



LE LIVRE NOIR DES BIG DATA

Les Big Data, une chance pour votre entreprise ?
A l'heure du buzz, poser la question est déjà presque
un crime. Pourtant, prises sous le mauvais angle, les Big
Data peuvent se transformer en piège.

« Sans
interprétation,
les Big data ne
produisent que
du bruit »

Brian Solis
prospectiviste



LES BIG DATA : UNE « VIEILLE INNOVATION » ?

Si l'on en croit le cabinet Gartner, les Big Data en tant que telles ne sont plus une innovation : les nouveautés résident dans leurs applications, comme l'analytique prédictive.

Chaque année, le cabinet d'étude Gartner publie son célèbre *hype cycle* des technologies émergentes, un schéma représentant le degré d'adoption des nouvelles technologies par l'industrie. En 2014, les Big Data entamaient leur descente du « pic des espérances exagérées » vers le « gouffre des déceptions ». C'est le passage critique avant l'accession à la maturité, lorsqu'une technologie est bien comprise et exploitée pour ce qu'elle est. Aujourd'hui, ne cherchez plus les Big Data sur le *hype cycle*. Tout simplement parce que Gartner considère que ce **n'est plus une technologie émergente***, mais une technologie entrée dans l'ère des usages pratiques.

Et ces usages, ils sont nombreux. Evoquons-en simplement trois, parmi tant d'autres, parce qu'ils changent réellement la manière de travailler des entreprises. L'analyse prédictive, pour commencer, permet d'identifier des traits caractéristiques (*patterns*) dans des masses de données, et trouve des applications aussi bien dans l'industrie (maintenance prédictive) que dans les services (prévention de l'attrition – ou *churn*) ; l'exploration des Big Data permet, quant à elle, de détecter

des signaux faibles, permettant une segmentation fine de sa base de clients et une personnalisation poussée de ses campagnes ; enfin, l'analyse de flux de données en temps réel offre une réactivité sans précédent, notamment au travers des outils de datavisualisation.

Dans ce contexte, il est surprenant de noter que de très nombreuses entreprises se lancent encore dans les Big Data au travers de POC (« *Proofs Of Concept* »). Dans cet ouvrage, nous allons montrer que le recours à un POC pour aborder les Big Data est au mieux une fausse bonne idée, et au pire le moyen le plus sûr de perdre la course à l'innovation. ●

*http://www.theregister.co.uk/2015/08/21/forget_big_data_hype_says_gartner_as_it_cans_its_hype_cycle/

285 M€

Marché des usages pros des Big
Data en France (source : IDC)

QU'ENTEND-ON PAR « POC » ?

En France, la preuve de concept reste trop souvent la norme en matière de Big Data. Or, si elle faisait sens à l'époque des premières expérimentations, elle n'est plus adaptée à l'heure des usages opérationnels.

Pour bien comprendre pourquoi les POC sont si mal adaptés aux projets Big Data, arrêtons-nous sur leur définition. Comme son nom l'indique, un « *Proof Of Concept* » sert à démontrer la faisabilité de quelque chose. Il s'agit de prouver qu'une idée (« concept ») peut prendre corps dans une réalisation concrète. Savoir si cette réalisation sera utilisable en situation réelle n'est pas l'objet de la démonstration, comme l'explique fort bien l'inventeur de l'expression (voir encadré ci-contre).

De cette définition, on peut déduire quatre caractéristiques principales d'un POC :

1. UN POC EST UNE RÉALISATION À PETITE ÉCHELLE

Faire un POC en Big Data, c'est, paradoxalement, se cantonner à une petite échelle : un pays, un département d'une entreprise, une gamme de produits... Pour un département Commercial, par exemple, on s'intéressera aux remontées de caisse sur une zone particulière de chalandise, en faisant abstraction de toutes les autres. C'est une bonne approche, permettant de valider des technologies, mais à condition de veiller à la cohérence de l'échantillon, faute de quoi l'on risque d'introduire

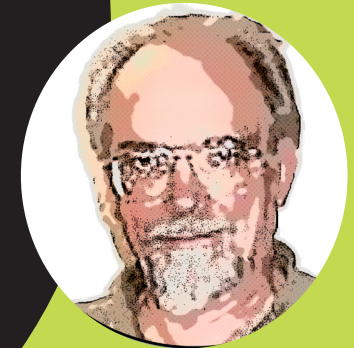
des biais considérables, pouvant complètement fausser l'interprétation des chiffres.

2. IL EST MENÉ DANS DES CONDITIONS « DE LABORATOIRE »

Un POC est conçu *in vitro*. On utilise des jeux de données particuliers, sur un périmètre limité, pour valider des hypothèses prédéfinies. Sur des technologies de rupture, sans impératif de production, cela peut être tout à fait pertinent, pour identifier les futurs changements de paradigme. Or le terme « Big Data » est malheureusement devenu un *buzz word*, et toutes les entreprises veulent avoir « leur » projet Big Data. C'est pourquoi les POC ne sont plus menés en mode exploratoire, avec une démarche scientifique, mais pour valider coûte que coûte le fait de se lancer dans les Big Data.

En informatique, les développeurs ont une expression pour désigner le test d'une solution dans les conditions optimales. Ils parlent de « *golden path* » ou de « *happy day scenario* », c'est-à-dire d'un scénario de test dans lequel l'utilisateur ne rencontre que des cas dont on sait à l'avance qu'ils ne posent pas de problèmes. Dans le monde des Big Data, un POC est souvent réalisé dans cette optique : il y a dès lors autant de différences entre

Bruce Carsten
Chercheur, ingénieur
et consultant*



« Prototype preuve de concept est un terme que j'ai inventé (me semble-t-il) en 1984. On l'utilisait pour désigner un circuit construit sur des lignes similaires à un prototype industriel, mais dont le but était uniquement de démontrer la faisabilité d'un nouveau circuit et/ou technique de fabrication, et dont le but n'était pas d'être une première version d'un design de production ».

*Carsten's Corner. Power Conversion and Intelligent Motion, November 1989

un POC et la réalité qu'entre un circuit de voitures télécommandées et la circulation automobile sur route. Pour qu'un POC permette de refléter ce qui va se passer dans le monde réel, il faudrait que celui-ci puisse être totalement modélisé à l'avance, en évacuant tout aléa, ce qui est la plupart du temps impossible. Par ailleurs, lorsqu'elles réalisent un POC, les équipes ont également tendance à fonctionner dans un état d'esprit « R&D » : le POC est bâti par itérations, en mode « essais-erreurs ». Or, si c'est une bonne pratique pour un POC, elle s'avère inadaptée pour un projet devant fonctionner en situation réelle.

3. UN POC EST UNE RÉALISATION À FAIBLES MOYENS / FAIBLE BUDGET

Une entreprise ou un département dans une entreprise se lance dans un POC lorsqu'il est en situation de doute vis-à-vis d'une technologie. Dans la balance, la crainte de dépenser de l'argent à fonds perdus l'emporte sur le désir suscité par les promesses de la technologie. C'est pourquoi les budgets consacrés à un POC sont généralement très limités. D'autant plus qu'ils sont souvent réalisés en dehors des circuits budgétaires habituels, par exemple par une direction métier souhaitant « pousser »

Parce qu'il est empreint d'un fort caractère « R&D », le POC reste à mille lieues des situations réelles.

un projet et convaincre soit la DSI soit la Direction Financière. Cette faiblesse des ressources conduit à son tour à se cantonner à certains cas d'usages sur un périmètre restreint : la boucle est bouclée, mais le POC est alors à mille lieues des situations réelles qui seront rencontrées.

4. UN POC N'APPORTE AUCUNE GARANTIE DE MONTÉE EN CHARGE

Compte tenu des limitations budgétaires inhérentes aux POC, ceux-ci privilégient les solutions matérielles et logicielles *low cost*, sans tenir compte des contraintes de production. Or les environnements de production répondent à des cahiers des charges parfaitement normés, que ce soit en termes de formats et de nomenclatures de fichiers, de systèmes de bases de données, de technologies réseaux, etc. Un système en production doit aussi répondre à des exigences strictes en matière de disponibilité, de qualité de service, de sécurité, etc. Le POC n'intègre pas toutes ces dimensions, tout simplement parce que, par définition, il est généralement déconnecté du reste du SI. Une fois que le POC a démontré ce qu'il devait démontrer, il a rendu le service qu'on attendait de lui, et on le remise sur une étagère. Si l'on sait que le « concept » est réalisable, il va falloir concevoir une architecture différente pour qu'il fonctionne en production, avec les solutions déjà en place. En termes de développement, on repart donc quasiment de zéro. Du point de vue financier, le POC constitue une perte sèche, et il n'a pas permis d'avancer sur le chemin qui mène à une solution pérenne en conditions réelles. ●



Abdoul Seck
Responsable Big Data
et Analytics

TESTER UNE SOLUTION POUR LE MONDE RÉEL AVEC UN POC : UN CONTRESENS

« Pour beaucoup d'entreprises françaises, se lancer dans les Big Data consiste avant tout à déterminer si une technologie Big Data est une solution applicable ou non à leurs usages. Pour cela, elles utilisent les POC qui ont pour objectif de démontrer qu'il est possible de mettre en œuvre des technologies Big Data, et de traiter des *use cases*. Les entreprises utilisent donc ces programmes pour **répondre à leurs problématiques d'industrialisation**. Mais si un POC permet bel et bien de monter en compétence techniquement, dans son usage même, il ne permet pas d'adresser la question du passage en production. Ce qui fait défaut actuellement, ce sont des projets qui permettent un retour d'expérience, bien plus profitable ».

LES EFFETS PERVERS DES POC

Alors que les Big Data sont déjà utilisées couramment par de nombreux acteurs, notamment aux Etats-Unis, le POC n'est pas seulement inadapté : il est contre-productif.

Gardez bien ceci en tête : pendant que vous faites un POC, vos concurrents lancent des projets pilotes. En la matière, l'avance des Etats-Unis est particulièrement significative. Un éditeur français nous confiait avoir réalisé 100 K€ de CA en France sur les Big Data en 2014, contre 2 M€ aux USA. Certes, la taille des marchés est différente, mais la tendance est la même si l'on s'intéresse au montant moyen des contrats : 18 K€ en France contre 140 K€ aux Etats-Unis. Ces chiffres révèlent une chose : de ce côté-ci de l'Atlantique, les entreprises se contentent encore trop souvent de tests (POC), alors que leurs concurrentes américaines investissent déjà dans de véritables projets. Autrement dit, ces dernières sont en train d'accumuler de l'expérience et de se mettre en ordre de marche pour conquérir le marché des nouveaux usages permis par les Big Data. En rester au stade des POC, c'est pour les entreprises françaises le risque de prendre un retard qui sera difficile voire impossible à rattraper.

LE POC, UN PRÉTEXTE POUR NE PAS INNOVER

Quelle entreprise ne veut pas se présenter comme innovante ? Mais il y a deux façons de considérer l'innova-

tion. La première, c'est de développer des choses nouvelles, jamais vues jusqu'alors. En automobile, ce sera par exemple un *concept car*. Le but est de démontrer sa capacité à faire du neuf, mais aussi d'impressionner, de créer un « effet Waow ». A ce stade, on ne se préoccupe pas de savoir comment industrialiser le produit.

La deuxième, c'est **d'introduire du neuf dans quelque chose qui a un caractère bien établi***. Là, c'est beaucoup plus compliqué. Innover avec les Big Data, ce sera pour une entreprise introduire du neuf dans ses processus ou dans ses offres grâce à une nouvelle approche des données.

Malheureusement, un POC correspond bien souvent à la première vision de l'innovation. Un budget innovation lui est consacré, et le POC sert de pièce à conviction pour prouver que l'on a bien innové. Quelquefois, le POC est exposé, comme on expose un *concept car* ; il fait l'objet d'un article dans le magazine interne, voire obtient des retombées presse. Et puis, on le range, et on continue à faire du *business as usual*.

Autrement dit, le POC, loin de permettre à l'entreprise de gagner des parts de marché grâce à des innovations



Nicolas Laroche
Président JEMS datafactory

GAGNER EN EXPÉRIENCE CONCRÈTE, LE MEILLEUR MOYEN D'AVANCER

« Bien souvent, une entreprise s'interroge d'abord sur les technologies à employer, avant de savoir comment utiliser cet investissement. Elle prend alors un risque, car, dans bien des cas, elle s'engage dans un procédé en totale rupture avec ce qu'elle connaît, sans savoir comment l'appréhender car elle ignore son fonctionnement et ses bonnes pratiques. Pour éviter de se laisser distancer par la concurrence, nous expliquons à nos clients que le mieux est de se lancer dans un projet pilote qui servira de base pour un projet de plus grande ampleur. **Un pilote permet de se forger une expérience**, en se confrontant à la réalité de la production.

Derrière le problème des POC se cache celui de la méthodologie. Il arrive que des entreprises soient déçues de la mise en place de solutions Big Data, qu'elles jugent non performantes et ne répondant pas à leurs attentes. Mais en réalité, il leur manque des bonnes pratiques et de la méthodologie.

Les clients sont réfractaires aux risques, c'est une raison qui explique ces tests de technologies avant de les lancer en production. Mais ce n'est pas une bonne démarche. A l'inverse, nous les accompagnons pour leur proposer une vraie méthodologie de mise en place de solutions ».

*Trésor de la Langue Française informatisé. <http://www.cnrtl.fr/lexicographie/innover>

produits (ou services), est souvent utilisé comme une excuse pour continuer à faire comme avant.

LE POC, VECTEUR D'INCOMPRÉHENSION ENTRE L'IT ET LES MÉTIERS

Les échelles de temps entre les directions fonctionnelles, notamment le Marketing, et les DSI sont différentes. Le Marketing veut pouvoir réagir le plus vite possible aux évolutions des marchés, alors qu'un changement d'infrastructure IT prend nécessairement du temps.

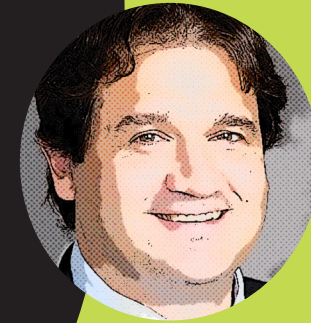
Le POC peut être un moyen d'apporter une mauvaise réponse à ce problème. La tentation sera, pour le Marketing, de réaliser lui-même un POC Big Data, dans un cloud public, à partir de son propre jeu de données, sans impliquer l'IT ; puis de mettre cette dernière dans la boucle, en fin de course, en lui demandant de passer le POC en production.

Comment s'étonner que l'IT oppose généralement une fin de non-recevoir à ce genre de demande ? Les informaticiens ne peuvent tout simplement pas passer le POC en production tel quel, car il est souvent incompatible avec les SI réels, voire pourrait compromettre la stabilité des solutions en place. S'en suit une série de récriminations de part et d'autre, le Marketing et l'IT se renvoyant la balle de l'impossibilité d'innover dans l'entreprise.

EN CONCLUSION

Un POC présente bien des limites et peut engendrer, nous l'avons vu, des effets pervers. Mais ceux-ci se produisent quand on fait du POC un mauvais usage. En matière de SI, les preuves de concept ne sont pas mauvaises en soi. Elles sont même très utiles en phase de R&D, pour explorer de nouvelles possibilités. Ce que nous avons voulu montrer dans ce livre noir, c'est que cette phase est désormais largement derrière nous en ce qui concerne les Big Data. L'étape du POC est devenue superflue et fait perdre du temps. Par ailleurs, c'est une erreur de croire qu'elle pourrait être la première marche avant de passer en production : les univers de l'expérimentation et de l'exploitation en conditions réelles sont trop éloignés.

Pour démarrer dans les Big Data, il est légitime de vouloir y aller progressivement : mais c'est par un projet pilote, bien différent d'un POC, qu'il faut commencer. C'est ce que nous allons voir dans le « livre blanc des Big Data ». ●



Cyrille Coqueret
Directeur technique
JEMS datafactory

LE POC, UN PRÉTEXTE POUR NE PAS INNOVER

« Trop peu d'entreprises françaises franchissent le pas de l'investissement dans les Big Data, incertaines quant au ROI qu'elles pourraient en attendre. En effet, les clients voient encore trop souvent l'usage des technologies Big Data comme un nouveau marché d'équipement qui coûte cher, alors qu'en fait, il s'agit d'un nouveau modèle qui va leur permettre de détecter de nouvelles opportunités. Pourtant, le nombre de communications sur des use cases Big Data est en constante augmentation, notamment dans le domaine de la Finance. Cela peut paraître paradoxal, sauf si l'on s'intéresse aux projets réellement mis en production. Bien souvent, les use cases Big Data dont on parle ne sont que des faire-valoir de l'innovation, plutôt qu'un réel vecteur d'innovation. En production, les sujets sont encore trop traités par des technologies classiques, alors que les technologies Big Data correspondraient mieux aux besoins, sans pour cela tout révolutionner. Par exemple, si la banque et l'industrie font déjà de la détection des fraudes, ce sont clairement les technologies autour du Big Data qui sont les mieux à même de résoudre ce type de problématiques. Enfin, il apparaît clair que les directeurs de l'innovation au sein des groupes éprouvent des difficultés à coordonner les initiatives Big Data pour leur donner corps ».



DÉCOUVREZ LA SUITE DANS...

LE LIVRE BLANC DES BIG DATA

bit.ly/livre-noir-blanc