

Documentation technique

Avant-propos

Et si traiter ses données n'était plus une difficulté ?

C'est le défi que s'est lancé [Data Players](#) une société de services créée par une communauté d'experts et de passionnés.

Data Players intervient dans l'étude et le développement d'outils d'interopérabilité, de valorisation, et d'agrégation des données. Depuis 2017, nous distribuons le logiciel open-source [Bus Sémantique](#) géré par l'[Assemblée Virtuelle](#) et maintenu par une communauté dont nous faisons partie.

Ce partenariat est né d'un objectif commun qui est d'outiller les acteurs du monde de la transition. Fort de ses succès et de l'intérêt qu'il suscite, nous avons conçu une solution professionnelle afin d'apporter à nos clients un traitement accessible et ergonomique, capable de répondre aux nouveaux usages.

Notre ambition est de faire avancer la recherche et le développement de solution innovante pour faciliter la coopération numérique.

Ce document a pour but de présenter le fonctionnement et les fonctionnalités de [Grappe](#) qui est l'adaptation en SAAS du logiciel Bus Sémantique rendant le logiciel accessible sans devoir gérer son installation.

Sommaire

Avant-propos	2
Le Workflow	4
Les composants	5
Connecteurs d'entrée	5
Gestionnaires de flux	6
Stockage	6
Connecteurs de sortie	6
Détails des composants	6
Le fonctionnement d'un Workflow	8
Cas d'usage	9
Les fonctionnalités d'un Workflow	11
Partager	11
Paramétrer	11
Superviser	12
Consulter	12
Les offres Grappe	13
Grappe.io	13
Offre de bienvenu	13
Grappe Pro - On-Premise	14
Nos services	14
Formation	14
Accompagnement	14
Evolution	14
Assistance téléphonique	14
Nous contacter	14

Le Workflow

Le Workflow est un graphe de tâches conçu pour designer, déclarer et paramétrer un traitement dans le Bus Sémantique.



Interface du Workflow

Le Bus Sémantique hybride les propriétés d'un Bus logiciel (ESB) et d'un ETL, rendant possible l'interopérabilité entre plusieurs systèmes hétérogènes, techniquement et sémantiquement.

Cette technologie exploite la puissance et la polyvalence du traitement des graphes d'objets afin de raisonner au plus près des données telles qu'elles sont, sans réduire leurs complexités d'un bout à l'autre du flux.

Le Workflow facilite les procédés de préparation de données, d'interopérabilité et d'agrégation.

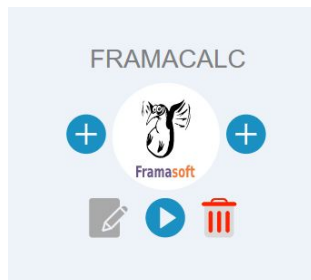
Il est composé de connecteurs et de gestionnaires de flux appelés "composants". Ils aident à interagir et connecter diverses données quelles que soient leurs natures et leurs sources.

De multiples usages sont réalisables tels que :

- Piloter la communication entre deux ou plusieurs systèmes, appareils ou éléments informatiques.
- Optimiser et gérer des systèmes d'information pour gagner en efficacité et performance.
- Modeler des données en fonction des usages.
- Exposer et consulter des données sur tous types de plateformes en temps réel.
- Faciliter l'accès aux données pour permettre leur interprétation.
- Appeler et mettre à disposition des web services internes (micro traitement spécialisé et réutilisable) et externe à Grappe.

Les composants

Les composants sont différentes tâches qui interagissent sur les données et entre eux. Leurs usages facilitent la création et la compréhension d'un traitement.



Composant FramaCalc

Il en existe 4 types : les connecteurs d'entrées, de sorties, les gestionnaires de flux et de stockage. Les composants disposent de deux connecteurs permettant de diriger le flux des données d'un composant à l'autre.

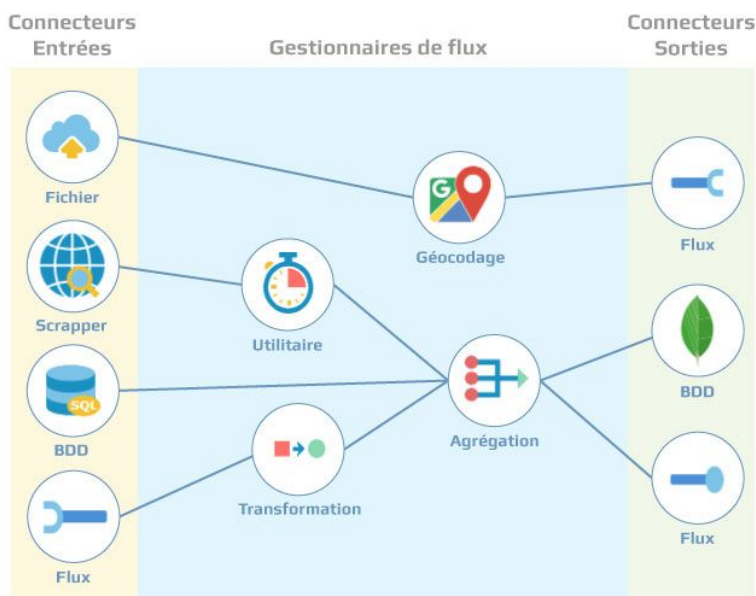


Illustration des différents types de composants.

Connecteurs d'entrée

Permet de connecter une source de données sur le Workflow de plusieurs manières :

- Charger le contenu d'un fichier
- Consommer ou pousser des API (web service)
- Connecter une base de données
- Récupérer des données sur des documents en ligne (Google Sheets, FrameCalc,..)
- Interpréter des flux
- Scrapper des pages web

Gestionnaires de flux

Ces composants exécutent des tâches variées et complexes sur les données (transformer, agréger, joindre, filtrer, géolocaliser,...) et acheminent le flux traité d'un composant à l'autre.

Stockage

Ce composant permet de stocker les données traitées dans une base de données (cache).

Connecteurs de sortie

Ces composants réceptionnent et convertissent le flux traité pour adapter les données au(x) format(s) de sortie(s). Le résultat du composant de sortie peut être poussé via une API ou sur un système tabulaire (BDD, fichier externe,...) mais peut être également appelé par un autre Workflow.



Ce Workflow localise les entrées d'un fichier pour les mettre à disposition sur une API.

Détails des composants

Composants	Descriptions
Flow consumer	Interroger une API REST avec une requête GET qui fournit un flux JSON, XML. <ul style="list-style-type: none"> flux tiré flux supporté : JSON, XML
Transform	Transformer un objet par mapping grâce à un objet transformation. <ul style="list-style-type: none"> Ajout de la capacité d'évaluation javascript <ul style="list-style-type: none"> comparable aux fonctions dans Excel activable avec le signe '=' les datas entrantes doivent être exprimées sous la forme '{\$.attribut}'
Google Sheets	Interroger une feuille de calcul Google Sheets qui fournit un flux JSON.
Aggregate	Agréger plusieurs flux pour n'en former qu'un seul.
Google geocoding	Interroger l'API Google geocode pour trouver une adresse avec la latitude et la longitude.
Cache NoSql	Sauvegarder un flux et le réutiliser sans avoir besoin de requêter la source.
data.gouv reverse geocoding	Interroger l'API adresse.data.gouv.fr pour trouver la latitude et la longitude avec une adresse.
Get provider	Exposer un flux de données sur une API HTTP GET.
Post provider	Déclencher un flux de données sur une API http POST. <ul style="list-style-type: none"> flux poussé

	<ul style="list-style-type: none"> flux supporté JSON
Framacalc	Interroger une feuille de calcul Framacalc qui fournit un flux CSV.
data.gouv.fr geocoding	Interroger l'API adresse.data.gouv.fr pour trouver une adresse avec la latitude et la longitude.
Join	Compléter un flux par un second en se basant sur un champ du 1er et un identifiant du 2nd.
Deeper Focus	Début de traitement d'un niveau de profondeur du flux.
Filter	<p>Filter les données pour n'en traiter qu'une partie.</p> <ul style="list-style-type: none"> basé sur sift syntaxe de requête mongodb pas de possibilité de projection
Upload	<p>Importer un fichier:</p> <ul style="list-style-type: none"> json xml csv excel ods rdf / owl / ttl
Scrapper	<p>Scrapper une page HTML. Extraire des données contenues dans une page web.</p> <ul style="list-style-type: none"> Utilisation de scénario WebDriverIO Couplage possible avec le composant Focus pour réaliser du Crawling <p>Attention aux CGU et aux droits des contenus (Data Players décline toute responsabilité)</p> <ul style="list-style-type: none"> Tarifs en fonction du compte utilisé
File consumer	Interroger un fichier mis à disposition sur une API REST avec une requête GET.
SQL connector	<p>Interroger une base de données SQL. Toutes les bases de données supportées par Sequelize</p> <ul style="list-style-type: none"> PostgreSQL MySQL SQLite Microsoft SQL Server Requête SQL
Mongo	<p>Interroger une base de données Mongo.</p> <ul style="list-style-type: none"> NoSQL <ul style="list-style-type: none"> Connexion à MongoDB Requête Mongo
Sparql	Requêter en SPARQL sur un fichier JSON Id.
Value Mapping	<p>Remplacer les valeurs d'une propriété par une autre.</p> <ul style="list-style-type: none"> Comparable à un traducteur d'une valeur source dans une valeur cible <ul style="list-style-type: none"> tableau de correspondance une valeur source peut être traduite en plusieurs valeurs cibles une valeur cible peut être référencée par plusieurs valeurs sources
Timer	<p>Déclencher un traitement à intervalle régulier par minute.</p> <ul style="list-style-type: none"> Permet de pré calculer des données et d'optimiser les performances Ce composant est utile uniquement s'il existe un conteneur de données pour stocker les données en sortie de flux
Params transform	Créer des paramètres de requête dans le flux.
Root from path	Extraire une valeur par son chemin.
Unicity	Structurer les données en vérifiant l'unicité par champ et répartir les valeurs par source.
Property matrix	Reconstruire des objets à partir de plusieurs propriétés en liste.
Post consumer	Envoyer les données en POST vers une URL externe.

Key To Array	Transformer les clefs d'un objet en tableau.
SFTP consumer	Interroger un fichier mis à disposition sur un serveur FTP.

Le fonctionnement d'un Workflow

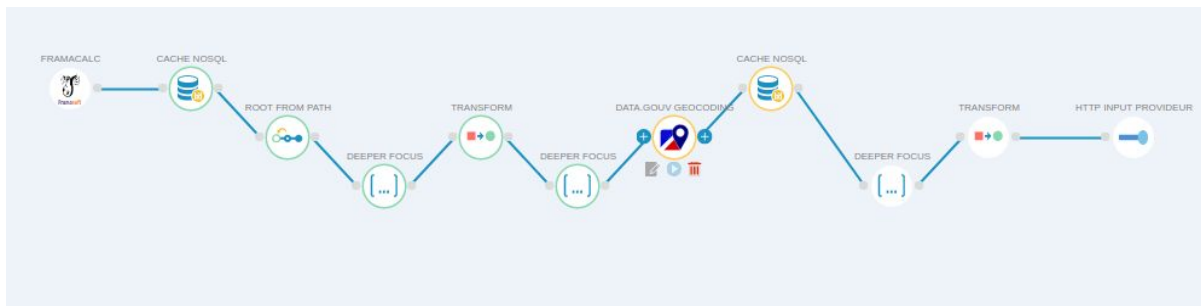
Les composants sont disposés sur un graphe ce qui permet de comprendre plus facilement l'usage fait et les tâches requises pour arriver aux résultats.

Lors de l'exécution du Workflow, une couleur verte, orange ou rouge apparaît autour du composant qui permet de suivre le flux et le statut de la tâche.

Vert : exécution terminée, Orange : exécution en cours, Rouge : exécution en erreur.

Afin d'assister à la création du Workflow, il est possible consulter le résultat sur chaque tâche à partir du moment où le traitement est terminé sur le composant.

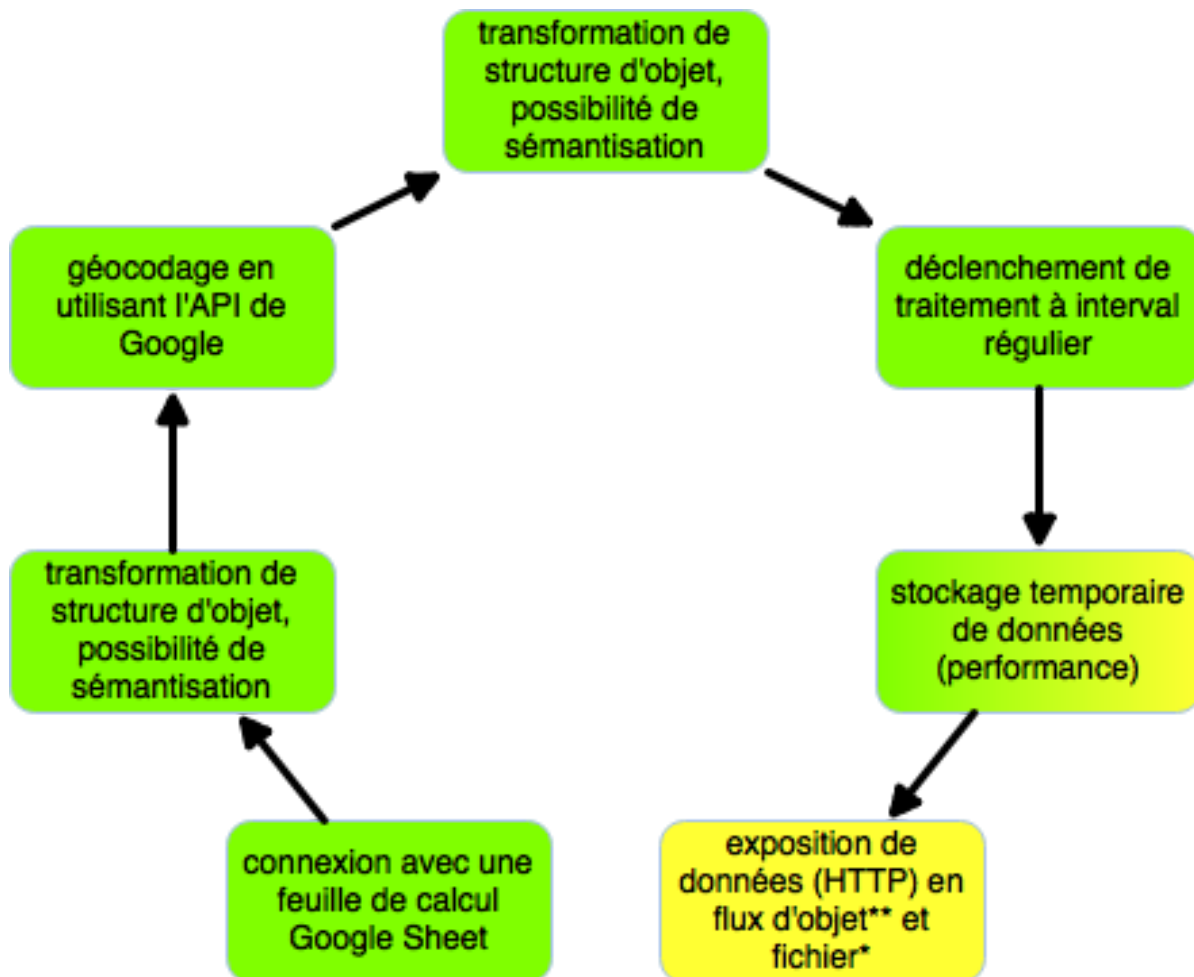
Pour consulter le flux obtenu par le composant, l'option persistance du process doit être cochée dans le paramétrage du composant. La persistance du process active une sauvegarde des données traitées par le composant.



L'exécution du Workflow peut être lancée depuis n'importe quels composants connectés.

Cas d'usage

Ce cas d'usage consiste à requêter des organisations sur un Google Sheets, les géocoder à partir de leur adresse, transformer les données en GeoJson (format standard pour la cartographie) et mettre à disposition le résultat sur une API HTTP pour un affichage sur un outil de cartographie comme Umap.



Explication étape par étape :

Capture du Workflow présenté

Google Sheets : connexion à une feuille de calcul Google Sheets

Ce composant est à paramétrer avec l'adresse d'une feuille de calcul Google Sheet. Il interroge l'API de Google pour récupérer les données et les transforme en objet.

Les données contiennent des organisations avec leur adresse postale.

Transform : transformation de structure d'objet, possibilité de sémantisation

Ce composant est capable de transformer la structure des données. Il se paramètre grâce à un objet de transformation qui décrit comment transformer la structure.

Cette étape structure les données qui arrivent de Google qui sont peu claires et peu adaptées à leur traitement ultérieur.

Google geocoding : géocodage en utilisant l'API de Google

Ce composant permet d'interroger l'API de Google pour trouver la latitude et la longitude à partir de l'adresse de chaque enregistrement. Il se paramètre en précisant quelle propriété contient quel type d'information (rue, ville, pays...) et quelles propriétés vont être renseignées avec la latitude et la longitude.

Transform : transformation de structure d'objet, possibilité de sémantisation

Cette étape permet de transformer la structure des données en GeoJson qui est un format standard pour être lu par les outils de cartographie comme Umap.

Timer : déclenchement de traitement à intervalles réguliers

Ce composant va déclencher à intervalles de temps réguliers le traitement des données afin de maintenir des données à jour. L'intervalle de temps est paramétrable. Pour ce cas, nous considérerons un rafraîchissement une fois par jour.

Cache NoSQL : stockage temporaire de données. (Performance)

Cela permet d'avoir toujours des données finalisées à disposition sans avoir besoin de refaire tous les calculs. L'intérêt est un gain substantiel de performance. Ce stockage n'est pas considéré comme une base de données "métier" mais uniquement un stockage technique.

Get provider : exposition de données (HTTP) en flux d'objet et fichier

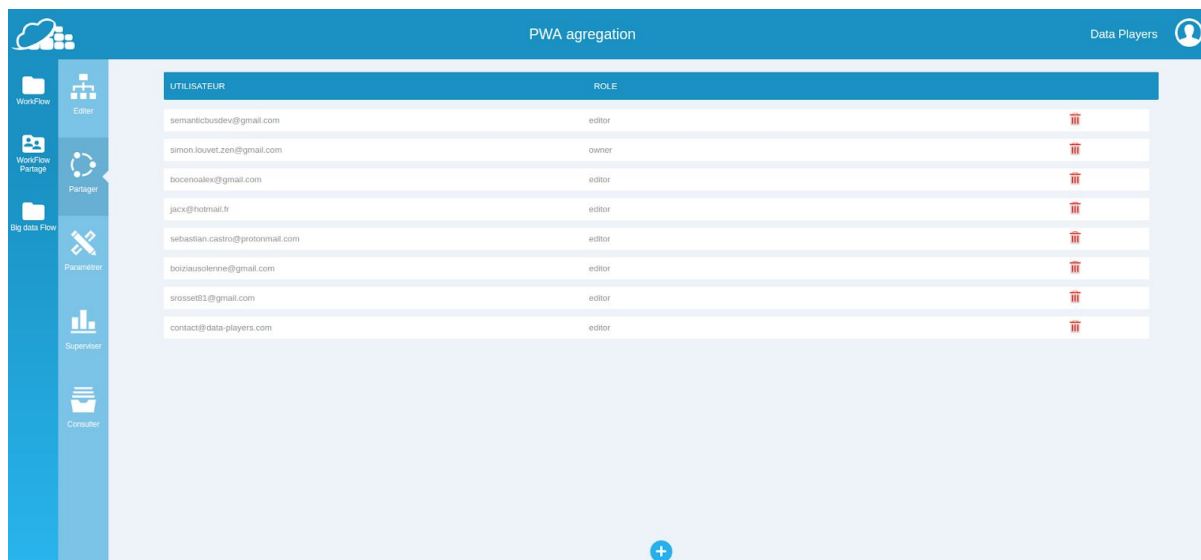
Des outils de cartographie comme Umap peuvent interroger une API HTTP pour obtenir les données à afficher. Il est donc nécessaire de créer cette "porte de sortie" pour qu'un outil de cartographie puisse accéder aux données qui sont à disposition dans le cache. Il se paramètre avec l'url à interroger pour obtenir les données.

Les fonctionnalités d'un Workflow

Chaque Workflow est composé de plusieurs fonctionnalités :

Partager

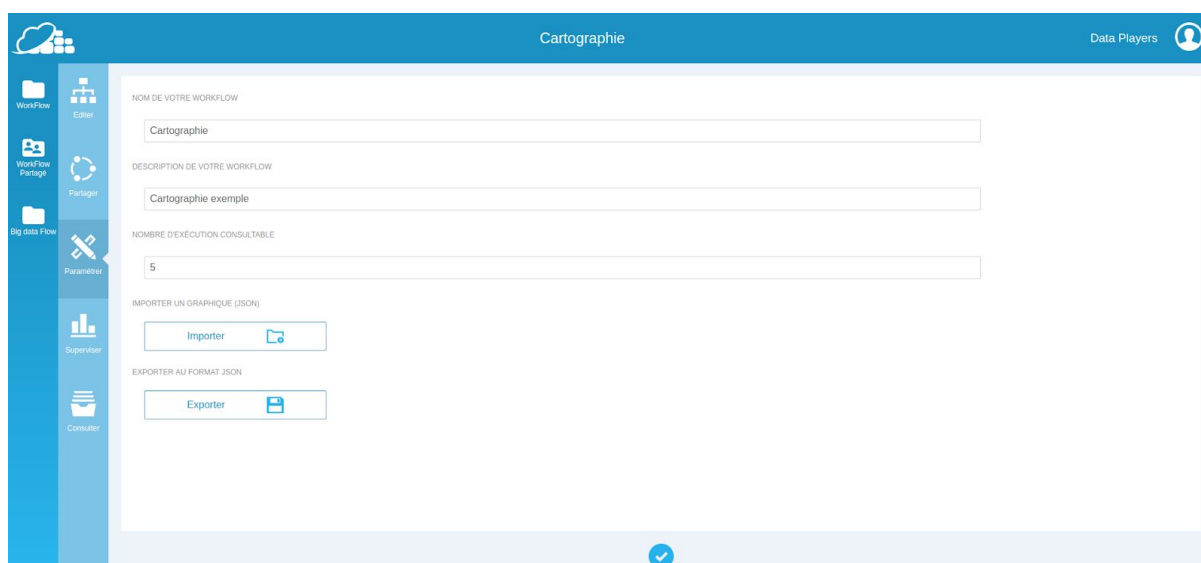
Permet de donner l'accès à d'autres utilisateurs Grappe au Workflow et à ses fonctionnalités.



UTILISATEUR	ROLE
semanicbu.dev@gmail.com	editor
simon.louvet.zen@gmail.com	owner
bocenoalex@gmail.com	editor
jack@hotmail.fr	editor
sebastian.castro@protonmail.com	editor
bolzausolenne@gmail.com	editor
srosset1@gmail.com	editor
contact@data-players.com	editor

Paramétrer

Permet d'administrer les informations du Workflow, le nombre d'exécutions consultable et d'exporter le Workflow ou d'importer un Workflow.



NOM DE VOTRE WORKFLOW
Cartographie

DESCRIPTION DE VOTRE WORKFLOW
Cartographie exemple

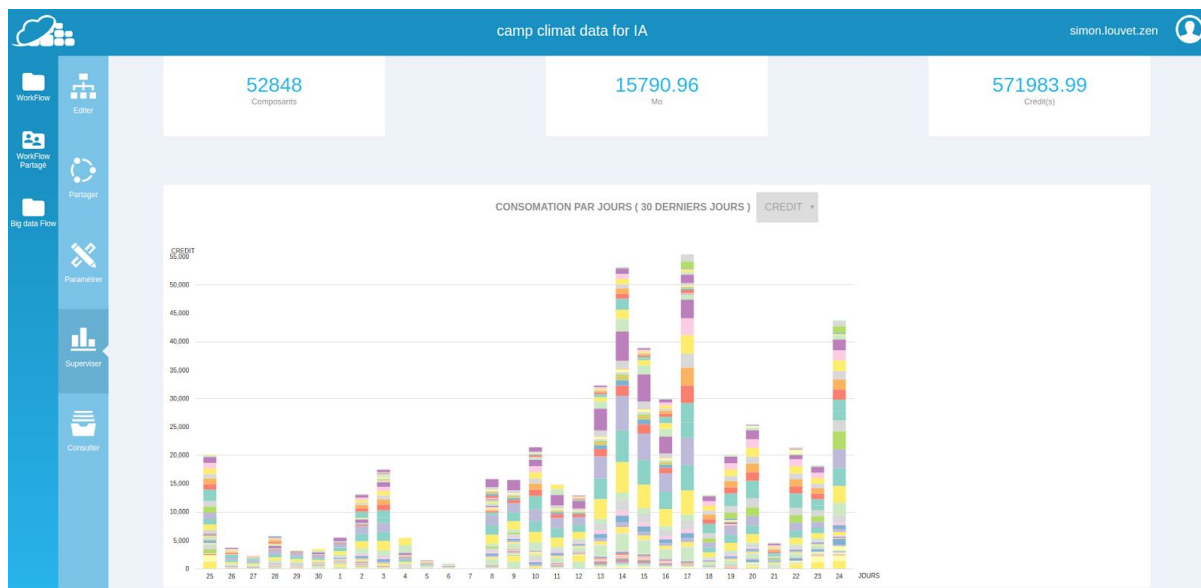
NOMBRE D'EXÉCUTION CONSULTABLE
5

IMPORTER UN GRAPHIQUE (JSON)
Importer

EXPORTER AU FORMAT JSON
Exporter

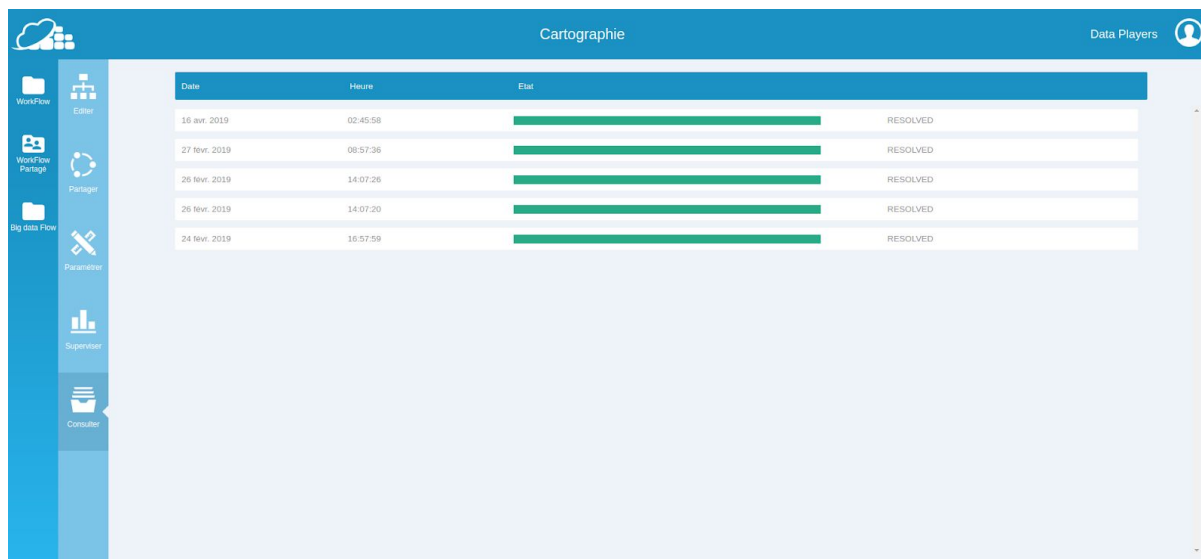
Superviser

Permet de visualiser la consommation du Workflow en crédit ou en Mo par composants sur une période de 30 jours.



Consulter

Permet de consulter un historique des exécutions du Workflow et de visualiser son état et les données persistées lors de l'exécution sélectionnée.



Date	Heure	Etat
16 avr. 2019	02:45:58	RESOLVED
27 févr. 2019	08:57:36	RESOLVED
26 févr. 2019	14:07:26	RESOLVED
26 févr. 2019	14:07:20	RESOLVED
24 févr. 2019	16:57:59	RESOLVED

Les offres Grappe

Grappe.io

Cette solution est entièrement hébergée en SaaS (Software as a Service). Elle est accessible en ligne depuis votre navigateur, sans installation.

L'inscription est gratuite, sans abonnement, ni licence. Seul le traitement et le stockage de vos données seront facturés au volume.

Les tarifs pratiqués par Data Players sont basés sur le volume de traitement de données. Toute donnée occupe une taille de mémoire exprimée en octet.

Pour simplifier la gestion financière du compte de l'utilisateur, nous exprimons nos coûts en crédit. 1 Euro correspond à 1000 crédits.

Le coût est calculé à chaque fois qu'un flux de données est traité par un composant.

Le compte de l'utilisateur est alors décrémenté du nombre de crédits consommés à chaque traitement.

Chaque composant a un tarif exprimé en crédits par Mo de traitement.

Le stockage possède un tarif propre dissocié du tarif de traitement. Le fait d'enregistrer ou de lire des données dans le stockage coûte des ressources mais le fait de conserver ces données dans le temps également.

Si le solde atteint la limite de 0 crédit, les flux de données seront alors bloqués.

Le calcul du prix s'effectue au 1er de chaque mois avec le volume de données stocké à cet instant.

Acheter des crédits

Le solde de crédit est consultable et rechargeable par carte bancaire via la page profil sur Grappe.

Offre de bienvenu

Lors de l'ouverture d'un nouveau compte sur Grappe, Data Players offre un solde de 100 crédits puis 1900 crédits une fois l'adresse e-mail du compte validée. Soit un total de 2€.

Cette offre couvre les frais d'exécution des composants d'un traitement "standard".

Chaque mois, Data Players crédite le solde du compte afin qu'il ne soit pas inférieur à 2000 crédits. Cette somme n'est pas cumulable et ne peut être débitée.

Grappe Pro - On-Premise

Il s'agit de la version On-Premise de Grappe adaptée pour les professionnels.

Cette offre est à 440€ HT /mois minimum, elle n'intègre pas le système de crédit.

Elle comprend :

- L'installation et la configuration d'un moteur de calcul avec une capacité de traitement de 10Go de données simultanément maximum.
- La maintenance et la mise à jour du service.

Le montant de l'offre varie selon le nombre de moteurs de calcul (nœud) installés sur le serveur hébergeur. L'offre est soumise à un engagement d'une durée de 3 mois minimum.

Nos services

Data Players met à votre service son expertise

Formation

Diffusion des pratiques autour de l'interopérabilité et mis en pratique avec Grappe.

Accompagnement

Notre expertise dans l'usage et le paramétrage de Grappe est à votre service.

Evolution

Ajouter de nouveaux composants et de nouvelles fonctionnalités sur Grappe.

Assistance téléphonique

Service disponible uniquement pour Grappe Pro uniquement si l'installation a été réalisée par Data Players.

Demandez un devis sur notre site internet <https://data-players.com/>

Nous contacter

Par mail :

contact@data-players.com

Par courrier :

DATA PLAYERS
6 RUE RENE VIVIANI
44200 NANTES
FRANCE