

Consultant Big Data

SYNTHESE DE COMPETENCES

❖ Savoir-faire :

- Recueil du besoin métier / Spécification des besoins utilisateurs
- Validation et Organisation avec les équipes métiers
- Maîtrise des concepts décisionnels et de la modélisation de bases de données
- Maîtrise des architectures distribuées et des infrastructures Cloud
- Méthodologie de gestion de projet agile
- Management d'une équipe de data engineers
- Mise en œuvre et déploiement de plateforme Data Analytics type Big Data.
- Proximité avec les équipes Datascientist et Business
- Tech Lead et chef de projet.
- Mise en place des différentes briques d'architectures basées sur un ETL et outil de restitution.
- Passionné de data et de machine learning : formation universitaire, python et Dataiku.
- Bonne connaissance de Google Cloud PlatForm et d'Amazon AWS.
- Appétences pour appréhender des nouvelles directions métiers.
- Secteurs/Métiers : Industrie, Voyageur, Crédit, Media, Retail, CRM, Marketing.

❖ Savoir être :

- Capacité de rédaction et à fédérer les forces vives
- Bon relationnel et sens de la communication
- Curiosité, Réactivité, initiative
- Méthode, rigueur, autonomie

❖ Compétences techniques :

- **BigData** : Hadoop, MapR (FS, DB), MapReduce, Spark, Hive, Hbase, hdfs, Hue, BigQuery.
- **Bases de données** : NETEZZA, MySQL, SQL Server, Oracle, RedShift, BigQuery.
- **Outils Décisionnels** : Talend Enterprise Data Integration, Talend Enterprise Big Data, Talend Administration Center, Tableau, R, Google Cloud DataLab.
- **Google Cloud Platform** : BigQuery, Google Cloud Storage, Cloud DataFlow, Datalab.
- **Systèmes** : Linux, Mac OS, Windows
- **Outil de développement** : Eclipse, svn, ant, maven, git, jenkins, trac, mantis, redmain.
- **Méthodologies** : UML, méthode agile : Scrum
- **Data Sciences** : Dataiku, Python, Modèle prédictif, Classification, Clustering, Statistiques.



EXPERIENCES PROFESSIONNELLES

Depuis Avril 2017 : SEPHORA

Chef de Projet Big Data

Projet : Vision Client 360°

Descriptif mission : Déploiement d'une plateforme BigData : Datalake.

- ❖ Accompagnement à l'organisation des données dans le DataLake
 - Aide à la mise en œuvre de la méthode SCRUM pour la gestion du projet datalake.
 - Animation des ateliers métiers et recueil du besoin
 - Rédaction des users stories ou des spécifications fonctionnelles selon les directions métiers
 - Construction du backlog avec l'équipe de développement
 - Conception et réalisation de différents traitements correspondant à la construction des couches de données (cleansed data, référentiels data, cosssdata et reporting data).
 - Conceptions et réalisation d'objets métiers (couche datable).
 - Conception et réalisation de différentes dataviews pour différents projets BI.
 - Arbitrage sur les choix de conception et de modélisation des KPI
 - Conception et réalisation des flux d'alimentation du datalake depuis les différentes sources (couche rawdata).
 - Conception et réalisation des flux temps réel Saprk-Streaming : Consumers et Producers.
 - Pilotage, encadrement et formations des consultants BigData/Talend/Dataiku.
 - Accompagnement à la recette et la mise en production des briques applicatives.
 - Accompagnement de la TMA et de l'équipe exploitation.
 - Rédaction de compte-rendu de comités et des cérémoniaux.
- ❖ Référent Dataïku et Accompagnement des DataScientists
 - Référent technique sur l'outil Dataïku
 - Pilotage du projet industrialisation des flux/jobs Dataïku
 - Accompagnement des data-scientistes dans la réalisation/optimisation sur différents projets data
 - Préparation des données pour les DataScientists à travers les Objets Métiers et des tables cross-data.
 - Réalisation des documents de conceptions techniques, de livraisons et du dossier techniques.

Environnement technique: Dataïku, Python, Jupiter, Scala, Talend Real-time Big Data Platform 6.2.1/6.4.1, Platform BigData Mapr : Mapr-Stream, Spark, Spark Streaming, Spark SQL, Drill, Hive, MaprDB, Parquet, MCS, Hue. Java, JMS, Linux, Control-M. Jira, Scrum.

Environnement fonctionnel : Retail



Janvier 2016 – Mars 2017 : VEOLIA EAU D'ILE DE FRANCE (VEDIF)

Tech Lead Data

Descriptif mission :

- ❖ **Projet Datalake VEDIF** : POC & Industrialisation du Datalake VEDIF sur la plateforme BI Google BigQuery.
 - Animation des ateliers techniques et fonctionnels
 - Identification des ruptures technologiques et proposition d'architecture
 - Réalisation de benchmark sur l'extraction des données à partir des bases Oracle et leur chargement dans BigQuery
 - Test d'un ensemble de requêtes complexes dans BigQuery.
 - Tunning de la solution d'extraction afin de mieux répondre au besoin
 - Test des différentes solutions d'intégration Talend, java-API Google dans BigQuery.
 - Choix de la solution ODI pour l'extraction et java-API Google pour l'intégration dans BigQuery.
 - Implémentation des solutions et tests de requêtes complexes sur plusieurs tables.
 - Validation de l'architecture et des solutions cibles avant industrialisation de la plateforme Big Data
 - Déploiement des chaînes d'extraction, intégration des données de Smart Metering
 - Mise en place de la chaîne des calculs et d'extraction/export à partir du Datalake.
 - Déploiement de la chaîne Aide au Diagnostic du Réseau (ADR)
 - Aide à la mise en production des chaînes sous l'orchestrateur Control-M.
 - Rédaction de la documentation technique
 - Formation et transfert de compétences

- ❖ **Déploiement d'une solution BI Google BigQuery** pour l'exploitation des données de Smart Metering avec le calcul des débits mensuelles.
 - Développement complète de la solution de l'extraction des données de mesures.
 - Déploiement du projet d'intégration des flux extraits : index, consommations et patrimoine dans le BigQuery.
 - Déploiement de la chaîne de calcul et d'export des calculs des débits mensuelles.
 - Aide à la mise en production sous l'ordonnanceur Control-M
 - Documentations.

- ❖ **Mise en place d'un nouveau Tag** dans la base de mesures (Big Data PI OSISOFT)
 - Ajout du nouveau Tag dans l'archive de PI OSISOFT.
 - Mise à jour de tous les scripts, job et web services pour prendre en charge le nouveau Tag.



- Déploiement du Tag dans les différents environnements
- Aide à la recette et la mise en production.
- ❖ POC sur l'extraction des données de mesures à partir d'un système de la base de mesures PI OSISOFT et leur chargement dans BigQuery.
 - Rechercher la solution pour l'extraction des données de mesures index et conso.
 - Implémentation d'une solution d'extraction des mesures.
 - Rechercher la solution d'intégration des données extraites dans la BigQuery de Google Cloud Platform.
 - Faire le choix de la solution et implémentation de la solution sur des exemples d'extraction.
 - Présentation des solutions.

Environnement technique : Google Cloud Platform : BigQuery, Google Storage, SQL Serveur, Oracle, PI OSISOFT, C#/SDK-SF, JAVA API Bigquery, GSON, ODI, ControlM, Scrum.

Environnement fonctionnel : Industrie.

Novembre – Décembre 2015 : SSII – Projet Interne

Expert Talend / Python

Descriptif mission : Data management Plateforme (DMP) pour exploiter toutes les données clients éparpillées sur les canaux digitaux - site Web et application mobiles.

- ❖ Optimisation du temps de traitement des logs de navigation des sites clients DMP
 - Analyse des chaînes de traitement des logs de navigation déjà en place.
 - Développement de nouveaux jobs de traitement parallèle des logs de navigation des différents comptes, par date et compte pour optimiser le temps de traitements de toute la chaîne de la data Factory.
 - Tests et mise en recette des nouveaux jobs.
- ❖ Qualité
 - Analyse du parcours de navigation des utilisateurs des sites clients, détection d'anomalie et correction au niveau des chaînes de traitements des logs.
- ❖ Application spécifique de mapping et de transformation de flux.
 - Développement en Python d'une application pour des transformations paramétrables des flux d'une API en format JSON vers un format CSV dont le contenu passe par un mapping avec tous types de transformations.

Environnement technique : Java, Talend 5.6.1 BIG DATA, Big Data cluster Hadoop, Hive, Hue, RedShift, Python, Mysql, S3, Google cloud Storage, Jenkins, Scrum

Environnement fonctionnel : e-marketing, ciblage comportemental.



Septembre – Octobre 2015 : OXYBUL

Expert Talend

Descriptif mission : Accompagnement à la mise en place des nouveau flux Talend

- Validation et recette des nouveau flux talend développés dans le cadre de la nouvelle version du site Okaidi avec une prise en charge multi-langue pour un même pays (La Belgique).
- Modification et correction ponctuelles sur les flux
- Mise en production de tous les flux Talend avec la mise en ligne du nouveau site Okaidi.
- Recommandation.

Environnement technique : Talend 5.6.1, Mysql, Magento, Linux, Job Scheduler.

Environnement fonctionnel : e-commerce.

Juillet – Août 2015 : MEDIAPRISM (Groupe La Poste)

Expert Talend

Descriptif mission : Déploiement d'une plateforme BigData de gestion des contacts.

- ❖ Expertise Talend pour la mise en place des premiers flux Talend génériques et spécifiques d'intégration des fichiers des partenaires de Mediaprism
 - Expertise Technique Talend / Big-Data.
 - Conception et développement des différents flux Talend génériques et spécifiques d'intégration des fichiers partenaires.
 - Rédaction d'une documentation Technique et de recommandation.

Juillet 2015 : OXYBUL

Expert Talend

Descriptif mission : Expertise technique TALEND et formation

- ❖ Accompagnement à la mise en place de la version 2 des flux Talend entre le site e-commerce Okaidi et les différents systèmes du si.
 - Validation des évolutions faites par une compétence interne sur les flux Talend de la V1 avant la mise en recette/production
 - Initiation de la version 2 des différents Flux Talend à travers le paramétrage des jobs via une base de données technique synchronisée avec des données de la base magenta du site e-commerce.
 - Développement du nouveau flux des ventes dans le cadre de la V2.
 - Documentation de recommandations pour la suite des développements.

Environnement technique : Talend 5.6.1, Mysql, Magento, Linux, Job Scheduler.

Environnement fonctionnel : e-commerce.



Mai – Juin 2015 : PRET D'UNION

Expertise Talend - Consultant Décisionnel

Descriptif mission : Mise en place d'une plateforme technique d'intégration des données autour des clients/prospects dans la Data Management Platform (DMP). Campagnes de ciblage comportemental.

- Analyse des spécifications fonctionnelles,
- Analyse des différentes sources de données et conception de la base cible.
- Développement des flux des données entrants Prêt d'Union vers une base RedShift.
- Développement des flux des données entrants API Eulerian vers une base RedShift.
- Développement des flux des données entrants MailPerformance vers une base RedShift.
- Développement des flux sortants de la base RedShift vers la DMP et MailPerformance,
- Rédaction de Documentations techniques,
- Mise en production des différents jobs sur une machine EC2 d'Amazon

Environnement technique : Talend 5.6.1, Amazon Redshift, S3 et AWSCLI.

Environnement fonctionnel : Crédit entre particuliers

Février – Avril 2015 : BAYARD PRESSE

Expertise Talend - Consultant Décisionnel

Descriptif mission : Mise en place d'un Datamart Marketing sur un serveur Redshift.

- Analyse des différentes sources de données et modélisation du schéma cible
- Rédaction des spécifications techniques.
- Développement du flux de données : base de production vers le datamart avec pivotement du modèle EAV vers des champs en colonnes.
- Développement du flux : ancienne base Marketing vers le datamart.
- Développement du flux retour Néolane vers le datamart.
- Rédaction d'une documentation technique : cahier d'exploitation des flux Talend.
- Accompagnement à la mise en production des flux sur une machine EC2 d'Amazon.

Environnement technique : Talend 5.6.1, Oracle SGBDR, Amazon Redshift, S3 et AWSCLI..

Environnement fonctionnel : Crédit entre Particuliers.



Novembre – Décembre 2014 : OXYBUL

Expertise Talend

Descriptif mission : Expertise technique TALEND et formation

- ❖ Refactorisation et optimisation de tous les flux Talend entre le site e-commerce Okaidi et les différents systèmes du SI.
 - Encadrement et formation d'une ressource interne à Talend Data Intégration.
 - Analyse de l'ensemble des séquenceurs Talend existants V1,
 - Mise en œuvre des correctifs/optimisations nécessaires dans le respect des best practices.
 - Refactorisation de tous les séquenceurs Talend (Ventes, Catalogue, Client, Facturation, Commande, Expédition, Stocks et Avoirs),
 - Accompagnement à la recette,
 - Mise en place des logs sur l'ensemble du projet.
 - Rédaction d'une documentation technique.
- ❖ Formation d'une compétence interne à Talend DI.

Environnement technique : Talend 5.6.1, Mysql, Magento, Linux, Job Scheduler.

Environnement fonctionnel : e-commerce.

Septembre – Octobre 2014 – DARTY

Expertise Talend Big Data

Descriptif mission : Expertise technique TALEND et formation Projet BIG DATA Wynn : Système de visualisation des ventes de tous les canaux (internet, magasins et ventes par téléphone).

- ❖ Passage en lignes de ventes et intégration des ventes et des entrées magasins dans Wynn.
 - Compréhension de l'architecture BIG DATA, du métier et du fonctionnel de Wynn.
 - Mise en place du passage en lignes de ventes.
 - Mise en place de la visualisation du nombre des entrées magasins avec la quantité des ventes à l'instant t de la journée.
 - Expertise Talend BIG DATA.
 - Conception et réalisation du job Talend permettant d'alimenter le serveur ElasticSearch. Ce dernier est directement requêté par l'application de visualisation nommée Wynn.
 - Conception et développement de flux d'alimentation.
 - Accompagnement à la recette.
 - Maintenance évolutive des flux Talend existants.
 - Documentation explicative des développements.

Environnement technique : Talend Big Data, Hadoop MAPR, Hive, Hbase, Elasticsearch, IBM MQ.

Environnement fonctionnel : Retail.



Mars à Juillet 2014 – GROUPE EXPRESSE ROULARTA : L'EXPRESS

Consultant Senior BI / CRM

Descriptif mission : Système d'intégration automatique des fichiers partenaires (contacts) et des retours de campagnes dans le CRM.

- ❖ Mise en œuvre d'un système permettant de collecter et d'intégrer les données des contacts avec la supervision des imports via une application Qlikview.
 - Expertise Talend, SugarCRM.
 - Spécifications techniques.
 - Conception de l'application et réalisation des différents jobs Talend.
 - Mise en place d'une base de suivi de l'intégration des fichiers, interrogée par une application Qlikview.
 - Mise en œuvre d'un système pour intégrer les retours campagne MailPerformance dans la base CRM.
 - Accompagnement à la recette et mise en production des jobs.
 - Documentations techniques, formation et passage de connaissances.
- ❖ TMA
 - Optimisation et harmonisation des anciens jobs Talend à travers un seul projet Talend avec externalisation du fichier de contexte.
 - Mise en place de nouvelles méthodes de web services permettant la création et la consultation des contacts/abonnements enregistrés dans CRM.

Environnement technique : Méthode Agile : Scrum, Kagilum, Mysql, Talend 5.4.2, SQL, PHP, shell, Confluence (wiki).

Environnement fonctionnel : Presses.

Mars 2012 à Février 2014 : PAGES JAUNES

Responsable Technique projet

Descriptif mission : Back Office User RelationShip Management: Infrastructure technique permettant de collecter et d'exploiter les données de navigation des utilisateurs des médias mobiles et fixes du Groupe Pages Jaunes.

- ❖ Back Office User RelationShip Management.
 - Spécifications techniques et estimation des charges des différentes briques applicatives.
 - Expertise Talend.



- Choix des solutions et de l'architecture technique
 - Conception des entrepôts de données.
 - Encadrement de deux développeurs Talend.
 - Conception et réalisation des différentes briques applicatives.
 - Accompagnement à la recette et mise en production des chaînes applicatives dans les délais impartis.
 - Présentations des solutions techniques, démonstrations et participation à l'animation des réunions.
 - Documentations, présentations, contrats d'interfaces, cahier d'exploitation des chaînes en production...
 - Exploitation des chaînes de production : Création de segments pour ciblage publicitaires, livraison des évolutions à la DSI, rattrapages et corrections des données...
 - MCE sur les différents projets.
 - Administration des différents projets sous TAC (Talend Administration Center).
 - Formation et passage de connaissances.
- ❖ Déploiement du Score de valeur et d'engagement (calcul automatique du scoring utilisateurs des sites pages jaunes fixe et mobiles).
- Conception et réalisation de la chaîne scoring.
 - Aide à la recette.
 - Mise en production
- ❖ Déploiement des référentiels utilisateurs pour la personnalisation du display sur le site pagesjaunes.fr.
- ❖ Backup du responsable technique du projet Yield Management (visibilité sur la vente des espaces publicitaires). Et aide à l'ordonnancement de la chaîne Yield Management et à sa mise en pré-production.
- ❖ Participation à l'évaluation de l'outil SPLUNK pour analyser les différents types de logs Pages Jaunes.
- ❖ Suivi du projet BIG DATA-KPI et aide à mise en place des KPIs pour juger de l'état de santé des services Pages Jaunes.

Environnement technique :

Netezza, Talend, JAVA, SVN, SQL, VTOM, Script shell, Cognos, Hadoop Cloudera, Splunk, Scrum, VDOC (outil de livraison), Confluence (wiki).

Environnement fonctionnel : Marketing.

Juin 2010 à Février 2012 : MARMARA (TUI France)**Chef de projet CRM et flux.****Descriptif mission :**

Déploiement de deux CRM avec automatisation des flux entrants et sortants.

❖ Mise en place des deux CRM B2C et B2B de Marmara.

- Recueil des besoins et spécifications fonctionnelles auprès des équipes métier.
- Encadrement des développements/customisation des deux CRM(SugarCRM).
- Automatisation de la mise à jour des données des CRM : agences, clients, ventes et prospects.
- Intégration des différents flux complétant les données clients/prospects et génération de flux sortants.
- Profilage et amélioration de la qualité des données des bases CRM,
- Accompagnement des équipes développements front pour la mise en place de la rubrique espace client sur le site Marmara.
- Accompagnement des développements des applications QlikView pour la partie rapports,
- Accompagnement des équipes Marketing et Commerciales.

Animation et présentation des applications à des utilisateurs (Direction commerciale et e-commerce, directions régionales, commerciaux,...).

Environnement technique : Talend Open Studio, Talend Open Profiler, Qlikview, SugarCRM, JAVA, SQL Server, SQL, procédures stockées, IIS, PHP, C#

Environnement fonctionnel : Tour opérateur (voyagiste)

Février à Mai 2010 : LEPAPE.COM**Consultant e-commerce et Décisionnel****Descriptif mission :** Développement du nouveau site e-commerce lepape.com

❖ Migration du catalogue de l'ancienne base lepape.com vers le nouveau site sous Magento.

Environnement technique : Talend, JAVA, Mysql, PHP

Environnement fonctionnel : e-commerce.



Septembre 2008 – Décembre 2009 : OPODO**Consultant JAVA/J2EE**

Descriptif mission : Refonte du site web Opodo.

- ❖ Développement des nouvelles applications de réservations des packages et des Hôtels.
- ❖ Refonte du site Opodo.

Environnement technique : Java : web services (Metro), hibernate, httpunit, Quartz, oscache,..., Struts2, JSP, servlet, XML/XSL, DWR, JQuery, Ajax JavaScript, MySQL.

Environnement fonctionnel : Agence de voyage en ligne (e-tourisme).

Août – Septembre 2008 – CAFE.COM**Consultant JAVA/J2EE**

Descriptif mission : Plateforme web de jeux multi-joueurs en ligne

- ❖ Développeur Java.

Environnement technique : Java-EE, Jetty, AJAX/COMET, MySQL

Environnement fonctionnel : Jeux en ligne.

Juillet 2008 – TVTRIP.COM**Consultant JAVA/J2EE**

Descriptif mission : Plateforme de recherche d'hôtels dans le monde avec vidéos dont le modèle de revenus est basée sur l'affiliation.

- ❖ Développement du nouveau site tvtrip.com.
- ❖ Intégration des données/flux des partenaires dans la base de données du site.
- ❖ Développement d'un back office pour la gestion des vidéos et photos.

Environnement technique : Java JEE: Hibernate, JAXB, Quartz, Log4j, Jtidy... XPath, JXL, Junit, apache cocoon, Web services. MySQL, XML, XSLT, XSP, JXTemplate, Flowscript, html, JQuery, Ajax, Shell UNIX

Environnement fonctionnel : Hôtellerie (e-tourisme).

FORMATION/ CERTIFICATIONSNELLES

- ❖ **2018 - Dataiku.** Préparation des données, conception des flux et machine learning.
- ❖ **2016 - Google Cloud Platform :** Produits Google, Import et export dans BigQuery, Streaming, Analyse des données (SQL Analytique) et optimisation des requêtes, Google Cloud Storage, gsutils, Interfaçage de BigQuery avec d'autres Outils, Analyse et visualisation via Google Cloud DataLab.



- ❖ **2015 - Big data** : Les Fondements du Big Data et son écosystème.
- ❖ **2014 - Hadoop Mapr** : Installation et administration d'un cluster Hadoop Mapr.
- ❖ **2014- Introduction aux statistiques avec R**
- ❖ **2006 – Master Professionnel Génie de l'Information et des Systèmes**
- ❖ **2005 – Maîtrise Génie de l'Information et des Systèmes, Spécialité traitement de l'information**

