



BIG DATA

POURQUOI IL FAUT MONTER UN DATA HUB EN 2022



Signe que l'exploitation des données passe désormais dans une phase d'industrialisation, de nombreuses entreprises déploient un data hub. Analyse et retours d'expérience.

par Alain Clapaud.

Si les années 2010 ont vu de nombreuses entreprises mettre en place une infrastructure big data, cette approche reste limitée à une population d'analystes et de data scientists. En 2022, il est temps de voir plus grand et d'aller vers ce que l'on appelle désormais des data hub. Véritable plaque tournante de la donnée dans l'entreprise, un data hub n'est ni un ETL, ni

un outil de gouvernance de la donnée, ni une passerelle d'API, mais tout cela à la fois.

Parmi les éditeurs à se positionner sur ce marché, Dell Boomi propose une approche iPaaS (Integration Platform as a Service) avec Boomi AtomSphere. Cette solution a été choisie par le groupe Les Mousquetaires qui l'a mise en œuvre pour sa collecte de tickets de caisse dans son enseigne Bricomarché. Boomi remonte en central les tickets en temps réel, pour les traiter et les analyser. L'interface Low Code, d'AtomSphere, leur a permis une mise en place rapide et une adoption par les équipes facilitée par cette approche. Ce premier projet a été réalisé en trois mois et 73 flux métiers qui ont été déployés sur la plateforme.

73

flux métiers déployés sur le data hub du groupe Les Mousquetaires.



L'architecture type d'un data hub. Ici, celui proposé par l'éditeur MarkLogic.

Acteur historique de l'intégration de données, Talend propose Data Fabric, avec l'ambition de couvrir au maximum tous les cycles de vie de la donnée sur une plateforme unique. Certaines grandes entreprises, comme Kiloutou, recourent à la Talend Data Fabric pour structurer leurs processus d'intégration de données, les mettre à disposition des utilisateurs métiers et gagner en agilité. C'est aussi le cas du réseau de crèches Babilou qui a fait le choix de la plateforme Talend.

natures très différentes. Ainsi, la solution de data hub de l'éditeur Semarchy a été choisie par La Redoute pour gérer sa donnée produit et par l'industriel Imerys pour alimenter sa plateforme de Business Intelligence. "Nous réalisons déjà 30 % de notre chiffre d'affaires à l'international, mais nous avons pour objectif de le doubler d'ici à 2024 pour passer de 1 à 2 milliards €. Notre plateforme produit doit être en capacité de diffuser nos produits marketplace à l'international", explique Alexandre Dupont, chargé du projet à La Redoute. Le distributeur vient de déployer la nouvelle version de Semarchy xDM. "Nous voulions un data hub, c'est-à-dire une plateforme capable de faire converger des données issues de multiples sources, travailler sur cette donnée et l'exposer. Avec un data hub, nous voulions un seul point de vérité qui puisse être exploité par les consommateurs directement", détaille-t-il. Pour Imerys, leader mondial des spécialités minérales pour l'industrie, la problématique était tout autre : "La politique d'acquisitions engagée dans les années 2000 fait que



UNE PLATEFORME UNIQUE POUR DE MULTIPLES CAS

Interconnexion de systèmes, publications d'API, Business Intelligence, intelligence artificielle, les cas d'usage d'un même data hub sont divers et de

Nous voulions une plateforme capable de faire converger des données issues de multiples sources, travailler sur cette donnée et l'exposer. Avec un data hub, nous avons un seul point de vérité qui peut être exploité par les consommateurs directement."

Alexandre Dupont, IT Domain manager for Marketplace & Product Data Management à La Redoute.

le groupe compte actuellement 80 ERP différents, générant une complexité. Aujourd'hui, la volonté, c'est d'aller vers une simplification et une rationalisation de ces moyens. Le groupe s'est réorganisé en cinq activités et évolue vers un modèle plus centralisé, avec un suivi de la performance avec une plateforme de Business Intelligence transverse au groupe." Le projet de BI global est lancé au troisième trimestre 2019, mais se heurte rapidement au manque d'alignement des données entre ces différents SI. On décide donc de lancer a posteriori un projet de data management au deuxième trimestre 2020. Le data hub est maintenant considéré comme un accélérateur pour le futur déploiement de l'ERP groupe.

UN MOTEUR DE CROISSANCE DANS LE DIGITAL

Le data hub peut faciliter les échanges de données dans des organisations complexes, et aider le management en alimentant les plateformes décisionnelles. Il peut aussi approvisionner les algorithmes d'IA et porter de nouveaux business models. C'est notamment le cas de Schneider Electric qui s'est doté d'une data factory s'appuyant sur la solution Databricks. "La data factory Schneider Electric est articulée autour d'Azure Databricks et une architecture Lakehouse dont les bénéfices sont de rationaliser et simplifier l'infrastructure pour héberger les cas d'utilisation et les data pipeline", souligne Herveig Lecuyer, Data Factory Program Lead, Schneider Electric. "Cette plateforme favorise la collaboration entre les métiers de la donnée, qu'il s'agisse des ingénieurs de la donnée,



Florian Petitjean, Head of Data Flow Architecture de Babilou Family, réseau de creches privées présent dans 12 pays

"L'activité de Babilou a commencé en 2003 et, comme toute entreprise qui démarre, nous avons eu besoin d'équiper rapidement les directions métiers et de déployer de multiples outils. L'étape suivante a été de faire communiquer ces outils entre eux. Très vite, nous avons créé des liaisons point à point et des dépendances fortes entre applications. Nous avons souhaité passer de multiples SI locaux à un SI groupe. La méthode : faire converger et interconnecter tous ces SI locaux qui ont des systèmes hétérogènes. Pour adresser cette partie interconnexion et le volet Data Management, nous nous appuyons sur Talend. Notre vision cible est d'offrir un panel de services qui permet à n'importe quelle application de venir se connecter pour consommer et exploiter notre data. Des API Rest ont été mises en place à destination de nos partenaires, de nos prestataires, et de nos hubs internationaux dans une approche "API First". L'idée est de pouvoir très rapidement brancher un nouveau partenaire sur nos connecteurs en mode plug & play."

5 MILLIONS
d'assets industriels
alimentent le data hub
de Schneider Electric.

les data analysts, les data scientists." Le data hub est alimenté par les offres digitales de Schneider Electric, notamment les cinq millions d'assets industriels connectés à son cloud et des solutions tierces qui collectent et ingèrent de la donnée. En sortie, le data hub fournit des données transformées et valorisées qui sont délivrées aux offres digitales ou mises à disposition sous forme DaaS (Data as a Service) sur le portail Schneider Electric Exchange. Herveig Lecuyer conclut : "Un bénéfice important de la Data Factory est de nous abstraire la complexité de l'infrastructure avec une solution SaaS qui nous permet de nous concentrer sur les logiques métiers et notre cœur de métier."



Un data hub moderne doit disposer d'interfaces utilisateur suffisamment simples – ici, celle de Boomi –, pour être utilisées par les experts métiers amenés à travailler sur leurs données.