



OPENVALUE

L'**ACCÉLÉRATEUR** DE VOS PROJETS
DATA & INTELLIGENCE ARTIFICIELLE

PARTENAIRE  databricks  Azure



Accélère les projets IA des grands comptes



Pernod Ricard



MoëtHennessy



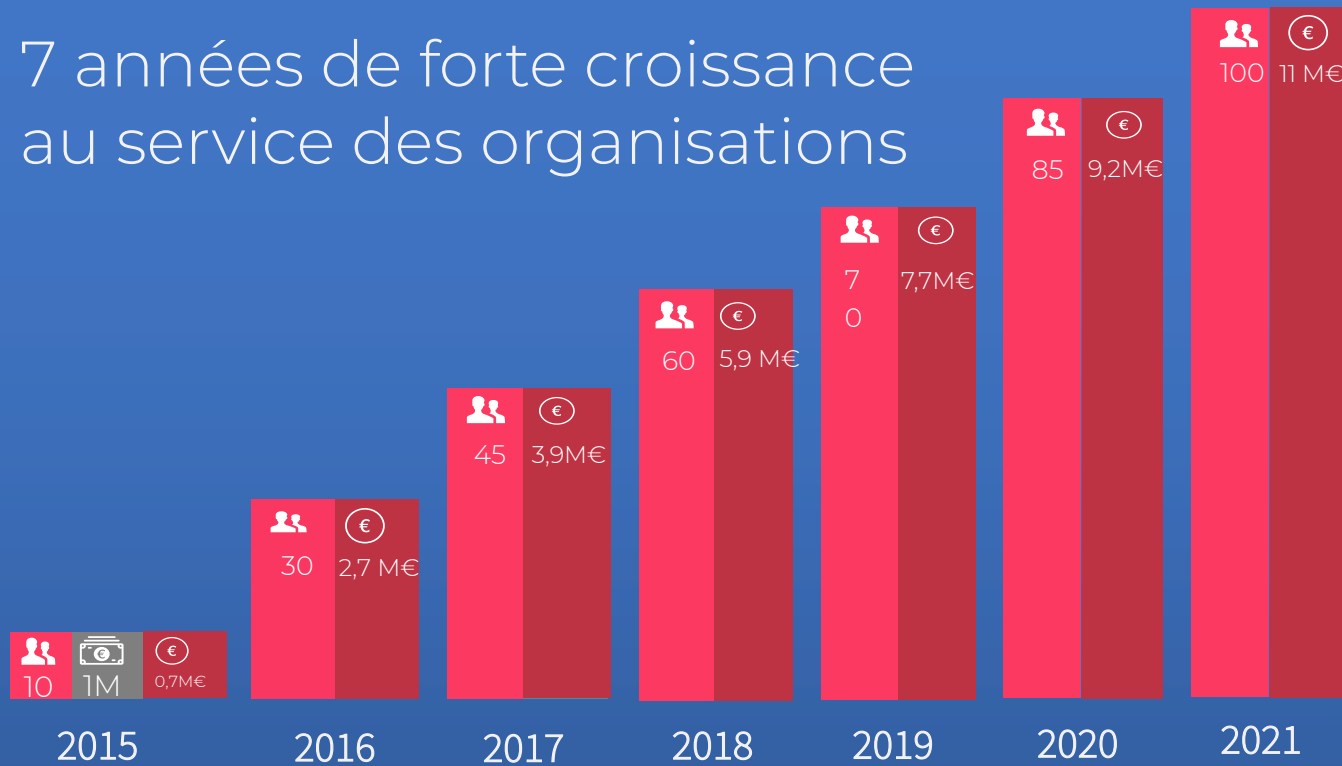
Christian Dior
COUTURE

FNAC DARTY

faurecia



7 années de forte croissance au service des organisations



70%

Projets en production

30+

Clients grands-comptes

100+

Collaborateurs



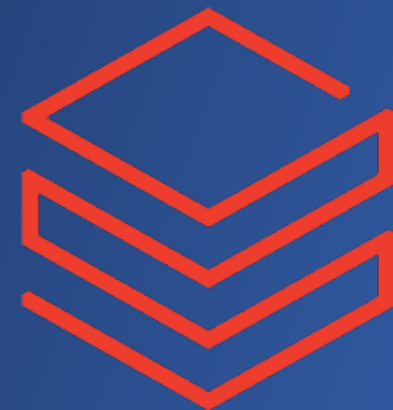
Des partenariats stratégiques solides



Des équipes certifiées des dernières technologies



L'OFFRE OPENVALUE : IA & DATA



L'Offre d'Openvalue :



Apporter au client tous les apports de l'IA pour une entreprise

Machine Learning

Deep Learning

NLP

Computer Vision

MLOps & DataOps

Agilité

Intégration continue

Move To Cloud

Industrialisation

Trajectoire et gouvernance des Données

Cas
d'usage

ROI

Cycle de l'exploitation de la data

5

Principaux usages



Applications possibles

- Analytiques** (recherche, analyse d'information, optimisation...)
- Textes** (émotions, avis, analyse des sentiments, chatbots...)
- Images/médias** (reconnaissance, suggestions, ...)
- Voix/son** (nouveaux usages, skills ...)

4

Restitution de la donnée (UX)



Interfaces

- UX & UI (interface Homme / Machine)
- IHM robotique

3

Traitement de la donnée (IA)



Intelligence artificielle

- Machine learning
- Réseaux de neurones
- Deep learning

2

Outils transport et transformation de la donnée (ETL)



Récupération
Extract



Préparation
Transform



Stockage
Load

1

Génération de la donnée



Data



Plateformes digitales



Machines, capteurs

CAS D'USAGE SECTORIELS




Finance / Assurance	Luxe / CPJ	Ind. / Construction	Transport	Services aux entreprises	Media	Retail
<ul style="list-style-type: none"> Conception et création d'une solution end-to-end de lutte contre la fraude sur les moyens de paiements scripturaux hors CB (virements, chèques, retraits d'espèces) Capitalisation sur l'acquis afin de définir un catalogue de données 	<ul style="list-style-type: none"> Mise à disposition d'une application qui permet de détecter ou d'exploiter n'importe quel élément d'une image (bouteille, personne...), mais aussi de détecter des ambiances dans l'image (extérieur, fête...) ou encore les textes 	<ul style="list-style-type: none"> Segmentation des transports de matériaux pour identifier les différents clusters et optimiser par la suite l'utilisation des camions pour la livraison des matériaux 	<ul style="list-style-type: none"> Reconnaissance de caractères manuscrits (OCR) de bulletins de conduite (métro) et intégration des informations automatiquement dans le SI 	<ul style="list-style-type: none"> Détection et qualification de la maturité de bourgeons de roses à partir d'un Ortho Photo Mosaïque (OPM) 	<ul style="list-style-type: none"> Segmentation des utilisateurs d'une plateforme de replay (catalogue de vidéos) en fonction de leur habitude de consommation, du contenu et de leur réponse à des fins de sollicitations commerciales 	<ul style="list-style-type: none"> Optimisation de la performance commerciale (segmentation et scoring clients et distributeurs, aide à la décision pour les commerciaux)
		<ul style="list-style-type: none"> Reconstitution d'une chaîne d'informations sur la qualité de produits à partir de données temporelles 	<ul style="list-style-type: none"> Création d'une application de priorisation des alarmes d'incidents sur une ligne du métro 	<ul style="list-style-type: none"> Détection et qualification de corrosion à partir de photos 		<ul style="list-style-type: none"> Création d'une plateforme RH conforme à la RGPD
<ul style="list-style-type: none"> Segmentation de clients de réseaux bancaires pour répondre à l'apparition de nouveaux canaux digitaux 	<ul style="list-style-type: none"> Classification de points de ventes en intégrant les données des réseaux sociaux 	<ul style="list-style-type: none"> Détection des dégradations (avec géolocalisation et catégorisation) sur les routes (vidéos prises par des smartphones) 	<ul style="list-style-type: none"> Optimisation de la consommation d'énergie sur une ligne du métro (étude des parcours des trains et données de consommations énergétiques des composants électriques) 	<ul style="list-style-type: none"> Création d'un outil de recherche de réglementation à partir de commentaires experts 	<ul style="list-style-type: none"> Création de moteurs de recommandation de vidéos (multi-supports) 	<ul style="list-style-type: none"> Analyse d'un SAV avec la mise en place d'un clustering des symptômes et diagnostics des pannes, de prédictions sur les besoins en pièces de rechange et les codes travaux techniques à réaliser
<ul style="list-style-type: none"> Création d'un outil de scoring pour (in)valider l'octroi de nano-crédits 	<ul style="list-style-type: none"> Création d'une plateforme clients BtoC conforme à la RGPD 	<ul style="list-style-type: none"> Création d'une plateforme RH conforme à RGPD 		<ul style="list-style-type: none"> Relevé de tableaux électriques à partir de photos (OCR) 		

PRÉDICTION DES VENTES – DISTILLERIE (IRLANDE)



Objectifs

- Prédire les ventes pour la distillerie sur le marché irlandais
- Analyser des données temporelles (timeseries)
- Utiliser des données d'économétrie et d'indicateurs économiques publics




Réalisation:

- 1 Data Scientist : création de deux modèles :
 - AutoML Azure pour les ventes intra produits
 - LSTM Keras pour les séries temporelles et indicateurs économiques



Déploiement : en production

- Projet en production qui tourne une fois par mois



Planning : 6 mois, 2019 - 2020



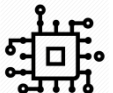
Livrables

- Code source
- Documentation de la solution
- Modèle en production



Prérequis

- Accès aux données



Stack technique

- **Environnement** : Azure Cloud (PaaS)
- **Plateformes** : Databricks & AKS (Kubernetes)
- **Technos** : Spark, AutoML, Azure LogicApp
- **Langages** : Python (Keras) & PySpark
- **Modèles** : LSTM (Keras), AutoML Azure, Voting model (for final result)




MOTEURS DE RECOMMANDATION – PLATEFORME DE REPLAY



Objectifs

- Création de 5 moteurs de recommandation de vidéo sur une plateforme de replay en fonction de la connaissance de l'utilisateur, l'environnement (web, app, TV), besoins de temps réel...
- Mise en place d'APIs pour exposer les modèles avec une inférence en temps réel pour certains
- Prise en compte des règles métier pour intégrer les directives éditoriales et règles liées à la RGPD



Réalisation

- 2 Data Engineers : collecte de données, chaîne CI/CD, déploiement, APIs en Spark Scala, Azure Functions, Cassandra, Airflow, Snowflake
- 2 Data Scientists : nettoyage des données, création des variables explicatives, entraînement des modèles, inférence avec PySpark, PyTorch, fastText



Déploiement : **en production**

- Sur la majorité des environnements de visionnage de la plateforme : web, applications mobiles iOS & Android, IPTV (tous les opérateurs sauf Orange)
- Disponibilité des recommandations sur la homepage et dans le player vidéo
- < 100ms de latence pour des charges jusqu'à 1000 appels/sec



Planning : **2019, 2020, 2021**




Livrables

- Code source
- Documentation de la solution
- APIs exposant les recommandations pour un utilisateur dans un contexte donné



Prérequis

- Compréhension métier des données
- Intégration des règles liées aux métiers



Stack technique

- **Environnement** : Azure Cloud (PaaS)
- **Plateforme** : Databricks & AKS (kubernetes)
- **Technos** : Spark, FastAPI, Kubernetes, Airflow, Azure EventHub, Azure Functions, Snowflake
- **Langages** : Python (PyTorch) & Scala (Spark)
- **Modèles** : RBM (réseau de neurones), ALS (filtrage collaboratif), fastText (NLP), clustering



DATA LAKE GLOBAL – MULTI-USAGES

Objectifs

- Création d'un Datalake dans le Cloud (France et International) pour couvrir des besoins de business intelligence, API et Data Science / Machine Learning
- Traitement en temps réel des rapports d'incidents et presque accidents sur les chantiers
- Mise en place de la chaîne CI/CD



Réalisation

- 3 Data Engineers : mise en place de la chaîne CI/CD et développement de nombreux sujets / données (ressources humaines, équipements /engins, usines, référentiels...)
- 1 Data Solution Architect : création de l'environnement global et choix des technologies
- 2 Data Scientists : création de l'environnement global et choix des technologies



Déploiement : **en production**

- Chaîne complète d'intégration des développements mise en place permettant des déploiements en production simples et testés



Planning : **2019, 2020, 2021**



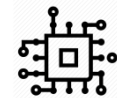
Livrables

- Plateforme en production avec monitoring
- Documentation de la solution (DAT, DEX)
- Ingestion de plus de 30 flux de données depuis plus de 400 serveurs différents



Prérequis

- Accès aux données brutes
- Compréhension des métiers



Stack technique

- **Environnement** : Azure
- **Plateforme** : Databricks
- **Technos** : Spark
- **Outils** : Azure DevOps, Azure Data Factory, Azure EventHubs, CosmosDB, SQL Database
- **Langage** : Scala



COMPUTER VISION – DÉTECTION DE DÉGÂTS

Objectifs

- Développer une application de détection de dégâts sur les routes à partir de flux vidéo ou images
- Application à divers *devices* d'acquisition (GoPro, Garmin, Smartphones etc...)
- Masquer automatiquement toutes données personnelles (véhicule, piéton...) en lien avec la RGPD

Réalisation

- 2 Data Scientists : mise en place de la plateforme de labélisation, utilisation d'un modèle open source pré-entraîné pour l'anonymisation, fine tuning d'un 2ème modèle pré-entraîné pour la détection des dégâts et entraînement sur des machines GPU en utilisant la librairie Detectron2 et le framework PyTorch
- 2 Experts métiers : labélisation d'images nécessaire à l'entraînement des modèles

Déploiement : **en production**

- Déclenchement automatique du job d'anonymisation et de détection des dégâts dès la mise à disposition d'images ou de fichiers vidéo
- Chaîne complète d'intégration avec association des dégâts détectés aux points GPS pour un affichage sur un logiciel de cartographie

Planning : **5 mois, 2020**

Livrables

- Code source de la solution
- Documentation de la solution
- Anonymisation & prédiction des dégâts sur des flux vidéo ou des images

Prérequis

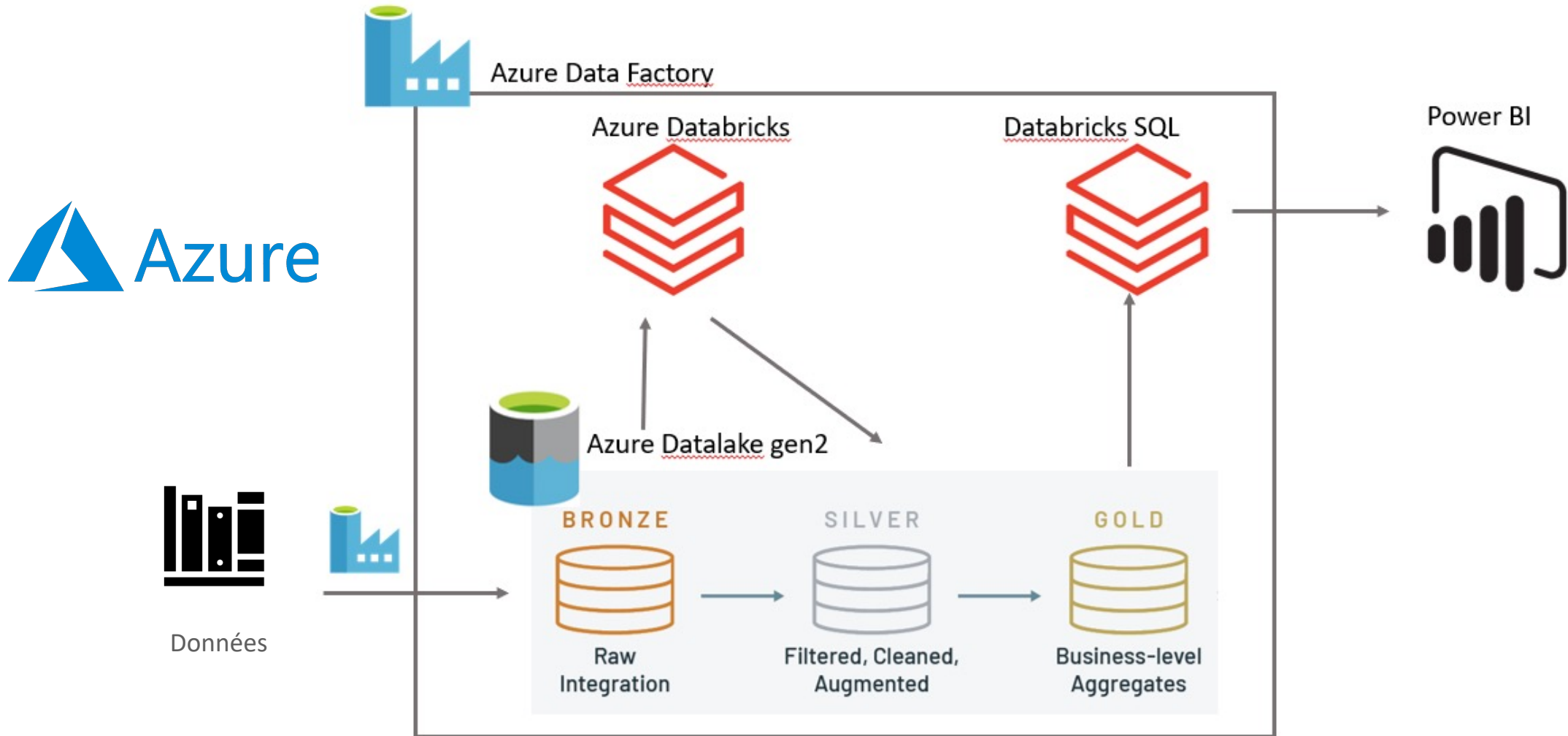
- Mise à disposition d'un jeu de données
- Labélisation des données
- Interlocuteurs métiers disponibles
- Ressources machines adéquates

Stack technique

- **Environnement** : Azure
- **Plateforme** : Databricks GPU
- **Technos** : PyTorch
- **Outils** : Azure Blob Storage, Azure Data Factory
- **Langage** : Python



EXEMPLE D'ARCHITECTURE DATABRICKS – DATALAKE & HISTORY MANAGEMENT



PARTENARIAT DATABRICKS





Clients en commun

- Bouygues
 - TF1 Data
 - TF1 Pub
 - Colas Digital Solutions
 - Bouygues Immobilier
- TotalEnergies
- Legrand
- Clarins
- Pernod Ricard
- Sodexo
- Fnac Darty
- Covéa
- Axa IM
- Apave
- Groupama
- LVMH



Formation

- + de 15 consultants certifiés
- 2 consultants certifiés dans le Champions Program
- Formateur agréé (en process)



Récompenses

- Partenaire de l'année 2019 SEMEA
- Openvalue : SI Customer Impact SEMEA 2020

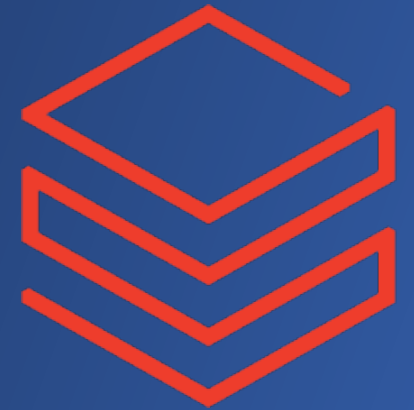


Les avantages identifiés par l'équipe du projet de la solution Databricks sont multiples :

- **Création d'un cluster de calcul simplifié et rapide** : interface de déploiement simple, création d'un cluster auto-scalable rapide, extinction automatique en cas d'inactivité
- **Support de plusieurs langages et mise à jour** : les notebooks Databricks permettent l'utilisation de plusieurs langages (Scala, Python, R...) et l'import simplifié de bibliothèques tierces. De plus, Databricks étant le principal contributeur à Spark, les dernières nouveautés sont disponibles à chaque mise à jour
- **Partage des ressources** : un même cluster peut être utilisé pour exécuter plusieurs jobs simultanément, la gestion des ressources étant effectuée automatiquement
- **Travail collaboratif** : plusieurs collaborateurs peuvent développer simultanément sur le même notebook, simplifiant grandement le processus de découverte
- **Dashboard** : intégration native d'une solution de dashboarding simple et partageable (one cell ↔ one graph)
- **Intégration et automatisation** : intégration avec Git pour les différentes versions du code, scheduling des différents jobs de manière très simple et rapide (interface graphique avec possibilité d'enchaînement de jobs, alertes mails ...)
- **Plateforme simple, ouverte et collaborative**



FOCUS OPENVALUE



Selon un sondage Gartner de 2017

15%

seulement des projets Big Data sont
déployés dans les entreprises

vs

70%

des projets Openvalue pour le compte de
nos clients

Comment?



De la stratégie au projet IA

Data Driven Strategy

Co-définition de la stratégie d'acculturation data et utilisation des 3 outils

OPENVALUE
Data Lab

OPENVALUE
As a service

OPENVALUE
Data Governance

- Créer les conditions de l'innovation en :
- Vous accompagnant sur le plan « organisationnel »
 - Facilitant les temps de travail collaboratifs
 - Créant un espace physique dédié à l'accélération des projets

- Des services sur étagère
 - Infrastructure
 - Data engineering
 - Data science
- Des services progressivement internalisables dans votre organisation

- Un service complet de data Governance
- Mise en place des outils adaptés
- Accompagnement au changement

Use case project

Workshops

OPENVALUE
Studio

SPRINT

- Priorisation des tâches
- Stand-up meetings
- Delivery régulier
- Mise en production
- Copro / copil

Développement

Services managés

Ux

Data

Infra.

- Comprendre les utilisateurs finaux, leurs points de douleur et leurs attentes
- Décliner ces dimensions en fonctionnalités
- Concevoir en conséquence les parcours utilisateurs
- Comprendre les attentes des utilisateurs
- Trouver les leviers de valorisation de la donnée
- Préparer la restitution
- Co-construction du socle technique
- Partage sur les usages
- Maintenance de l'infrastructure
- Maintenance applicative
 - Curative
 - Proactive
 - Evolutive



OPENVALUE

L'ACCÉLÉRATEUR DE VOS PROJETS
DATA & INTELLIGENCE ARTIFICIELLE