

# Améliorer la conception de moteurs et de véhicules avec **MATLAB** pour le Data Analytics sur les grands jeux de données d'ingénierie

## Introduction

Depuis de nombreuses années, les ingénieurs de l'industrie automobile analysent leurs données dans MATLAB® afin de gagner en performance et en efficacité. Les ingénieurs automobiles ont appris à utiliser MATLAB pour analyser des jeux de données collectées depuis des moteurs, transmission, châssis et pratiquement toutes les composantes d'un véhicule. Mais que se passe-t-il quand les données collectées atteignent des volumes de plusieurs téraoctets voire davantage, comme c'est souvent le cas pour les données issues de flotte de véhicules ?

MATLAB peut-il évoluer afin de répondre aux exigences de plus en plus élevées en matière d'analytique sur les données d'une flotte ? Est-il possible de développer des algorithmes d'analytique sur de grands volumes de données d'ingénierie et de les déployer dans un environnement de production à l'attention des data scientists, ingénieurs produits et analystes d'affaires ? La réponse à ces questions est tout simplement « Oui ».

Ce document explique comment MATLAB peut être utilisé pour extraire des informations clés à partir des données d'ingénierie en vue d'améliorer la conception des moteurs et des véhicules. Plus précisément, il décrit un système concret élaboré pour démontrer les capacités d'analytique sur de grands volumes de données d'ingénierie (« Big Engineering Data ») de MATLAB. Nous avons élaboré ce système pour collecter nos propres données, afin de pouvoir mieux rationaliser et perfectionner davantage les techniques d'application des analytiques MATLAB au Big Data. Ce système comprend une infrastructure de collecte des données afférentes à une flotte, ainsi qu'une application orientée Web d'analytique des données de flotte, intégrée et déployée avec MATLAB. Ce document se termine par des exemples d'enseignements qui peuvent être extraits à partir d'un tel environnement.

## Qu'est-ce que le Big Engineering Data et pourquoi est-ce difficile d'en retirer des informations exploitables ?

Tous secteurs d'activité confondus, le Big Data est généralement décrit en fonction des trois V : volume, variété et vitesse. (Un quatrième V, véracité, est fréquemment pris en compte. Ce point sera abordé plus tard dans le présent document.) Dans les applications d'ingénierie, nous utilisons le terme « Big Engineering Data » pour caractériser les données qui reflètent ces qualités tout en présentant un ensemble de problèmes subtilement différents. Les données sont généralement mieux structurées et plus variées, et elles présentent le potentiel d'évoluer beaucoup plus rapidement que le Big Data dans d'autres secteurs. Par exemple, dans le cas de données afférentes à une flotte :

- Le volume se rapporte à l'échelle des données. Pour les constructeurs automobiles, il est fréquent de travailler avec des jeux de données pouvant atteindre 20 To à la fois et de collecter plus de 30 Go de données par voiture et par jour. À ce rythme, les données peuvent facilement atteindre des volumes de l'ordre de plusieurs pétaoctets, les rendant impossibles à analyser en une fois. Même le système de démonstration le plus relativement modeste que nous ayons élaboré était capable de collecter 25 Mo de données de capteurs par jour et par voiture, et il a produit près de 2 Go de données sur plusieurs mois en s'intéressant à quelques conducteurs seulement. Ces chiffres ne comprennent pas les données vidéo, couramment utilisées dans les systèmes d'aide à la conduite, susceptibles d'amasser des données au rythme de plusieurs gigaoctets par heure.
- La variété reflète cette idée partagée par les analystes que la prise en compte de données provenant de nombreuses sources peut conduire à des enseignements plus intéressants et inattendus. Les véhicules d'aujourd'hui sont équipés de dizaines de capteurs, et les véhicules d'une flotte équipés d'instruments de suivi en comportent bien plus, tous générant une multitude de signaux,

notamment concernant la vitesse, la consommation de carburant et la température. Différents véhicules produisent différents types de données. Par exemple, les véhicules hybrides et électriques peuvent communiquer des informations sur la charge de la batterie plutôt que sur le débit de carburant. Dans notre système, les données du véhicule sont combinées et en phase chronologiquement en phase avec des données provenant d'autres capteurs, y compris les appareils GPS. Enfin, toutes ces données de séries chronologiques peuvent être complétées par des données de simulation, audio, vidéo, messages CAN, et autres.

- La vitesse, en termes de Big Data, décrit la rapidité à laquelle les données sont accumulées. Le système de démonstration MathWorks a collecté chaque seconde des données provenant de divers capteurs. En pratique, les capteurs automobiles avec des taux d'échantillonnage de 10 millisecondes génèrent une centaine de points de données par seconde. Les données émanant des multiples capteurs sur les nombreux véhicules font rapidement augmenter le volume collecté pour atteindre des gigaoctets ou des téraoctets d'informations.

Les entreprises qui cherchent à extraire de la valeur à partir du Big Engineering Data doivent disposer de systèmes capables de gérer ces données en termes de volume, de variété et de vitesse. Ces systèmes doivent posséder des algorithmes d'analytique flexibles et évolutifs, déployés dans un environnement de production, de sorte que les data scientists, les ingénieurs produits et les analystes d'affaires puissent tirer des enseignements de ces données.

## Élaboration d'un système d'analytiques Big Engineering Data à l'aide de MATLAB

L'équipe des services de conseil MathWorks, qui a travaillé avec de nombreux clients à la réalisation d'applications analytiques, voulait un cadre pour la collecte et l'analyse de Big Engineering Data que nous pourrions expérimenter, explorer, étudier et partager.

Il existe de nombreuses options d'architecture pour les applications MATLAB d'analytique de données. Le système que nous avons construit (Figure 1) commence par un adaptateur qui se raccorde au port OBD2 d'un véhicule et qui transmet les données sans fil via Bluetooth à un smartphone. Le téléphone transmet les données à une application Cloud s'exécutant sur une pile LAMR (Linux-Apache-MongoDB-Rails) sur Amazon EC2. Dans ce système, les véhicules agissent en tant que nœuds périphériques dans une solution d'Internet des objets (IoT), générant des données qui sont collectées et envoyées vers un agrégateur de données. Le système utilise l'infrastructure logicielle Hadoop basée sur Java pour simplifier le stockage et le traitement distribués de Big Engineering Data sur des clusters de calcul.

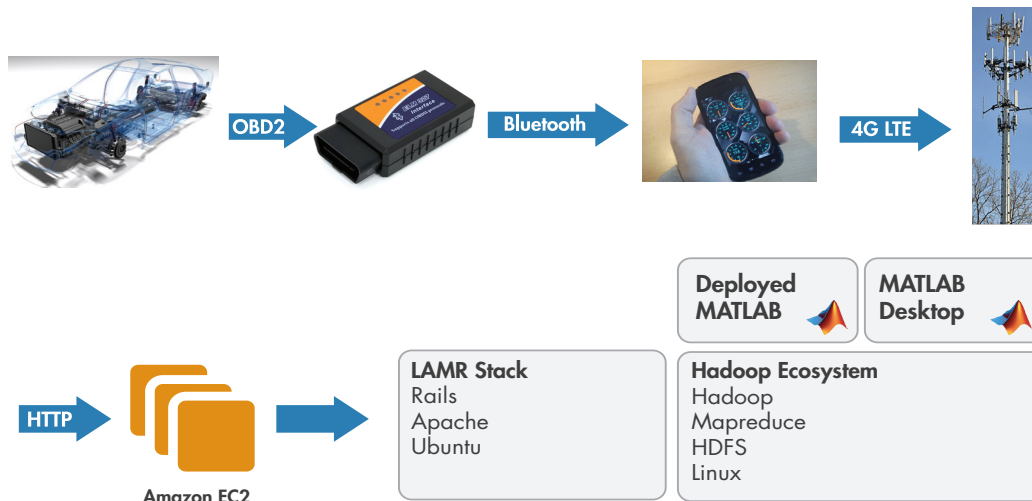


Figure 1. Infrastructure matérielle et logicielle permettant de collecter les données d'une flotte et d'effectuer des analytiques sur les données avec MATLAB et Hadoop.

Forts de cette infrastructure de collecte de données, nous avons utilisé un processus de travail basé sur MATLAB qui commence par des techniques d'analyse de données standard sur un seul ordinateur et s'achève avec le déploiement d'une application Web de production. Nous avons utilisé MATLAB sur ordinateur pour explorer les données que nous avons collectées et pour élaborer, tester et visualiser des idées concernant leur traitement. Suite à ce travail sur ordinateur, nous avons développé et déployé une application Web d'analyse des données d'une flotte qui associe la fonction analytique intégrée de MATLAB avec Hadoop, ainsi que d'autres technologies Web afin de permettre la visualisation, l'optimisation et l'analyse des caractéristiques de performances d'une flotte de véhicules (Figure 2).

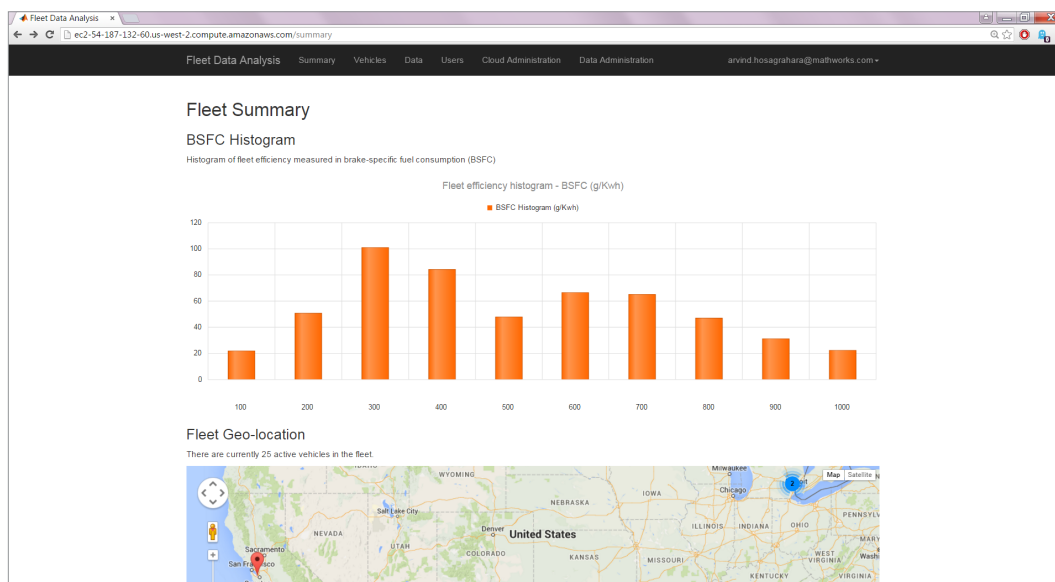


Figure 2. Application Web de Fleet Data Analytics basée sur MATLAB.

### MapReduce avec MATLAB et Hadoop

Le traitement de jeux de données Big Engineering Data impose plusieurs défis, notamment le fait qu'ils sont trop volumineux pour tenir dans la mémoire disponible et que leur analyse via un seul processeur prend trop de temps. La technique de programmation MapReduce répond à ces défis en traitant les données par petits blocs qui, individuellement, tiennent en mémoire ce qui permet de générer des résultats intermédiaires. Il suffit ensuite d'agréger ces résultats intermédiaires pour produire un résultat final. Hadoop MapReduce est une application populaire qui fonctionne avec le HDFS (Hadoop Distributed File System). MATLAB fournit sa propre implémentation de la technique de MapReduce.

Les analystes peuvent interactivement développer et déboguer des algorithmes à l'aide de l'association MapReduce/MATLAB sur leur ordinateur. Ils disposent alors de deux options pour exécuter leurs algorithmes, sans aucune modification du code, à l'aide de Hadoop. Ils peuvent utiliser MATLAB Distributed Computing Server™ pour exécuter les algorithmes basés sur MATLAB MapReduce dans Hadoop MapReduce sur les données stockées et gérées dans Hadoop. Ils peuvent également utiliser MATLAB Compiler™ pour créer des applications basées sur MATLAB MapReduce en vue d'un déploiement dans les instances de production de Hadoop.

Les techniques d'analyse de données traditionnelles impliquent généralement de transférer les données dans un environnement informatique pour y être analysées. Cette approche consistant à transférer les données vers les outils d'analyse peut ne pas être réalisable dans le cas du Big Engineering Data. Le support de Hadoop dans MATLAB permet aux équipes d'amener les outils d'analytique au niveau des données et de laisser les données Big Data là où elles sont stockées.

### Comment tirer des enseignements à partir des analytiques de Big Engineering Data

Les entreprises désireuses d'utiliser le Big Data pour générer de réelles améliorations dans la conception des moteurs et des véhicules devraient permettre à chacun de leurs salariés de pouvoir tirer des enseignements à partir des données qui ont été accumulées et stockées. Cette section décrit trois exemples d'enseignements obtenus suite aux analytiques réalisées sur des données à l'aide de MATLAB.

#### Compréhension de la consommation spécifique de freinage réelle

Les ingénieurs automobiles savent que même un moteur bien conçu ne peut pas fonctionner à son rendement maximal lorsqu'il est couplé à une transmission qui n'est pas réglée pour fonctionner avec ce moteur. Les mesures de la consommation spécifique de freinage (BSFC) réelle, qui reflètent non seulement l'efficacité énergétique, mais également la qualité du réglage d'un moteur, servent à optimiser les paramètres de la boîte de vitesses, les changements de rapports et autres paramètres de la transmission. Bien que la BSFC soit souvent considérée comme relevant d'un problème de simulation-optimisation, une approche axée sur les données fournit une perspective différente, donnant un aperçu des économies de carburant en fonction des habitudes de conduite réelles.

Nous avons utilisé les fonctions de Data Analytics de MATLAB pour étudier la BSFC sur environ 25 véhicules de la flotte échantillon, conduits par des personnes physiques au cours de leurs trajets quotidiens et autres déplacements. Dans MATLAB, nous avons tracé le couple en fonction du régime moteur et ajouté des hyperboles de puissance constante pour créer la représentation BSFC illustrée à la Figure 3.

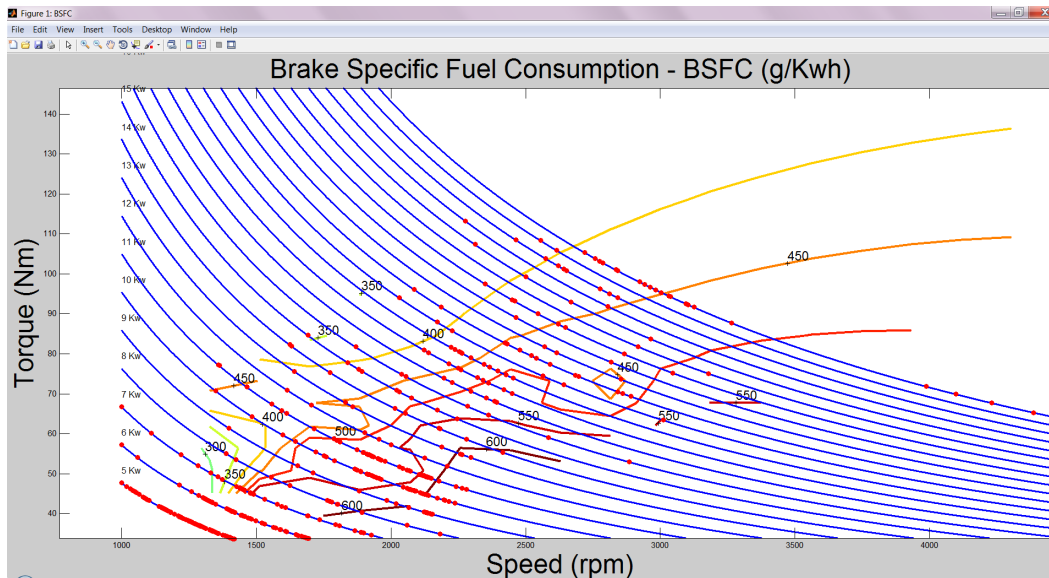


Figure 3. Représentation BSFC réalisée dans MATLAB à partir de données réelles d'une flotte.

Les points de données concrets utilisés pour créer la représentation BSFC comportent d'autres informations que simplement le régime moteur et le couple. Des données supplémentaires peuvent inclure la latitude, la longitude et l'altitude, afin que nous puissions déterminer l'endroit où se trouve n'importe quel véhicule de la flotte à un instant T alors même que le rendement de son moteur affiche un équilibre parfait. Les données supplémentaires peuvent également inclure des mesures de bruit et de vibration. Elles peuvent permettre de mieux comprendre comment les mesures visant à améliorer le rendement énergétique peuvent être entravées par des contraintes de bruit, vibration et rudesse (NVH).

#### Évaluation des modifications de l'infrastructure

Les ingénieurs experts en circulation routière peuvent utiliser les fonctions de Data Analytics de MATLAB sur les mêmes données de flotte pour identifier des modèles de trafic problématiques et évaluer des moyens de les résoudre. Par exemple, nous avons développé une fonction d'analyse pour déterminer les secteurs dans lesquels les véhicules de la flotte consommaient le plus de carburant. Nous avons intégré les analytiques de MATLAB à un logiciel tiers d'informatique décisionnelle pour réaliser rapidement un tableau de bord Web interactif (Figure 4). Les interactions de l'utilisateur avec le tableau de bord déclenchent des appels à du code MATLAB pour recalculer les résultats, qui sont ensuite utilisés pour mettre à jour les composantes visuelles du tableau de bord.



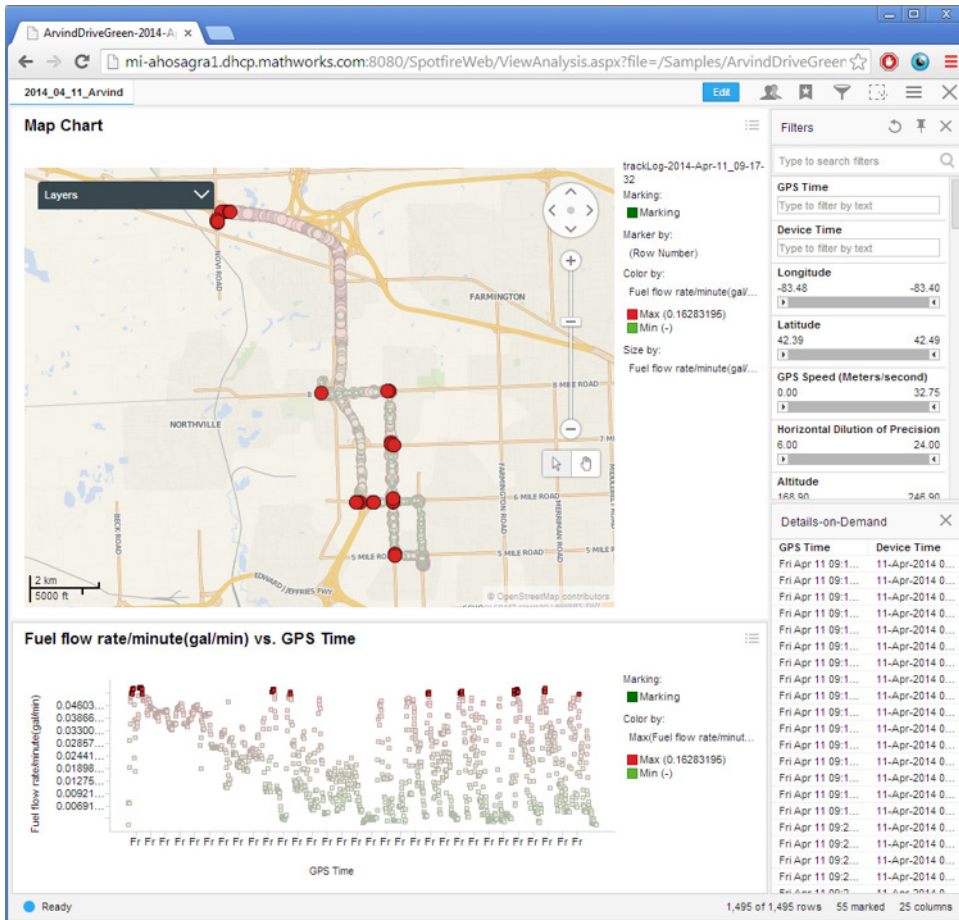


Figure 4. Tableau de bord Web interactif créé et dynamisé par MATLAB et un logiciel d'informatique décisionnelle.

Ce tableau de bord a mis en évidence plusieurs intersections où de longs moments d'inactivité ont conduit à une consommation de carburant particulièrement inefficace. En analysant plus avant les données dans MATLAB, nous avons déterminé la quantité de carburant consommée par un véhicule patientant près de 90 secondes à l'une des intersections posant problème. Sur la base du résultat observé (0,13 litres) par véhicule, on a estimé que le remplacement de l'intersection par un rond-point permettrait d'économiser plus de 450 litres de carburant par jour pour les conducteurs locaux et de réduire les émissions de CO<sub>2</sub> d'environ 20 400 tonnes par an. Pour un constructeur automobile, une telle analyse pourrait servir aux ingénieurs à optimiser les caractéristiques des véhicules, par exemple dans le cas d'un système Start/Stop, en s'appuyant sur des modèles de trafic concrets.

### Compréhension de la véricité

Le volume, la variété et la vitesse sont des notions essentielles dans le cadre de toute discussion sur le Big Data. Les entreprises ont reconnu que la compréhension de la véricité de leurs données est primordiale afin d'éviter de tirer des conclusions erronées à partir de l'analytique des données. Considérons, par exemple, une équipe d'ingénieurs automobiles qui crée des modèles de pronostics en les enrichissant à partir des données de flottes. Si les données utilisées pour enrichir les modèles sont erronées, alors les décisions métiers basées sur ces modèles seront sans fondement. De même,

une équipe chargée d'alimenter des réseaux pour des applications de Machine Learning serait confrontée à un défi comparable, car un modèle de classificateur enrichi à partir de données erronées ne fonctionnera pas comme prévu.

L'un des conducteurs de notre flotte avait demandé une modification du logiciel après-vente qui permette un couple maximal plus élevé, ainsi qu'une pression de suralimentation augmentée et un temps d'allumage optimisé. À la conduite, l'augmentation du couple occasionnée par la modification apportée était manifeste en comparaison avec les paramètres de base. À l'examen des données, nous n'avons cependant pas constaté de différence notable entre les paramètres. Nous avons réalisé que le logiciel après-vente conditionnait les données : autrement dit, nos données nous mentaient. La découverte que certaines données collectées sur le terrain peuvent être fausses constitue en soi un précieux enseignement, car elle peut permettre aux entreprises d'éviter de prendre des décisions de design biaisées car fondées sur ces données erronées.

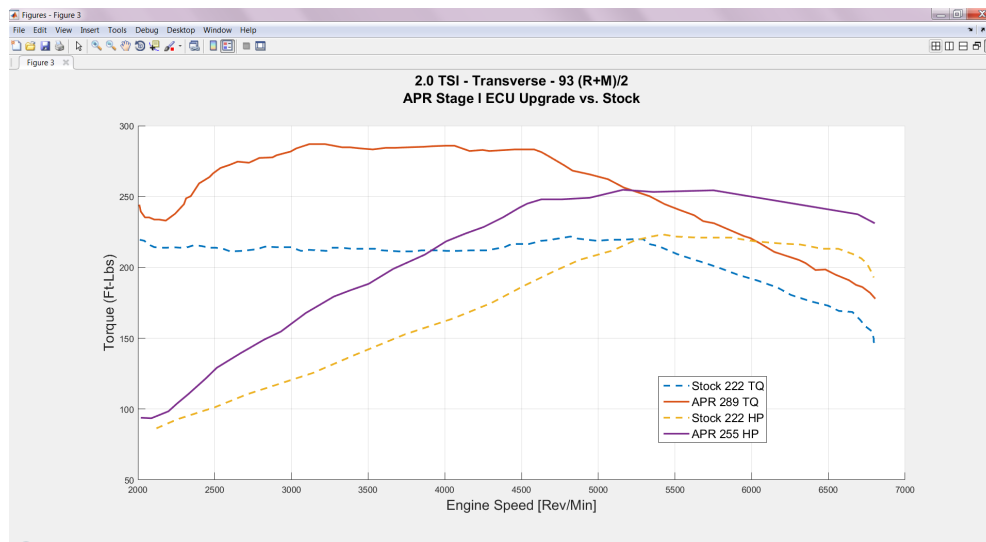


Figure 5. Estimation du gain de performance entre les paramètres de base et les paramètres optimisés de l'ECU

## Conclusion

Dans un large éventail de cas d'usage dans le domaine automobile, MATLAB fournit une plate-forme ouverte et extensible de Data Analytics qui s'appuie sur un processus de travail éprouvé, contribuant à rapidement développer, affiner et déployer des applications d'analytiques de données. MathWorks continue d'investir dans le Big Data et ses analytiques et de s'appuyer sur les interactions existantes entre MATLAB et les serveurs de données, MapReduce, Hadoop et les technologies connexes. Pour quiconque recherche des conseils d'experts et à être rapidement opérationnel, l'équipe Consulting de MathWorks est disponible pour aider les entreprises à améliorer leur capacité à extraire de précieux enseignements à partir des analytiques de du Big Engineering Data.

À voir Arvind Hosagrahara, consultant MathWorks, résume les résultats d'un projet de recherche mené sur toute l'année visant à mettre en œuvre un système d'analytiques de données d'une flotte de véhicules à titre de test, afin de dégager des enseignements intéressants à partir de données techniques pour différents cas d'usage.