



Expert Big Data

SYNTHESE DE COMPETENCES

Compétences fonctionnelles :

- ✚ **Big Data** : Maîtrise des enjeux de la volumétrie, la complexité des calculs et variété des données. Résolution de problèmes de performance et de montée en charge. Méthodologie de construction de data lake.
- ✚ **DataScience** : A la pointe des méthodes de machine Learning et d'indexation.
- ✚ **Développement et tests** : Maîtrise du cycle de vie de développement logiciel. Intégration continue. Synchronisation entre les différentes équipes et gestion de la dépendance modulaire. Client/serveur, 3Tiers.
- ✚ **Gestion de projets** : Architecture modulaire ou mise en place d'ESB pour le découplage et l'intégration d'applications hétérogènes, gestion des tâches, des attentes, des solutions, méthodes classiques et agiles. Conception et mise en œuvre rapide des techniques, démarche méthodologique.
- ✚ **Environnement** : Parfaite connaissance des systèmes d'exploitation.
- ✚ **Sécurité** : Maîtrise des solutions liées à l'authentification, la confidentialité et d'audit.
- ✚ **Décisionnel** : Architecture des données, Data Warehouse, MDM.
- ✚ **Langues** : Anglais, Arabe

Compétences Techniques :

- ✚ **Big Data** : Hadoop: (HDFS, Yarn, MapReduce, Cloudera Manager/Ambari, Hue, Hive/Tez, Pig/Tez, HBase), Impala, MongoDB, Cassandra, Spark SQL, Solr, ELK
- ✚ **Distributions Hadoop**: HortonWorks, MapR et Cloudera (Expert)
- ✚ **DataScience** : Spark ML
- ✚ **Développement et tests** : Scala, C++, Java, PHP, Python, JUNIT/ScalaTest, JENKINS
- ✚ **Gestion de projets** : MS PROJECT, JIRA, ClearQuest
- ✚ **ETL** : Genio, Talend, Pentaho, Sqoop, Flume
- ✚ **Platform BI**: Business Objects BI 4.x, QlikView, Spotfire
- ✚ **Systèmes D'exploitation** : Linux (Redhat, CentOS, Ubuntu), Windows
- ✚ **Sécurité** : Kerberos, Knox, Ranger, Cryptage symétrique/asymétrique (RSA, DES...)





EXPERIENCES PROFESSIONNELLES

Depuis Octobre 2015 : Jems group

Expertise Big data et techlead d'équipe spark Scala : BNP FF (depuis avril 2018)

- ✚ Audit de la plateforme BigData (HDP 2.6) et des développements Spark 2 scala 2.x.
- ✚ Mise en place de standards et bonnes pratiques de construction du datalake.
- ✚ Développement de nouvelles applications spark 2 scala, intégrant des jar java existants et passant par devOps (IntelliJ, SonarQube, gitLab, Maven, Nexus, Jenkins et autres outils standards internes à PF).
- ✚ Expertise de mise en place de kerberos, habilitation avec Ranger et gestion des clusters via ambari.

Expertise et conseils en Big Data : RCI Banque (2 mois)

- ✚ Architecture et mise en place d'une plateforme Big Data **CDH 5.11.1** afin de porter la vision 360° du client
- ✚ Construction et structuration d'un Data Lake selon le modèle trois couches (données provenant d'Oracle et SQL Server)
- ✚ Alimentation dans Hive et HBase
- ✚ Requêtes via **Impala** de données liées aux clients (contrats, leads etc.)
- ✚ Gestion des développements Spark/Scala via JIRA, MAVEN, GIT

Expertise et conseils en Big Data : Renault Digital (9 mois)

- ✚ Architecture et mise en place de la data gouvernance du data lake
- ✚ Mise en place de standard de développement d'ingestion des données dans le data lake (spark scala + archetype maven) et dé-normalisation. Modélisation et optimisation Hive
- ✚ Industrialisation des workflows et coordinateurs oozie via knox.
- ✚ Amélioration de la CI/CD
- ✚ Support technique aux équipes projets utilisant le data lake. Diffusion et vulgarisation de la culture bigdata.

Expertise et conseils en Big Data : Renault (6 mois)

- ✚ Supervision des développements spark sql/scala. Garant du respect des standard et le bonnes pratiques des livrables.
- ✚ Modélisation Hive avec Partitioning et Bucketing sur le format ORC.
- ✚ Indexation ES.
- ✚ Workflow/coordonateur oozie
- ✚ Formation d'autres consultants sur les briques big data.

Expertise et conseils en Big Data : Société Générale (9 mois)





- ✚ Mise en place de configuration big data (Data lake, MongoDB, ElasticSearch)
- ✚ Collecter et filtrer les logs avec logstash => kafka
- ✚ Indexation des logs dans ElasticSearch (un cluster de 59 noeuds)
- ✚ Créer des dashboards avec Kibana4
- ✚ Stocker les erreurs d'elasticsearch (les événements mal formatés) dans hadoop
- ✚ Analyser des erreurs avec spark en streaming
- ✚ Archiver les index (par ans) dans hadoop
- ✚ Proposition d'une solution de sécurisation big data basée sur ranger et knox
- ✚ Supervision fonctionnelle de l'infrastructure
- ✚ Mise en place de config
- ✚ Rédaction de propositions techniques big data pour différents clients (banques, transports, produits de beauté, industrie automobile)

Juin 2015 - Septembre 2015 : Business&Decision

Cas d'usages métiers sur Hadoop HortonWorks : CREDIT AGRICOLE TS (4 mois)

- ✚ Gestion de flux provenant de DB2 MVS vers Teradata puis vers le data lake
- ✚ Architecture de données permettant le découplage et la gestion du versionning entre SIO et SID.
Proposition de la mise en place d'ESB (Talend)
- ✚ Mise en place de critères d'éligibilité et de de qualité des données alimentant le DataWarehouse en se basant sur les anomalies existantes.
- ✚ Etude comparative entre MongoDB et Cassandra.
- ✚ Mise en place d'un cluster Hadoop HortonWorks.
- ✚ Cas d'usages métier pour le pôle marketing.

Septembre 2013 - Mai 2015 : Bluestone

Expertise et conseils en Big Data : R&D (3 mois)

- ✚ Expertise autours de Hadoop ecosytem (Cloudera 5.4.x)
- ✚ Formation et coach Big Data notamment avec Cloudera, HDP
- ✚ Proposition d'architecture, conseils et chiffrage
- ✚ Installation, Maintenance et optimisation de la distribution Cloudera
- ✚ Scraping/Crawling avec python/Beautiful Soupe/MongoDB
- ✚ Configuration de VirtualHosts dans Apache

Résolution des problèmes de lenteur par application de solutions Big Data : CACEIS (2 mois)

- ✚ Analyse des données retenues afin d'en contrôler la cohérence et d'en faire le nettoyage.
- ✚ Optimisation des chaines de traitement et de chargement





- ✚ Modification des modèles de données utilisés pour répondre aux exigences de l'environnement et performances attendues
- ✚ DataManagement utilisant **Hive** et **Spark SQL**. Division par 10 du temps de traitement.
- ✚ Restitution et DataViz avec QlikView.

Responsable de la qualité logiciel et développement Eclipse RCP sur EMF : THALES TGS (7 mois)

- ✚ Développement d'un framework de test automatique avec **Junit** et Windowstester/SWTBot (13000tests)
- ✚ Mise à jour et surveillance des jobs JENKINS sur les JUNIT
- ✚ Mise en place de la méthodologie SCRUM.
- ✚ Développement d'un mini ESB entre JENKINS et Orchestra - IVVQ
- ✚ ClearCase, pour la gestion des sources
- ✚ Doors, pour la gestion des exigences
- ✚ Clearquest, pour le suivi des bugs/nouvelles fonctionnalités

Déploiement Distribution Hadoop Cloudera pour commercialisation d'une DMP Analytic (7 mois)

- ✚ Installation de systèmes Linux CentOS 6.4
- ✚ Configuration HBase.
- ✚ Déploiement et mise en place de clusters Cloudera
- ✚ Mise en place de programmes R en vue d'analyses statistiques et représentation via un portail ShinyServer
- ✚ Développement d'un package de déploiement automatique de nœuds (script shell)

Janvier – Août 2013 – HOP! BRITAIR : DEPARTEMENT SI

Mise en place d'un data warehouse : Déploiement du Designer BO XI 3.1 (8 mois)

- ✚ Création de jointure externe
- ✚ Création et optimisation d'univers sur Oracle 10g et PostgreSQL 8.3
- ✚ Administration et requêtes via Oracle Sql Developer et PgAdmin III
- ✚ Développement d'une application via le SDK SAP
- ✚ Étude et mise en place de DW pour la DSI de Brit Air
- ✚ Utilisation de BO QueryBuilder
- ✚ Développement de Modules et process Genio 7.1 pour l'alimentation des DW
- ✚ Mise en place de DataSet Genio avec des requêtes SQL complexes.
- ✚ Développement et maintenance de projets **Talend** OS DI 5.x avec PostgreSQL 8.3 pour alimenter Oracle 10g et inversement. Ce projet a été développé afin d'éviter l'achat d'un nouveau connecteur GENIO.
- ✚ Génération à partir de **Talend** d'applications exécutables utilisables comme process GENIO.
- ✚ Développement de scripts unix pour l'automatisation (ssh, scp, smbclient, cron, ...).
- ✚ Utilisation de Zabbix pour la surveillance des processus.





- ✚ Test et maintenance.

Environnement technique : BO XI 3.1, Oracle 10g, PostgreSQL 8.3, Sql Developer, Genio 7.1, Talend DI 5.x, Scripts unix.

2000-2012 – SAP BUSINESS OBJECTS

Déploiement de BO pour les comptes des partenaires JP MORGAN, PSA PEUGEOT, FRANCE TELECOM, AIR FRANCE, ...

EXPERT BI

Déploiement « Information Designer Tool » : SAP (6 mois)

- ✚ Création de connections SSU et MSU
- ✚ Résolution des problématiques liées aux multi-sources
- ✚ Création d'univers sur des cubes Essbase 9

Déploiement GENIO & Talend : SAP (3 mois)

- ✚ Définition d'un projet Talend d'alimentation oracle 10g à partir d'un jeu de données (vue Excel et Posgres)
- ✚ Définition d'un projet Genio d'alimentation oracle 10g à partir d'un jeu de données (vues Excel et Posgres)
- ✚ Création d'univers sur le résultat fourni par Java (Talend) et Genio.
- ✚ Création d'un univers sur des tables dérivées (vues au sens Désigner)
- ✚ Développement de rapports sur les trois univers
- ✚ Rapport de l'étude

Création de rapports complexes : Air France (10 mois)

- ✚ Création de connections SSU
- ✚ Création d'univers avec plusieurs centaines de classes d'objets.
- ✚ Création d'objets utilisant des fonctions avancées oracles 10g
- ✚ Résolution des problèmes liés aux valeurs « null » des objets d'agrégats
- ✚ Détection des contextes automatique et adaptation manuel
- ✚ Création de rapports WEBI/DESKI complexes multi-requêtes et utilisant la fusion BO
- ✚ Création de variables et de filtres configurables dans WEBI

Déploiement du Designer : France Télécom (4 mois)

- ✚ Création de DNS
- ✚ Création de connexions





- ✚ Architecture d'un data-warehouse en étoile sur oracle 10g et Teradata Database 13.
- ✚ Développement de stored proc en PL/SQL
- ✚ Création d'univers complexes avec des contexts.
- ✚ Résolution des problématiques de performances et des objets utilisant les agrégats.
- ✚ Import/Export vers et du repository BO

Développement d'un ESB de drivers d'accès aux différents SGBDR : SAP (8 mois)

- ✚ Développement d'une couche de service encapsulant les drivers native/ODBC d'accès aux DB (Oracle OCI/ODBC, Informix,...)
- ✚ Définition du cahier de charges
- ✚ Suivi de livrables en MS Project
- ✚ Méthodologie Scrum
- ✚ Test des drivers

Mise en place de la méthode SCRUM : SAP (8 mois)

- ✚ Organisation du travail de l'équipe dans un backlog (post-it, jira)
- ✚ Assurer la produire des versions à chaque sprint
- ✚ Décomposition de user stories
- ✚ Projection des user stories en tasks
- ✚ Définition et exécution des tests d'acceptation

CONSULTANT BI ET DEVELOPPEUR JAVA

Développement Java eclipse : SAP (14 mois)

- ✚ Déploiement du Designer
- ✚ Migration vers BO XI 3.1 avec ImportWizard et utilisation du format Biar
- ✚ Développement Eclipse, application RCP reposant sur le modèle Ecore EMF
- ✚ Méthode SCRUM/LEAN
- ✚ Test et maintenance

Modélisation de l'application BO 4.0 IDT : SAP (6 mois)

- ✚ Développement de plusieurs fonctionnalités coté serveur en C++
- ✚ Développement en JAVA Eclipse (RCP, SWT, EMF)
- ✚ Définition du méta modèle dans Ecore : graph.ecore
- ✚ Génération de l'éditeur arborescent
- ✚ Génération du code Java
- ✚ Implémentation des fonctions du modèle
- ✚ Démonstration du sprint
- ✚ Test Unitaires et d'intégration

Développement J2EE (EJB/JSP/JSF/JDBC/JavaMail) : SAP (4 mois)

- ✚ Développement de plusieurs projets JEE





- ✚ Création de pages jsp de nos serveurs d'application Webi (BO)
- ✚ Utilisation de JDBC pour l'accès au DB
- ✚ Utilisation de JNDI pour l'utilisation de LDAP
- ✚ JavaMail pour la diffusion des emails
- ✚ Test et maintenance

Développement JAVA de framework de test UI : SAP (7 mois)

- ✚ Développement de tests fonctionnels UI avec WindowTester sous Eclipse 3.4 (Ganymede)
- ✚ Génération et factorisation de tests UI

CONSULTANT BI ET DEVELOPPEUR WEB :

Développement WEB PHP 5.3, APACHE 2, MySQL 5.5 : SAP (2 mois)

- ✚ Application Client/Server : Gestion d'une bibliothèque online
- ✚ Test et maintenance de l'application

Déploiement du SDK : France Télécom (1 mois)

- ✚ Export des objets de l'univers BO (Classes, Objects, Tables, Columns, Joins, Contexts) via le SDK Designer en utilisation le VBA d'Excel.
- ✚ Développement d'application de la gestion des licences en jsp.

Etude comparative entre Deski et QlikView : SAP (1 mois)

- ✚ Installation et utilisation de QlikView,
- ✚ Installation de BO Deski
- ✚ Document comparatif

Développement front web: SAP (1 mois)

- ✚ Installation et déploiement de LAMP/WAMP
- ✚ Développement HTML (htmledit, notepad+), XML, JavaScript (Ajax, JQuery)

Développement dans l'application BO designer: SAP (2 mois)

- ✚ Développement de plusieurs fonctionnalités coté serveur en C++
- ✚ Création de plusieurs univers sur les univers de démos
- ✚ Développement en JAVA Eclipse (RCP, SWT,)
- ✚ Test Unitaires et d'intégration

CONSULTANT BI ET DEVELOPPEUR C++ :

Développement du Check Integrity» pour assurer l'intégrité du «repository» BO : SAP (8 mois)

- ✚ Définition du modèle et implémentation de la grammaire
- ✚ Implémentation des règles de la grammaire
- ✚ Déploiement sur des environnements clients
- ✚ Test et maintenance

Participation au développement du Supervisor BO : SAP (9 mois)

- ✚ Développement de plusieurs fonctionnalités
- ✚ Développement intensif en Eclipse SWT et Visual Studio
- ✚ Test Unitaires et d'intégration





Développement Multitask : SAP (11 mois)

- ✚ Développement d'une librairie Multithread de gestion du cache
- ✚ Cache pour améliorer les performances des serveurs BO suite
- ✚ Test de la lib et mise en production

Développement et correction de bugs : SAP (12 mois)

- ✚ Développement de nouvelles features et correction de bugs
- ✚ Installation et configuration de visual studio
- ✚ Installation et configuration de clearcase
- ✚ Test et livraison

Développement transversal : SAP (4 mois)

- ✚ Développement de nouvelles features en java (swing)
- ✚ Participation au développement des scripts perl d'intégration
- ✚ Développement et MAJ de scripts de compilation et gestion des sources (perforce)
- ✚ Test

Environnements : BO, Genio, Talend, SQL, OS Windooow/Unix/Linux, Apache/Tomcat, IIS/PHP, Apache/PHP, MySQL, MyPHPAdmin, EasyPhp, Eclipse, Perforce/ClearCase, VMWare, MSProject et d'autres outils

1998-1999 – OPENTEXT /Humming Bird /Leonard Logic

- ✚ Bug fixing C++/VB sous Unix/Windows pour l'ETL Genio
- ✚ Déploiement de l'ETL avec tous les connecteurs.

1994- 1997 – Service d'informatique du CHU et clinique Drevon de Dijon

- ✚ Cryptage et Chaînage des informations : Logiciel en C
- ✚ Réalisation d'un logiciel en FoxPro/VB, de gestion des dossiers des malades
- ✚ Réalisation d'un logiciel en Turbo Pascal de traitement de signal

FORMATION

- ✚ 2017 : Certification HDP 2.4
- ✚ 2013 : Formations BigData (CentOs, Hadoop, HortonWorks, Cloudera, MapR)
- ✚ 2013 : Formation SAP : FI CO SD MM
- ✚ 2012 : Formation Talend
- ✚ 1995 – 1998 : Doctorat d'Informatique de l'Université de Bourgogne : sécurité informatique
- ✚ 1993 – 1994 : D.E.A. d'informatique au CRISS de Grenoble
- ✚ 1987 – 1992 : Ingénieur en Informatique à l'institut d'informatique d'Annaba

