

Contact : Halima SOUARES
Mail : halima.souares@sunlog.fr
Mobile : 07 69 95 16 18

DOSSIER DE COMPETENCES

G.M
Développeur / Lead Big Data

SYNTHESE GENERALE DU PROFIL

- Product Owner
- Lead Dev
- Développement pipeline Data
- Architecture Logicielle/Technique
- Architecture d'entreprise
- Traitement de données
- Capacités d'analyse
- Autonomie
- Capacités rédactionnelles
- Formation interne
- Études / Ingénierie
- Test et intégration
- Gestion de projet
- Recueil de besoins
- Peer to peer code reviewing
- Mise en place de tableau de bord pour le Monitoring
- Coordination et animation d'ateliers/réunions
- Gestion de projet Agile

Compétences techniques :

- Hadoop
- Spark
- Python, SQL
- Machine Learning : Sklearn, PyTorch
- Notebook : Jupyter, Zeppelin
- Hive, Elasticsearch, MongoDB
- Git : Gitlab, Github
- Puppet, Ansible
- Kubernetes
- OVH, AWS (S3, EC2, EMR)
- JIRA

Secteurs d'activité :

- Bancaire - Finance
- Énergie
- Retail

Langues :

Anglais (courant)

FORMATIONS

Formation initiale :

2016 : Institut AGROPARITECH – Université Paris-Dauphine Ingénierie Big Data et Data Science

2015 : EFREI _ Ingénierie Business Intelligence

2012 : Université PARIS-EST MARNE LA VALLÉE _ Licence Mathématiques informatiques

Contact : Halima SOUARES

Mail : halima.souares@sunlog.fr

Mobile : 07 69 95 16 18

SUNLOG - RCS Nanterre Siege : 51, Avenue du Maréchal Joffre- 9200 Nanterre

EXPERIENCES PROFESSIONNELLES

* L'ensemble des données sont des données fictives et ne doivent pas être utilisées à des fins commerciales ou financières. Les données sont destinées à l'usage exclusif de SUNLOG GROUP et ne doivent pas être divulguées à des tiers sans la permission écrite de SUNLOG GROUP.

Poste : Lead Dev python Big Data reporting

Date : Novembre 2020 – Aujourd'hui

Client : Ingénico

Contexte : Service Shared Data Services Développements python dans un environnement Datalake MapR Hadoop pour le reporting des données de transactions sur les terminaux de paiements (1 à 50 Millions d'enregistrement par jour), en suivant la méthodologie agile.

- Amélioration de la pipeline d'ingestion de données (enrichissement des métriques, optimisation du stockage et des requêtes des données)
- Tests de charge.
- Rédaction de documentation. Fonctionnelles, Présentations techniques.
- Mise en place de rapports d'analyse des données collectées générés automatiquement.
- Participation au maintien du bon fonctionnement du code sur les différents déploiements (monitoring, analyse de dysfonctionnements, analyse des performances)
- Support technique

Environnement technique : MapR Hadoop Cluster, Linux ; Outils de déploiement de la plateforme : GitLab CI, Python, Grafana, Git, puppet

Poste : Lead PoC Big Data, Data Architect, Pre-Sales Engineer

Date : Février 2020 – Octobre 2020

Client : Nucleon-Data

Contexte : Rédaction de l'offre de services data, participation aux réunions d'avant-vente, réalisation de PoC sur de nouvelles technologies.

Missions :

- Encadrement technique et support des développeurs
- Mise en place de cluster Hadoop Standalone pour tester les nouveaux Frameworks
- Rédaction de notes techniques.
- Mise au point d'images d'OS Linux pour les développeurs
- Veille technologique (Delta Lake, Semarchy...)

Environnement technique : Python, pyspark, Delta Lake, Linux, OVH Cloud, API REST, Debian, Git

Poste : Product Owner ; Leader technique DATA

Date : Avril 2017 – Janvier 2020

Client : Société Générale

Contexte :

Dans le cadre du programme de transformation digitale du groupe Société Générale, j'ai contribué d'abord en tant que chargé de développements digitaux et finances auprès de la direction du contrôle permanent Financier du groupe, puis en tant que Product owner de la plateforme data du contrôle permanent auprès de la direction générale du programme de la transformation digitale et finalement en tant que leader technique en data Science et data engineering dans le cadre de l'implémentation du Datalab de la direction financière du groupe

Missions :

- PySpark : Création de job de traitement et de préparation des données
- Configuration générique Spark pour l'équipe du Datalab
- Création d'environnement standard pour les projets du DataLab
- Création de projet DataScience avec architecture Kedro
- Suivi des projets datascience en respectant la méthodologie CRISP

En tant que Product Owner de l'application de monitoring et d'exécution du contrôle permanent de la banque au sein du programme PCT "Permanent Control Transformation" du groupe Société Générale. Cette application est utilisée par les directions des finances, des risques et de la conformité par tous les départements de contrôle permanent. Le périmètre comprend toutes les filiales du groupe Société Générale. Cette application a été développée en suivant la méthodologie agile.

Les activités pour assurer le rôle de Product Owner :

- Participation aux ateliers de sprint design
- Planification de la roadmap de développement de l'application
- Contribution à l'élaboration de l'architecture fonctionnelle de l'application
- Analyse et synthèse des documents réglementaires (Recommandations de la BCE) pour l'alimentation du cadrage fonctionnel
- Élaboration des spécifications fonctionnelles avec toutes les équipes de contrôle permanent du monde.
- Supervision et suivi des développements en collaboration avec la Scrum Team.
- Suivi des livraisons et réalisation des démonstrations
- Pilotage de la performance Scrum
- Organisation des ateliers de co-working avec l'ensemble des proxy product owner et les contrôleurs des filiales pour affiner les besoins et alimenter la backlog de la Scrum Team
- Participation au comité de projet pour défendre les avancées de développement et la pertinence de l'application
- Intégration de l'application dans l'écosystème technique du programme PCT "Permanent Control Transformation"

En tant que Data Scientist Chargé de projet digitaux au sein de la direction du contrôle permanent financier du groupe Société Générale j'étais :

- En charge de la cellule Digitale de l'équipe Monitoring
- En charge de l'automatisation et rationalisation des processus de contrôle permanent au sein de la Direction Financière
- En charge de l'accompagnement des chantiers d'architecture d'entreprise et d'architecture SI dans le cadre du programme de transformation du contrôle permanent
- En charge du Développement des outils tactiques et des programmes impératifs pour l'analyse de Data et pour l'automatisation de différents traitements de données (Consolidation, Rapprochement, Reporting)
- En charge de l'élaboration de la structure des données et construction de la base de données des plans annuels du contrôle permanent
- Contributeur au collège DataScience du programme
- En charge d'études et identifications de useCases pour des PoCs en DataScience/ Big Data
- Contributeur en tant que référent métier aux séminaires des évolutions des process métiers

Environnement technique : Linux (Debian, Ubuntu), HDP, Python, Spark, Kedro, JIRA, Elasticsearch, Parquet, pandas, Dask, JSON, GitLab, Shell, Github, Spotfire, VBA, EXCEL

Poste : Ingénieur Développement Big Data**Date : Mars 2016 – avril 2017****Client : Ministère de l'économie et des finances****Contexte :**

Consultant pour GFI client Ministère des finances.

Développement de moniteur de stress sur le Cluster hadoop du service Tracfin.

Missions :

- Cadrage des besoins
- Développement en Java de générateur de csv aléatoire
- Conception et développement de parseur de logs génériques
- Configuration et stockage dans ElasticSearch
- Développement de traitements et d'analyse des logs
- Création de dashboard de visualisation de KPI de performance à l'aide de Kibana

Environnement technique : Linux, Java, Git, Hadoop**Poste : Ingénieur Business Intelligence****Date : Juillet 2015 – Décembre 2015****Client : Enedis****Contexte :** Direction du développement et des concessions en ile de France,**Missions :**

- Traitement de données en Hive, Pig et en SQL
- Manipulations des données patrimoines.
- Traitement de fichiers et de mails avec pièces jointes.
- Test de traitement sur la colonne C5.
- Documentation
- Collecte du besoin
- Edition du cahier des charges et des spécifications fonctionnelles.
- Modélisation et élaboration de la structure de la base de données mère.
- Choix de la solution technologique.
- Extraction des données des SI métiers de l'entreprise (SAP).
- Transformation et mise en forme des données.
- Conception des manipulations sur les données (SQL, VBA).
- Contrôle et validation des données.
- Contrôle et test des capacités machines clientes.
- Mise en place de la connexion réseau entre les bases de données.

Environnement technique : Hadoop, Python, Linux, Shell, Hive, Pig, MS ACCESS, SQL, SAP, VBA