

Curriculum Vitae

Hafsa Emira

INGENIEUR BIG DATA

COMPETENCES

Compétences

Techniques

- BigData : Hadoop (MapReduce, Yarn, HDFS), Hive, Hbase, Kafka, Spark, Sqoop.
- Langages : Java, Scala.
- Bases de données : MySQL.
- Développement web : HTML, Spring MVC, Angularjs.
- Indexation de données : Elasticsearch.
- Intégration continue : Maven, Git.

Formation

- Méthodologie Agile SCRUM

Septembre 2020 à Aujourd'hui

Contexte	BFM 3A – mission chez EDF
Rôle	Ingénieur BIG DATA – Projet MyHr
Mission	<p><u>Contexte:</u></p> <p>Mise en place de la chaine myhr qui permet de traiter les données Ressources Humaines à forte volumétrie envoyés par Talent Soft sous format fichiers CSV et les stocker dans des tables historiques Hive.</p> <p><u>Réalisations :</u></p> <ul style="list-style-type: none">- Collecte des données : Réception des données brutes dans des fichiers CSV envoyés par Talent Soft.- Traitement des données : Développement de jobs batch par Spark et scala pour traiter et nettoyer les fichiers reçus : supprimer les doublons et les nulls, définir les séparateurs de colonnes, définir une structure de donnée figée des fichiers, traiter les différents formats des données et générer des logs.- Préparation des données : Chargement des données stockées dans des fichiers csv dans des tables externes dans Hive.- Ingestion des données : stockage des données dans des tables de préparation Hive afin d'effectuer les différentes jointures pour finalement les stocker dans des tables finales.- Implémentation des workflows de récupération et d'analyse de données.- Configuration et développement de Coordinateurs/Workflows Oozie pour chaque action et pour les lancements des jobs Spark.- Mise en place des scripts Shell pour automatiser les tests unitaires.
Technologies	Spark, scala, Hive, Oozie, Hadoop, Git, Jenkins, nexus.

DU Décembre 2018 AU Mars 2020

Contexte	Lansrod – mission chez EDF
Rôle	Ingénieur BIG DATA – Projet Hypster
Mission	<p><u>Contexte:</u></p> <p>Migration d’une plateforme de monitoring de l’activité des téléconseillers sur une architecture temps réel basée sur Kafka, Spark et Elasticsearch. L’implémentation repose sur des jobs Spark Streaming pour l’ingestion des messages Kafka et sur Kibana comme Web UI pour l’interface utilisateurs. Les données proviennent de l’opérateur télécom SFR sur plusieurs topics Kafka selon la maille d’analyse et d’autres topics Kafka sont alimentés en interne chez EDF et concerne les planning et les objectifs des téléconseillers. Finalement, une base à froid Oracle sert de référentiel utilisé pour enrichir et/ou valider les éléments provenant de Kafka.</p> <p><u>Réalisations :</u></p> <ul style="list-style-type: none"> - Développement d’un module Core servant comme package commun pour tous les jobs du projet, le module contient les mécaniques de configuration des jobs, l’accès au référentiel Oracle (mis à disposition sur HDFS) ainsi que différents connecteurs. - Développement de jobs Streaming pour la consommation des topics Kafka et alimentation des index Elastic, un projet est créé par job afin de séparer les processus de déploiement et de lancement. - Configuration et développement de Coordinateurs/Workflows Oozie pour chaque job . - Développement d’un batch journalier qui copie les tables référentiel de la base Oracle vers HDFS. - Mise en place de plusieurs Scripts dans le module Core (création des index, Script de submit génériques pour le lancement des jobs Spark) . - Création de plusieurs Scriptes Fields pour les indicateur Kibana . <p>Assistance de l’équipe MCO dans le déploiement des jobs via Jenkins et Ansible.</p>
Technologies	Kafka, HDFS, Spark, Elasticsearch, java, Kibana, gitlab, Jenkins, nexus, shell, Oracle, Kerberos.

Contexte	Sopra HR Software
Rôle	Ingénieure - Java
Mission	<p><u>Contexte:</u></p> <p>migration d'une partie d'une solution RH complète parfaitement adaptée aux besoins des Directions des Ressources Humaines du langage cobol au langage java. L'implémentation repose sur une api java qui lit des données provenant des fichiers XML puis affichées dans des fichiers excels grâce à la bibliothèque Apach POI.</p> <p><u>Réalisations:</u></p> <p>Participer aux différentes phases de conception, personnalisation, développement et validation des solutions. Elaborer des cahiers de tests et des guides utilisateurs. Stabiliser les environnements de développements.</p> <ul style="list-style-type: none"> - Préparation des données : lire des données stockées à partir des fichiers XML. - Nettoyage des données : Suppression des doublons et des valeurs nulles. - Analyse des données. - Chargement des données dans des fichiers EXCEL. - Installation des Kits sur les environnements QA. - Création et installation des Kits et des patchs sur les environnements de production des clients. - Maintenance corrective, évolutive.
Technologies	Design center, SQL, java7, Apach poi, Filezilla, Putty.

Février 2017 à Juillet 2017

Contexte	Advanced Directive Solution - ADS
Rôle	Ingénieure – Java JEE - Stage de fin d'études
Mission	<p><u>Contexte:</u></p> <p>Conception et développement d'un système de contrôle interne de l'entreprise ADS qui permet de faire interagir les employés, de faire circuler l'information, de suivre et de contrôler les processus métiers de l'entreprise.</p>
Technologies	StarUml, Eclipse, Java7, Spring MVC, Spring Data, Spring JPA, Sprint Security, Web Service REST, Workflow, Camunda Modeler.

Diplômes et Formations

Année	Libellé
De 2014 à 2017	Diplôme d'Ingénieur en Informatique
2012	Baccalaureate scientifique