

Systeme et coherence

En amont de la réflexion en termes de système, les sciences et les technologies de l'information, ont adopté les deux américanismes de *datawarehouse* et de *datamining*. La première partie de ce texte s'étend sur les spécificités de ces deux termes, ainsi que leur complémentarité. La seconde partie se consacrera alors à la genèse et à l'actualité du concept de cohérence dans la philosophie du système.

*

Datawarehouse : volet énonciatif de l'information, souvent traduit par « entrepôt de données », le concept de datawarehouse a été créé par Bill Inmon, un expert étasunien en bases de données¹. Devenu usuel dans le jargon quotidien des statisticiens francophones, le datawarehouse désigne un ensemble de fichiers statistiques et textuels,- de données brutes et de synthèse,- qui embrassent les différents aspects d'un domaine.

Bien qu'en règle générale, le datawarehouse concerne l'entreprise, l'informatisation des sociétés fait qu'on parle désormais de datawarehouse dans toute gestion rationnelle d'un ensemble de fichiers de renseignements.

Le datawarehouse, ou entrepôt de données, tend à remplacer ce qui était jadis désigné par « *banque de données* ». Une des raisons plausibles de ce nouveau vocabulaire est sans doute l'émergence de systèmes de gestion des bases de données plus puissants qu'autrefois. De manière concomitante, les coûts de stockage ont diminué en symbiose avec la baisse des prix des mémoires.

Ces facilités « matérielles » ont permis aux Institutions publiques de conserver des informations qui se répètent régulièrement (Comme, par exemple, en France, l'enquête sur le recensement général de la population.). Voilà qui a considérablement enrichi le contenu des « banques de données », donnant aussi les moyens d'un suivi à travers le temps. Ces transformations ont été accompagnées de la naissance du verbe « *historiser* » avec bien de formes conjuguées. Quant au substantif « *historisation* », il est entré de plein pied dans la définition de l'entrepôt de données².

Datamining : volet exécutif du traitement de l'information, le datamining se traduit usuellement par « fouille de données ». Tout aussi bien passé dans le jargon des statisticiens francophones que son confrère le datawarehouse, ce mot désigne l'ensemble des logiciels qui se destinent à analyser, traiter et synthétiser les éléments d'un entrepôt de données.

Ce travail incorpore le calcul numérique autour duquel la statistique mathématique avait développé de nombreux outils, dès l'émergence des langages de programmation ; dans les années 1960.

Le datamining, ou instruments de fouille des données, tend aujourd'hui à remplacer ce qui, jadis, était désigné par « bibliothèque de programmes ». Sans doute le besoin d'une nouvelle manière de parler de l'analyse a-t-il été stimulé par la naissance de langages de programmation plus brefs, plus

¹ Cf. http://www.technologytransfer.eu/speaker/125/bill_inmon.html

² Cf. Encyclopédie libre Wikipédia http://fr.wikipedia.org/wiki/Entrep%C3%B4t_de_donn%C3%A9es

puissants et surtout mieux intégrés dans un système qui autorise tout à la fois la gestion des fichiers et l'analyse de leur contenu.

Voilà pourquoi, en même temps qu'il a été question de datamining, l'inflation des mots a porté sur le « contrôle de qualité » et les « systèmes informatiques d'aide à la décision », ou plus brièvement de « Système d'informatique décisionnelle » ; voir « informatique de décision ». Le numéro 1 mondial de ce domaine est l'Institut étasunien SAS³.

Système : bien que les concepts de datamining et de datawarehouse soient distincts, aucun d'eux n'a de sens que par l'autre. Aussi, quand nous rédigeons un système d'aide à la décision, nous devons aussi bien réfléchir sur le recueil des données et le choix de leur contenu que sur leur organisation et leur analyse.

Dans l'entreprise, à la maison, ou dans les Institutions de formation, pour qu'un entrepôt de données remplisse ses fonctions d'une manière efficace, sa préparation, sa mise à jour et son suivi doivent suivre des règles protocolaires qui optimisent l'analyse du contenu par les logiciels de fouille.

Enfin, et quel que soit le vocabulaire utilisé, les deux piliers du traitement de l'information que sont l'entrepôt et la fouille de données demeurent l'énoncé et l'analyse : L'énoncé de la problématique analysée doit reposer sur des données exactes et complètes. En effet, aucun juge ne peut prononcer son verdict sur la base d'approximations et de oui-dire. A ma connaissance, c'est le Livre d'Esaïe (situé au 7^{ème} siècle AC) qui représente le plus ancien témoignage d'un idéal qui exclut les oui-dire (Esaïe 11 :3), tout au moins en est-il ainsi en Méditerranée.

Quant à l'analyse, elle doit s'appuyer sur de justes raisonnements. En effet, même si l'énoncé d'une question est clair, aucun observateur sérieux ne peut plonger dans ses formules en laissant au vestiaire le patrimoine de la logique. Ce principe trouve dans la classification des sciences (إحصاء العلوم) d'Al-Farabi (871- 951) la fleur des méditations. Le second maître de logique après Aristote y explique en effet qu'aucune spéculation mathématique ne saurait éjecter la logique. A fortiori pourrions-nous le faire aujourd'hui pour une informatique de décision qui s'appuie essentiellement sur les mathématiques !

*

J'en arrive à la deuxième partie de l'exposé où je voudrais traiter de la cohérence dans le concept de système.

Il est admis que la logique du système germe et évolue en langue grecque, avec les stoïciens, globalement entre 322 AC et 180 PC, c'est-à-dire au cours de ces cinq siècles qui vont de la vie de Zénon de Chypre à celle de l'empereur Marc Aurèle. Pierre Aubenque a expliqué que, par l'idée de système, les anciens stoïciens (orientaux en majorité) voulaient signifier que :

« (...) la philosophie est un tout, que l'on peut certes diviser en parties pour les besoins de l'enseignement, mais à condition d'apercevoir que chaque partie est solidaire des

³ Cf. <http://www.sas.com/>

autres et que l'abandon d'une seule partie ou d'une partie de partie entraînera la ruine de l'ensemble »⁴

Cette définition reste valable pour l'écriture d'un système informatisé. En effet, les différents modules (programmes ou ensemble de programmes) d'un système informatique peuvent être divisés en parties pour les besoins propres d'une tâche, mais la mise en œuvre d'un produit fini requiert des conventions protocolaires et/ou des programmes interfaces qui rendent les modules solidaires d'une même problématique.

Autrement dit, et comme dans les propos d'un discours, les modules d'un système doivent adhérer ensemble sans qu'aucun d'eux ne contredise un autre.

Ce principe incorpore l'idée de « cohérence ». Fondamentale dans l'écriture du système comme dans l'expression parlée, la cohérence a cependant fait l'objet de merveilleux écrits au cours de la période stoïcienne, et ce,- non par un stoïcien,- mais par l'excellent ennemi des stoïciens que fut Plutarque (46 – 125).

Etant donné qu'on ne peut pas construire un raisonnement cohérent à partir d'idées qui heurtent le bon sens, Plutarque a démontré l'incohérence de la fameuse aporie de Zénon (non pas le stoïcien Zénon de Chypre, mais Zénon d'Elée, cinquième siècle AC). D'après ce paradoxe, Achille n'atteindra jamais la tortue partie avant lui (Cf. encadré : « Aporie de Zénon d'Elée »).

Aporie de Zénon d'Elée

Selon Zénon, une tortue qui est allée courir avant Achille ne sera jamais rattrapée par ce dernier. En effet :

Le temps qu'Achille comble son retard, la tortue aura avancé d'une distance de ε_1
 Le temps qu'Achille avance de ε_1 , la tortue aura avancé de ε_2
 Le temps qu'Achille avance de ε_2 , la tortue aura avancé de ε_3
 Etc. etc.

Ces propos sont parfois cités comme ayant été comme l'un des piliers de la discipline actuellement dite « topologie », science des lieux et des objets qui s'y posent. C'est fort possible, mais il convient sans doute d'écouter la critique que Plutarque oppose à ce raisonnement. A cet effet, considérons la phrase Ph :

Ph = Si un être géométrique E jouit de la propriété P, les possibilités de l'objet O, qui s'y meut, ne sont pas réductibles à P.

Posons :

E = un segment de droite

⁴ Texte de Pierre Aubenque à la page 193 de « *La philosophie païenne* ». Par Pierre Aubenque, Jean Bernhardt, François Châtelet. Hachette Littérature. 1972.

P = divisibilité à l'infini

O = Achille (ou le rapide cheval d'Adraste, peu importe)

Nous obtenons Ph' :

Ph' = Si un segment de droite est divisible à l'infini, les possibilités d'Achille qui s'y meut ne sont pas réductibles à la divisibilité à l'infini.

Il en découle que les caractéristiques du mouvement d'Achille,- lequel mouvement se pose sur un segment de droite,- ne sont pas réductibles à la divisibilité à l'infini comme c'est le cas du segment de droite lui-même.

C'est donc seulement parce que Zénon a restreint son horizon d'observation (du mouvement d'Achille) à un segment de droite qu'il a confondu la divisibilité à l'infini avec l'incapacité d'Achille à dépasser la tortue.

L'erreur de Zénon est sans doute plus récurrente que ne le pensent certains modernes. Cette erreur est même, me semble-t-il, de nature à confondre certains topologistes. C'est pourquoi, il me paraît utile de commémorer, sans retouches, un texte majeur de l'histoire de la topologie ; comme de la logique.

***Au berceau de la topologie
Plutarque. Des notions communes⁵***

*Certes les Stoïciens accusent Épicure de créer de misérables fantômes et de violenter les notions communes en faisant mouvoir tous les corps d'une égale vitesse sans admettre que l'on aille plus vite que l'autre. Ce qui est encore beaucoup plus misérable et nous écarte davantage des notions communes, c'est de dire que rien n'est rejoint par rien, pas même si, comme le dit le proverbe, « le rapide cheval d'Adraste poursuit par derrière » une tortue. C'est une conséquence nécessaire si, comme ils le veulent, les mouvements ont lieu selon l'avant et l'après, et si les intervalles qu'ils parcourent sont divisibles à l'infini. Si, en effet la tortue dépasse le cheval seulement de cent pieds et si cet espace est divisible à l'infini tandis que le mouvement a lieu selon l'avant et l'après, jamais ils n'amèneront le mobile le plus rapide jusqu'au plus lent, puisque le plus lent est toujours en avant d'un intervalle qui se divise à l'infini. Que de l'eau qui s'écoule d'un vase ou d'une coupe ne s'épuisera jamais, n'est-ce pas contraire au sens commun ? Mais comment ne serait-ce **pas la conséquence de leurs thèses** ? Car on ne peut imaginer comment un mouvement qui a lieu suivant l'avant et l'après dans un espace divisible à l'infini pourrait être conçu comme s'achevant complètement ; il est toujours en arrière d'une division sur l'écoulement complet, sur le glissement et le déversement total du liquide, sur le mouvement total du solide, sur la chute d'un poids lâché qui n'est jamais achevée.*

Adib G. Hathout
Décembre 2010
Adoube2003@yahoo.fr

⁵ Page 177 de « *Les Stoïciens* ». Traduction par Emile Bréhier, revue par V. Goldschmidt. Bibliothèque de la Pléiade. Gallimard. 1962.