# Compte-rendu de la présentation d’IBM du 15/05/2017 au CRA-W (par Thomas Goossens)

## Résumé

IBM fournit une offre de *cloud-computing* appelée « *Bluemix*[[1]](#footnote-1) ». Ce type de plateforme vise à faciliter le développement d’applications exigeantes en terme de capacité de calcul, stockage et disponibilité du réseau en permettant de déporter la configuration, la maintenance et la mise-à-jour à la volée des serveurs et logiciels (base de données, outils de traitement numériques, sécurité, etc.) nécessaires au fonctionnemet de l’application, auprès du prestataire de service – IBM dans le cas de *Bluemix*. Ces « *PaaS* » (pour « Platform as a service »), permettent théoriquement de libérer le développeur de la contrainte de conception de l’architecture nécessaire à faire tourner ses applications afin qu’ils puissent se concentrer sur le cœur de son métier : la conception de logiciels. Dans le cas de *Bluemix*, la *PaaS* est basée sur une instance du logiciel open-source « *Cloud Foundry* »[[2]](#footnote-2).

Outre ces services d’infrastructure de déploiement, le produit *Bluemix* intègre un ensemble d’outils divers parmi lesquels une API d’intelligence artificielle (*Watson*) ainsi qu’une collection de base de données interrogeables en temps réel dont « *Weather Company Data* [[3]](#footnote-3)» qui fournit des données météorologiques historiques, d’observation et prévision à très haute résolution spatiale (500 m) et temporelle (mises-à-jour tous les quarts d’heure) en tout point du globe. La tarification[[4]](#footnote-4) quant à l’utilisation de cette base de données météorologiques est basée sur le nombre d’appels à l’API par unité de temps. Ces données sont fournies par « The Weather Company », une société rachetée en 2016 par IBM, et sont accessibles également en dehors de la plateforme de *cloud-computing* *Bluemix*. *The Weather Company* s’approvisionne elle-même en données auprès des instituts nationaux de météorologie et de particuliers disposant de stations individuelles connectées.

Les aspects techniques quant aux données météorologiques spatialisées (origines des données, traitements géo-statistiques, incertitude, grid utilisé, etc) sortaient du cadre de la présentation et le site web de *The Weather Company* n’en dévoile pas plus. Cependant, l’opportunité d’organiser une première réunion technique par visio-conférence avec les météorologistes anglais de *The Weather Company* impliqués dans la gestion de ces données nous a été proposée.

## Premières réactions

Tant la présentation qui nous a été donnée que le site web de *Bluemix* réussissent à vendre le haut potentiel du *cloud-computing* open-source[[5]](#footnote-5) d’IBM pour AGROMET et plus généralement pour les divers projets du CRA-W. Le site web de *The Weather Company* présente (bien évidemment) les avantages qu’il y a à travailler avec leurs services dans des secteurs d’activité variés dont l’agriculture[[6]](#footnote-6). Une liste des packages[[7]](#footnote-7) de données météorologiques disponibles est présentée. On y trouve également un exemple d’étude de cas dans le secteur de l’agriculture : *« The Poseidon Project :  Making the most of the world’s water resources through smarter irrigation [[8]](#footnote-8)*»

Ainsi présentées, ces solutions ont l’air prometteuses, mais il s’agit de rester critique et d’établir une analyse des risques quant à l’utilisation de tels outils dont la maitrise peut nous échapper, contrairement à l’utilisation de produits entièrement « faits maison ». Cette vigilance ne doit cependant pas nous empêcher d’investiguer quelles utilisations rationnelles de ceux-ci pourraient apporter une réelle plus-value au projet AGROMET. Pensons notamment :

* À la disponibilité très rapide et assurée des données météorologiques spatialisées (utilisable dans une première phase ? En solution de fallback ?)
* La possibilité de comparer leurs données de spatialisation avec les futures nôtres
* Accord de collaboration gagnat-gagnant : nous leur fournissons les data de PAMESEB et en échange nous accédons aux données spatialisées (?)

## A faire prochainement

Tout d’abord, il s’agira d’investiguer en profondeur quelles pourraient-être les avantages et inconvénients (techniques, éthiques, financiers, humains, etc) qu’il y aurait à travailler avec les solutions vendues par IBM.

Ensuite, il parait évident que comprendre ce que sont les données de The Weather Company est la condition préalable à leur potentielle utilisation. Il y aura donc lieu de préparer un ensemble de questions à leur soumettre concernant l’origine, le traitement, la qualité et la diffusion de leurs données météorologiques. En vrac, une première liste de questions à soumettre :

* Quelles-sont vos sources de données (satellites, stations et types de stations, data-providers, etc) pour :
  + Température de l’air
  + Température au sol
  + Humidité relative
  + Précipitations
* Quels-sont les algorithmes de spatialisation (kriging, co-kriging, Multiple-régressions, Artificial neural Networks) utilisés pour générer le dataset griddé à 500m ? Peut-on avoir accès au code ?
* Quelles-sont les éventuelles variables indépendantes (pentes, expositions, latitude, longitude, altitude, etc) utilisées pour la spatialisation ?
* Quelles-sont les techniques de validation utilisée pour la spatialisation et quelles-en sont les résultats ?
* Quelles-sont les incertitudes sur les divers jeux de données disponibles :
  + Data observation spatialisées
  + Data prévision spatialisées (aux différentes échelles de temps)
* Quelles-sont vos règles de « nettoyage » des données pour ce qui concerne la diffusion en temps réel ?
* Quelles-sont vos règles de « nettoyage » des données historiques ?
* Comment certifiez-vous que vos données ne sont pas intentionnellement modifiées ?
* Pourriez-vous nous fournir un détail commenté du calcul de prix mensuel pour la fourniture :
  + De données d’observations spatialisées sur un grid de 500 m couvrant toute la Région Wallonne avec mise-à-jour horaire
  + De données de prévisions à 48h spatialisées sur un grid de 500 m couvrant toute la Région Wallonne avec mise-à-jour horaire
  + De données de prévisions pour les 3,5, 7 et 10 prochaines jours spatialisées sur un grid de 500 m couvrant toute la Région Wallonne avec mise-à-jour horaire
  + Pour le jour J, des données historiques au jour j-1.
* Pourriez-vous nous communiquer le prix de la récupération des données historiques disponibles avant la date du jour ?
* Pourriez-vous nous indiquer en détail nos droits quant à la réutilisation et diffusion des données que l’on récupère chez vous ?
* Comment pourrions-nous intégrer notre réseau de station pro à vos algorithmes de traitement de données météorologiques et comment pourrions-nous négocier un éventuel accord de prix sur base de cette réciprocité ?

1. <https://www.ibm.com/cloud-computing/bluemix/fr> et <https://www.ibm.com/cloud-computing/bluemix/fr/what-is-bluemix> [↑](#footnote-ref-1)
2. <https://www.cloudfoundry.org/platform/> [↑](#footnote-ref-2)
3. <https://console.ng.bluemix.net/catalog/services/weather-company-data?env_id=ibm:yp:us-south> [↑](#footnote-ref-3)
4. <https://console.ng.bluemix.net/docs/services/Weather/weather_overview.html#pricing_models> [↑](#footnote-ref-4)
5. <https://www.ibm.com/cloud-computing/bluemix/fr/open-source> [↑](#footnote-ref-5)
6. <https://business.weather.com/industry-solutions/agriculture> [↑](#footnote-ref-6)
7. <https://business.weather.com/products/weather-data-packages> [↑](#footnote-ref-7)
8. <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?subtype=AB&infotype=PM&htmlfid=WWC12354USEN&attachment=WWC12354USEN.PDF> [↑](#footnote-ref-8)