

Big Data

La numérisation du monde

SI7

Partie 3

Le Big Data en économie

Rappels

- * www.OECD.org : Lab and trends
- * http://kpcbweb2.s3.amazonaws.com/files/85/Internet_Trends_2014_vFINAL_-_05_28_14-_PDF.pdf?1401286773
- * Big Data : transversalité disciplinaires
- * Est-ce qu'il y a de la valeur derrière le Big Data ?
- * Est-ce seulement du battage médiatique ?
- * Y-a-t-il de l'argent derrière ce concept ?

Rappels

* 1 out of 5, « game changers » in USA.

1. Big Data : leadership
2. Energie : gaz de schiste, indépendance vers 2020 et exportation !
3. Commerce : grâce au Big Data
4. Education : classement de Shanghai : <http://www.shanghairanking.com>
5. Infrastructures : projets innovants : Hyperloop, marché du tourisme spatial, ...

Rappels

- * Les EU possède 1/3 des données mondiales disponibles.
- * Ils disposent des centres de recherche, mais aussi des start-up.

1. Où sont les sources ?

- * Le web : réseaux sociaux, photos, vidéos, documents en lignes, ...
- * Informations commerciales : BD de clients, chaîne de facturation, chaîne statistique, supply chain, historique, ...
- * Informations personnelles : dossier médical d'une personne, profil d'une personne, mails, ...
- * Données scientifiques : physique, astronomie, climat, ...
- * Données ouvertes : open data, <http://www.data.gouv.fr>

2. Les bonnes pratiques

- * Mise en place progressive, souplesse, tests, ... pas d'urgence.
- * Mise en place d'équipes dédiées.
- * Travailler sur vos données, pas celles de la concurrence.
- * Eviter les corrélations trop évidentes ...



2. Les bonnes pratiques

- * Mettre en place une solution de « dataviz »
- * Kibana lié à Elastisc Search ...

2. Les bonnes pratiques

- * 1869 Charles Joseph Minard : compagnie de Russie 1812-1813

- ## 2. Les bonnes pratiques
- * 1869 Charles Joseph Minard : compagnie de Russie
1812-1813



3. Data Scientist

- * Mise en place de formation (NYU).
- * Transversalité nécessaire du métier :
 1. Mathématicien, statisticien
 2. Informaticien
 3. Expert métier

3. Data Scientist

- * 4 juillet 2013, Fleur Pellerin lance un plan pour donner une impulsion à la filière Big Data en France :
 1. Ouverture d'une formation Data Scientist
 2. Ouverture d'un centre de ressources technologiques
 3. Lancement d'un fond d'amorçage dédié au Big Data
 4. Animation de différents écosystèmes permettant la rencontre entre les nouvelles technologies et les entreprises.

4. Limites du Big Data

- * Hyper connexion, infobésité, nécessité de décrocher
- * Nous ne sommes pas des machines ...
- * Problématique du Big Data :
 1. Problème de communications, de profils, de technologies
 2. Problème de disponibilité des données
 3. Problème de l'analyse des données
 4. Problème sur la communication des données, mise en perspective

4. Limites du Big Data

- * Paradigme nouveau : non structuration, informations catégorielles
- * Comment se rapprocher du « nowcasting » : limite
- * Le contexte : nécessaire à une analyse, pour lui faire prendre du sens.
- * Bruit parasite : données brutes/données nettoyées

4. Limites du Big Data

- * Manque de spécialistes : 200000 aux USA (2015)
- * 1,5 millions de managers avec une expertise Big Data



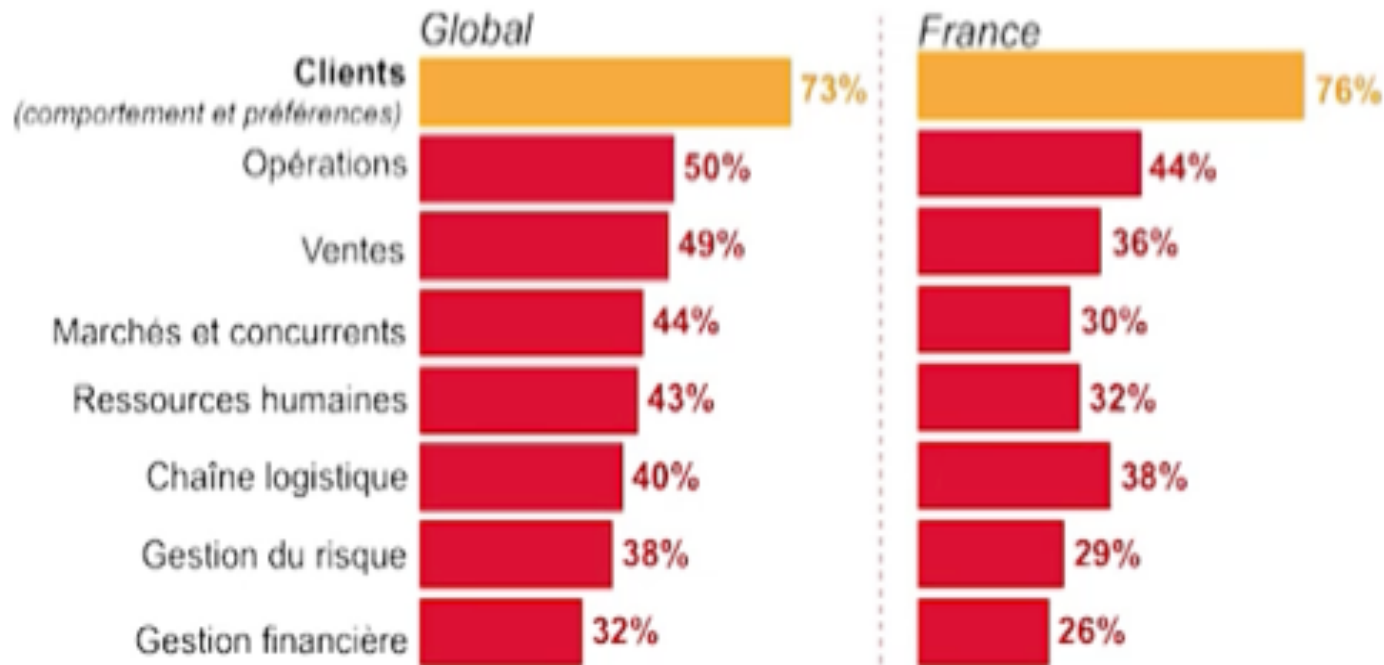
4. Limites du Big Data

* Transversalité :

1. Architectes logiciels (maîtrise des outils et leur mode de déploiement) du Cloud au mobile
2. Mathématiciens/statisticiens
3. Ergonomes et designers (data visualisation), ce nouveau métier d'ailleurs ne s'apprend pas encore.
4. Juristes (vie privée, données anonymes, propriétés intellectuelles)
5. Forces de vente (modèle commercial)
6. Stratégie (partenariat)

5. Applications économiques

- * Thales : projet Babel (dataviz), mur multitouch



5. Applications économiques

- * Big Analytics : x2,2 valorisation boursière
- * VESTAS : éoliennes (16 Po)
- * Université Ontario + IBM : maladies nosocomiales
- * IBM : watson
- * TerrasEcho : ondes lumineuses
- * Problème du business model, aide du Big Data

5. Applications économiques

- * «To make money, you've got to predict two things : what's going to happen and what people think is going to happen.»

Hal Varian - chief economist at Google.

6. Point de vue juridique

- * L'utilisateur doit savoir quand se fait la collecte de données, et son utilité.
- * La jurisprudence qui s'applique est celle de l'état où est implantée l'entreprise (USA)
- * Les requêtes sont systématiquement rejetées de l'autre côté de l'atlantique
- * Nécessité d'une jurisprudence pour la donnée
- * La suppression de 1 Go de données nominatives coûte 18 000 dollars (recherche, suppression, perte financière, ...).

6. Point de vue juridique

- * Skype n'est pas un opérateur de téléphonie publique : le droit en matière d'écoute est donc différent ...
- * La vente de MP3 « d'occasion » n'est pas autorisée aux USA (droits d'auteur pas arrivés à terme).
- * Le Big Data soulève deux questions juridiques :
 1. Protection juridique des données traitées.
 2. Usage qui est fait des traitements.

6. Point de vue juridique

- * Un fichier ne présente pas toujours une menace.
- * C'est l'interconnexion de fichiers, le croisement de données qui pose problème.
- * USA : la protection d'une BD se fait par le droit d'auteur, qui est caractérisé par l'originalité (effort créatif).
- * Au final, une base de données peut appartenir à tout le monde
- * En Europe : Lorsque l'on produit une BD sur laquelle on peut justifier des investissements matériels, humains, et financiers, la société est alors protégée.
- * Au final, elle ne peut pas appartenir à tout le monde

6. Point de vue juridique

- * Dommages et intérêts : en cas d'astreinte, une société accusée, qui gagne plus que ce qu'il faut payer, ne reconnaîtra pas sa faute.
- * Danger = Watson : en solution Cloud, exploitable dans tous les pays, avec aucun recours.
- * Avec le Cloud, plus aucune loi ne peut s'appliquer
- * Purger un système de données avant un contrôle est considéré comme une fraude.
- * Le conseil juridique, c'est donc de purger régulièrement vos systèmes de collecte ...

7. Tendance verte

- * Consomme autant que les autres activités ...
- * Peut apporter des économies dans les déplacements, dans les système de transport d'électricité (Smart Grid)
- * Le télétravail reste marginal ...
- * Tous les centres de traitements numériques dans le monde consomment l'équivalent de 30 centrales nucléaires
- * Utilisation de groupes électrogènes diesels d'appoint !
- * Utilisation d'un nombre incalculable de batteries !
- * Aucune substitution ne s'opère.

7. Tendance verte

- * La consommation mondiale de papier reste stable
- * Objectif zéro papier des années 1980 ?

7. Tendance verte

- * Obésité croissante des logiciels : obsolescence matérielle
- * Obésité croissante du web : augmentation de la consommation des infrastructures
- * De 0,8% en 2005 -> 2% en 2012 (consommation mondiale des infrastructures)
- * Dans 25 ans : consommation équivalente à celle de 2008 au niveau mondiale !!!
- * Il est nécessaire de prendre des mesures !

7. Tendance verte

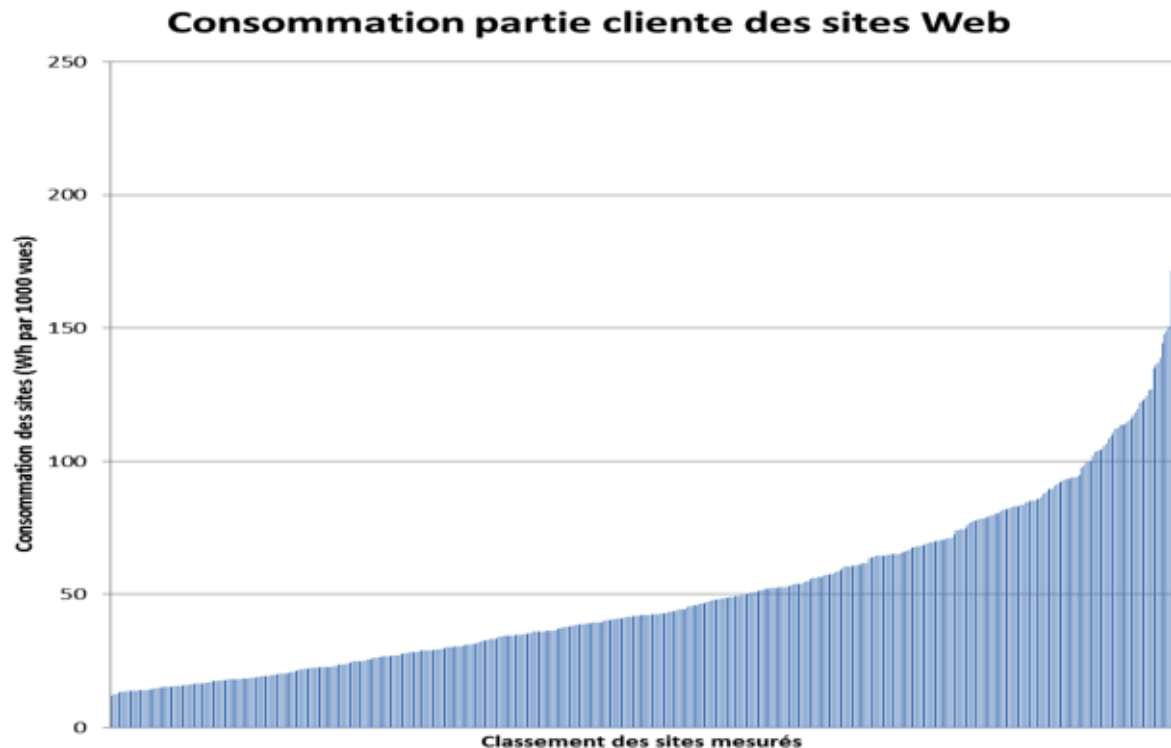
- * « Selon l'ADEME, l'essentiel des impacts environnementaux liés à une page web (épuisement des ressources non renouvelables, pollutions des sols et de l'air, eutrophisation de l'eau, etc.) sont notamment corrélés au temps passé par l'internaute devant son ordinateur et à la durée de vie active de cet ordinateur ».

7. Tendance verte

- * Eteignez vos « devices » s'ils ne sont pas utilisés ...
- * Limiter le nombre de requêtes inutiles ...
- * Utiliser des marques pages ... (cela évite également la collecte vous concernant sur vos habitudes)
- * Utiliser des proxy pour jouer le rôle de cache ...
- * Privilégiez les « devices » aux « desktops »
- * Simplifiez vos infrastructures, limitez les redondances inutiles (stockage)
- * Pensez que l'électricité, ce sont des énergies fossiles le plus souvent (pour l'instant), ou des batteries ...

7. Tendance verte

- * Du point de vue du client qui utilise un site web :
- * Consommation : La moyenne est de 50,1 Wh. La médiane est de 42,2 Wh.



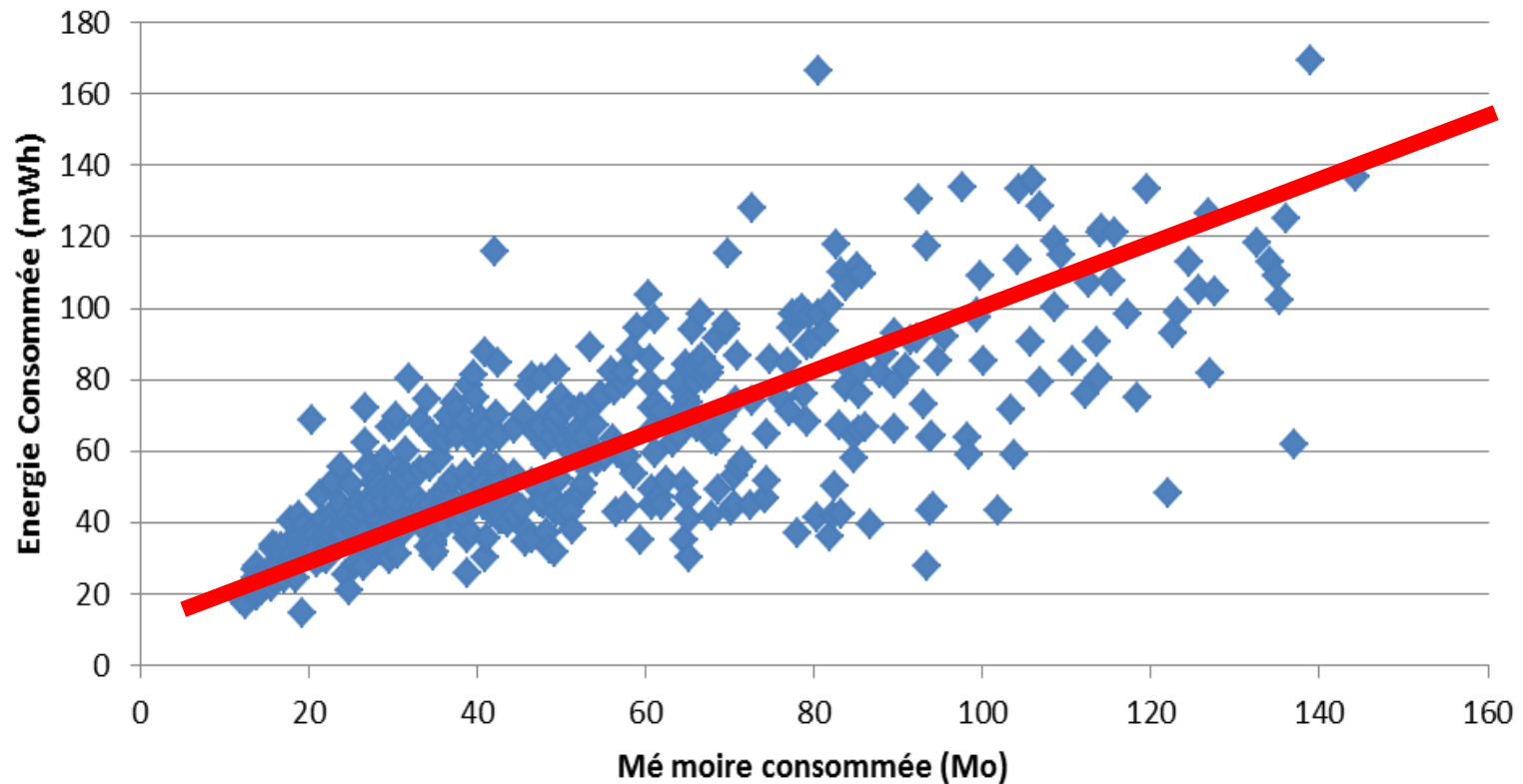
7. Tendance verte

* Corrélation en différentes mesures : ratio

Energie / mémoire	Energie / Page speed	Energie / Nb Req	Energie / Chargement	Energie / Taille	Mémoire / page speed	Mémoire / Nombre de req	Mémoire / Chargement	Mémoire / Taille
0,77	0,17	0,69	0,30	0,50	0,23	0,73	0,39	0,6666

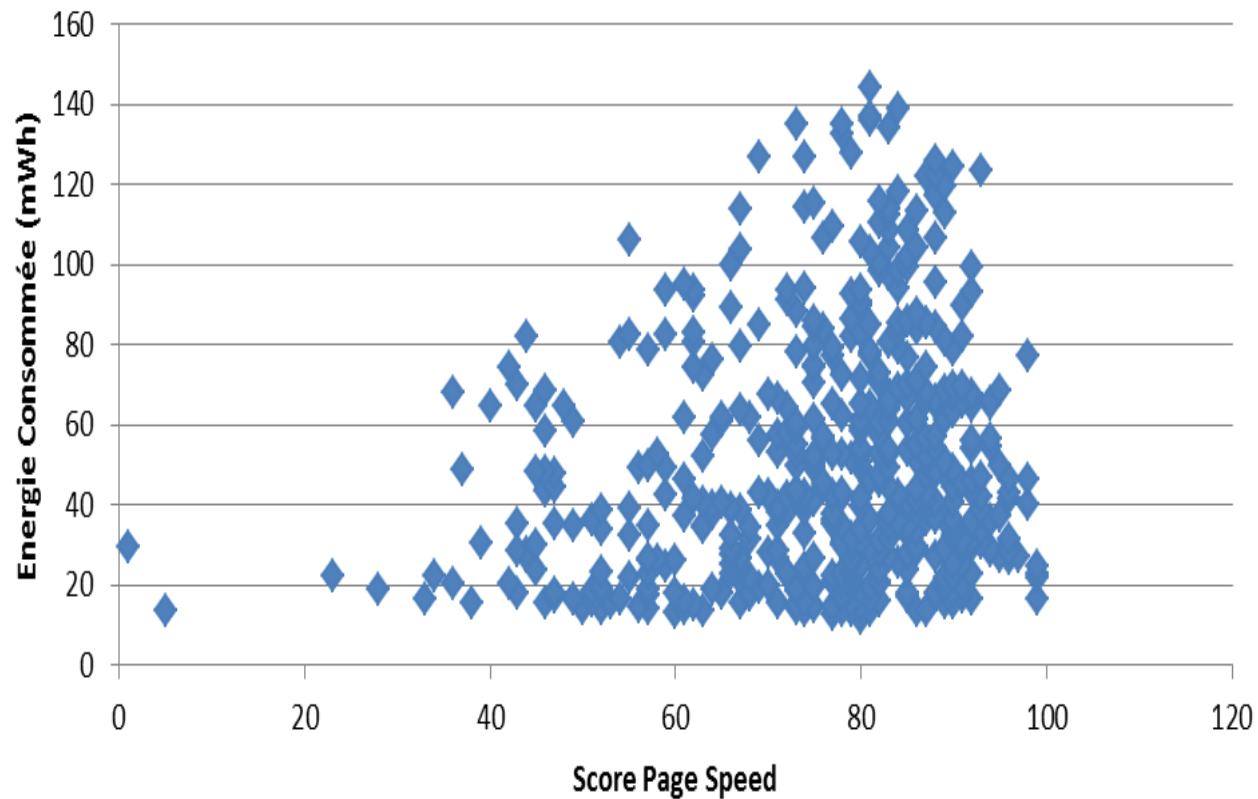
7. Tendance verte

Energie Vs Mémoire



7. Tendance verte

Energie Consommée Vs Score Google Page Speed

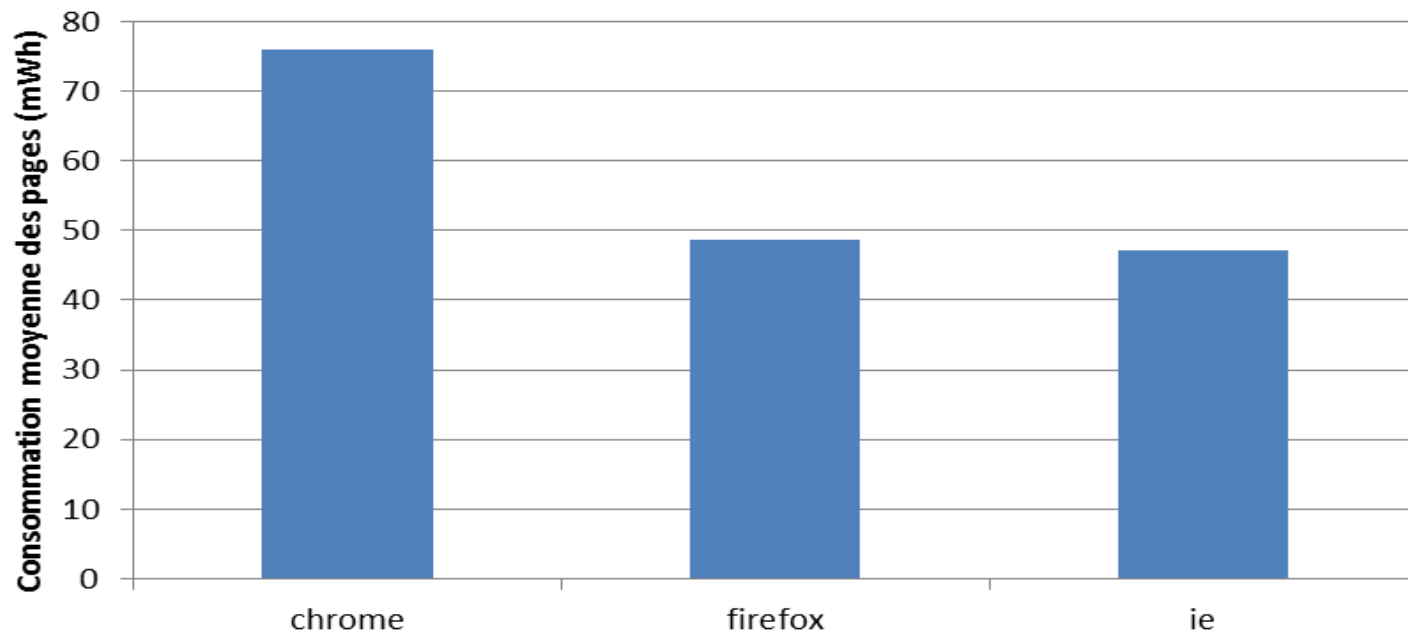


7. Tendance verte

- * Google Page Speed : <https://developers.google.com/speed/pagespeed/insights/>
- * Conseils d'optimisation mobile/PC
- * Indice de chargement
- * Expérience utilisateur

7. Tendance verte

Consommation des sites en fonction des navigateurs



7. Tendance verte

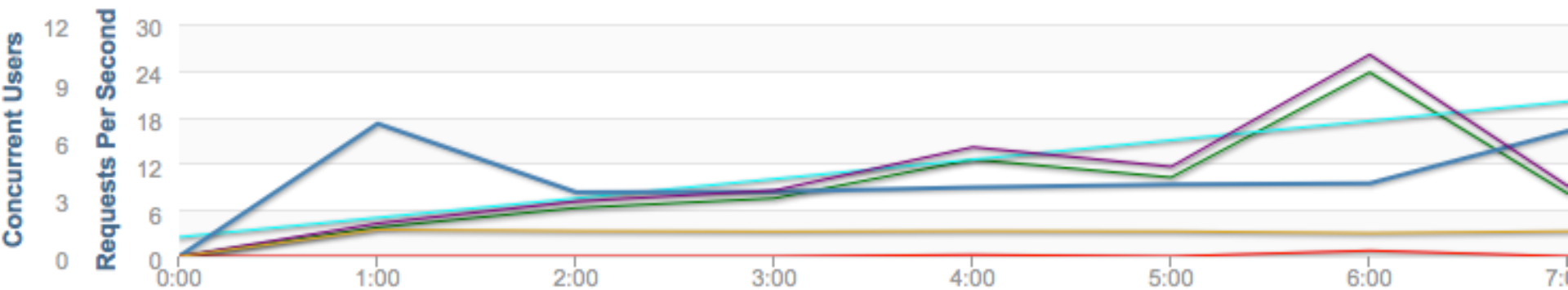
- * consommation moyenne requête http = 28 uWh
- * Une page = 1 à plusieurs dizaine de requêtes
- * Résultat = moyenne de 2,7 mWh avec une médiane de 1,7 mWh pour un panel de 100 sites web le plus utilisés.
- * lemonde.fr= 3,3 mWh

7. Tendance verte

- * 100 sites les plus visités (France) = consommation de 68 GWh soit la consommation de 25 400 foyers (France) !
- * Consommation : partie PC + partie serveur + partie infrastructure
- * La consommation via smartphone est moindre
- * On peut d'ailleurs faire des estimations sur le coût au niveau batterie.

7. Tendance verte

* <http://loadstorm.com> : Charge temps réel



* <http://www.webpagetest.org> : impact de chargement d'une page web en temps

