

Un Entrepôt de Données pour l'Analyse de la Recharge des Véhicules Electriques: un Retour d'Expérience

Kevin Royer*, Ladjel Bellatreche**
Anne Le-Mouel*, Gilbert Schmitt*

* EDF-R&D
Département Eco-Efficacité et Procédés Industriels
Site des Renardières
77818 Moret sur Loing
kevin.royer,anne.le-mouel,gilbert.schmitt@edf.fr,
**LIAS/ENSMA
Téléport 2 - 1 Avenue Clément Ader
BP 40109
86961 Futuroscope Chasseneuil cedex
bellatreche@ensma.fr

Résumé. La technologie de l'entreposage de données a contribué largement au développement des entreprises de grande distribution (comme Walmart, Ohlinger (2006)). Dans cet article, nous présentons un retour d'expérience de développement d'entrepôt stockant des données issues des expérimentations véhicules électriques au sein de l'entreprise Electricité de France (EDF). Cet entrepôt constitue le coeur de la plateforme EDF-MINERVE¹ © chargée de collecter, stocker et d'analyser les données des expérimentations. Premièrement, nous présentons la chaîne d'acquisition des données et sa validation, puis nous détaillons nos attentes concernant l'exploitation des données. Finalement, nous exposons les raisons qui nous ont amenés à choisir la solution des entrepôts de données ainsi que sa mise en œuvre pour répondre à la problématique d'analyse temporelle multi-critères des données de charge.

1 Introduction

Le développement du véhicule électrique présente de nombreux enjeux, et requiert la collaboration de divers acteurs. EDF en tant que fournisseur d'électricité est l'un de ces acteurs majeurs. Le véhicule électrique est aujourd'hui, en France et dans certains pays d'Europe (EDF-R&D (2011)), à un stade expérimental.

Afin d'anticiper l'utilisation massive de ces véhicules, EDF R&D a pour mission d'analyser l'impact de l'usage des véhicules électriques sur la courbe de charge nationale, de façon à faciliter par la suite l'adéquation entre la demande et la production d'électricité. La difficulté

1. MINERVE : Moyens d'INformations dédiés aux ExpéRimentations de Véhicules Electriques.

Entrepôt de données et recharge des véhicules électriques.

de cette mission réside dans la méthode de capitalisation des informations accumulées pendant les expérimentations. Or cette capitalisation est cruciale pour comprendre et prédire le comportement des utilisateurs vis-à-vis de la recharge des véhicules. Ceci justifie l'intérêt que porte EDF R&D aux entrepôts de données, et particulièrement à leur composante temporelle, qui offre la perspective nécessaire pour mener à bien cette mission.

Ce papier présente la démarche adoptée pour concevoir et construire un entrepôt adapté au domaine d'étude. Il est organisé en six sections. La section 2 présente un état des lieux sur l'utilisation des entrepôts de données dans le monde industriel et académique. La section 3 introduit le domaine d'étude, les données et les fonctionnalités attendues de l'entrepôt. La section 4 présente les objectifs de la démarche. La section 5 développe la méthodologie employée qui nous a amenés à la solution entrepôt de données pour la plateforme EDF-MINERVE©. La section 6 conclut le papier.

2 Etat de l'art

Entrepôts de données. Les entrepôts de données sont apparus dans le monde industriel, parmi de grands distributeurs (e.g., Coca-Cola, Walmart). Leur développement a permis de donner un nouvel essor à la notion de *business intelligence*. Le développement de cette technique d'entreposage s'est ensuite poursuivi tant dans l'industrie qu'en recherche universitaire.

Longtemps les entrepôts de données sont restés l'apanage des problématiques du commerce et de certains services, comme les assurances. Depuis plusieurs années les entrepôts commencent à investir des nouveaux domaines comme la gestion de flot de données Bauzer-Medeiros et al. (2006) ou encore l'analyse de comportements plutôt que de produits.

Ce nouvel usage des entrepôts apporte de nouvelles problématiques. En effet, contrairement aux besoins des distributeurs, le choix des données à étudier et à agréger est moins immédiat dans la mesure où elles concernent davantage des concepts que des produits. Cette réflexion est importante car le degré de précision souhaité par les décideurs va influencer la structure de l'entrepôt et également son efficacité. Une manière de détailler un concept est de ramener un fait ou une dimension à un document XML (étude de Boukraâ D. et O. (2006)).

La recherche universitaire sur les entrepôts de données a apporté de nombreuses techniques de conception Romero et Abello (2010), de construction Golfarelli et Rizzi (2009), d'optimisation Chaudhuri (2004) et, plus récemment, des travaux sur l'importance des utilisateurs et la prise en compte de leurs attentes Golfarelli et al. (2011), Romero et al. (2011). La surabondance de données peut donc être gérée par des techniques de manipulation et de stockage des données. Toutefois, la construction d'un entrepôt à partir des besoins des utilisateurs permet de se passer de certaines de ces techniques, parfois coûteuses à utiliser, ce qui tend à démocratiser la construction d'entrepôts dans des contextes industriels ou académiques.

Positionnement. Ce travail sur la problématique de conceptualisation des faits par rapport aux besoins et le développement de l'entrepôt de données, se fait dans l'idée de mettre en place du contrôle-commande. Le schéma fonctionnel de ce procédé est illustré par la figure 1. De ce point de vue les entrepôts de données impliquent d'établir des prédictions afin d'introduire la notion de consigne.

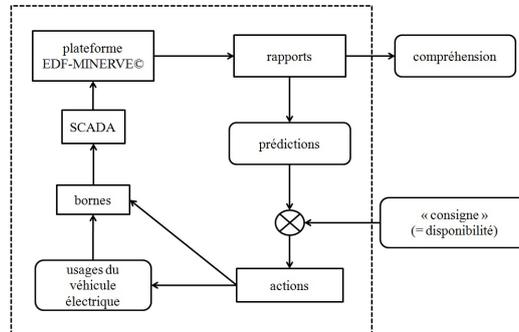


FIG. 1 – Illustration du contrôle-commande dans l'entrepôt de données de MINERVE.

3 Données et applications

Description du domaine. Pour conceptualiser les problématiques liées à la recharge des véhicules électriques différentes informations sont nécessaires :

- L'information principale relative à la recharge est l'énergie, exprimée en kWh, elle est mesurée par la borne servant à la recharge du véhicule grâce à un compteur.
- Cette information d'énergie doit être couplée à un repère temporel : les dates de début et de fin de charge.

Ces trois informations sont cruciales car toutes les analyses reposeront en premier lieu sur ce triplet <énergie, date de début de charge, date de fin de charge>. Ces triplets permettent d'établir une information stratégique : la courbe de charge liée à l'usage des véhicules électriques ainsi que des informations relatives au CO_2 émis par cet usage. Cette notion est de plus en plus médiatisée, notamment lors des pics de consommation en hiver. L'une des missions d'EDF étant de contribuer à maintenir l'équilibre entre l'offre et la demande d'électricité, la connaissance, la compréhension et, *a fortiori*, la possibilité d'influencer la courbe de charge nationale sont des concepts qui s'ancrent au cœur de ce travail.

Afin de poursuivre les analyses, d'autres informations sont requises :

- Le type d'usage du véhicule : les travaux menés par Mme Magali PIERRE, sociologue d'EDF R&D, ont permis de mettre en évidence plusieurs types de comportement.
- De la même façon, la connaissance de l'emplacement et des modalités d'accès à une borne permet d'affiner les analyses. On distingue les bornes à domicile des bornes sur la voie publique ou encore des bornes situées dans un parking d'entreprise.

Les études sociologiques (Pierre (2011) et Pierre M. et N. (2011)) ont permis d'établir une typologie regroupant les principaux usages du véhicule électrique. On différencie ainsi :

- les particuliers,
- les mono-utilisateurs d'un véhicule de service ou d'un véhicule de fonction,
- les multi-utilisateurs d'un véhicule de service ou d'un véhicule de fonction.

Entrepôt de données et recharge des véhicules électriques.

Recueil des données. Actuellement, deux modes d'acquisition sont utilisés. Le premier consiste à acquérir les données automatiquement depuis des serveurs (figure 2), le deuxième à charger, dans la base de données, des fichiers plats.

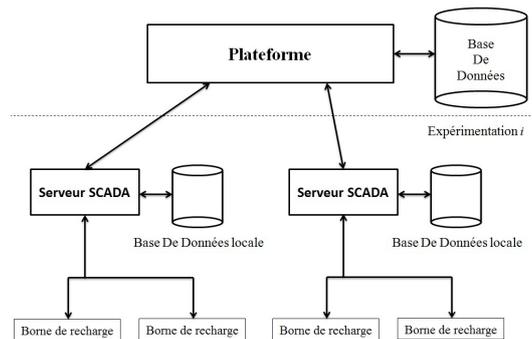


FIG. 2 – Schéma de récupération optimale des données depuis les bornes jusqu'à la plateforme.

L'objectif à terme est de n'utiliser que le premier mode d'acquisition. La plateforme EDF-MINERVE© chargée de centraliser les données exploite la technologie CubicWeb®. CubicWeb® est une plateforme logicielle de développement d'application Web, publiée sous la licence LGPL, qui permet aux développeurs de construire efficacement des applications Web en réutilisant des composants (appelés cubes) tout en respectant les principes de conception de la programmation orientée objet. On distingue plusieurs types d'interrogations mais toutes sont périodiques (avec différentes périodes) et se basent sur des services web utilisant un formalisme XML défini par EDF R&D, en attendant d'adopter une norme standard.

Validation de chaîne d'acquisition. Avant de basculer en mode d'acquisition automatique (sans chargement de fichiers plats), la chaîne d'acquisition doit être validée, depuis la borne de recharge jusqu'à la plateforme.

Une première série de tests consiste à raccorder à une borne de recharge un équipement dont on connaît la puissance à un moment précis : on connaîtra alors les horodates de branchement, de la mise en route, de l'arrêt et du débranchement. Cette série de tests permet de suivre la chaîne d'acquisition, dans le cadre d'un fonctionnement normal, depuis la borne jusqu'à la plateforme, afin de s'assurer, d'une part, que les grandeurs mesurées (énergies, horodates, ...) sont correctes et, d'autre part, qu'elles sont correctement acheminées jusqu'à la plateforme où elles seront stockées.

Une seconde série de tests s'attache à valider la chaîne de communication. La méthode consiste à interrompre la communication entre deux entités et, normalement, lorsque les données n'arrivent pas à la plateforme, celle-ci doit renouveler sa demande de remontée de données jusqu'au rétablissement de la communication. Ces tests doivent permettre de valider le comportement de la chaîne d'acquisition dans deux situations :

- la perte momentanée de communication qui empêche toute remontée de données,

- l’interruption soudaine des communications qui peut détériorer les données contenues dans les trames d’informations transmises à cet instant.

Pour remédier à ces deux situations, avant d’être transmises, les données sont stockées dans les bornes (stockage à la source), ce qui permet de les récupérer lors d’une transmission différée, qui intervient généralement quand la communication est rétablie. Pour finir, la plateforme vérifie la cohérence entre les trames d’informations reçues et le formalisme XML défini, ce qui permet d’écarter les trames incomplètes ou erronées.

Tous ces tests sont très importants, car, à l’usage, on remarque des pertes fréquentes de communication et parfois des erreurs de paramétrage des bornes. Il est donc essentiel, pour bénéficier de données de qualité, de s’assurer qu’elles seront bien mesurées et bien remontées à la plateforme. Gartner mentionne que 50% des projets d’entreposage de données échouent à cause de la mauvaise qualité des données (Gartner (2005)).

Tri. L’objectif de la plateforme étant d’obtenir des analyses pertinentes, la sélection des données est au centre des préoccupations. La validité de la chaîne d’acquisition n’est pas suffisante pour exploiter les données remontées, certaines données peuvent être douteuses (charge trop élevée ou trop faible par exemple). Il est nécessaire d’écarter de la plateforme les données suspectes. Pour cela, cette dernière dispose de deux types de tri :

- le premier vérifie la cohérence des données lorsqu’elles sont redondantes (par exemple : durée de charge = date de fin de charge - date de début de charge),
- le deuxième est défini à partir de notre retour d’expérience sur les premières expérimentations, ce tri permet d’écarter les charges jugées anormales.

4 Objectifs

4.1 Analyses

La capitalisation d’un grand nombre de données permet de poser diverses questions *via* des requêtes. La compréhension de l’usage du parc de véhicules électriques est la première étape de la démarche. Cette première analyse doit pouvoir se faire en tenant compte de critères sociologiques, tel que défini plus haut (type d’usage) et des infrastructures (bornes publiques, bornes de particuliers, bornes de parkings privés, etc.). Pour comprendre les différents comportements il faut pouvoir les étudier dans le temps, d’abord avec une granularité épaisse (par saisons, par mois) puis de plus en plus fine (par semaine, journée).

En plus des analyses ci-dessus, il est essentiel pour EDF de pouvoir construire la courbe de charge liée à l’usage du véhicule électrique (puissance appelée sur une période par les recharges avec un pas aussi fin que possible) et d’affiner l’analyse jusqu’à construire des courbes de charge par catégories d’usage. A titre d’exemple la figure 3 montre une courbe de charge nationale.

A ces considérations directement liées au cas d’étude on peut rajouter d’autres informations utiles pour la compréhension des comportements : des informations climatiques (température extérieure, pluviométrie), les conditions tarifaires, les kilomètres parcourus, etc. Ces informations permettront d’affiner les analyses effectuées.

Entrepôt de données et recharge des véhicules électriques.

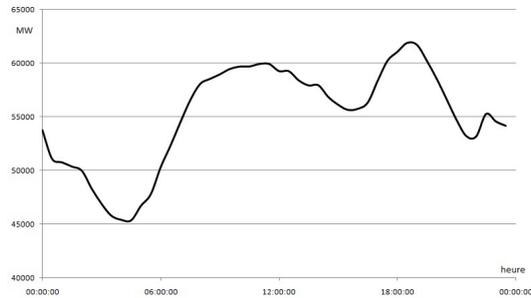


FIG. 3 – Courbe de charge nationale - extrait de www.rte-france.com, décembre 2011 à janvier 2012.

4.2 Prédiction

Après ces analyses, EDF R&D souhaite établir des modèles permettant de prédire le comportement de recharge des usagers des véhicules électriques et par conséquent leurs besoins énergétiques. Grâce à ces modèles prédictifs il sera possible d'envisager des moyens pour influencer le comportement de recharge des usagers des véhicules électriques. Si les expérimentations ont lieu, pour le moment, dans des zones où le réseau est robuste, le développement attendu du véhicule électrique ne bénéficiera pas d'un tel contexte au niveau national. On imagine facilement que si de nombreux foyers rechargent leur véhicule électrique sur le même nœud du réseau au même moment cela posera des problèmes de distribution. Les moyens qui devront être déployés auront pour but de lisser la courbe de charge : éviter les pics de consommation ou la participation aux pics tout en respectant les contraintes imposées par l'utilisateur (une voiture chargée à 100% le matin à 7h par exemple) ainsi que les caractéristiques des véhicules. Pour reprendre l'exemple de la courbe de charge nationale, plutôt que d'augmenter les pics il serait souhaitable de remplir les creux (la figure 4 illustre le résultat souhaité).

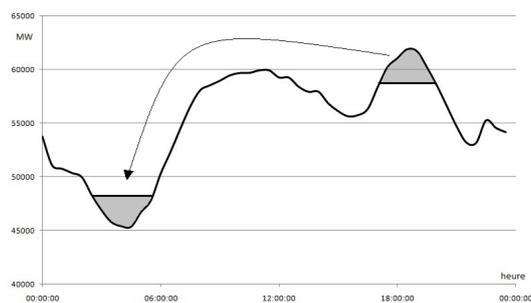


FIG. 4 – Illustration du résultat visé.

Dans la compréhension et la prédiction des comportements il y aura un volet important de compréhension *a posteriori*. Etant donnée la possibilité de garder une trace des évènements, la

mémoire de l'occurrence d'un évènement particulier impactant les comportements étudiés (par exemple : un évènement sportif important, une météo particulière. . .) permettra d'expliquer les comportements observés et d'améliorer les modèles de prévision.

5 Démarche

5.1 Entrepôt de données

Le concept d'entrepôt de données a été formalisé pour la première fois par Bill Inmon en 1990. Il s'agissait de constituer une base de données "orientée sujet, intégrée et contenant des informations historisées, non volatiles et exclusivement destinées aux processus d'aide à la décision."

La prise de décision dans une entreprise ne peut pas se limiter à la connaissance du présent. Le passé et les interactions doivent également être mémorisés. Ceci est une première problématique à laquelle doit répondre un entrepôt de données, une seconde problématique concerne la surabondance et l'éparpillement des données. Dans ce contexte il est nécessaire de rassembler et d'homogénéiser les données pour pouvoir les exploiter.

La modélisation dimensionnelle (ou modélisation OLAP²) se présente comme une alternative à la modélisation relationnelle car elle est plus proche des attentes des décideurs. Elle permet de visualiser une masse importante de données complexes. On distingue plusieurs raisons d'être aux entrepôts de données :

- rassembler les données en un même lieu, sans surcharger les bases de données,
- permettre un accès universel à diverses sources de données et assurer la qualité des données,
- extraire, filtrer et intégrer les informations pertinentes à l'avance pour des requêtes ultérieures,
- dégager des connaissances et faire un apprentissage sur l'entreprise, le marché et l'environnement.

La proximité de ces raisons avec nos objectifs nous a orientés vers le développement d'un entrepôt de données (figure 5), dont les caractéristiques sont décrites dans les paragraphes suivants.

5.2 Application

Données. Les données relatives aux premières expérimentations étaient stockées dans des fichiers plats ou dans des bases de données traditionnelles. Ces données étaient difficiles d'accès et leur capitalisation était complexe, lourde à mettre en œuvre et l'augmentation du volume de données devenait problématique. C'est pour cela que nous avons opté pour le développement d'un entrepôt de données. D'autre part, les mécanismes d'agrégation proposés dans les entrepôts de données sont parfaitement adaptés aux problèmes d'historisation des données.

Dans ce contexte, l'application des méthodes de construction et d'exploitation d'un entrepôt de données est bien adaptée, tant au niveau de l'agrégation des données qu'au niveau de la visualisation des indicateurs.

2. On-Line Analytical Processing

Entrepôt de données et recharge des véhicules électriques.

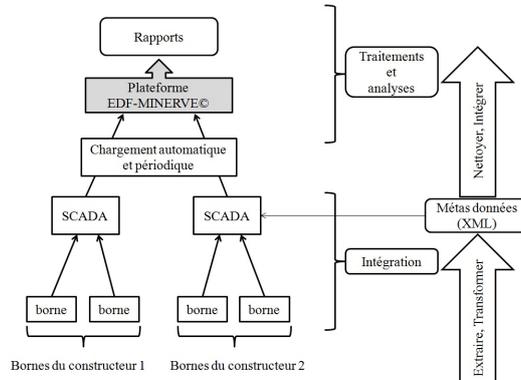


FIG. 5 – Peuplement de l’entrepôt de données.

Choix des dimensions. Le choix des dimensions va conditionner l’efficacité de l’entrepôt, sachant que le fait de base est la recharge et les paramètres et mesures qui lui sont associés. On peut distinguer trois dimensions majeures :

- La première dimension est le temps : l’entrepôt doit disposer d’un calendrier. En plus du découpage année/mois/jour il faut pouvoir discerner les jours de la semaine, les samedis, dimanches et jours fériés, ou encore les dates de : départs/retours de vacances, événements particuliers, etc.
- La deuxième dimension caractérise les types d’usage mis en évidence par les études sociologiques.
- La troisième dimension correspond à l’emplacement, aux modalités d’accès et la disponibilité des infrastructures de recharge.

La plateforme logicielle sur laquelle nous travaillons évolue rapidement. La figure 6 représente une partie du modèle de données. Cette vue partielle représente le champ d’action actuel de l’entrepôt.

Difficultés rencontrées. Le déploiement de cet entrepôt n’a pas été aisé, plusieurs problèmes sont apparus :

- grande hétérogénéité des sources,
- certaines informations sont lacunaires par rapport au modèle de données.

Vers les objectifs. A terme l’entrepôt de données devra permettre :

- de comprendre les comportements des utilisateurs des véhicules électriques dans tous les scénarios possibles,
- d’établir des modèles prédictifs de comportement,
- et d’affiner ces modèles en les alimentant, au fil de l’eau, par des données de qualité les plus nombreuses possibles.

Actuellement, le premier point est abordé *via* des requêtes réalisées directement sur les données (nombre de faits < 25 000). Toutefois des problématiques d’optimisation vont ap-

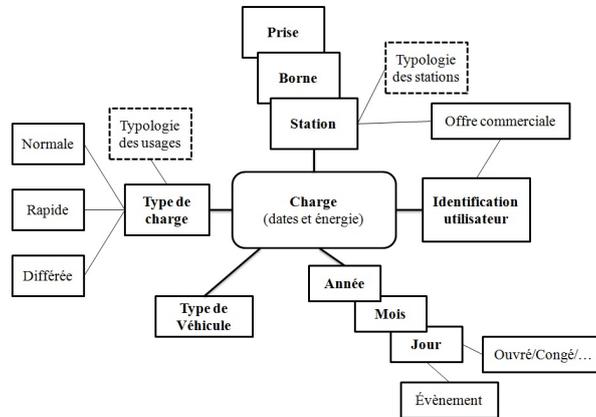


FIG. 6 – Champ d'action de l'entrepôt de données.

paraître avec l'augmentation du nombre de données. L'entrepôt permet d'anticiper cette problématique grâce aux techniques d'optimisation qu'il est possible de mettre en œuvre. Parmi ces techniques on peut citer : les vues matérialisées et les index. Nous allons mettre en place des vues matérialisées correspondant aux requêtes les plus fréquentes (les courbes de charges par usage par exemple). Pour le moment, nous consacrons nos efforts à l'analyse des données car la plateforme réagit rapidement à toutes les formes de requête, simples ou complexes, sur l'ensemble des données (quelques secondes pour les requêtes les plus complexes).

6 Conclusion et perspectives

Dans cet article nous avons développé notre démarche depuis la description et la formalisation de notre domaine d'étude jusqu'au choix de la structure de l'entrepôt. Les fonctionnalités attendues de la plateforme EDF-MINERVE© ont permis de définir précisément le domaine d'étude. Ces mêmes attentes nous ont permis de choisir une structure pour mener nos travaux d'analyse et de prédiction. On constate que la connaissance au préalable des besoins des utilisateurs de la plateforme permet de s'orienter rapidement vers une solution, dans notre cas un entrepôt de données, et que la définition précise du domaine facilite la création de la structure cet entrepôt.

La méthode de développement depuis nos objectifs vers la réalisation d'un entrepôt de données nous a permis de disposer d'un outil, la plateforme EDF-MINERVE©, performant pour analyser des données et pour comprendre et prédire des comportements.

L'exploitation de l'entrepôt de données sera amenée à revêtir différents aspects. La partie prédiction va nécessiter une analyse en profondeur et détaillée de l'entrepôt afin de mettre en place des modèles prédictifs. Ce travail est dans la suite logique de la conception et de la construction de l'entrepôt. Suite à ces travaux, nous nous intéresserons à la manière d'établir des modèles prédictifs globaux, en combinant les modèles prédictifs de consommation issus

Entrepôt de données et recharge des véhicules électriques.

des premières études de l'entrepôt avec des modèles d'affaires dont disposerait EDF pour influencer la recharge des véhicules électriques.

Références

- Bauzer-Medeiros, C., O. Carles, F. Devuyst, B. Hugueney, M. Joliveau, G. Jomier, M. Manouvrier, Y. Naija, G. Scemama, et L. Steffan (2006). Vers un entrepôt de données pour le trafic routier. In *2ème journée francophone sur les Entrepôts de Données et l'Analyse en ligne (EDA 2006)*, Versailles, Volume B-2, pp. 119–138.
- Boukraâ D., B. M. R. et B. O. (2006). Proposition d'un modèle physique pour les entrepôts xml. *Premier Atelier des Systèmes Décisionnels*.
- Chaudhuri, S. (2004). Index selection for databases : A hardness study and a principled heuristic solution. *16(11)*, 1313–1323.
- EDF-R&D (2011). EDF et la mobilité électrique. <http://innovation.edf.com/>.
- Gartner (2005). Gartner report release. <http://www.gartner.com/it/page.jsp?id=492112>.
- Golfarelli, M. et S. Rizzi (2009). *Data Warehouse Design : Modern Principles and Methodologies*. McGraw Hill.
- Golfarelli, M., S. Rizzi, et P. Biondi (2011). myolap : An approach to express and evaluate olap preferences. *IEEE TKDE 23(7)*, 1050–1064.
- Ohlinger, P. (2006). Walmart's data warehouse. <http://derbaum.com/tu/WalMarts%20DWH.pdf>.
- Pierre, M. (2011). Expérimenter le véhicule hybride rechargeable, une reconfiguration du système de mobilité ? *Congrès de l'Association Française de Sociologie*.
- Pierre M., J. C. et L. N. (2011). Driving an electric vehicle, a sociological analysis on pioneer users. *Energy Efficiency 4*, 511–522.
- Romero, O. et A. Abello (2010). Automatic validation of requirements to support multidimensional design. *Data Knowl. Eng. DKE 69(9)*, 917–942.
- Romero, O., P. Marcel, A. Abelló, V. Peralta, et L. Bellatreche (2011). Describing analytical sessions using a multidimensional algebra. In *DaWaK*, pp. 224–239.

Summary

After introducing the field of investigation (electrical vehicles and their charging needs) this paper exposes our approach from acquiring data to the definition of the expected functionalities and the architecture of the EDF-MINERVE© platform in charge of collecting and processing the data. The data acquisition system is presented, along with its validation. A focus is made on the data analysis expectations looked-for. This paper exposes our motivations to choose a data warehouse solution and its application to a time based and multiple criteria analysis of electrical vehicles charging events.