



TOUTE L'ACTUALITÉ / LOGICIEL / BIG DATA

La SNCF traque ses dépenses énergétiques

Dominique Filippone, publié le 12 Novembre 2021

Dans le cadre de son projet d'exploitation industrielle des données, la **SNCF** a mis en place un outil de management de l'énergie (OME) pour traiter et gérer les données liées à ses dépenses énergétiques. Avec à la clé des économies significatives.



« Chaque année la **SNCF** dépense plus d'un milliard d'euros en achat d'énergie au niveau du groupe », explique Benoît Muller, responsable Factory Big Data & IA Itnovem/SNCF. (crédit : **SNCF**)

Dans la galaxie des sociétés appartenant à la **SNCF**, Itnovem est la filiale technologique du groupe en charge de l'intégration des services numériques dédiée à la gestion de l'infrastructure ferroviaire. Cette dernière travaille notamment pour sa direction Energie, dont la mission est d'acheter les contrats de consommation d'électricité sur les marchés financiers. Parmi les enjeux à adresser : définir une feuille de route autour de la transition carbone et des moyens à mettre en oeuvre pour l'appliquer. Des bonnes pratiques sont ainsi poussées auprès des agents de la **SNCF** pour adopter les gestes d'écoconduite des rames motrices afin de dégager des économies. Objectif : maîtriser et optimiser la consommation énergétique du groupe pour accompagner sa transition vers un modèle décarboné vertueux débouchant aussi sur des économies financières.

Pour accompagner le déploiement de son outil de management de l'énergie (OME) pour traiter et gérer les données liées à ses dépenses énergétiques, la **SNCF** s'est appuyée sur l'entité Factory Big Data & IA. L'enjeu est très important pour la **SNCF** qui consomme chaque année 3% de la consommation nationale totale d'électricité en France. Sur la globalité de sa





propre consommation d'électricité, ce sont les trains qui constituent la part la plus importante (60%), le solde étant issu de ses gares, bâtiments, sites de maintenance... Un volume d'électricité qui a un coût : « Chaque année, nous dépensons plus d'un milliard d'euros en achat d'énergie au niveau du groupe », explique Benoît Muller, responsable Factory Big Data & IA Itnovem/[SNCF](#). Pour répondre à son triple objectif (réduction de l'empreinte carbone, de sa consommation énergétique et des coûts), la [SNCF](#) s'est donc attelée à avoir une bonne visibilité des équipements, matériels, services et processus les plus énergivores. « Il y a besoin d'avoir une cartographie fine et fiable de cette consommation pour être capable de concevoir et mettre en oeuvre des plans d'action et d'optimisation », poursuit Benoît Muller.

100 millions de lignes traitées par mois

C'est fin 2017 que la DSI de la direction Energie de la [SNCF](#) a mobilisé son entité Factory Data pour son projet d'OEM. À cette époque, les systèmes d'information en place n'étaient pas conçus pour répondre à cet enjeu de connaissance et de maîtrise des consommations. La nécessité de se doter d'un système ad hoc capable de traiter les données avec un niveau élevé de volumétrie et de complexité est donc apparue essentielle. « On avait des bases de données, mais pas d'Hadoop, l'enjeu a été de réussir à tirer parti des technologies big data pour répondre à la problématique big data », indique Benoît Muller. Au début, l'outil de management de l'énergie reposait sur une architecture Microsoft Azure articulée autour de HD Insight, de l'Hadoop IaaS pour le traitement, et SQL Datawarehouse pour le stockage et Qlik pour la couche datavisualisation. « On avait deux principaux composants principaux qui coûtaient assez cher pour une raison de volumétrie importante avec quasiment 100 millions de lignes à traiter chaque fin de mois nécessitant agrégation et retraitement chaque mois en tenant compte de beaucoup de paramètres customisables pour calculer les prix énergétiques », explique Alexandre Bergère, data architect Factory Big Data & IA Itnovem/[SNCF](#). « Il fallait rejouer l'historique sur 6 mois à un an et cela mettait beaucoup de temps avec une recette métier assez longue »

En parallèle, un projet d'infrastructure as code a été mené avec passage sur Teraform pour faciliter l'écriture de code et simplifier toute l'infrastructure sous-jacente pour répondre au besoin d'évolution d'OME pour recréer des environnements assez rapidement et mettre de l'autoscale pour répondre à la problématique de volumétrie importante à venir à laquelle l'infrastructure alors en place ne pouvait pas répondre. « On a choisi Databricks avec comme gros avantage d'enlever HD Insight pour migrer tout le code en Scala et utiliser le metastore Hive intégré sur lequel Qlik peut se connecter et lire directement les données pour accéder au datalake », poursuit Alexandre Bergère.

Un coût coupé en deux





Au final, le groupe a pu mettre un terme à une double facturation en termes de stockage, à la fois sur le datalake et la base de données BI et faire baisser la facture passée de 55 000 à 22 000 euros avec Databricks. « Cette réduction de coût a permis de mettre en place et payer notre passage vers l'infrastructure as code », indique Alexandre Bergère. En termes de volumétrie, près de 100 millions de lignes sont traités par mois par Databricks soit un total de plus d'un milliard depuis trois ans. Parmi les autres bénéfices, des gains pour les métiers ressortent également. Si l'application est complexe, les métiers vont pouvoir accéder plus rapidement aux calculs des coûts énergétiques qui passent de 6 heures auparavant à environ 1h30. Jusqu'alors, l'outil de management de l'énergie utilisait un cluster Databricks couplé au datastore Hive pour alimenter Qlik, mais des évolutions ont été pensées que ce dernier se charge beaucoup plus rapidement. Comment ? Via des nouveaux clusters SQL Analytics optimisés pour la partie BI afin de faire face aux futures croissances des usages de la plateforme.

Avant d'atteindre un tel niveau de maturité, des efforts ont été déployés et des difficultés ont dû être relevées. « Un premier point de complexité était lié pas tant à ce sujet de l'énergie, mais à la transformation numérique du groupe SNCF qui en était au début de l'histoire sur le fait de travailler sur le cloud Azure de Microsoft, l'utilisation de technologies big data et de recourir à des métiers en data engineer. Les 18 premiers mois du projet, nous avons fait face à des difficultés en étant collectivement en phase d'apprentissage », se rappelle Benoît Muller. Un exemple ? Lors de la mise en place de la plateforme Azure un premier flux de données a été mis en place pour être connecté au SI source ce qui n'était pas industrialisé à l'époque et a nécessité plusieurs mois avec notamment des problèmes de sécurité à résoudre. « Le fait d'avoir des data engineers qui comprennent mieux le métier de l'énergie en croisant les données avec les engins, derrière c'est du code, mais cela nécessite de comprendre bien les métiers et dialoguer facilement avec eux ce qui n'était pas gagné au début du projet et a mis du temps à se mettre en place », conclut Benoît Muller.

