

Panda 4.2 : une mise à jour officielle mais insaisissable ?



Par Philippe Yonnet

Domaine :	Recherche	Référencement
Niveau :	Pour tous	Avancé

Les webmasters dont le site avait été victime du filtre Panda l'attendaient depuis longtemps : au cours du week end du 18/19 juillet 2015, Google a déployé une nouvelle itération du filtre Panda, la version 4.2. Et a en même temps communiqué sur les caractéristiques surprenantes de cette mise à jour, notamment un déploiement extrêmement lent : il doit durer plusieurs mois. Ce comportement totalement différent signifie-t-il qu'il s'agit d'un tout nouvel algorithme (est-ce un Panda 5.0, voire vraiment encore un filtre "Panda") ? Quel est l'impact de cette itération de l'algorithme ? Et pourquoi est-il si lent à se déployer ? Autant de questionnements que nous allons nous efforcer d'éclaircir dans cet article...

Une mise à jour qui s'est fait attendre et toujours repoussée

La dernière itération de Panda officialisée, Panda 4.1, avait été déployée le 25 septembre 2014. Il a donc fallu attendre plus de 10 longs mois pour que le filtre soit enfin mis à jour. 10 mois qui ont dû paraître interminables aux webmasters « pandalisés » par les versions précédentes, car, rappelons-le, on ne récupère la visibilité perdue par un site victime de Panda qu'à l'occasion de l'une de ces itérations du filtre qui permettent une réévaluation du site.

Cette mise à jour avait été annoncée dès début juin par Gary Illyes, qui l'annonçait lors du SMX Advanced, « pour dans deux à quatre semaines ». Mais un mois plus tard, rien ne s'était passé.

Entre fin juin et le 18 juillet, les porte-parole de Google ont été régulièrement interrogés sur les raisons de ce retard. John Mueller a invoqué des « raisons techniques » pour l'expliquer.

Quant au délai assez long entre les deux mises à jour, Gary Illyes a donné quelques éléments d'explications au cours du SMX Advanced : selon lui, le filtre classe les sites à Pandaliser à partir de données collectées sur ces sites pendant des semaines, voire des mois.

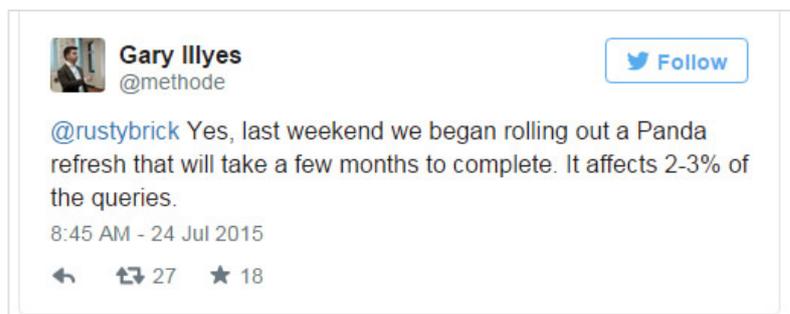


Fig.1. Le 24 juillet, Gary Illyes a confirmé officiellement sur Twitter le déploiement d'un « Panda refresh ». L'information avait été communiquée par Barry Schwartz dès le 22/07.

Mais ces données peuvent contenir du « bruit », des informations parasites. Dans ce cas, la seule solution pour « nettoyer » ces données, c'est d'élargir la collecte sur une période de temps encore plus longue.

Est-ce l'explication du délai important entre ces deux mises à jour : on peut supposer que oui...

Un impact faible, et une mise à jour présentée comme une simple itération

Cette version est qualifiée de « refresh », il s'agit donc d'une simple itération, et pas d'une nouvelle version du filtre comme en mai 2014. Logiquement, elle a donc été numérotée Panda 4.2, et non Panda 5.0. Mais si c'est bien une itération, elle est tellement atypique que ce qualificatif de « refresh » laisse perplexe.

L'impact annoncé est de 2 à 3% des requêtes, ce qui est plutôt dans la fourchette basse de l'impact des itérations de Panda. Depuis Panda 4.0, on est bien entré dans l'ère des « softer updates », des mise à jours plus « douces », qui sélectionnent plus finement les sites impactés, génèrent moins de faux positifs (mais en génèrent toujours...) et ont donc un impact moins massif sur les pages de résultats.

Un déploiement étalé sur ... plusieurs mois

L'une des caractéristiques étonnantes de cette itération est que pour la première fois, son déploiement est annoncé comme devant durer plusieurs mois. D'habitude, le filtre est déployé rapidement, en quelques heures, voire quelques jours (on soupçonne néanmoins Panda 4.1 d'avoir mis dix jours à être complètement déployé).

A la date de rédaction de cet article (mi-septembre), on ne savait toujours pas si le déploiement était achevé. Le 8 septembre, dans un hangout, John Mueller a avoué qu'il pensait que non, mais qu'en fait, il avait demandé récemment des nouvelles à ce sujet, mais n'avait pas encore eu de réponse.

Pourquoi un déploiement aussi lent ? Un changement d'infrastructure de crawl et d'indexation ?

On peut émettre une hypothèse pour expliquer la lenteur du déploiement : quelques jours après l'annonce du déploiement de Panda 4.2, Gary Illyes a confirmé que Google pouvait crawler plus d'urls dorénavant. Il a lâché cette information sur Twitter en réponse à une question de Glenn Gabe qui avait remarqué un pic de crawl sur les sites de ses clients (fig.2).

Un peu plus tard, Gary Illyes a confirmé que ce changement de comportement de crawl était lié à une augmentation des capacités de la capacité de traitement de Google (fig.3).



Fig.2. Le tweet de Glenn Gabe.



Fig.3. La réponse de Gary Illyes.

Dans le passé, Google avait tendance à communiquer sur ces changements d'infrastructure de façon beaucoup plus officielle, notamment à l'occasion de Caffeine ou d'Hummingbird (on se souvient également de l'infrastructure surnommée Big Daddy dans le passé). Ces changements ont toujours été décrits comme lents et délicats...

Est-ce ce changement d'infrastructure qui a autant retardé le déploiement de cette nouvelle itération de Panda ? Cela semble très probable. Le déploiement du filtre sur la nouvelle infrastructure, et le nouveau comportement de crawl, peuvent expliquer bien des choses.

Est-ce à cause de cette nouvelle infrastructure que Panda 4.2 se propage lentement ? Peut-être aussi, mais il est plus difficile de relier les deux événements faute d'infos précises. Ce qui est clair, c'est que John Mueller a

évoqué de vagues « problèmes techniques » et des « raisons techniques » à plusieurs reprises pour expliquer les retards et ce comportement particulier (notamment dans un hangout le 31 juillet).

Un impact progressif sur les positions occupées par vos pages

La lenteur du déploiement se caractérise, d'après Google, par un impact progressif du filtre sur une page de votre site après l'autre, un impact qui peut donc mettre des mois à affecter l'ensemble des pages du site...

Donc, si votre site est victime du filtre, les pages vont chuter dans les classements, les unes après les autres, jour après jour, jusqu'au déploiement final, qui interviendra plusieurs mois après le 18 juillet...

Google Panda impacte-t-il toujours l'ensemble des pages d'un site ?

En dépit de ce nouveau comportement, qui conduit pendant la phase de déploiement à des situations où des pages peuvent être impactées par le filtre, tandis que d'autres ne le sont pas encore, le filtre reste appliqué à l'intégralité du site... Ce point a été confirmé à plusieurs reprises par Google.

Par contre, l'impact du filtre n'est jamais uniforme : le « malus » apporté par Panda aux notes d'une page

individuelle conduit à des changements plus ou moins visibles en fonction du niveau des autres signaux : certaines pages sont déclassées, mais pas toutes. Et cela change d'une requête à une autre. Pas facile...

Comment savoir si j'ai été touché par Panda 4.2 ? Ou si je suis sorti du filtre Panda ?

Panda, dans ses premières versions, était déployé très rapidement. Le déploiement le plus long a duré une dizaine de jours, mais les sites impactés ont toujours vu leur visibilité SEO être impactée rapidement. Pour savoir si un site était impacté par Panda, il suffisait :

- D'éliminer toutes les autres causes possibles (problème technique par exemple, ou changement sur le site au même moment ;
- Et de vérifier que le décrochage collait bien à la date de déploiement de Panda.

Pour la version 4.2, on nous annonce un déploiement très long (plusieurs mois) avec un impact progressif sur les pages des sites impactés au fur et à mesure du crawl...

De telles caractéristiques : déploiement long, impact progressif, rendent en fait très difficile l'attribution d'une baisse de visibilité SEO à Panda, car toutes les conditions soient réunies pour que l'impact « passe sous le radar », ou qu'il soit impossible d'attribuer la baisse à cette cause plutôt qu'une autre.

Le paradoxe est que certains webmasters ont quand même observé des chutes radicales post Panda 4.2. Et aussi des sorties de Panda. Est-ce bien dû à cette nouvelle mouture du filtre Panda ? Peut-être, ou peut-être pas. Il existe une contradiction entre les nouvelles caractéristiques annoncées de Panda, et les courbes observées par ces webmasters, qui sont conformes à l'ancien fonctionnement.

D'habitude, il était possible d'analyser grâce à des outils comme Searchmetrics ou Sistrix, une semaine après le déploiement de Panda, la liste des principales chutes et hausses de visibilité pour identifier parmi les sites présents dans ces listes ceux qui avaient probablement été impactés par le filtre.



Fig.4. Un exemple de sortie de Panda, consécutive au déploiement de Panda 4.0 : la reprise était toujours brutale, extrêmement visible, car l'ensemble des pages retrouvaient leur ancienne visibilité en quelques heures, quelques jours au maximum !

Pour Panda 4.2 , de l'aveu même de Marcus Tober de Searchmetrics, ses données ne montraient rien de notable fin juillet :

<http://blog.searchmetrics.com/us/2015/07/24/google-panda-update-4-2-winner-loser/>

Quelques cas suspects

Malgré cela, quelques observateurs ont attribué à Panda 4.2 certaines chutes ou hausses suspectes. Et effectivement, certains cas sont troublants. L'évolution est moins brutale, mais pour ces sites, l'impact reste spectaculaire (fig. 5 et 6).

Panda 4.2 a-t-il permis à des sites Pandalisés à l'automne 2014 (ou avant) de récupérer leur visibilité ?

Cette itération est aussi atypique, car on cherche toujours des exemples vérifiables de sites clairement « Pandalisés » qui auraient retrouvé leur visibilité dans les moteurs de recherche à la suite de cette itération.

Certes, certains ont cru voir un gain significatif fin juillet, mais ces gains ont été annulés au bout de deux ou trois semaines ! Panda 4.2 est donc une itération dont on voit difficilement l'impact, à la hausse, ou à la baisse ...

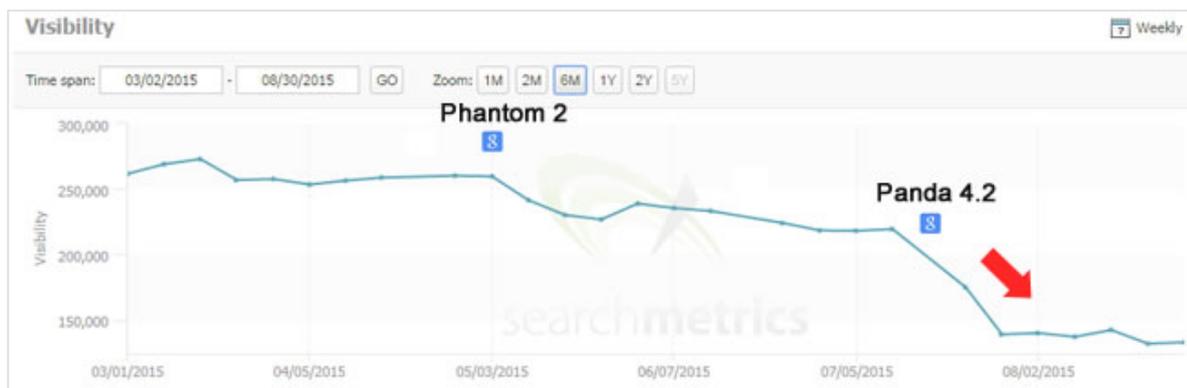


Fig.5. Un exemple de baisse de trafic fourni par Glenn Gabe, de GSquared Interactive



Fig.6. Exemple "suspect" sur un site français.



Fig.7. La courbe de trafic du site Seroundtable.com communiquée par Barry Schwartz : le boost « post Panda » a duré deux semaines, au bout d'un mois, le trafic était revenu au niveau précédent.



Fig.8. Un exemple, donné par Glenn Gabe de G Squared Interactive, de site qui pensait avoir été Pandalisé fin juillet, alors qu'il s'agissait d'un simple transfert de trafic ...

Pourquoi certains webmasters pensent que Google est revenu en arrière ?

Parmi ceux qui ont cru voir l'impact de Panda 4.2 sur son site, on compte Barry Schwartz, l'éditeur du site Seroundtable.com. Selon lui, son site présentait tous les signes d'une sortie de Panda après le week-end du 19 juillet.

Puis, début septembre, il a constaté que la hausse de trafic ne s'était pas confirmée au fil des jours et des semaines, et que tout se passait comme si « les effets de Panda 4.2 » étaient annihilés (fig.7). D'autres webmasters ont confirmé avoir observé le même phénomène : une hausse fin juillet, suivi d'un retour aux niveaux précédents quelques semaines plus tard.

Barry Schwartz a émis l'hypothèse que Google avait peut-être annulé le déploiement de Panda et effectué un roll back. Le problème est que la hausse qu'il constatait fin juillet est peut-être due à autre chose, comme le retour à des niveaux normaux (fig.8).

Que faire si mon site a été impacté négativement par le filtre 4.2 ?

Dans l'hypothèse où vous avez clairement pu associer une baisse sensible de visibilité au filtre Panda 4.2 (ce qui, répétons-le, est particulièrement complexe avec cette version), c'est le moment de régler tous les problèmes de « qualité » qui affectent votre site et son contenu.

Les critères de qualité, spécifiques à Panda, sont définis dans un post d'Amit Singhal sur le blog de Google pour les webmasters :

<http://googlewebmastercentral.blogspot.fr/2011/05/more-guidance-on-building-high-quality.html>

On trouve d'autres consignes dans les guidelines de Google :

https://support.google.com/webmasters/answer/35769?hl=fr#quality_guidelines

En réalité, le filtre sélectionne également les sites à « Pandaliser » sur d'autres critères : nous vous invitons à relire à ce sujet notre article de juin 2014 sur Panda 4.0, sur cette même lettre professionnelle d'Abondance (<http://recherche-referencement.abondance.com/2014/06/panda-40-anatomie-de-la-derniere.html>).

Quelles conclusions en tirer pour les futures itérations de Panda ?

Finalement, rien n'aura été normal dans le déploiement et les caractéristiques de cette itération 4.2. La communication confuse, parfois contradictoire, de Google n'a pas arrangé les choses. La lenteur du déploiement et le caractère progressif de l'impact nous empêchent peut-être d'analyser ce qui s'est passé efficacement à la date d'aujourd'hui.

La prochaine itération de Panda sera-t-elle semblable, ou reviendra-t-on à un schéma plus classique ? Difficile à dire. Quand interviendra-t-elle ? Impossible de le prédire, car même en interne chez Google, on voit bien que les porte-paroles « officiels » se sont trompés sur la date effective de lancement de la dernière version.

Et il y'aura-t-il réellement un Panda 4.3 ? Et si cette version était en fait la version « mise à jour continue » promise depuis longtemps ? Google a en fait déclaré qu'il ne s'agissait que d'un premier pas dans cette direction...

Bref, on se perd en conjectures. Et disons-le, Google entretient notre confusion, volontairement ou non. Nous venons de connaître une séquence étrange, avec une mise à jour non officielle (la « Phantom update » de mai 2015) dont l'impact était bien visible, et une mise à jour officielle dont on n'arrive pas à mesurer les effets... Le SEO n'est pas un métier facile...

Webographie

L'analyse de Glenn Gabe sur Panda 4.2 :

<http://www.hmtweb.com/marketing-blog/panda-4-2-analysis-and-findings>

Le post d'annonce de Google Panda en février 2011 :

<http://googleblog.blogspot.com/2011/02/finding-more-high-quality-sites-in.html>

Le billet d'Amit Singhal de Google sur les consignes pour créer du contenu de qualité :

<http://googlewebmastercentral.blogspot.fr/2011/05/more-guidance-on-building-high-quality.html>

L'interview de « Google » sur Panda 4.2 publiée par Barry Schwartz sur Searchengineland :

<http://searchengineland.com/google-panda-4-2-faqs-we-interviewed-google-on-the-latest-panda-update-227065>



Philippe YONNET, *Directeur Général de l'agence Search-Foresight, groupe My Media*
(<http://www.search-foresight.com>)