



Sana KHIARI
Consultante BIG DATA
3 ans d'esperienze

Compétences

Techniques

- Big Data: Apache Hadoop, Apache OOOIE,Apache Hive, HBase, Spark, MongoDB et Redis
- Qualité logicielle : SOLID, Design patterns, Jenkins, Maven, Firebase, Github, Selenium et Docker
- Base de données : MySQL (certifié Microsoft Technology Associate in Database)
- Langage de programmation : Python, R, Nodejs, Angular4, Java, C# et Shell
- Système d'exploitation : Windows, Linux (certifié LPIC 101)
- Méthodologies : Agile (SCRUM).
- Architectures : REST, SOA

Fonctionnelles

- Accompagnement des experts métier dans la conception et leur expression du besoin fonctionnel
- Traduction des règles de gestion fonctionnelle en règles techniques
- Analyse des anomalies des règles métier des données EDF de ressources humaines
- Rédaction des spécifications techniques détaillées des demandes clients
- Réalisation de tests unitaires et d'intégration sur l'environnement de recette et production et la rédaction des fiches de tests

Expériences professionnelles

Novembre 2020 à Aujourd'hui

Contexte	FEATWAY – mission chez EDF
Rôle	Pilote du projet Big Data des flux sortants d'un Datalake RH
Mission	<ul style="list-style-type: none"> Assurer les échanges avec le client EDF au travers d'instances de suivi du chantier. Réaliser la planification mensuelle des tâches et le suivi des livraisons des membres de l'équipe. Piloter les mises en production pour chaque lot de livraison et le suivi des incidents de production. Créer de nouveaux flux sur le projet intitulé « plateforme flux d'échange ». Le projet est un ensemble de flux sortants d'un Datalake de données RH vers les systèmes d'information et les web services externes à EDF. Réaliser des REX sur les incidents en production. Réaliser des optimisations de code avec Hive/Spark. Accompagner les nouveaux membres de l'équipe dans leurs intégrations et la montée en compétence sur les technologies BIG DATA. Mener des ateliers de conception avec les experts métiers et big data de EDF afin de trouver des solutions aux problèmes survenues notamment l'optimisation des traitements sur HADOOP.
Technologies	OOZIE, HIVE, GPG, OpenSSL, SFTP, SPARK, Excel

De Janvier 2020 à Octobre 2020

Contexte	FEATWAY – mission chez EDF
Rôle	Consultante Big Data sur le projet des flux sortants d'un Datalake RH
Mission	<ul style="list-style-type: none"> Concevoir et développer les briques d'un Datalake Big Data. Les solutions développées portent bien sur les données entrantes dans le Datalake ainsi que les données sortantes. Stocker les données partitionnées avec des tables Hive. Réaliser des scripts Shell de chiffrement des données sortantes du Datalake avec le protocole GPG ou OpenSSL. Nettoyer les données avec Spark. Déchiffrer les données et leurs insertions dans des vues Hive. Créer un étage de données spécifiques aux flux sortants : une base de données contenant seulement des vues. Réaliser des ordonnancements de workflow avec OOZIE. Créer deux coordinateurs OOZIE par flux sortant : un coordinateur qui assure la transformation d'une vue en un fichier chiffré sur HDFS et un

	<p>coordinateur qui assure l'envoi sécurisé de ce fichier avec SFTP vers un système d'information externe.</p> <ul style="list-style-type: none"> • Réaliser la spécification technique de chaque flux sortant. • Réaliser la gestion du code sur GIT et les MERGE avant chaque mise en production. • Mise en production des évolutions de code sur l'environnement de production. • Assurer le suivi des coordinateurs lancés en production avec l'interface OOZIE. • Assurer la gestion DEVOPS (intégration continu) avec Jenkins depuis GITLAB sur l'environnement de production, recette et intégration.
Technologies	Jenkins, Hive, Spark, HBase, Hadoop, Mobaxterm, visual studio code, gitlab, scoop

De Mars 2019 à Janvier 2020

Contexte	FEATWAY – mission chez EDF
Rôle	Consultante Big Data sur le projet des flux entrants d'un Datalake RH
Mission	<ul style="list-style-type: none"> • Concevoir et développer des ordonnanceurs OOZIE servant à l'intégration des données envoyées par les clients externes à EDF dans le Datalake. • Gérer et optimiser le stockage des données dans des tables Hive partitionnées. • Réaliser des scénarios de retour arrière automatisés des données dans des dossiers HDFS en cas de panne ou perte. • Stocker les données dans des tables HBASE qui servent au web services de EDF. • Créer deux étages (deux bases de données Hive) pour stocker les différents types de données : étage 0 pour stocker les données brutes et étage 1 pour stocker les données transformées. • Réaliser des tests unitaires et d'intégration sur l'environnement d'intégration et recette. • Mise en production des évolutions de code sur l'environnement de production.
Technologies	Jenkins, Hive, Spark, HBase, Hadoop, Mobaxterm, visual studio code, gitlab, scoop

Février 2018 à Septembre 2018

Contexte	KONVERGENCE Business & technologies
Rôle	Stagiaire Data
Mission	<ul style="list-style-type: none"> • Implémenter une API de données avec le protocole oData • Exposer les données dans une interface DATAVIZ en utilisant D3js et javascript. • Implémenter une API se basant sur une architecture de micro-service et les technologies Spring Boot, Spring Cloud, Netflix service de découverte eureka, kafka et redis. • Stocker les données dans un conteneur Docker géré par Rancher. • Connecter l'API de données aux logiciels de BI comme PowerBI et Tableau pour rendre les données disponibles en temps réel pour les clients. Les données sont de types financiers.
Technologies	PowerBI, Tableau, Spring Boot, Spring Cloud, Kafka, Redis et Docker

Juillet 2017 à Aout 2017

Contexte	SMART INNOVATION TECHNOLOGY STARTUP
Rôle	Stagiaire en science de données
Mission	<ul style="list-style-type: none"> • Développer un agrégateur et analyse de données des offres de travail (offres sur linkedin par exemple) • Réaliser l'extraction des offres de travail (Web Scraping) avec Selenium en Python. • Assurer le nettoyage avec Python et standardisation avec l'algorithme de Levenshtein. • Gérer le stockage des données dans Redis puis MongoDB. • Réaliser l'exploration des données avec R. • Réaliser un moteur de traitement et la REST API avec loopback en nodejs. • Réaliser une interface Dataviz avec des graphes avec les statistiques des tendances technologiques avec Angular 4 et D3js.
Technologies	Python, MongoDB, Nodejs, Angular 4 et D3js

Décembre 2016

Contexte	Projet personnel de création d'une plateforme d'intégration continue MEANSTACK
Rôle	Développeur Big Data
Mission	<ul style="list-style-type: none"> • Combiner les outils OpenSource pour faciliter le processus d'intégration continue pour les développeurs javascript, nodejs et Angular. • Utiliser Git pour le versionning du code comme base de début. • Valider la correction syntaxique du code avec l'outil de build Travis CI. • Tester le code avec les outils de test Mocha et CHAI. • Ajouter les outils GULP, JSHINT et JSLINT pour garantir la qualité du code. • Implémenter le déploiement du code avec Heroku et Firebase.
Technologies	Git, Travis CI, Mocha, CHAI, GULP, JSHINT, JSLINT, Heroku et Firebase

Avril 2016

Contexte	Projet au sein d'une compétition à l'université de New york - création d'un POC de plateforme pour les réfugiés
Rôle	Développeur Angular
Mission	<p>Cette plateforme est conçue afin de réunir les réfugiés en Seattle, USA souhaitant compenser leurs années d'éducation perdues on les mettant en contact avec les étudiants ou professeurs les plus proches se portant volontaires pour les enseigner :</p> <ul style="list-style-type: none"> • Développer l'application avec Angular. • Stocker les données dans Firebase. • Assurer le contact entre les volontaires les plus proches en utilisant Google Maps Api.
Technologies	Angular, Firebase, Google Maps Api

Novembre 2016

Contexte	Projet personnel : POC d'une montre intelligente - IoT et fouille de données
Rôle	Développeur IA
Mission	<ul style="list-style-type: none"> • Créer une montre intelligente pour les personnes âgées détectant leurs malaise ou insécurité avec l'outil IBM WATSON. • Réaliser la simulation des capteurs avec l'outil BLUEMIX et les outils Big Data (SparkR + MongoDB). • Implémenter des algorithmes d'apprentissage non supervise avec Random forest et SVM.
Technologies	IBM WATSON, BLUEMIX, SparkR, MongoDB, Random forest, SVM

Septembre 2016

Contexte	Projet personnel : Réseau social pour les sportifs
Rôle	Développeur Nodejs/Big Data
Mission	<ul style="list-style-type: none"> • Réaliser un projet dans le contexte Big Data : une application mobile pour les amoureux de tous types de sport. • Implémenter l'application mobile avec ReactNative et Nodejs. • Stocker les données dans MongoDB sur HADOOP. • Traiter les données avec Spark et python. • Réaliser l'application avec une architecture REST.
Technologies	ReactNative, Nodejs, MongoDB, Spark, Hadoop HDFS, Elastic Search, python, REST.

Diplômes et Formations

Année	Libellé
De 2013 à 2018	Ingénieur en Génie Logiciel Spécialité management des systèmes d'information