mise en place d'un datalake pour l'AMF*

- Mise en place d'un datalake pour le stockage et l'analyse de toutes les transactions de l'AMF
- Contrôle des transactions et génération des alertes
- Etude de l'historique des transactions

Environnement: MapR, MapR-FS, MapR-DB, MongoDB, Drill, Spark, Talend, Hue, Python

# INTERVENANT : H.M.

## Expériences

---

**INA (Institut National de l'Audiovisuel ) – Data Engineer (08/2015 – 09/2016)**

*Projet : Mise en place d'un DataLake (personnalisé) pour l'INA*

- Mise en place d'une couche de stockage
  - Installer MongoDB (Sharding, replica)
  - Installer d'ElasticSearch (Cluster)
  - Mise en place d'openLink Virtuoso
  - Sauvegarde et restauration de MongoDB et ElasticSearch
  - Mise en place des interfaces d'administration de MongoDB
  - Etude multi-site dataLake (Actif/passif)
- Mise en place d'une couche de traitement
  - Talend DI et Talend for Big Data (Spark)
  - Développement Job Talend pour charger des documents dans MongoDB, ElasticSearch et openLink Virtuoso
  - Tests de charge MongoDB et ElasticSearch avec YCSB (Yahoo Cloud Service Benchmark)
- Mise en place d'une couche de supervision
  - Développement Java : un contrôleur (aiguilleur) point d'entrée du lake.
  - Développement Java : un ordonnanceur (Quartz) pour ordonnanceur les différents traitements (synchrone et asynchrone).
  - Automatisation des tests d'intégration et de déploiement continue
  - Sécurisation du Brokers (Message Oriented Middleware)
  - Création des dashboards Kibana pour la supervision du datalake.

Environnement: MongoDB, ElasticSearch, OpenLink Virtuoso, LogStash, Kibana, Talend, Spark, Java, ActiveMQ, Sparql, Jenkins, SoapUi, Nexus, Chef, Yscb, Jmeter,