

## Hadoop Par La Pratique

Des bases pour la performance et le Big Data En quelques années, le volume des données brassées par les entreprises a considérablement augmenté. Émanant de sources diverses (transactions, comportements, réseaux sociaux, géolocalisation...), elles sont souvent structurées autour d'un seul point d'entrée, la clé, et susceptibles de croître très rapidement. Autant de caractéristiques qui les rendent très difficiles à traiter avec des outils classiques de gestion de données. Par ailleurs, l'analyse de grands volumes de données, ce qu'on appelle le Big Data, défie également les moteurs de bases de données traditionnels. C'est pour répondre à ces différentes problématiques que sont nées les bases de données NoSQL (Not Only SQL), sous l'impulsion de grands acteurs du Web comme Facebook ou Google, qui les avaient développées à l'origine pour leurs besoins propres. Grâce à leur flexibilité et leur souplesse, ces bases non relationnelles permettent en effet de gérer de gros volumes de données hétérogènes sur un ensemble de serveurs de stockage distribués, avec une capacité de montée en charge très élevée. Elles peuvent aussi fournir des accès de paires clé-valeur en mémoire avec une très grande célérité. Réservées jusqu'à peu à une minorité, elles tendent aujourd'hui à se poser en complément du modèle relationnel qui dominait le marché depuis plus de 30 ans. Du choix de la base NoSQL à sa mise en oeuvre Cet ouvrage d'une grande clarté dresse un panorama complet des bases de données NoSQL, en analysant en toute objectivité leurs avantages et inconvénients. Dans une première partie, il présente les grands principes de ces bases non relationnelles : interface avec le code client, architecture distribuée, paradigme MapReduce, etc. Il détaille ensuite dans une deuxième partie les principales solutions existantes (les solutions de Big Data autour de Hadoop, MongoDB, Cassandra, Couchbase Server...), en précisant spécificités, forces et faiblesses de chacune. Complétée par une étude de cas réel, la dernière partie du livre est consacrée au déploiement concret de ces bases : dans quel cas passer au NoSQL ? quelle base adopter selon ses besoins ? quelles données basculer en NoSQL ? comment mettre en place une telle base ? comment la maintenir et superviser ses performances ? À qui s'adresse cet ouvrage ? Aux experts en bases de données, architectes logiciels, développeurs... Aux chefs de projet qui s'interrogent sur le passage au NoSQL

Un ouvrage de référence pour les (futurs) data scientists Les bibliothèques, les frameworks, les modules et les boîtes à outils sont parfaits pour faire de la data science. Ils sont aussi un bon moyen de plonger dans cette discipline sans la comprendre. Dans cet ouvrage, vous apprendrez comment fonctionnent les outils et algorithmes les plus fondamentaux de la data science, en les réalisant à partir de zéro. Si vous êtes fort en maths et que vous connaissez la programmation, l'auteur, Joel Grus, vous aidera à vous familiariser avec les maths et les statistiques qui sont au coeur de la data science et avec les compétences informatiques indispensables pour démarrer comme data scientist. La profusion des données d'aujourd'hui contient les réponses aux questions que personne n'a encore pensé à poser. Ce livre vous enseigne comment obtenir ces réponses. La deuxième édition, revue et augmentée, de cet ouvrage comporte des codes et exemples associés entièrement réécrits en Python 3.6 et intègre de nouveaux chapitres sur l'apprentissage profond (deep learning), les statistiques et le traitement en langage naturel. Suivez un cours accéléré de Python. Apprenez les fondamentaux de l'algèbre linéaire, des statistiques et des probabilités, et comprenez comment et quand les utiliser en data science. Collectez, explorez, nettoyez, bricolez et manipulez les données. Plongez dans les bases de l'apprentissage automatique. Implémentez des modèles comme les k plus proches voisins, la classification naïve bayésienne, les régressions linéaire ou logistique, les arbres de décision, les réseaux neuronaux et le clustering. Explorez les systèmes de recommandation, le traitement du langage naturel, l'analyse de réseau, MapReduce et les bases de données. À qui s'adresse cet ouvrage ? Aux développeurs, statisticiens, étudiants et chefs de projet ayant à résoudre des problèmes de data science. Aux data scientists, mais aussi à toute personne curieuse d'avoir une vue d'ensemble de l'état de l'art de ce métier du futur.

Written in a friendly, example-driven Beginner's Guide format, there are plenty of step-by-step instructions and examples that are designed to help you get started with RavenDB. If you are a .NET developer, new to document-oriented databases, and you wish to learn how to build applications using NoSQL databases, then this book is for you. Experience with relational database systems will be helpful, but not necessary.

Although interest in machine learning has reached a high point, lofty expectations often scuttle projects before they get very far. How can machine learning—especially deep neural networks—make a real difference in your organization? This hands-on guide not only provides the most practical information available on the subject, but also helps you get started building efficient deep learning networks. Authors Adam Gibson and Josh Patterson provide theory on deep learning before introducing their open-source Deeplearning4j (DL4J) library for developing production-class workflows. Through real-world examples, you'll learn methods and strategies for training deep network architectures and running deep learning workflows on Spark and Hadoop with DL4J. Dive into machine learning concepts in general, as well as deep learning in particular Understand how deep networks evolved from neural network fundamentals Explore the major deep network architectures, including Convolutional and Recurrent Learn how to map specific deep networks to the right problem Walk through the fundamentals of tuning general neural networks and specific deep network architectures Use vectorization techniques for different data types with DataVec, DL4J's workflow tool Learn how to use DL4J natively on Spark and Hadoop

Un ouvrage de référence pour les (futurs) data scientists Les bibliothèques, les frameworks, les modules et les boîtes à outils sont parfaits pour faire de la data science. Ils sont aussi un bon moyen de plonger dans la discipline sans comprendre la data science. Dans cet ouvrage, vous apprendrez comment fonctionnent les outils et algorithmes les plus fondamentaux de la data science, en les réalisant à partir de zéro. Si vous êtes fort en maths et que vous connaissez la programmation, l'auteur, Joël Grus, vous aidera à vous familiariser avec les maths et les statistiques qui sont au coeur de la data science et à acquérir les compétences informatiques indispensables pour démarrer comme data scientist. La profusion des données d'aujourd'hui contient les réponses aux questions que personne n'a encore pensé à poser. Ce livre vous enseigne comment obtenir ces réponses. Suivez un cours accéléré de Python Apprenez les fondamentaux de l'algèbre linéaire, des statistiques et des probabilités, et comprenez comment et quand les utiliser en data science Collectez, explorez, nettoyez, bricolez et manipulez les données Plongez dans les bases de l'apprentissage automatique Implémentez des modèles comme les k plus proches voisins, le Bayes naïf, les régressions linéaire ou logistique, les arbres de décision, les réseaux neuronaux et le clustering Explorez les systèmes de recommandation, le traitement du langage naturel, l'analyse de réseau, MapReduce et les bases de données À qui s'adresse cet ouvrage ? Aux développeurs, statisticiens, étudiants et chefs de projet ayant à résoudre des problèmes de data science. Aux data scientists, mais aussi à toute personne curieuse d'avoir une vue d'ensemble de l'état de l'art de ce métier du futur.

Depuis 2012, la notion de Big Data ne cesse de se populariser, reflétant la réalité des entreprises confrontées à une masse de données croissante, qui représentent un enjeu commercial et marketing. Comment récolter ces données ? Comment les traiter et les analyser ? Quels outils techniques mettent en place ? Comment s'en servir ensuite ? Comment les monétiser et les rentabiliser ? Telles sont les questions auxquelles répond cette boîte à outils concrète et pratique.

Le pouvoir change de mains. Qu'il s'agisse de la famille, des marchés, de l'entreprise ou du gouvernement, notre société se métamorphose. Notre façon de communiquer modifie en profondeur nos comportements, notre conception de la liberté, de l'identité ou de la hiérarchie. Toutes les générations sont devenues « digitales ». Stéphane Richard nous propose ici une analyse de ces transformations technologiques et de leurs conséquences sociales, économiques et anthropologiques.

Le développement spectaculaire d'internet, des réseaux sociaux, de la technologie mobile et la multiplication des capteurs provoquent une croissance exponentielle des données à laquelle les entreprises doivent faire face : c'est le phénomène du Big Data. Ses enjeux sont

considérables. Au-delà de la simple question technique du stockage, il offre la possibilité de tirer profit du contenu de ces nouvelles sources d'information. Les solutions décisionnelles classiques laissent progressivement place au Business Analytics et aux méthodes prédictives, transformant l'avalanche de données en valeur ajoutée. La technologie est aujourd'hui disponible, les bases de données traditionnelles ont évolué et les solutions dédiées à l'exploitation des données massives, telles que Hadoop, sont désormais opérationnelles. S'appuyant sur différents cas pratiques, Enjeux et usages du Big Data met l'accent sur les méthodes, les techniques et les ressources nécessaires pour permettre aux entreprises d'entrer avec succès dans l'ère de l'information à grande échelle.

If you're considering R for statistical computing and data visualization, this book provides a quick and practical guide to just about everything you can do with the open source R language and software environment. You'll learn how to write R functions and use R packages to help you prepare, visualize, and analyze data. Author Joseph Adler illustrates each process with a wealth of examples from medicine, business, and sports. Updated for R 2.14 and 2.15, this second edition includes new and expanded chapters on R performance, the ggplot2 data visualization package, and parallel R computing with Hadoop. Get started quickly with an R tutorial and hundreds of examples Explore R syntax, objects, and other language details Find thousands of user-contributed R packages online, including Bioconductor Learn how to use R to prepare data for analysis Visualize your data with R's graphics, lattice, and ggplot2 packages Use R to calculate statistical tests, fit models, and compute probability distributions Speed up intensive computations by writing parallel R programs for Hadoop Get a complete desktop reference to R

Data et publicité : où en êtes-vous ? J'utilise des données de différentes natures pour accroître la performance de ma publicité digitale. J'automatise une partie de mes investissements display en passant par des plateformes programmatiques. Je tire parti des fonctionnalités de différentes briques ad tech pour toucher les bons consommateurs au bon moment et au meilleur coût. Je fais disparaître les frontières entre les logiques publicitaires et celles héritées du CRM. Vous avez coché moins de deux propositions ? Vous voulez vous lancer ? Ce livre est fait pour vous ! Ad tech et programmation révolutionnent la publicité digitale. Dans cet ouvrage, 15 professionnels et universitaires travaillant au sein d'organismes leaders dans le domaine démystifient cette révolution. Ils en livrent les grands principes et les recettes opérationnelles dans des termes accessibles, afin d'améliorer l'efficacité de vos campagnes publicitaires.

Les data sciences constituent une discipline fascinante, qui vous servira à transformer des données brutes en idées et connaissances aisément compréhensibles. L'objectif de ce livre est de vous aider à maîtriser les outils essentiels qui vous permettront d'utiliser R dans la pratique des data sciences. Après l'avoir lu, vous pourrez faire face à la plupart des situations que vous rencontrerez dans vos projets, en exploitant au mieux les fonctionnalités de R. Mais avant de devenir un expert en R, vous devrez tout d'abord importer vos données, c'est-à-dire les lire, en général depuis un fichier, une base de données ou une API web, et les charger dans un cadre de données dans R. Si vous ne pouvez pas transférer vos données dans R, vous ne pourrez pas les analyser ! Une fois vos données importées, vous gagnerez beaucoup à les ranger. Une fois vos données rangées, vous passerez bien souvent par une étape de transformation. Une fois vos données rangées avec les variables dont vous avez besoin, la génération de connaissances reposera principalement sur deux moteurs : la visualisation et la modélisation. Leurs forces et faiblesses sont complémentaires, et toute bonne analyse doit tenir compte des deux. Inutile d'être un programmeur expert pour être un bon analyste de données, mais apprendre à programmer

Le phénomène Big Data bouleverse l'univers économique. Rien qu'en 2012, plus de 2,8 zettaoctets (des milliards de gigaoctets !) de volumes de données auraient été échangés ! Seules les entreprises capables de comprendre et de maîtriser une telle quantité de données dans leur complexité prendront l'avantage. Car au-delà du volume, ce qui va être décisif, est la capacité à structurer, analyser et transformer des données en connaissances, en innovations et en valeurs commerciales pour l'organisation. Au-delà du buzz qu'il génère, quelles applications concrètes le Big Data peut-il apporter au monde des affaires ? Comment mettre en place une stratégie Big Data dans l'entreprise et comment tirer profit de cette puissante ressource ? Cet ouvrage permet d'identifier les atouts du Big Data et montre notamment comment : • améliorer la prise de décision • consolider les relations client • créer de la valeur ajoutée • enrichir l'expérience utilisateur • amplifier le retour sur investissement • réduire les coûts et les temps de réalisation • développer de nouvelles offres Un guide pratique enrichi de nombreux cas concrets (Google, LinkedIn, eBay, Amazon, GE, UPS) et de plans d'action, qui permettra à tous les managers de tirer parti de cette formidable opportunité. Archiver est devenu un processus de plus en plus complexe. Le défi n'est plus de pouvoir stocker les données, mais de les conserver de façon intelligente afin de les exploiter dans le temps, tout en maintenant leur intégrité et leur authenticité. Les technologies numériques entraînent de grandes transformations, que ce soit au niveau des types de documents versés et conservés aux services d'archives, des comportements et des pratiques des sciences humaines et sociales (humanités numériques), mais aussi du volume des données et des capacités technologiques de gestion et de conservation des archives (Big Data). Les archives à l'ère du numérique met en valeur ces différentes transformations numériques sur les archives et examine comment le droit de mémoire et d'information des générations futures est confronté au droit à l'oubli, prérogative numérique qui garantit aux individus leurs vies privées et leurs libertés.

Docker est une technologie open source qui a pour but de faciliter le déploiement d'une application ainsi que la gestion de l'infrastructure sous-jacente. L'application est embarquée dans un conteneur virtuel qui pourra s'exécuter sur n'importe quel serveur (Linux ou Windows). Le but de cet ouvrage est d'offrir une approche pratique à Docker et à l'écosystème technologique qui l'accompagne. Autour du Docker Engine qui exécute les conteneurs il existe plusieurs outils complémentaires pour conditionner, assembler et orchestrer les applications (CoreOS, Apache Mesos, RKT, Kubernetes...).

Explore big data concepts, platforms, analytics, and their applications using the power of Hadoop 3 Key Features Learn Hadoop 3 to build effective big data analytics solutions on-premise and on cloud Integrate Hadoop with other big data tools such as R, Python, Apache Spark, and Apache Flink Exploit big data using Hadoop 3 with real-world examples Book Description Apache Hadoop is the most popular platform for big data processing, and can be combined with a host of other big data tools to build powerful analytics solutions. Big Data Analytics with Hadoop 3 shows you how to do just that, by providing insights into the software as well as its benefits with the help of practical examples. Once you have taken a tour of Hadoop 3's latest features, you will get an overview of HDFS, MapReduce, and YARN, and how they enable faster, more efficient big data processing. You will then move on to learning how to integrate Hadoop with the open source tools, such as Python and R, to analyze and visualize data and perform statistical computing on big data. As you get acquainted with all this, you will explore how to use Hadoop 3 with Apache Spark and Apache Flink for real-time data analytics and stream processing. In addition to this, you will understand how to use Hadoop to build analytics solutions on the cloud and an end-to-end pipeline to perform big data analysis using practical use cases. By the end of this book, you will be well-versed with the analytical capabilities of the Hadoop ecosystem. You will be able to



build powerful solutions to perform big data analytics and get insight effortlessly. What you will learn Explore the new features of Hadoop 3 along with HDFS, YARN, and MapReduce Get well-versed with the analytical capabilities of Hadoop ecosystem using practical examples Integrate Hadoop with R and Python for more efficient big data processing Learn to use Hadoop with Apache Spark and Apache Flink for real-time data analytics Set up a Hadoop cluster on AWS cloud Perform big data analytics on AWS using Elastic Map Reduce Who this book is for Big Data Analytics with Hadoop 3 is for you if you are looking to build high-performance analytics solutions for your enterprise or business using Hadoop 3's powerful features, or you're new to big data analytics. A basic understanding of the Java programming language is required.

Cet ouvrage s'adresse à tous ceux qui cherchent à tirer parti de l'énorme potentiel des « technologies Big Data », qu'ils soient data scientists, DSI, chefs de projets ou spécialistes métier. Le Big Data s'est imposé comme une innovation majeure pour toutes les entreprises qui cherchent à construire un avantage concurrentiel grâce à l'exploitation de leurs données clients, fournisseurs, produits, processus, machines, etc. Mais quelle solution technique choisir ? Quelles compétences métier développer au sein de la DSI ? Ce livre est un guide pour comprendre les enjeux d'un projet Big Data, en appréhender les concepts sous-jacents (en particulier le Machine Learning) et acquérir les compétences nécessaires à la mise en place d'un data lab. Il combine la présentation : • de notions théoriques (traitement statistique des données, calcul distribué...) ; • des outils les plus répandus (écosystème Hadoop, Storm...) ; • d'exemples d'applications ; • d'une organisation typique d'un projet de data science. Les ajouts de cette troisième édition concernent principalement la vision d'architecture d'entreprise, nécessaire pour intégrer les innovations du Big Data au sein des organisations, et le Deep Learning pour le NLP (Natural Language Processing, qui est l'un des domaines de l'intelligence artificielle qui a le plus progressé récemment).

Le manuel d'apprentissage de référence Cet ouvrage est un manuel d'apprentissage technique qui a été rédigé pour toute personne souhaitant développer des compétences sur une ou plusieurs technologie(s) de l'écosystème Hadoop. Il permet d'utiliser de façon professionnelle 18 technologies clés de l'écosystème Hadoop : Spark, Hive, Pig, Impala, Elasticsearch, HBase, Lucene, HAWQ, MapReduce, Mahout, HAMA, Tez, Phoenix, YARN, ZooKeeper, Storm, Oozie et Sqoop. L'ouvrage permet d'initier les débutants pour les emmener vers une utilisation professionnelle de ces technologies. Pour faciliter la compréhension de l'ouvrage, chaque chapitre s'achève par un rappel des points clés et un guide d'étude qui permettent au lecteur de consolider ses acquis. Des compléments web sont également disponibles en téléchargement sur le site [www.editions-eyrolles.com/dl/0067478](http://www.editions-eyrolles.com/dl/0067478). Au fil de la lecture de cet ouvrage, vous allez comprendre les approches conceptuelles de chacune de ces technologies pour rendre vos compétences indépendantes de l'évolution d'Hadoop. Vous serez finalement capable d'identifier les portées fonctionnelle, stratégique et managériale de chacune de ces technologies. À qui cet ouvrage s'adresse-t-il ? Aux consultants BI/big data, data scientists, chargés d'études et chefs de projets data Aux étudiants désireux de s'orienter vers le big data Plus généralement, à tout professionnel souhaitant prendre le virage du big data ou souhaitant valoriser les données de son entreprise

The best-selling author of Big Data is back, this time with a unique and in-depth insight into how specific companies use big data. Big data is on the tip of everyone's tongue. Everyone understands its power and importance, but many fail to grasp the actionable steps and resources required to utilise it effectively. This book fills the knowledge gap by showing how major companies are using big data every day, from an up-close, on-the-ground perspective. From technology, media and retail, to sport teams, government agencies and financial institutions, learn the actual strategies and processes being used to learn about customers, improve manufacturing, spur innovation, improve safety and so much more. Organised for easy dip-in navigation, each chapter follows the same structure to give you the information you need quickly. For each company profiled, learn what data was used, what problem it solved and the processes put it place to make it practical, as well as the technical details, challenges and lessons learned from each unique scenario. Learn how predictive analytics helps Amazon, Target, John Deere and Apple understand their customers Discover how big data is behind the success of Walmart, LinkedIn, Microsoft and more Learn how big data is changing medicine, law enforcement, hospitality, fashion, science and banking Develop your own big data strategy by accessing additional reading materials at the end of each chapter

Ce manuel de cours comprend à chaque chapitre : un cours structuré et illustré d'exemples, tableaux et schémas une mise en avant des compétences - nouveauté du programme des synthèses des applications simples Le chapitre final est consacré : aux applications plus poussées avec sujets type examen, corrigés et commentaires détaillés à la méthodologie et aux conseils Sur le site : tous les corrigés largement commentés et des exercices complémentaires, des mises à jour (en cours d'année) en fonction de l'actualité réglementaire Un ouvrage ultra complet pour se préparer et réussir l'épreuve !

Un livre à la fois théorique et pratique Cet ouvrage a pour ambition de couvrir un large spectre du domaine des data sciences. Il va plus loin qu'un simple tour sur les algorithmes d'apprentissage automatique et s'attaque aux autres aspects, malheureusement négligés mais fondamentaux pour tout data scientist : concepts généraux mais poussés, dont la maîtrise est indispensable ; algorithmes d'apprentissage automatique les plus connus ; aspects liés à l'exploration des données ; mesures de performances et d'autres métriques utilisées par les algorithmes ; différents concepts fondamentaux en mathématiques à connaître pour mieux explorer et comprendre les données ; notions importantes des big data ; études de cas pratiques en langage Python. La première partie de ce livre est théorique et adopte un format questions/réponses qui présente plusieurs avantages, dont la possibilité pour le lecteur de lire distinctement chacune des questions pour parfaire son savoir. Un autre point fort de cette structure est qu'elle incite à entrer dans un dialogue. Ainsi, grâce aux questions posées, le lecteur est poussé vers une réflexion où il confronte ses réponses à celles données par le livre. La deuxième partie est pratique et propose deux exemples d'implémentation de modèles d'apprentissage automatique. Vous y trouverez des codes écrits en Python et un aperçu de différentes difficultés que peut rencontrer un spécialiste lors de l'exercice de son métier. À qui s'adresse cet ouvrage ? Le présent ouvrage est adapté à toute personne ayant une certaine maîtrise de la data science et du Machine Learning. Il aidera notamment à se rappeler des concepts importants, mais suppose que le lecteur soit initié sur le sujet. Il sera particulièrement utile à ceux qui veulent se préparer pour un concours, un examen ou un entretien.

I- Les chapitres du livre : INFORMATIONS GÉNÉRALES -- AVANT-PROPOS -- INTRODUCTION GÉNÉRALE -- BUSINESS INTELLIGENCE -- BIG DATA -- ARCHITECTURE BIG DATA -- COHABITATION BI ET BIG DATA -- PROJET BIG DATA de A à Z -- CONCLUSION.

\*\*\*\*\* II-

Résumé : Aujourd'hui, beaucoup d'entreprises disposent d'une infrastructure de Business Intelligence encore appelée

infrastructure d'aide à la décision. Cependant depuis quelques années avec l'avènement des objets connectés; jamais les entreprises et les organisations n'ont eu autant de données à analyser pour faire progresser leurs activités. Alors toutes aujourd'hui se posent la question à savoir si elles doivent abandonner leurs infrastructures Business Intelligence existantes pour une infrastructure Big Data telle que Hadoop. Il s'agit d'une question tout à fait légitime au vu des opportunités d'affaires qu'offre l'utilisation des Big Data aujourd'hui. Ce pendant d'un point de vue technique et pragmatique la réalité est tout autre. En effet les Big Data, plus précisément les technologies Big Data ne remplacent pas la Business Intelligence qui signifie aide à la décision, mais reste plutôt un prolongement de celle-ci. Grâce aujourd'hui aux technologies Big Data, les analystes et data scientists peuvent désormais analyser les immenses quantités de données que nous produisons au quotidien et ainsi développer de nouvelles applications d'aide à la décision; ce qui n'était pas possible jusqu'ici avec les technologies BI. Dans cet ouvrage, nous vous montrerons comment intégrer les technologies Big Data dans vos infrastructures décisionnelles existantes. A cet effet, nous parlerons plutôt de «Advanced BI ou BI avancée» pour faire référence à ce que certains appellent couramment Big Data, car le mot est mal approprié pour qualifier ce qui reste et demeure l'aide à la décision (BI) basée sur une grosse volumétrie de données. \*\*\*\*\*

\*\*\*\*\*  
\*\*\*\*\* III-

Informations générales : Le livre s'adresse d'une part aux décideurs, qu'ils soient techniciens (responsable informatique, data engineer, architecte Big Data, spécialiste de la Business Intelligence...) ou fonctionnels (responsable de la stratégie, directeur général, MOA...), et d'autre part au monde informatique, ou étudiants qui veulent comprendre de façon simple pourquoi hier on parlait de Business Intelligence et aujourd'hui de «Big Data». Quelles peuvent être leurs places dans une organisation disposant d'une architecture Big Data et Business Intelligence? Dans les chapitres concernés, nous allons démystifier le Big Data et rompre les limites inexistantes entre la Business Intelligence abrégée BI et le Big Data, par un exemple pratique et des termes techniques limités au strict minimum. Une bonne compréhension des quatre premiers chapitres devrait vous faciliter la lecture des deux derniers chapitres de ce livre consacré à la pratique.

Un livre pratique pour faire son diagnostic et comprendre les enjeux liés au numérique pour chaque aspect de la stratégie d'entreprise. Le modèle est utilisable tant par les TPE et PME que les grandes entreprises, les organisations et les administrations, quel que soit leur secteur d'activité.

Spark est un moteur de traitement de données rapide dédié au big data. Très en vogue depuis quelques années il permet de traiter de gros volumes de données de manière distribuée. Grâce à sa vitesse, à sa simplicité d'usage et à la variété de ses bibliothèques d'algorithmes, il est de plus en plus utilisé pour les applications de machine learning. Ce manuel de prise en main présente quatre des cinq modules de Spark. L'auteur a fait le choix du langage Python qui convient mieux aux débutants. Son objectif est de permettre au lecteur d'installer et d'utiliser Spark de manière autonome, de comprendre les concepts du machine learning et d'en maîtriser les bonnes pratiques. L'ouvrage est complété par de nombreux compléments en ligne accessibles sur GitHub.

Les Big Data ou données massives sont le nouveau carburant de l'ère numérique. Comment les entreprises peuvent-elles exploiter ces données volumineuses et créer de la valeur ajoutée nécessaire à leur développement ? Vous êtes en train de poster sur Facebook une photo de vos vacances en Bretagne. Au même moment, votre banque constate qu'à Marseille, quelqu'un d'autre est en train d'utiliser abusivement votre carte bancaire. Grâce aux technologies de Big Data, votre banque va pouvoir confronter instantanément ces deux données personnelles et va vous éviter le pire. Les médias sociaux et les données du Web en général permettent aux entreprises d'acquérir et d'analyser de gigantesques quantités de données et ainsi de mieux connaître leurs clients, leurs concurrents et leur marché tout entier. Les entreprises qui maîtrisent les Big Data peuvent améliorer leurs produits, enrichir l'expérience utilisateur de leurs clients, créer de la valeur ajoutée et amplifier ainsi le retour sur investissement. La mise en œuvre des technologies de Big Data est complexe. Ce guide pratique est une excellente première étape à accomplir pour tirer parti des Big Data. En effet, le livre de Bill Schmarzo s'appuie sur des techniques pratiques, des méthodologies concrètes et des exemples réels. Dans ce livre : - Les promesses et les opportunités des Big Data - D'où viennent les Big Data et ce qu'il y avait avant - Evaluer l'impact sur l'activité et s'organiser - Créer la stratégie Big Data - Comprendre le processus de création de valeur - Le cas des données liées à l'expérience utilisateur et aux habitudes de consommation - Mise en œuvre et architecture nécessaire.

Le Big Data s'est imposé comme une innovation majeure pour toutes les entreprises qui cherchent à construire un avantage concurrentiel grâce à l'exploitation de leurs données clients, fournisseurs, produits, processus, machines, etc. Ce livre est un guide pour comprendre les enjeux d'un projet Big Data, en appréhender les concepts sous-jacents (en particulier le machine learning) et acquérir les compétences nécessaires à la mise en place d'un data lab. Il combine la présentation de notions théoriques (traitement statistique des données, calcul distribué...), d'outils (écosystème Hadoop, Storm...) et d'exemples de machine learning ; Cette deuxième édition comporte des ajouts sur le deep learning et les réseaux de neurones, ainsi que des compléments et des mises à jour sur les moteurs de recommandations et Spark. Les compléments en ligne seront enrichis de nouveaux jeux de données pour un début de mise en pratique.

Pratique et basé sur l'expérience, ce guide opérationnel complet propose 91 fiches qui abordent de façon exhaustive l'univers des réseaux sociaux pour une utilisation adaptée à la pratique professionnelle. Comprendre, maîtriser et utiliser les principaux réseaux sociaux Faire du marketing par l'image Blog ou site web ? Veille, animation, promotion Bâtir sa stratégie sur les réseaux sociaux Formaliser la politique de conduite du changement Lancer un nouveau produit ou un projet

Plongez au coeur du Deep Learning Ce livre a été écrit pour tous ceux qui souhaitent s'initier au Deep Learning (apprentissage profond). Il est la suite logique du titre " Le Machine learning avec Python" paru en février 2018. Le Deep Learning est une technologie nouvelle qui évolue très rapidement. Ce livre en présente les bases principales de cette technologie. Au coeur de celle-ci on trouve les réseaux de neurones profonds, permettant de modéliser tous types de données et les réseaux de convolution, capables de traiter des images. Et enfin, cette technologie de plus en plus utilisée dans les applications d'intelligence artificielle introduit le notion de Reinforcement Learning (apprentissage par renforcement) qui permet d'optimiser les prises de décision par exemple pour le fonctionnement d'un robot. Au programme : La genèse du Deep Learning Les réseaux de neurones Les bases des réseaux de type Deep learning L'architecture réseau Créer un réseau type Adapter le réseau à des besoins propres Les architectures spécifiques La vectorisation Le Deep Learning et DL4J sur Spark Au coeur de l'intelligence artificielle RL4J et Reinforcement Learning



Written in Ron Cody's signature informal, tutorial style, this book develops and demonstrates data cleaning programs and macros that you can use as written or modify which will make your job of data cleaning easier, faster, and more efficient. --

Plus de 500 exemples pour apprendre en pratiquant La plate-forme Node.js est passée du statut d'OVNI à celui d'incontournable. En 2018, elle fait fonctionner des applications web, de l'outillage front-end, de la distribution de fichiers pair-à-pair, des objets connectés et la NASA l'utilise même pour analyser les données télémétriques de ses sondes spatiales. L'intention de cet ouvrage est d'être votre base de départ pour comprendre ce qu'il se passe dans cet écosystème en perpétuel mouvement. Il devrait vous permettre de faire des choix techniques durables et de vous rendre opérationnelle sur LA plate-forme JavaScript, le langage le plus populaire du monde selon GitHub. Nous allons parcourir ensemble l'écosystème Node.js en partant de son historique et de sa gouvernance jusqu'au déploiement en continu de nos applications. Nous (re)partirons sur des bases solides de JavaScript pour ensuite découvrir petit à petit le potentiel des modules Node.js. Ces connaissances sont les fondations pour explorer la conception, le développement et le déploiement d'applications web, d'utilitaires en lignes de commandes et d'applications front-end modernes. Une sélection variée de modules complémentaires vous permettra d'y voir plus clair parmi les centaines de milliers de contributions de la communauté Node.js. Plusieurs centaines d'exemples ponctuent l'ouvrage pour mieux cerner les concepts et leurs résultats. Ils sont exécutables dans un terminal grâce à un module npm ou dans un navigateur web grâce à une surcouche interactive. À qui cet ouvrage s'adresse-t-il ? Aux développeurs et développeuses qui veulent se reconvertir depuis un autre langage de programmation comme Péri, PHP ou Ruby Aux développeurs et développeuses qui cherchent à tirer parti de l'outillage npm front-end À toute personne autodidacte et désireuse de renforcer ses compétences en programmation web

Développer une gamme de produits, suivre et analyser le marché, former la force de vente, contrôler les budgets... Découvrez toutes les compétences clés du chef de produit marketing à travers : Un portrait complet de la fonction : définition, responsabilités, tâches au quotidien. Les outils du savoir-faire et du savoir-être : cycles de travail, organisation des réunions, management... Les éléments de la gestion de carrière : organigrammes, structures d'entreprise, chefs de produits spéciaux... Cette sixième édition met l'accent sur le digital marketing et les nouvelles technologies. Débutants dans la fonction ou étudiants en école de commerce, voici le guide professionnel dont vous avez besoin pour entamer et réussir votre carrière. LES + > Modèles de documents (check-list, mind map, revue de marque et d'enseigne, brief d'agence, plan marketing...) > Lexique français/anglais : 500 mots du métier > 85 citations pour animer une présentation

Collins Easy Learning French Verbs offers beginners a clear and easy-to-understand guide to French verbs.

Building models is a small part of the story when it comes to deploying machine learning applications. The entire process involves developing, orchestrating, deploying, and running scalable and portable machine learning workloads--a process Kubeflow makes much easier. This practical book shows data scientists, data engineers, and platform architects how to plan and execute a Kubeflow project to make their Kubernetes workflows portable and scalable. Authors Josh Patterson, Michael Katzenellenbogen, and Austin Harris demonstrate how this open source platform orchestrates workflows by managing machine learning pipelines. You'll learn how to plan and execute a Kubeflow platform that can support workflows from on-premises to cloud providers including Google, Amazon, and Microsoft. Dive into Kubeflow architecture and learn best practices for using the platform Understand the process of planning your Kubeflow deployment Install Kubeflow on an existing on-premises Kubernetes cluster Deploy Kubeflow on Google Cloud Platform step-by-step from the command line Use the managed Amazon Elastic Kubernetes Service (EKS) to deploy Kubeflow on AWS Deploy and manage Kubeflow across a network of Azure cloud data centers around the world Use KFServing to develop and deploy machine learning models

Une introduction aux bases de l'administration Linux Ce guide pratique et ludique répond à toutes les questions que vous vous posez sur les bases de l'administration Linux, en explorant les distributions GNU/Linux les plus populaires, notamment CentOS 7 et la famille Red Hat Enterprise Linux. Il vous permet d'installer, de configurer et de gérer des serveurs sous Linux. Conçu étape par étape, il suit une progression pédagogique cohérente, en détaillant tout ce qu'il faut savoir sur les bases de l'administration système Linux (shell et ligne de commande, droits d'accès, processus, etc.). 100 % pratique et compatible RHEL/CentOS Les configurations fournies en exemple sont immédiatement utilisables, telles quelles. Aucun prérequis externe n'est nécessaire puisque les techniques enseignées permettent à un administrateur novice de RHEL/CentOS de maîtriser les bases du système, tout en lui montrant les bons réflexes pour se documenter et apprendre en autonomie par la suite. À qui s'adresse cet ouvrage ? • Aux administrateurs de serveurs qui souhaitent maîtriser leurs systèmes Linux au quotidien. • Aux adeptes de Linux désireux d'approfondir leurs connaissances. • Aux professionnels qui préparent la certification RHCSA ou RHCE.

Depuis 2005, la conférence francophone sur Entrepôts de Données et Analyse en ligne (EDA) offre un cadre de rencontres régulières aux chercheurs, industriels et utilisateurs intéressés par les dernières avancées scientifiques et technologiques. Ce numéro spécial est basé sur les actes de la conférence EDA 2019 : 15èmes Journées Business Intelligence & Big Data. La conférence s'est tenue en octobre 2019 à Montpellier en France.

Cet ouvrage s'adresse à tous ceux qui réfléchissent à la meilleure utilisation possible des données au sein de l'entreprise, qu'ils soient data scientists, DSI, chefs de projets ou spécialistes métier. Le Big Data s'est imposé comme une innovation majeure pour toutes les entreprises qui cherchent à construire un avantage concurrentiel grâce à l'exploitation de leurs données clients, fournisseurs, produits, processus, machines, etc. Mais quelle solution technique choisir ? Quelles compétences métier développer au sein de la DSI ? Ce livre est un guide pour comprendre les enjeux d'un projet Big Data, en appréhender les concepts sous-jacents (en particulier le machine learning) et acquérir les compétences nécessaires à la mise en place d'un data lab. Il combine la présentation : • de notions théoriques (traitement statistique des données, calcul distribué...) ; • d'outils (écosystème Hadoop, Storm...) ; • d'exemples de machine learning ; • d'une organisation typique d'un projet de data science.

Hadoop par la pratique

Archiving has become an increasingly complex process. The challenge is no longer how to store the data but how to store it intelligently, in order to exploit it over time, while maintaining its integrity and authenticity. Digital technologies bring about major transformations, not only in terms of the types of documents that are transferred to and stored in archives, in the behaviors and practices of the humanities and social sciences (digital humanities), but also in terms of the volume of data and the technological capacity for managing and preserving archives (Big Data). Archives in The Digital Age focuses on the impact of these various digital transformations on archives, and examines how the right to memory and the information of future generations is confronted with the right to be forgotten; a digital prerogative that guarantees individuals their private lives and freedoms.

As more corporations turn to Hadoop to store and process their most valuable data, the risk of a potential breach of those systems increases exponentially. This practical book not only shows Hadoop administrators and security architects how to protect Hadoop data from unauthorized access, it also shows how to limit the ability of an attacker to corrupt or modify data in the event of a security breach. Authors Ben Spivey and Joey Echeverria provide in-depth information about the security features available in Hadoop, and organize them according to common computer security concepts. You'll also get real-world examples that demonstrate how you can apply these concepts to your use cases. Understand the challenges of securing distributed systems, particularly Hadoop Use best practices for preparing Hadoop cluster hardware as securely as possible Get an overview of the Kerberos network authentication protocol Delve into authorization and accounting principles as they apply to Hadoop Learn how to use mechanisms to protect data in a Hadoop cluster, both in transit and at rest Integrate Hadoop data ingest into enterprise-wide security architecture Ensure that security architecture reaches all the way to end-user access

Le projet Business Intelligence, clés en main Les tableaux de bord sont au coeur du processus de management de la performance. La qualité et la cohérence du système d'information décisionnel conditionnent la réussite de la stratégie déployée. Encore faut-il que les tableaux de bord soient conçus méthodiquement, qu'ils mesurent toutes les formes de performance et qu'ils assistent du mieux possible les décideurs en situation. Ce livre, best-seller depuis plusieurs années, s'est imposé auprès des concepteurs de système de business intelligence, des consultants et des chefs de projet. Centré sur la problématique du décideur en action, cet ouvrage régulièrement mis à jour explique comment : Réaliser le projet de business intelligence dans sa totalité, de la conception à la mise en action. Assurer le déploiement de la stratégie gagnante. Faciliter la mise en pratique de la gouvernance. Composer les tableaux de bord pour assister la prise de décision. Choisir les indicateurs de performance les mieux adaptés à chaque situation. Fiabiliser les informations dès la collecte des données. Comprendre et exploiter le "Big Data". Manager les risques à tous les stades du processus décisionnel. Choisir les outils de business intelligence, méthode de sélection détaillée. L'ouvrage développe une démarche coopérative et illustrée en dix étapes qui insiste sur la sélection des objectifs selon la stratégie, le choix des indicateurs de performance, la mise en oeuvre et le déploiement. Chacune des dix étapes de la méthode est directement reliée au site [www.piloter.org](http://www.piloter.org) associé à ce livre. Vous pourrez y consulter les compléments, des dossiers en téléchargement libre, des références (livres et sites) ainsi que les mises à jour.

Configurer les services les plus courants Dans la continuité du premier tome, "Les fondamentaux de l'administration système", ce guide pratique et ludique répond à toutes les questions que vous vous posez sur la configuration des services les plus répandus sous Linux : authentification SSH, serveur DHCP serveur DNS, synchronisation NTR serveur de bases de données, serveur web, serveur mail, serveur de fichiers, serveur d'annuaire, serveur proxy, etc. Il traite également de plusieurs aspects utiles au quotidien et souvent négligés comme le peaufinage de votre environnement de travail ou l'installation de Linux sur du matériel spécifique comme les routerboards ou les serveurs dédiés dans les datacenters. Vous apprenez à héberger les applications web les plus populaires comme WordPress, Dolibarr, OwnCloud et Roundcube. Sans oublier les réflexes indispensables à adopter face à un serveur qui ne répond plus. Chaque chapitre est organisé sous forme d'atelier pratique. Conçu étape par étape, ce livre suit une progression pédagogique cohérente, en détaillant tout ce qu'il faut savoir pour configurer un serveur Linux immédiatement utilisable en production. Un focus sur la sécurité des serveurs La sécurité des configurations occupe une place centrale dès les premiers chapitres : configuration du pare-feu, protection contre les attaques en force brute, utilisation de SELinux (Security Enhanced Linux) en mode renforcé, audits de sécurité, lutte contre le spam, etc. À qui s'adresse cet ouvrage ? Aux administrateurs de serveurs qui souhaitent maîtriser leurs systèmes Linux au quotidien. Aux adeptes de Linux désireux d'approfondir leurs connaissances. Aux professionnels qui préparent la certification RHCSA ou RHCE.

Hadoop par la pratique s'articule autour d'explications approfondies et d'exemples de code. Chaque chapitre décrit un problème technique puis propose une solution découpée en étapes faciles à suivre. Au programme : • ETL, compression, sérialisation et importation/exportation de données ; • analyse simple et élaborée d'agrégats ; • analyse graphique ; • apprentissage automatique ; • dépannage et débogage ; • persistance évolutive ; • administration et configuration de cluster. Vous trouverez dans cet ouvrage : • un contenu direct et facile à suivre ; • une sélection des tâches et des problèmes les plus importants ; • des instructions parfaitement organisées pour résoudre efficacement les problèmes ; • des explications claires sur les actions réalisées ; • une application des solutions à d'autres situations. La lecture de cet ouvrage vous permettra de développer facilement des applications massivement parallèles et d'exploiter toute la puissance d'Hadoop.

[Copyright: 85d4b81a2135cc7925bce298a5175870](https://www.piloter.org/)