



# Data Science, Infrastructure et Big Data

Conseil et développement Open Source

[Introduction](#)

[Missions et Tâches](#)

[Data Science](#)

[Infrastructure Cloud](#)

[Outils de Gestion](#)

[Google Cloud Platform](#)

[Expérience](#)

[Banque / Finance](#)

[Data Science](#)

[Equipe](#)

[Partenaires \(Data Science\)](#)

[Accompagnement Client](#)

[Collaboration, Transfert de Connaissances](#)

[Documentation](#)

## Introduction

SenTai est une structure de conseil et développement informatique et scientifique qui réunit plusieurs [spécialistes](#) et [sociétés partenaires](#) aux compétences complémentaires. Chaque [membre de l'équipe SenTai](#) a plus de 10 ans d'expérience à haut niveau dans divers domaines tels que:

- Mathématiques appliquées, et la [Data Science](#),
- Plateformes de calcul optimisées sur le Cloud,
- Environnements, langages et technologies [Open Source](#) et [.Net](#),
- Virtualisation hardware ([Machines Virtuelles](#)) et la virtualisation d'applications ([Containers](#))
- Infrastructure Cloud (déploiement, optimisation, gestion, et automatisation de diverses technologies sur des clouds privés, publics ou hybrides).

- Technologies Web.

## Missions et Tâches

Sur la base de ces connaissances, nous sommes à même de conduire des missions de bout en bout sur une diversité de projets autour des technologies Open Source:

- **Open Source:** Utilisation des technologies Open Source pour développer de manière ultra-agile des solutions sur mesure.
- **Cloud:** Analyse de systèmes existants, transformation, adaptation et optimisation de ces systèmes sur le cloud (privé, public, ou hybride).
- **Infrastructure:** Analyse d'infrastructure existante, planification, déploiement, gestion et optimisation d'un parc de ressources diverses).
- **Big Data:** Utilisation des techniques et infrastructures Big Data (Hadoop, Spark, [Dataflow](#), [Dataproc](#), [BigQuery](#), [BigTable](#),...).
- **Data Science:** Conception de solutions prédictives ou de détections d'anomalies fondées sur l'[Intelligence Artificielle](#). Visualisation de données.
- **Web:** Développement d'interface web sur les dernières architectures et technologies ([SOA](#), AngularJS, CoffeeScript, ...).

## Data Science

En termes de connaissances en Data Science, nous maîtrisons les techniques suivantes:

- **[Data Science et Intelligence Artificielle](#):** Conception de solutions prédictives ou de détections d'anomalies fondées sur l'[Intelligence Artificielle](#) ([Apprentissage Automatique](#), [Réseaux de Neurones](#)):
  - [Régression logistique](#)
  - [Régression PLS](#) (Partial Least Square)
  - [Machine à vecteurs de support](#) (SVM)
  - [Réseaux de neurones](#)
  - [Forêts aléatoires](#) et méthodes d'ensemble
  - [Gradient adaptatif](#) ou [Gradient boosting](#)
- **Outils de Visualisation:** Tableau, Qlik, Matlab
- **Mathématiques Appliquées:** analyse des données, modélisation, méthodes numériques (l'[Algorithmique](#) et le l'[analyse numérique](#)).
- **Plateforme de Calcul de données en séries temporelles sur la base de Modèles Excel:** nous fournissons des outils qui facilitent l'interaction entre les besoins business de sociétés, souvent exprimés sous forme modèles financiers dans des tableurs Excel, et leur intégration dans des systèmes de calculs évolutifs dans le Cloud. Notre outil propriétaire permet de décrire un modèle de calcul dans un tableur

avec une syntaxe de formules “pseudo-excel” (i.e. proche d'Excel), et de convertir le modèle en Java grâce à un parser de langage ([Antlr](#)).

## Infrastructure Cloud

### Outils de Gestion

L'indépendance aux diverses plateformes Cloud (AWS, Azure, GCP, OVH, ..) et la gestion de multiples plateformes et ressources simultanées, privées ou publics, est essentiel pour garantir la pérennité et l'optimisation d'un Cloud d'entreprise. Cet agnosticisme permet un choix optimal et diversifié de gestion d'infrastructure, ainsi qu'une agilité garantie par des outils de gestion d'infrastructure cloud.

Notre outil de predilection pour cette tache est [Rancher.com](#), l'outil le plus avancé en terme de gestion Open Source basé sur des Containers

Nous nous spécialisons dans le déploiement de “Rancher Infrastructure Management” (R.I.M.) dans un cloud existant pour de grandes entreprises.

Voici quelques spécificités de Rancher::

- [Docker](#) Containers
- Hybrid Private-Public multi cloud ([AWS](#), [Azure](#), [GCP](#)...)
- Cloud Provider Independence (AWS, Azure, GCP)
- Infrastructure Management Independence
- Orchestrator independence ([Rancher](#), [Kubernetes](#), [Mesos](#), [Swarm](#))
- Highly secured container communication via IPSec (independently / per machine encrypted) Tunnels
- User friendly Dashboard (GUI) to manage, deploy, automate, monitor, highly customisable
- Automated Deployment, Monitoring
- Load Balancing
- [Jenkins](#) dev. environment facilitation for devs
- [NFS load balancing](#) ([GlusterFS](#))
- Shared Volumes: Persistent Volume Storage across environments with [Convoy-NFS](#)
- [Data Analytics Integration with ElasticSearch](#)
- DataDog Monitoring integration
- [Slack](#) integration

## Google Cloud Platform

SenTai est partenaire de [Google Cloud Platform](#) (GCP):. Nous avons une connaissance approfondie de leurs systèmes et sommes à même de juger de leur pertinence économique et technique afin de résoudre une problématique donnée:

- Nous utilisons leurs solutions [IAAS](#) et [PAAS](#) telles que leurs outils [Big Data](#):
  - [BigQuery](#): une BDD analytique à faibles coûts entièrement gérée (PAAS).
  - [DataLab](#): un outil interactif puissant pour explorer, analyser, visualiser les données.
  - [DataProc](#): des services gérés (PAAS) Hadoop MapReduce, Spark, Pig, et Hive pour le traitement de données Big Data à faibles coûts.
- Nous effectuons une veille permanente sur les nouvelles solutions dans le domaine du “Machine Learning”, par exemple [TensorFlow](#) de Google, sortie récemment. Cela nous permet de fournir des solutions prédictives, hautement évolutives et optimisées dans diverses industries.

## Expérience

### Banque / Finance

Nous sommes référents dans le domaine **Banque / Finance**: notre dernier projet est la re-conception de [Quest](#), une plateforme financière d’analyse fondamentale de +9000 sociétés cotées en bourse, pour la banque Canadienne [Canaccord Genuity](#):

**Calculs, Algorithmique et Cloud**: grâce à nos connaissances algorithmiques et technologiques, sur une infrastructure [Google Cloud Compute](#) haut de gamme, nous avons:

- apporté une solution plus souple,
- conçu un modèle altérable par les business analystes sans ajout de code,
- réduit le temps de calcul par un facteur de 100

**Web**: à partir de données provenant de [S&P](#) (annuelles et clôture journalière) sur chaque société, les résultats sont mis à disposition via une interface web réactive, pour consultation par des gérants de portefeuille institutionnels. Cette plateforme web contient de multiples tables et graphes.

### Étude de Cas:

- <http://sentai.eu/quest/>
- <https://cloud.google.com/customers/canaccord-genuity/>

## Data Science

Un document séparé est à disposition pour les divers techniques et études de cas.

## Equipe

Un document séparé est à disposition pour plus de détails des qualifications et expériences des membres de l'équipe SenTai et de ses partenaires Data Science.

Notre expérience nous permet d'effectuer des [développements très agiles](#) conjointement avec les équipes clientes.

## Partenaires (Data Science)

Nous combinons notre savoir faire Data science avec celui d'autres petites équipes très spécialisées en Data Science, qui maîtrisent certaines techniques et certaines industries.

Nous sommes capables de monter en charge et de diversifier la réflexion et le travail sur des techniques innovantes en Data Science. Nous travaillons particulièrement avec un partenaire Français et un partenaire Canadien:

- [Hadrian Advisors](#) (FR)
- [Aware Technologies](#) (Canada)

## Accompagnement Client

Chaque [membre de l'équipe](#) est capable d'interagir avec les diverses équipes (business, technique, etc..) du client, pour comprendre la problématique, et arriver de la manière la plus simple et rapide à une solution optimale.

### Collaboration, Transfert de Connaissances

Une collaboration étroite, par outils de visioconférence ou sur place avec le client permet un transfert de connaissance qui garantira l'indépendance de celui-ci, et la maîtrise de toute solution mise en oeuvre.

### Documentation

Une documentation est élaborée en continue pour permettre au client de reprendre la main et de poursuivre l'exploitation de chaque solution.