



Les technologies et enjeux du big data : Collecte, stockage, transmission et traitement des données numériques

The Technologies and Challenges of Big Data: Collection, Storage, Transmission, and Processing of Digital Data

Koné Abdoulaye

Docteur en Sciences Juridiques et Politiques
FSJES Marrakech
Université Cadi Ayyad – Marrakech - Maroc

Khalid Hermani

Doctorant en Droit Public
FSJES Mohammedia
Université Hassan 2 – Mohammedia - Maroc

Résumé : Cet article analyse les technologies et les défis liés au big data, en s'intéressant à la collecte, au stockage, à la transmission et au traitement des données numériques. Une méthodologie qualitative permet d'explorer divers cadres théoriques, dont les lois de Moore, de Gates, de Parkinson, de Koomey, de Nielsen, de Butter et de Metcalfe. L'étude s'appuie sur une analyse de documents variés, tels que les livres, les articles de revues et de journaux reconnus, afin de mieux comprendre les aspects associés au big data.

La collecte des données s'effectue principalement à travers l'internet mobile et l'internet des objets, un phénomène qui soulève des enjeux en matière de surveillance en ligne et de consentement des utilisateurs. Le stockage des données, concrétisé par les datacenters et le cloud computing, pose des questions de souveraineté numérique et de sécurité des données. La transmission des données repose sur les câbles sous-marins et les réseaux 4G, 5G et 6G, et est affectée par des enjeux d'espionnage et de confidentialité des données. Le traitement et l'exploitation des données, rendus possibles par la puissance de calcul et l'intelligence artificielle, entraînent des préoccupations en matière de consommation énergétique et de transparence des algorithmes.

Cette analyse révèle un écosystème complexe, où la gestion des données numériques requiert une attention accrue sur les défis techniques et éthiques présents dans la société contemporaine.

Mots-clés : Big data, internet des objets, datacenters, cloud computing, puissance de calcul, intelligence artificielle.

Digital Object Identifier (DOI): <https://doi.org/10.5281/zenodo.14243037>

1. Introduction

La révolution des données, catalysée par l'internet des objets, engendre une explosion d'informations sans précédent. Cette dynamique, bien que source d'infobésité, soulève des défis complexes en matière de collecte, de stockage, de transmission et de traitement des données. Toutefois, elle ouvre également la voie à des opportunités notables. Le big data désigne un ensemble de données numériques qui se caractérise par sa grande variété, son accroissement quantitatif rapide et sa circulation incessante, rendant nécessaire l'adoption de méthodes de traitement innovantes pour optimiser la compréhension et l'automatisation des processus (Gartner, 2021). Les outils d'analyse traditionnels s'avèrent inadaptes face à cette masse de données, et leur valorisation repose désormais sur l'utilisation de technologies avancées (Schmitt, 2018).

Une distinction se dessine entre les données massives (big data) et les données ouvertes (open data). Alors que les premières requièrent des compétences et des outils spécifiques pour être exploitées, les secondes bénéficient d'une accessibilité publique qui favorise leur utilisation et leur réutilisation par tous (Babinet & Orsenna, 2016, p. 257). Les données ouvertes, issues tant du secteur public que du secteur privé, permettent aux entreprises et aux citoyens d'accéder à des informations essentielles, ce qui renforce la transparence et la responsabilité gouvernementales et améliore la qualité des services offerts aux utilisateurs (Samuel & Mabi, 2014, pp. 81-91).

À mesure que l'utilisation du big data se développe, plusieurs enjeux majeurs se dessinent, notamment la sécurité, la confidentialité et la gouvernance des données numériques. **Dans ce contexte, comment les technologies de collecte, de stockage, de transmission et de traitement des données numériques influencent-elles les défis associés à la gestion du big data ?**

Cet article explore les enjeux des données massives à travers un plan structuré. Il aborde successivement : le cadre théorique et la méthode de recherche, les six V du big data, la collecte des données via l'internet mobile et l'internet des objets, le stockage dans le cloud computing, la transmission des données par l'intermédiaire des réseaux 4G, 5G et 6G, et enfin, le traitement et l'exploitation des données en lien avec la loi de Moore et l'intelligence artificielle. Chaque section traite brièvement des enjeux connexes.

2. Cadre théorique et méthode la recherche

La dimension technoscientifique du big data s'inscrit dans un cadre théorique, constitué par plusieurs lois qui décrivent l'évolution des technologies numériques et leur impact sur le traitement, le stockage et la transmission des données.

Parmi celles-ci, la loi de Moore constitue un pilier majeur concernant le traitement des données. Cette loi stipule que la capacité de calcul des microprocesseurs double tous les deux ans, créant ainsi un rythme de progrès technologique qui stimule le développement d'applications requérant une puissance de traitement toujours plus élevée (Moore, 1975). La loi de Moore, énoncée en 1965 et réévaluée en 1975, demeure pertinente pour comprendre les tendances actuelles de l'innovation technologique.

Cette loi joue un rôle essentiel dans l'expansion du big data. En effet, la loi de Moore a permis une augmentation exponentielle de la puissance de calcul disponible, ce qui a permis le traitement et l'analyse de volumes de données toujours plus importants. Par ailleurs, la loi de Gates, formulée en hommage à Bill Gates, observe que la rapidité d'exécution des logiciels diminue de moitié tous les dix-huit mois, ce qui pose un défi supplémentaire pour les technologies du big data (Techslang, 2021). Cependant, la loi de Moore reste un facteur clé en garantissant que les avancées en matière de puissance de calcul peuvent compenser cette baisse de performance logicielle. Dans le même ordre d'idée, la loi de Koomey stipule qu'à performance de calcul équivalente, la consommation d'énergie est divisée par deux tous les 18 mois. Cette évolution encourage l'utilisation croissante de l'informatique mobile et des appareils connectés, tout en diminuant les besoins en énergie et la chaleur produite. Parallèlement, la quantité de données à traiter et le temps de calcul nécessaires dans les centres de données continuent d'augmenter (Koomey, 2010).

Les lois relatives aux vitesses de transmission s'avèrent également déterminantes dans la dynamique du big data. Selon Greg Papadopoulos, directeur de la technologie de Sun Microsystems, la bande passante disponible sur Internet double tous les six mois. Ce phénomène révèle que le débit maximal d'information suit une courbe exponentielle, dépassant même la lenteur relative de la loi de Moore (Alberganti, 2000). Les « tuyaux » du réseau mondial s'élargissent trois fois plus rapidement que la capacité de calcul, ce qui souligne l'importance de l'infrastructure de communication dans la gestion des flux de données. Par ailleurs, en 1998, Jakob Nielsen, un expert en interactions homme-machine, a proposé l'hypothèse selon laquelle la bande passante des connexions Internet domestiques augmente de 50 % chaque année. Cette croissance rapide de la bande passante entraîne une augmentation de la consommation de données par les utilisateurs. En conséquence, les centres de données doivent améliorer leur capacité de traitement pour répondre à cette demande croissante. Ce phénomène crée un cycle auto-renforcé : à mesure que les utilisateurs consomment plus de données, les infrastructures réseau et les centres de données doivent s'adapter et se développer, ce qui, à son tour, permet des vitesses de connexion encore plus élevées et encourage une consommation de données encore plus importante. (Kalousdian, 2024). D'autre part, la loi de Butter indique que la capacité de transmission des données sur les réseaux de communication, notamment les lignes optiques, double environ tous les neuf mois (Vedin, 2010). Cela signifie que la quantité de données pouvant être transmise augmente rapidement, ce qui a des implications importantes pour le développement des infrastructures de communication et la gestion des données à l'échelle mondiale. Cette croissance exponentielle de la capacité de transmission permet de soutenir l'augmentation continue de la demande en bande passante due à des facteurs tels que l'essor du streaming vidéo, des services cloud et de l'internet des objets.

Diverses lois aident également à comprendre le stockage des données. La loi de Parkinson, formulée par Cyril Northcote Parkinson, stipule que « *le travail s'étend pour remplir le temps disponible* » (Klimek, Hanel, & Thurner, 2008, p. 7). En d'autres termes, plus on a de temps pour une tâche, plus elle s'allonge, même si elle ne nécessite pas autant de temps. Cette notion s'applique également au

stockage des données : lorsque la capacité de mémoire augmente, les systèmes deviennent plus gourmands en ressources, remplissant cette capacité avec des données supplémentaires. Ainsi, les besoins en mémoire continuent d'augmenter, doublant tous les dix-huit mois depuis 1985. Ce phénomène est partiellement compensé par la loi de Moore, qui indique que la puissance de calcul des ordinateurs double tous les deux ans, ce qui permet de répondre à cette demande croissante. La loi de Kryder contribue aussi à la compréhension de l'essor du big data. Celle-ci indique que la capacité de stockage des données double tous les 13 mois, ce qui permet de conserver et d'analyser d'importantes quantités d'informations à un coût de plus en plus réduit (Walter, 2005).

D'autres lois sont également utiles pour appréhender les facteurs économiques se rapportant aux technologies du big data. La loi de Machrone souligne que le prix d'un ordinateur est déterminé par le consommateur, peu importe l'augmentation de ses capacités techniques (Quoniam & Lucien, 2010). Ainsi, la valeur perçue par les utilisateurs influence les prix et l'adoption des technologies de big data. De son côté, la loi de Rock affirme que le coût des infrastructures pour fabriquer des semi-conducteurs double tous les quatre ans (Kanellos, 2003). Ces lois montrent les dynamiques économiques et technologiques qui impactent les infrastructures nécessaires pour traiter le big data. Enfin, la loi de Metcalfe, qui déclare que la valeur d'un réseau est proportionnelle au carré de son nombre d'utilisateurs, explore une dimension sociale du big data. Cette loi établit que l'augmentation du nombre d'utilisateurs accroît considérablement la valeur du réseau et la quantité de données disponibles (Koné, juin 2024).

Dans le cadre de cette recherche, une approche qualitative est adoptée. Les données sont collectées à partir d'ouvrages académiques, d'articles de revues spécialisées et de journaux, afin d'analyser les divers aspects théoriques et pratiques liés au big data et à la transformation numérique. Cette méthodologie permet de saisir l'interaction entre les technologies émergentes et les enjeux associés à la gestion des données numériques. La recherche se concentrera principalement sur les technologies et infrastructures liées au big data. Étant donné l'ampleur et la complexité du big data, cette étude ne prétend pas à l'exhaustivité, mais vise à éclairer des aspects clés de ce domaine en pleine expansion.

3. Les 6 V du big data

La notion de big data se caractérise fréquemment par l'usage des « 6 V » : volume, variété, vitesse, véracité, valeur et visualisation.

Le volume fait référence à la quantité de données produites et stockées. L'essor des technologies numériques a conduit à une génération inédite de données au quotidien. Des estimations révèlent des chiffres colossaux, tels que 347 222 tweets publiés chaque minute ou 20,8 millions de messages échangés sur WhatsApp (Excelacom, 2016).

Tableau 1 : Les activités numériques sur différentes plateformes en une minute, selon les estimations de 2016

Plateforme	Activité par minute
Twitter	347 222 tweets
WhatsApp	20 800 000 messages
Google	2 400 000 recherches
E-mails	150 000 000 e-mails
Uber	1 389 demandes de courses
Amazon	203 596 dollars de chiffre d'affaires
Apple App Store	51 000 applications téléchargées
YouTube	2 780 000 vidéos visionnées
Tinder	972 222 swipes
Facebook	701 389 connexions

Le tableau 1 est élaboré par les chercheurs, auteurs de cet article sur la base des informations disponibles sur : Excelacom Inc. (2016, 29 février). 2016 update: What happens in one internet minute. Excelacom. <https://www.excelacom.com/resources/insights/2016-update-what-happens-in-one-internet-minute.html>.

Le volume de données numériques s'est accru de manière exponentielle, culminant à environ 50 zettaoctets en 2020, avec des prévisions indiquant un potentiel de 2000 zettaoctets d'ici 2035 (Cazals, F. & C. Cazals, 2019, pp. 29-35 ; Gaudiaut, 2020). L'enjeu principal lié au volume réside dans la capacité des systèmes à stocker et analyser cette masse de données, car de nombreux outils traditionnels atteignent rapidement leurs limites.

La variété englobe les différents formats et types de données, allant des données structurées à des ensembles non structurés tels que des vidéos et des publications sur les réseaux sociaux. Chaque format nécessite des méthodologies d'analyse spécifiques, une réalité qui complexifie le traitement des données (Cazals & Cazals, 2019, pp. 29-35). Les entreprises doivent ainsi s'équiper de techniques appropriées pour tirer parti de cette diversité informationnelle, sous peine de ne pas exploiter pleinement le potentiel des données disponibles.

La vitesse désigne la rapidité avec laquelle les données sont générées, collectées et analysées. Dans le contexte actuel, des millions de tweets, des transactions financières et des données de capteurs sont produits en temps réel, ce qui rend essentielle la capacité à traiter ces informations instantanément (Bourany, 2018, pp. 27-31). Ce défi impose des exigences technologiques élevées pour assurer une prise de décision rapide et pertinente dans divers domaines, tels que la finance, la santé ou le marketing.

La véracité souligne l'importance de la qualité et de la fiabilité des données traitées. Avec un volume élevé et une variété immense, des problèmes liés à l'exactitude des données surviennent fréquemment. Les données peuvent s'avérer incorrectes, incomplètes ou biaisées, ce qui nécessite des méthodes

rigoureuses de nettoyage et de validation avant utilisation (Bourany, 2018, pp. 27-31). Les conséquences d'une mauvaise gestion de la véracité peuvent s'avérer critiques, comme la mauvaise décision stratégique à des pertes financières importantes.

La valeur fait référence à l'utilité et à l'importance des informations extraites des données. Il ne suffit pas de posséder de grandes quantités de données ; l'enjeu consiste à transformer cette masse d'informations en connaissances exploitables qui amélioreront la prise de décision et l'efficacité opérationnelle. L'identification et l'extraction de modèles et d'informations significatifs demeurent des défis majeurs dans l'analyse du big data (Cazals & Cazals, 2019, pp. 29-35).

La visualisation représente un aspect notable dans l'analyse des données massives. Elle permet une représentation graphique qui facilite l'accès et la compréhension des informations complexes. Cette technique aide à synthétiser les données afin de révéler des tendances et des relations importantes (Talend, 2022). La capacité à visualiser des informations sous des formats accessibles constitue un atout majeur pour favoriser une prise de décision appropriée.

Les 6 V du big data posent des défis majeurs sur le plan technologique, méthodologique et organisationnel. Chaque dimension appelle à des réflexions stratégiques et des solutions adaptées afin d'exploiter au mieux le potentiel des données massives.

4. L'atout économique et technologique du big data

La déclaration de Clive Humby en 2006, selon laquelle « Data is the new oil » (Charles, 2017), souligne l'émergence de la donnée en tant qu'actif économique central dans l'économie numérique. De la même manière que le pétrole est devenu un moteur majeur de l'économie mondiale et un enjeu géopolitique, la donnée se dresse aujourd'hui comme le fondement sur lequel reposent de nombreux modèles économiques. Néanmoins la comparaison entre la donnée et le pétrole fait débat. En effet, la production de la donnée, contrairement à celle du pétrole, génère de nouveaux ensembles de données, rendant cette ressource d'autant plus dynamique et évolutive (Chapelet & Ponal, 2022).

Le big data joue un rôle clé dans le phénomène d'ubérisation, qui se caractérise par la connexion instantanée entre l'offre et la demande. Cette dynamique est facilitée par l'internet des objets, le big data et l'intelligence artificielle (Koné, octobre 2024). L'exploitation des données massives permet ainsi d'optimiser les services et de répondre rapidement aux attentes des consommateurs.

L'apparition du web 2.0 a accentué la présence des données massives, lesquelles se créent continuellement à travers les interactions en ligne des utilisateurs. Chaque action sur Internet que ce soit une requête de recherche ou un clic sur un lien génère des données, ce qui place l'analyse des comportements en ligne au cœur des stratégies économiques des acteurs numériques (Colin, Nicolas, & al., 2015). Cependant, les données les plus récentes sont celles qui représentent une valeur inestimable pour la compréhension et la prédiction des comportements futurs. Comme l'indique John Battelle (2006), le web fonctionne désormais comme une vaste « base de données de nos intentions »,

démontrant à quel point nos interactions sont minutieusement suivies et interprétées par des algorithmes.

Les données personnelles, qui englobent des informations pouvant identifier un individu, sont devenues un atout dans le paysage économique actuel. Ces données, comprenant des éléments allant des coordonnées personnelles aux opinions sensibles, sont collectées par une multitude d'acteurs, notamment des entreprises et des organismes gouvernementaux. Elles sont principalement utilisées pour personnaliser les services, cibler les publicités et mener des recherches. Elles soulèvent cependant des questions éthiques et de protection de la vie privée. En Europe, des régulations comme le RGPD ont été mises en place pour garantir la sécurité des données personnelles, tandis que le Maroc applique la loi n° 09-08 pour protéger ses citoyens.

Par ailleurs, la gestion efficace des données se révèle être un levier stratégique essentiel pour les entreprises numériques. Par exemple, les GAFAM exploitent leurs vastes bases de données pour établir des profils détaillés de leurs utilisateurs. Cette approche permet des prestations de services hautement personnalisées. Par exemple, Amazon s'est distingué par sa capacité à anticiper les besoins de ses clients grâce à des données comportementales, tandis que Google a démontré son aptitude à traiter des flux de données massifs pour, entre autres, prédire des tendances sanitaires (Migot, 2019). Ces actions montrent l'émergence d'une économie centrée sur les données, où la valorisation de l'information transcende les actifs matériels.

Le rôle du big data dans le développement de l'intelligence artificielle mérite aussi une attention particulière. La performance des algorithmes d'apprentissage dépend du volume et de la qualité des données collectées. Des exemples de données correctement annotées, comme celles issues des fichiers bancaires ou des documents d'organisations internationales, montrent comment ces informations peuvent entraîner des modèles prédictifs et améliorer la traduction multilingue (Fondation Prospective et Innovation, 2021). Cependant, il existe également des modèles d'apprentissage qui ne nécessitent pas d'annotation, tels que les approches d'apprentissage non supervisé et d'apprentissage par renforcement. Ces modèles exploitent des données brutes pour découvrir des structures sous-jacentes ou optimiser des décisions sans intervention humaine directe.

Tableau 2 : Quelques applications et utilisations du big data

<i>Secteur</i>	Applications	Utilisations
<i>Santé</i>	Analyse des données médicales, recherche clinique	Amélioration des diagnostics, personnalisation des traitements
<i>Finance</i>	Détection de fraudes, analyse des risques	Gestion des portefeuilles, prévision des tendances économiques
<i>Marketing</i>	Segmentation des clients, analyse des sentiments	Campagnes ciblées, optimisation des prix
<i>Commerce</i>	Gestion des stocks, prévisions de ventes	Amélioration de l'expérience client, recommandations personnalisées
<i>Transport</i>	Optimisation des itinéraires, analyse du trafic	Réduction des coûts, amélioration de la logistique
<i>Énergie</i>	Surveillance des réseaux, prévisions de consommation	Gestion des ressources, optimisation des opérations
<i>Agriculture</i>	Analyse des données climatiques, suivi des cultures	Amélioration des rendements, gestion durable des ressources
<i>Sport</i>	Analyse des performances, suivi des joueurs	Amélioration des stratégies d'équipe, prévention des blessures
<i>Éducation</i>	Analyse des performances des étudiants	Personnalisation des parcours d'apprentissage, amélioration des méthodes pédagogiques
<i>Sécurité</i>	Surveillance des données, détection des menaces	Protection des systèmes, réponse rapide aux incidents

Le tableau 2 est élaboré par les chercheurs, auteurs de cet article.

5. La collecte des données : l'importance de l'internet des objets

Actuellement, d'énormes quantités de données sont collectées quotidiennement grâce à une multitude d'objets connectés. La collecte automatique des données s'effectue à travers des dispositifs variés, tels que des capteurs, des sismographes, des radiotélescopes, des moniteurs de fréquence cardiaque, des caméras et des microphones. En complément de cette approche automatisée, des méthodes manuelles, telles que le crowdsourcing, permettent également de rassembler des informations. Les techniques et méthodes de compilation des données se sont considérablement améliorées, rendant possible la collecte de données issues de tweets sur des sites de microblogging, de requêtes envoyées aux moteurs de recherche et de commentaires publiés sur des réseaux sociaux. Les applications telles que Waze enregistrent systématiquement les déplacements et les vitesses instantanées des utilisateurs. Elles centralisent ensuite ces informations pour optimiser la gestion du trafic routier et proposer des itinéraires personnalisés. L'avenir pourrait voir l'intégration de paramètres physiologiques, tels que la fréquence cardiaque ou la glycémie, connectés à des dispositifs de type montre ou bracelet, permettant un suivi continu des états de santé (Ganascia, Jean-Gabriel. 2017, p. 46).

Ainsi, la synergie entre le big data et l'internet des objets est manifeste. L'augmentation du nombre d'objets connectés à Internet entraîne la génération de volumes massifs de données. L'analyse de ces données facilite l'identification de tendances, ce qui permet de générer des informations pertinentes pour divers secteurs d'activité (L, Bastien. 30 août 2019).

Plusieurs exemples illustrent la relation entre le big data et l'internet des objets. Dans le domaine du transport, les données collectées par les capteurs et analysées à travers des systèmes d'intelligence artificielle contribuent à améliorer la rentabilité et l'efficacité opérationnelle. Les véhicules de livraison, équipés de capteurs, rapportent des informations sur des paramètres tels que l'état du moteur ou le parcours emprunté. UPS (United Parcel Service), par exemple, exploite le big data pour surveiller un parc de 60 000 véhicules grâce à des capteurs qui anticipent les pannes potentielles en détectant les signes d'usure ou de dysfonctionnement (Rifkin, Jeremy, & al. 2014, p. 27). Dans le secteur agricole, la collecte de données par les capteurs de terrain facilite la gestion de l'irrigation en déterminant les niveaux d'humidité optimaux pour les cultures (L, Bastien. 30 août 2019).

Tableau 3 : Les approches de la collecte des données à l'ère du big data

Moyen de collecte	Avantages	Inconvénients
Capteurs et IoT	- Collecte en temps réel - Informations précises	- Investissements élevés - Problèmes de confidentialité
Réseaux sociaux	- Compréhension des comportements - Données en temps réel	- Données biaisées - Problèmes de confidentialité
Enquêtes (Surveys)	- Informations spécifiques - Adaptabilité aux besoins	- Coût et temps élevés - Réponses biaisées
Web scraping	- Surveillance des activités en ligne	- Risques légaux - Données incomplètes
Données transactionnelles	- Infos sur les habitudes d'achat - Personnalisation des offres	- Limité aux transactions spécifiques - Problèmes de confidentialité
Données gouvernementales	- Infos démographiques et socio-économiques - Analyse des politiques publiques	- Données limitées par région ou période - Problèmes de confidentialité

Le tableau 3 est élaboré par les chercheurs, auteurs de cet article.

La collecte de données soulève des questions éthiques et sociétales majeures. Les dispositifs connectés, comme les caméras connectés utilisés en Chine pour la cybersurveillance et le système de crédit social, posent des risques pour la vie privée (Arsène, Séverine, janvier 2021). Cela alimente des inquiétudes sur la sécurité des données personnelles. Les organismes doivent naviguer entre l'exploitation des données à des fins d'innovation et le respect des droits des individus.

Un autre aspect important concerne le consentement des utilisateurs. Le RGPD, qui impose des obligations strictes sur la collecte et le traitement des données personnelles, garantit que les individus soient informés et qu'ils puissent contrôler l'utilisation de leurs données (The European Parliament & The Council of the European Union, 2016). Le Digital Services Act, quant à lui, vise à responsabiliser les plateformes en ligne en matière de transparence et de sécurité. Ce faisant, cette disposition protège les droits des utilisateurs face à des pratiques parfois intrusives (The European Parliament & The Council of the European Union, 2022).

6. Le stockage des données : les avancées du cloud computing

Le cloud computing repose sur l'utilisation des datacenters, infrastructures physiques chargées d'héberger et de gérer les serveurs, le stockage et les réseaux nécessaires au fonctionnement des services cloud. Ces datacenters fournissent l'espace, la puissance de calcul et la connectivité nécessaires pour héberger les services en nuage. Le déploiement des technologies de cloud computing, également connu sous le nom d'informatique en nuage ou d'informatique dématérialisée, représente une avancée majeure dans le domaine du stockage des données. Le cloud computing englobe toutes les technologies matérielles et logicielles permettant d'accéder à des services informatiques à travers Internet, tels que les logiciels, le stockage, les plateformes et les serveurs. Contrairement au stockage local, qui implique l'utilisation d'un disque dur pour stocker des données ou exécuter des programmes, le cloud computing implique une externalisation des capacités de stockage et de calcul. Cette externalisation s'avère nécessaire en raison de l'augmentation exponentielle de la quantité de données nécessitant d'énormes capacités de calcul et des espaces de stockage (Peter & Grance, 2011).

Le cloud computing se divise en plusieurs catégories de services qui répondent à des besoins variés des utilisateurs et des entreprises. Parmi ces services, on trouve le SaaS (Software as a Service) qui propose des applications accessibles directement via le web, payées selon un modèle d'abonnement ou de consommation. Le PaaS (Platform as a Service) constitue un environnement de développement en ligne qui fournit des outils et des ressources nécessaires pour créer des applications sans se soucier des infrastructures sous-jacentes. En outre, le IaaS (Infrastructure as a Service) offre des solutions d'hébergement et de gestion d'infrastructures informatiques à distance, ce qui permet aux utilisateurs de louer des ressources matérielles en fonction de leurs besoins (Tournier, Hubert & al. 2012, pp. 14-17). De même, le DaaS (Data as a Service) facilite l'accès et la gestion des données en ligne, offrant à la fois des services de stockage et des bases de données partagées. Enfin, le MBaaS (Mobile Backend as a Service) permet aux développeurs d'accéder à des services pour le développement d'applications mobiles tout en assurant un backend fiable.

Concernant l'usage du cloud computing, les entreprises se distinguent par leur approche, qui varie selon la configuration choisie : cloud privé, cloud public et cloud hybride. Le cloud privé se caractérise par son déploiement au sein même de l'organisation, avec un contrôle total sur les infrastructures et les ressources. Ce modèle peut également être hébergé chez un tiers mais toujours dans un environnement sécurisé et accessible uniquement aux acteurs désignés. À l'opposé, le cloud public offre des ressources accessibles par Internet, sous l'administration d'un fournisseur externe, dont l'infrastructure est partagée entre plusieurs clients. Par ailleurs, le cloud hybride combine les fonctionnalités des clouds public et privé, permettant une intégration fluide des données et des applications entre ces deux modèles, ce qui offre plus de flexibilité pour les entreprises (Syntec informatique, 2010).

Par ailleurs, la virtualisation optimise la gestion des ressources informatiques en permettant l'exécution de plusieurs systèmes d'exploitation sur un seul serveur physique. Cette approche réduit les coûts et la complexité. Elle se divise en virtualisation des serveurs, qui utilise un serveur pour plusieurs machines virtuelles, et virtualisation des réseaux, qui simule un réseau physique tout en offrant des avantages supplémentaires. Bien que distincts, la virtualisation et le cloud computing se complètent pour améliorer l'agilité et les solutions d'auto-service des entreprises, qui virtualisent souvent leurs serveurs avant d'adopter des solutions cloud (VMware, juin 2022).

Les applications du cloud computing sont multiples et variées. Cette technologie a permis d'établir des solutions de stockage de données et de puissance de calcul pour le big data, tout en bénéficiant aussi à l'internet des objets. Par exemple, les tablettes, les smartphones et les montres connectées peuvent accéder à des serveurs dotés de capacités quasi illimitées, ainsi qu'à des logiciels fiables, sécurisés et constamment mis à jour. Dans cette perspective, le développement du cloud computing s'inscrit dans la lignée des datacenters et accompagne l'expansion de l'internet des objets (VAYRE, 2016).

Les enjeux du stockage des données revêtent une importance considérable. Ce secteur constitue un aspect clé de la souveraineté numérique. La question de la territorialisation des données, c'est-à-dire de leur emplacement géographique, soulève des interrogations majeures sur cette souveraineté. Par exemple, des pays comme la Chine et la Russie imposent que les données soient hébergées sur leur territoire, dans des datacenters nationaux. Le Grand Firewall de la Chine a pour objectif d'empêcher la fuite des données à l'étranger (Martin & al. 4 juillet 2019).

L'Union européenne a mis en place le RGPD pour garantir la protection des données personnelles. Elle a également instauré la Data Governance Act pour encourager un partage sécurisé et éthique des données (Commission européenne, 1 mars 2024). Aux États-Unis, le Cloud Act suscite des préoccupations concernant l'accès des gouvernements aux données, ce qui alimente le débat sur la confidentialité et l'extraterritorialité des lois (Ablard-Chauvet, 25 juin 2021). Par ailleurs, le stockage des données soulève des questions de sécurité, notamment face aux cyberattaques et aux violations de données. Par exemple, le piratage de Sony Pictures en 2014 a révélé des informations sensibles et a causé des pertes financières importantes. De même, les attaques de WannaCry en 2017 ont affecté des milliers d'organisations dans le monde, exploitant une faille de sécurité pour chiffrer des données et demander une rançon. Enfin, NotPetya, également survenu en 2017, a été une attaque destructrice qui a ciblé des infrastructures critiques et a entraîné des perturbations massives ainsi que des coûts de réparation exorbitants (Tourny, juillet 2017).

7. La transmission des données : 4G, 5G et 6G

La transmission des données constitue un élément central des infrastructures numériques contemporaines, en particulier avec l'essor des technologies de communication mobile telles que la 4G, la 5G et la future 6G. Les câbles sous-marins, en tant qu'éléments fondamentaux, assurent la connectivité mondiale en facilitant des débits élevés et une latence réduite, nécessaires pour soutenir

ces réseaux avancés. Environ 99 % des données échangées à l'échelle mondiale transitent par des câbles sous-marins, tandis que seulement environ 1 % des données sont transmises via satellite (Besanger, Serge, 22 octobre 2021).

La 4G, qui a transformé la communication mobile, a introduit le très haut débit grâce à l'architecture LTE (Long Term Evolution). Avec un débit théorique atteignant 150 Mbit/s pour la 4G LTE et 1 Gbit/s pour la 4G+, cette technologie a permis une amélioration substantielle de l'expérience utilisateur, notamment pour le téléchargement de fichiers volumineux et le streaming de vidéos en haute résolution (Guaino, Magali. 30 mars 2023). La norme VoLTE (Voice over LTE) a également été intégrée, offrant une qualité audio supérieure aux appels vocaux, ce qui a contribué à l'essor des smartphones et des tablettes.

Face à l'augmentation exponentielle du nombre d'appareils connectés dans le cadre de l'internet des objets, la 5G a été développée pour répondre aux besoins de vitesse, de latence et de coût. Cette technologie permet des débits atteignant jusqu'à 10 Gbit/s, tout en offrant un temps de latence réduit à 1 milliseconde, propice à des applications en temps réel comme les véhicules autonomes et la réalité augmentée (Paul-Simon, Jean. 11 juin 2019 ; Guaino, Magali). De plus, la 5G favorise une consommation d'énergie optimisée, prolongeant la durée de vie des objets connectés, et permettant un déploiement massif de ces derniers (Thales Group, 28 mai 2021).

La 6G, qui succédera à la 5G, est prévue pour un déploiement aux alentours de 2030. Cette technologie promet des débits pouvant atteindre 1 téraoctet par seconde et une latence réduite à une microseconde (R., Ando. 22 septembre 2021). En intégrant à la fois le domaine physique et numérique, la 6G se caractérisera par une immersion accrue dans l'écosystème numérique, en facilitant des applications comme la chirurgie à distance et le streaming en 16K. La 6G sera également sensible à divers paramètres environnementaux. Cela permettra des interactions plus intelligentes et efficaces entre les appareils connectés et le réseau, tout en mettant l'accent sur la sécurité et la lutte contre les cyberattaques (Caulier, Sophy. 15 mars 2022).

Outre les avancées technologiques, la transmission des données soulève également des enjeux majeurs. La surveillance de masse, mise en lumière par des affaires comme celle de Snowden, expose les risques d'espionnage et d'atteintes à la vie privée. Des logiciels espions tels que Pegasus, qui peuvent infiltrer des smartphones, montrent les dangers liés à la collecte et à la transmission de données sans consentement (La Croix, 22 juillet 2021). Pour contrer ces menaces, des solutions comme le chiffrement des données, le contrôle d'accès rigoureux et l'utilisation de VPN sont essentielles pour tenter de protéger les informations personnelles contre les accès non autorisés.

8. Traitement et exploitation des données : la loi de Moore et la révolution de l'intelligence artificielle

Le traitement des données repose sur l'évolution exponentielle de la puissance de calcul, conformément à la loi de Moore, qui postule un doublement des transistors intégrés dans les circuits

intégrés tous les deux ans (Moore, 1975). Les avancées en miniaturisation, ont permis de multiplier le nombre de transistors sur les puces électroniques, passant de quelques milliers dans les premiers processeurs à plusieurs milliards dans les modèles récents.

En 2016, le supercalculateur chinois Sunway Taihulight a atteint un record de 93 millions de milliards d'opérations par seconde, tandis qu'en 2022, les États-Unis ont inauguré le premier supercalculateur exaflopique, Frontier, avec une puissance de calcul de 1,1 exaflop (Loukil, Ridha. 3 juin 2022). La généralisation des processeurs graphiques, qui permettent le traitement parallèle de grandes quantités d'éléments, a largement favorisé les avancées en intelligence artificielle. Les architectures de calcul basées sur les GPU, en comparaison avec les CPU traditionnels, ont permis des progrès conséquents dans le domaine des réseaux de neurones artificiels, lesquels nécessitent des ajustements massifs de paramètres durant l'apprentissage (Ascoli, Stéphane D. 2020, p. 25).

L'intelligence artificielle occupe un rôle central dans l'exploitation du big data. Les algorithmes d'apprentissage, alimentés par des volumes croissants de données, permettent d'extraire des informations et d'identifier des motifs complexes. Cette capacité contribue à la prise de décision dans divers secteurs tels que la santé, la finance et la logistique. Les techniques d'apprentissage supervisé, non supervisé et par renforcement améliorent les processus existants et favorisent des découvertes novatrices ainsi que des solutions adaptées à des problèmes spécifiques. Les modèles prédictifs ont un impact majeur sur les résultats d'affaires. L'intelligence artificielle s'avère efficace dans l'analyse des comportements des consommateurs, l'optimisation des chaînes d'approvisionnement et le développement de médicaments (Russell & Norvig, 2021).

Cependant, l'explosion de la puissance de traitement soulève également des enjeux majeurs. Parmi ces enjeux figurent la consommation énergétique associée aux infrastructures de calcul (datacenters) utilisées pour traiter une quantité toujours croissante de données, entraînant une empreinte carbone importante. À mesure que la demande pour des systèmes de calcul toujours plus puissants augmente, la nécessité d'optimiser l'utilisation de l'énergie et de développer des solutions durables devient nécessaire. La transparence des algorithmes et des modèles d'IA constitue un défi majeur. Le deep learning, en raison de son opacité, rend difficile la compréhension de ses processus décisionnels. En effet, il est impossible de retracer l'ensemble des calculs effectués ; on sait simplement que plus l'IA dispose de données, meilleures sont ses décisions (Barraud. 2020, p. 44). Cela soulève des enjeux majeurs par exemple lors de processus de recrutement où l'IA prend des décisions. Les questions d'équité, de biais algorithmiques et de responsabilité nécessitent une régulation pour garantir l'intégrité des décisions prises par ces systèmes.

Tableau 4 : Les enjeux associés à la gestion des données numériques

Approche	Enjeux
Collecte des données	- Cybersurveillance - Consentement des utilisateurs
Stockage des données	- Souveraineté numérique (territorialisation des données) - Sécurité des données (cyberattaques)
Transmission des données	- Cyberespionnage - Confidentialité des données (chiffrement)
Traitement et exploitation des données	- Consommation énergétique - Transparence algorithmique

Le tableau 4 est élaboré par les chercheurs, auteurs de cet article.

9. Conclusion

Les technologies et les enjeux du big data révèlent un paysage multidimensionnel. La collecte, le stockage, la transmission et le traitement des données numériques soulèvent des réflexions majeures qui dépassent les simples considérations technologiques. À l'échelle mondiale, l'émergence et la diffusion des technologies numériques ont propulsé la numérisation à des niveaux sans précédent, au point où l'on parle de « dataïsme », concept défini par Yuval Noah Harari. Ce phénomène postule un monde régi par les flux de données, où la navigation dans les océans d'informations détermine les trajectoires économiques, sociales et politiques.

L'infrastructure qui soutient cette révolution numérique est totalement interconnectée via l'internet mobile, l'internet des objets, ainsi que par des infrastructures et technologies avancées telles que les datacenters, le cloud computing et les câbles sous-marins. Ces éléments ont transformé la manière d'interagir avec le monde, créant une dépendance croissante à l'égard des réseaux de communication, notamment grâce aux évolutions des standards de télécommunication tels que la 4G, la 5G et bientôt 6G. Par ailleurs, l'augmentation des capacités de calcul, qu'elles soient réalisées par des semi-conducteurs performants ou des supercalculateurs, accompagnée des avancées en intelligence artificielle, permet une analyse des données en temps réel, ce qui optimise les processus décisionnels et favorise l'innovation dans divers secteurs.

Cependant, les innovations liées au big data ne se déploient pas sans soulever des défis structurels et éthiques. La cybersurveillance s'impose comme un phénomène constant, où le consentement des utilisateurs reste souvent ambigu. La question de la souveraineté numérique mérite également une attention particulière. La territorialisation des données devient un enjeu de sécurité nationale face aux menaces du cyberespionnage, de la confidentialité des données et de la sécurité des systèmes d'information. De plus, les préoccupations autour de la consommation énergétique des infrastructures de stockage et de traitement revêtent une importance croissante dans le cadre des discussions sur la durabilité.

Les évolutions réglementaires telles que le RGPD, le DSA et le Data Privacy Framework sont des réponses législatives visant à encadrer les pratiques autour du big data. Ces dispositifs se présentent comme des repères dans un environnement en constante mutation.

De plus, le big data a propulsé l'intelligence artificielle. Son évolution en a fait non seulement un outil d'analyse, mais aussi un levier pour prédire les comportements et les tendances. Néanmoins, cette capacité prédictive soulève des préoccupations éthiques importantes. D'une part, elle promet des avancées considérables dans des domaines tels que la santé, l'éducation et la gestion des ressources. D'autre part, elle peut faciliter la surveillance de masse, où la prédiction pourrait s'apparenter à un outil de contrôle social. À l'ère de l'internet des objets, la frontière entre prédiction et surveillance risque aussi de s'estomper. S'il est possible de surveiller la population à travers les milliards d'objets connectés, tels que les caméras connectés, quel pourrait être alors l'intérêt de la prédiction ?

La gouvernance du big data doit tenir compte des enjeux contemporains liés à l'innovation et à la protection des droits des individus. La régulation doit être suffisamment robuste pour garantir la sécurité et la confidentialité des données, tout en permettant aux entreprises d'exploiter le potentiel transformateur du big data, essentiel pour alimenter des modèles d'IA performants. Dans ce contexte, quelles solutions innovantes et inclusives peuvent être mises en œuvre pour transformer les flux de données en moteur de développement durable et équitable des sociétés ?

BIBLIOGRAPHIE

Ouvrages

- Ascoli, S. D'. (2020). *Comprendre la révolution de l'intelligence artificielle*. e-book, First.
- Babinet, G., & Orsenna, E. (2016). *Big Data, penser l'homme et le monde autrement*. e-book, Le Passeur.
- Barraud, Boris (2020). *L'intelligence Artificielle : Dans toutes ses dimensions*. e-book, L'Harmattan, 2020.
- Battelle, J. (2006). *La révolution Google*. Eyrolles.
- Cazals, F., & Cazals, C. (2019). *Intelligence artificielle : L'intelligence amplifiée par la technologie*. e-book, DE BOECK SUP.
- Fondation Prospective et Innovation. (2021). *Intelligence artificielle*. e-book, Ginkgo éditeur.
- Ganascia, J.-G. (2017). *Le mythe de la singularité. Faut-il craindre l'intelligence artificielle ?* e-book, SEUIL.
- Quoniam, L., & Lucien, A. (2010). *Intelligence compétitive 2.0, organisation, innovation et territoire* [PDF]. LAVOISIER. http://quoniam.info/competitive-intelligence/PDF/publications/2010/Intelligence_Compétitive_2_0.pdf
- Rifkin, J., & al. (2014). *La nouvelle société du coût marginal zéro : L'Internet des objets, l'émergence des communaux collaboratifs et l'éclipse du capitalisme*. e-book, Les Liens Qui Libèrent.

Russell, S., & Norvig, P. (2021). *Intelligence artificielle : Une approche moderne*. Pearson.

Syntec Informatique. (2010). *Le livre blanc du cloud computing : Tout ce que vous devez savoir sur l'informatique dans le nuage*.

Tournier, Hubert & al (2012). *Cloud computing : Décider - Concevoir - Piloter - Améliorer*. e-book, EYROLLES.

Articles de revues

Arsène, Séverine (2021, janvier). « Le système de crédit social en Chine : La discipline et la morale ». *Réseaux* (N° 225). pp. 55-86.

Bourany, T. (2018). Les 5V du big data. *Regards croisés sur l'économie*, 23(2), 27-31. <https://doi.org/10.3917/rce.023.0027>.

Colin, N., & al. (2015). Économie numérique. *Notes du conseil d'analyse économique*, 26(7), 1-12.

Goëta, S., & Mabi, C. (2014). L'open data peut-il (encore) servir les citoyens ? *Mouvements*, 79(3), 81-91.

Klimek, P., Hanel, R., & Thurner, S. (2008). Parkinson's law quantified: Three investigations on bureaucratic inefficiency. *Cornell University Library*. <https://arxiv.org/pdf/0808.1684>.

Koné, Abdoulaye. (2024, 14 juin). La convergence de l'internet des objets, du big data et de l'intelligence artificielle. *International Journal Of Applied Management And Economics*, 2(08), 103–131. <https://doi.org/10.5281/zenodo.11653470>.

Koné, Abdoulaye. (2024, 15 octobre). La structure des révolutions industrielles : Trajectoires énergétiques, communicationnelles et économiques de la vapeur à l'intelligence artificielle. *Revue Internationale de la Recherche Scientifique (Revue-IRS)*, 2(5), 2674–2694. <https://doi.org/10.5281/zenodo.13934672>.

Mell, P., & Grance, T. (2011, 1 janvier). The NIST definition of cloud computing. <https://doi.org/10.6028/nist.sp.800-145>.

Migot, J.-S. (2019). Big data : qui sont les maîtres de la donnée ? *La Revue des médias*. larevuedesmedias.ina.fr/big-data-qui-sont-les-maitres-de-la-donnee.

Moore, G. E. (1975). Progress in digital integrated electronics. *IEEE Text Speech*.

Paul-Simon, J. (2019, 11 juin). 5G : quels usages pour les médias ? *La Revue des Médias*. larevuedesmedias.ina.fr/5g-quels-usages-pour-les-medias.

Tourny, Ève (2017, juillet). Cyber sécurité et attaques informatiques : les leçons à tirer de Wanna Cry et Not Petya. *Paix et sécurité européenne et internationale*. halshs-03156241.

Thèse de doctorat

Schmitt, E. (2018). Explorer, visualiser, décider : Un paradigme méthodologique pour la production de connaissances à partir des big data (Thèse de doctorat). Université de Technologie de Compiègne.

Documents officiels

Commission européenne (1 mars 2024). « Loi européenne sur la gouvernance des données ». Bâtir L'avenir Numérique de L'Europe, digital-strategy.ec.europa.eu/fr/policies/data-governance-act.

The European Parliament & The Council of the European Union. (2016). General Data Protection Regulation. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.

The European Parliament & The Council of the European Union. (2022). Digital Services Act. Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>.

Articles de journaux et de web

Chapelet, L., & Ponal, C. (2022.). Dans la mobilité, is data really the new oil ? *Publications - PMP - Beyond Consulting*. <https://pmpstrategy.com/insights/publications/item/401-dans-la-mobilite-is-data-really-the-new-oil>.

Gartner. (2021). Definition of big data - *Gartner information technology glossary*. <https://www.gartner.com/en/information-technology/glossary/big-data>.

Talend. (2022). Visualisation de données : définition et guide pratique. <https://www.talend.com/fr/resources/visualisation-donnees/#:%7E:text=La%20visualisation>.

Untersinger, Martin & al. « Cyberspace : la guerre mondiale des données ». *Le Monde.fr*, 4 juillet 2019, www.lemonde.fr/international/visuel/2018/07/23/cyberspace-la-guerre-mondiale-des-donnees_5334901_3210.html.

Walter, C. (2005). Kryder's Law. *Scientific American*. <https://www.scientificamerican.com/article/kryders-law/>.

Besanger, Serge (22 octobre 2021). « 99 % du réseau internet passe par les câbles sous-marins : sommes-nous prêts à une coupure ? » *Le Journal Du Dimanche*, www.lejdd.fr/International/cables-sous-marins-ce-qu'il-faut-savoir-sur-ces-autoroutes-du-fond-des-mers-4072986.

VMware. (2022, juin). Qu'est-ce que la virtualisation ? <https://www.vmware.com/fr/solutions/virtualization.html>.

Kanellos, M. (2003, 22 janvier). Soaring costs of chipmaking recast industry. *Archive.today*. https://archive.is/20130119214941/http://news.com.com/Semi+survival/2009-1001_3-981418.html.

Excelacom Inc. (2016, 29 février). 2016 update: What happens in one internet minute. *Excelacom*. <https://www.excelacom.com/resources/insights/2016-update-what-happens-in-one-internet-minute.html>.

Caulier, S. (2022, 15 mars). La 6G promet de fusionner les mondes humains et numérique. *Polytechnique Insights*. <https://www.polytechnique-insights.com/dossiers/digital/5g-6g/la-6g-promet-de-fusionner-les-mondes-humains-et-numerique>.

Guaino, M. (2023, 30 mars). 5G VS 4G : quelles différences entre les deux réseaux ? *Echos du Net*. <https://www.echosdunet.net/dossiers/reseau-5g/comparaison-4g>.

Techslang. (2021, 14 avril). What is Wirth's Law? *Techslang — Tech Explained in Simple Terms*. <https://www.techslang.com/definition/what-is-wirths-law/>.

Thales Group. (2021, 28 mai). Qu'est-ce que la 5G ? www.thalesgroup.com/fr/europe/france/dis/mobile/inspiration/5g.

Loukil, R. (2022, 3 juin). Le premier supercalculateur exaflopique au monde est américain. *Usine Nouvelle*. <https://www.usinenouvelle.com/article/le-premier-supercalculateur-exaflopique-au-monde-est-americain.N2010342>.

Koomey, J. G. (2024, 14 juillet). Outperforming Moore's law. *IEEE Spectrum*. <https://spectrum.ieee.org/outperforming-moores-law>.

Croix, La, (2021, 22 juillet). « Pegasus : Tout comprendre à cette affaire de cyberespionnage ». *La Croix*, www.la-croix.com/Monde/Affaire-Pegasus-tout-comprendre-cette-affaire-cyberespionnage-2021-07-22-1201167430.

L, B. (2019, 30 août). IoT et Big Data : comprendre la relation entre ces deux technologies. *LeBigData.fr*. <https://www.lebigdata.fr/iot-big-data>.

Kalousdian, P. (2024, 19 septembre). Lois de Moore, Kryder et Nielsen. *ISLEAN*. <https://islean-consulting.fr/fr/transformation-digitale/lois-de-moore-kryder-et-nielsen/>.

R, Ando. (2021, 22 septembre). 6G : guide complet sur la sixième génération de réseau sans fil. *Objetconnecte.com*. <https://www.objetconnecte.com/6g-guide-complet>.

Alberganti, M. (2000, 10 novembre). Sun imagine un ordinateur réduit à un simple écran connecté à des réseaux extérieurs. *Le Monde.fr*. https://www.lemonde.fr/archives/article/2000/11/11/sun-imagine-un-ordinateur-reduit-a-un-simple-ecran-connecte-a-des-reseaux-exterieurs_3711445_1819218.html.

Vedin, B.-A. (2010, 30 novembre). Abiding the laws... Are there any for innovation? *Innovation Management*. <http://www.innovationmanagement.com/articles/abiding-the-laws-are-there-any-for-innovation/>.