
IFB-Biosphère, Infrastructure Nationale Multi-Cloud pour les Sciences de la Vie

Christophe Blanchet*¹, Matéo Boudet², Guillaume Brysbaert³, Micaël Calvas⁴, Stéphane Delmotte⁵, Hervé Gilquin⁶, Nadia Goue⁷, David Grimbichler⁸, Jean-François Guillaume⁹, Antoine Mahul⁷, Jérôme Pansanel¹⁰, Bruno Spataro¹¹, and Cyrille Toulet¹²

¹Institut Français de Bioinformatique – Commissariat à l'énergie atomique et aux énergies alternatives, Institut National de la Santé et de la Recherche Médicale, Centre National de la Recherche Scientifique, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement – France

²GenOuest – Université de Rennes, L'Institut National de Recherche en Informatique et en Automatique (INRIA), Centre National de la Recherche Scientifique - CNRS, IRISA/OBELI – GenOuest core facility Univ Rennes, Inria, CNRS, IRISA F-35000 Rennes, France, France

³UMR8576 (UGSF) – Centre National de la Recherche Scientifique - CNRS, Université Lille Nord (France) – France

⁴Centre Blaise Pascal – Ecole Normale Supérieure de Lyon – France

⁵Laboratoire de Biométrie et Biologie Evolutive (LBBE) – CNRS : UMR5558, Université Claude Bernard - Lyon I (UCBL), INRIA – 43 Bld du 11 Novembre 1918 69622 VILLEURBANNE CEDEX, France

⁶Unité de Mathématiques Pures et Appliquées (UMPA-ENSL) – CNRS : UMR5669, Ecole Normale Supérieure de Lyon – France

⁷Université de Clermont-Ferrand – Université de Clermont-Ferrand – France

⁸Mesocentre Clermont – Université de Clermont-Ferrand – France

⁹Nantes Université – Université de Nantes – France

¹⁰Institut Pluridisciplinaire Hubert Curien (IPHC) – CNRS : UMR7178, Université de Strasbourg – 23, rue du Loess - BP28 - 67037 STRASBOURG CEDEX 2, France

¹¹Laboratoire de Biométrie et Biologie Evolutive – CNRS : UMR5558, Université Claude Bernard - Lyon I (UCBL), INRIA – France

¹²Mésocentre de Lille – Université de Lille - Sciences et Technologies – France

Résumé

Le cloud de l'Institut Français de Bioinformatique - IFB-Biosphère (1) - fournit aux scientifiques de nombreux services numériques pour les applications en Biologie. Il leur permet de déployer à la demande des environnements numériques de recherche, sous la forme de machines virtuelles (VM) et conteneurs avec les outils bioinformatiques intégrés pour leurs analyses des données biologiques. L'infrastructure Biosphère propose des services cloud à différents niveaux : pour les utilisateurs scientifiques, pour les développeurs d'applications bioinformatiques, pour les administrateurs de cloud et pour les formateurs en sciences de la vie.

L'infrastructure Biosphère est une fédération nationale de 8 clouds localisés dans les plate-

*Intervenant

formes bioinformatiques de l'IFB et pour certains en collaboration avec des mésocentres ou datacenters nationaux. La configuration logicielle est basée sur les standards ouverts comme Openstack pour les machines virtuelles, Docker pour les conteneurs, CEPH ou ZFS couplés à Manila pour le stockage partagé entre les VMs. L'infrastructure regroupe plus de 9000 vCPU et différents modèles de GPU (2), et propose des gabarits de VMs variés, de quelques vCPU et quelques Go de mémoire, à des VM BigMem (jusqu'à 4To de RAM), et des VMs avec beaucoup de vCPU (jusqu'à 255 vCPU pour une VM).

Biosphère propose à ces utilisateurs, développeurs et administrateurs un portail web multi-cloud pour déployer les environnements scientifiques sur les clouds de l'IFB. Le portail Biosphère (3) fournit plusieurs services centraux : un catalogue RAINBio des environnements bioinformatiques reproductibles (4), un tableau de bord pour la gestion des VM et des données (publiques et projets), et un service d'authentification des utilisateurs basé sur la fédération d'identité Renater. Les environnements logiciels bioinformatiques sont publiés dans le catalogue RAINBio, regroupant des logiciels scientifiques communs, notamment R/Rstudio/Shiny, Jupyter, Nextflow et Snakemake, très utilisés par la communauté, et

d'autres outils plus spécifiques en collaboration avec les développeurs et experts du domaine. Les services aux développeurs permettent l'intégration des environnements scientifiques sous la forme de VM prédéfinies avec des recettes ou des images de VM et conteneur. Les recettes de déploiement suivent le principe 'Infrastructure as Code', pour intégrer des applications scientifiques dans des machines virtuelles et des conteneurs. Les développeurs et utilisateurs avancés peuvent ainsi réaliser la configuration des VMs et conteneurs avec des dépôts git (gitlab, github) ou des espaces drives en ligne (nextcloud, owncloud). Cela permet la délégation des droits et la gestion asynchrone des modifications des configurations par les développeurs sans intervention des administrateurs cloud. L'utilisation de dépôt git en ligne permet également de réaliser l'intégration continue des environnements sous la forme de conteneurs pré-définis et publiés dans un catalogue de conteneurs. L'utilisation de ces images de conteneurs permet alors un déploiement plus rapide dans les VMs Biosphère.

Les utilisateurs peuvent alors déployer de manière universelle les environnements bioinformatiques, disponibles dans le catalogue, sur les 8 clouds de la fédération. La connexion des utilisateurs au portail se fait avec les identifiants académiques usuels des organismes et universités (fédération d'identité Renater). Ils peuvent demander à rejoindre le groupe de leur unité d'affiliation, une formation ou un projet. Les accès aux VMs sont sécurisés avec la clé SSH de l'utilisateur pour un accès en ligne de commande, et avec des jetons uniques de session pour les environnement avec interface web, comme Jupyter, Rstudio et R-Shiny. Le cloud Biosphère fournit un soutien aux formations et projets scientifiques par une allocation de ressources CPU/RAM et stockage dédiés, et différents services spécifiques. Parmi ces services, les formateurs et responsables de projets peuvent gérer directement les membres du groupe, disposer d'un tableau de contrôle des ressources déployées, et de pages de monitoring pour leur groupe (quota, utilisateurs, apps). Ils pourront également créer leurs environnements dédiés (sur le modèle des services aux développeurs).

Le portail Biosphère propose également des services aux administrateurs des clouds avec un tableau de bord des ressources déployées sur leur cloud, des interfaces de monitoring des utilisateurs, des groupes et des VMs déployées. Ils peuvent également soutenir des formations et projets partenaires en leur allouant une partie des ressources de leur plateforme cloud.

Le cloud IFB-Biosphère est une infrastructure numérique nationale utilisée pour la recherche en Biologie, et pour les diverses formations scientifiques comme les écoles thématiques, les cursus universitaires, les ateliers techniques et les hackathons.

1. Cloud IFB-Biosphère : <https://www.france-bioinformatique.fr/cloud-afb/>
2. Tableau de bord des ressources cloud : <https://biosphere.france-bioinformatique.fr/cloud/systemstatus3>
3. Portail Biosphère : <https://biosphere.france-bioinformatique.fr>
4. Catalogue RAINBio : <https://biosphere.france-bioinformatique.fr/catalogue>