



HAL
open science

FG-Cloud : Cloud communautaire distribué à vocation scientifique

Mohamed Airaj, Catherine Biscarat, Cécile Cavet, Nicolas Clementin, Sébastien Geiger, Christine Gondrand, Vanessa Hamar, Michel Jouvin, Vincent Legoll, Sha Li, et al.

► To cite this version:

Mohamed Airaj, Catherine Biscarat, Cécile Cavet, Nicolas Clementin, Sébastien Geiger, et al.. FG-Cloud : Cloud communautaire distribué à vocation scientifique. JRES (Journées réseaux de l'enseignement et de la recherche) 2015, Renater, Dec 2015, Montpellier, France. hal-04805590

HAL Id: hal-04805590

<https://hal.science/hal-04805590v1>

Submitted on 26 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

FG-Cloud : Cloud communautaire distribué à vocation scientifique

Mohamed Airaj

Laboratoire de l'Accélérateur Linéaire
Centre Scientifique d'Orsay
Bâtiment 200 – BP 34
91898 Orsay Cedex

Catherine Biscarat

Laboratoire de Physique Subatomique et de Cosmologie
53, avenue des Martyrs
38026 Grenoble Cedex

Cécile Cavet

Université Paris Diderot-Paris 7
Laboratoire APC
Bâtiment Condorcet
Case 7020
75205 Paris Cedex 13

Nicolas Clementin

Laboratoire Univers et Particules de Montpellier
Place Eugène Bataillon – CC 72
34095 Montpellier Cedex 05

Sébastien Geiger

Institut Pluridisciplinaire Hubert Curien
23, rue du Loess – BP28
67037 Strasbourg Cedex 2

Christine Gondrand

Laboratoire de Physique Subatomique et de Cosmologie
53, avenue des Martyrs
38026 Grenoble Cedex

Vanessa Hamar

Centre de Calcul de l'IN2P3
21, avenue Pierre de Coubertin CS70202
69627 Villeurbanne Cedex

Michel Jouvin

Laboratoire de l'Accélérateur Linéaire
Centre Scientifique d'Orsay
Bâtiment 200 – BP 34
91898 Orsay Cedex

Vincent Legoll

Institut des Grilles et du Cloud
c/o Institut Pluridisciplinaire Hubert Curien
23, rue du Loess – BP28
67037 Strasbourg Cedex 2

Sha Li

Laboratoire de Physique Subatomique et de Cosmologie
53, avenue des Martyrs
38026 Grenoble Cedex

Charles Loomis

SixSq
rue du Bois-du-Lan, 8
1217, Meyrin
Genève, Suisse

Matthieu Marquillie

Service Calcul Scientifique Intensif
Université Lille 1 - Bâtiment M4
59655 Villeneuve d'Ascq Cedex

Gilles Mathieu

INSERM/DSI
13 rue Watt
75013 Paris

Jérôme Pansanel

Institut Pluridisciplinaire Hubert Curien
23, rue du Loess – BP28
67037 Strasbourg Cedex 2

Guillaume Philippon

Laboratoire de l'Accélérateur Linéaire
Centre Scientifique d'Orsay
Bâtiment 200 - BP 34
91898 ORSAY Cedex

Jean-Marc Pierson

Institut de Recherche en Informatique de Toulouse
118, route de Narbonne
31062 Toulouse Cedex 9

Mattieu Puel

Centre de Calcul de l'IN2P3
21, avenue Pierre de Coubertin CS70202
69627 Villeurbanne Cedex

Geneviève Romier

Institut des Grilles et du Cloud
c/o Centre de Calcul de l'IN2P3
21, avenue Pierre de Coubertin CS70202
69627 Villeurbanne Cedex

François Thiebolt

Institut de Recherche en Informatique de Toulouse
118, route de Narbonne
31062 Toulouse Cedex 9

Andrej Tsaregorodtsev

Centre de Physique des Particules de Marseille
163, avenue de Luminy – Case 902
13288 Marseille Cedex 09

Résumé

Le GIS France Grilles, initialement focalisé sur les grilles de calcul et de stockage, propose désormais une offre de services basée sur le « Cloud Computing » IaaS.

Il s'agit, au travers de l'intégration des dernières avancées technologiques, de proposer un service répondant aux évolutions des besoins de traitement des données scientifiques, notamment en termes d'élasticité et de flexibilité.

France Grilles a donc développé, toujours sur le principe de partage et de mutualisation d'expertise, une fédération de Clouds IaaS, rassemblant des ressources (calcul, stockage, réseau et humaines) fournies librement par les différents sites participants en leur imposant le minimum de contraintes.

En première partie, nous présenterons les caractéristiques de l'infrastructure : les sites et le réseau humain à la base de l'infrastructure, les technologies utilisées, les services en production, l'accueil de nouveaux sites participant, l'intégration à l'échelle européenne et la certification EGI.

La deuxième partie détaillera quelques exemples d'utilisation de l'infrastructure, et en particulier les leçons que nous avons pu tirer du Cloud Challenge, test grandeur nature de l'infrastructure par un utilisateur expert.

Nous finirons par un rappel sur les modalités d'accès au Cloud France Grilles et les différents services offerts aux utilisateurs : instanciation de clusters à la demande, accompagnement, documentation, développements, passerelle vers le Cloud fédéré EGI.

Mots-clefs

cloud, fédération, France Grilles, EGI, services, infrastructure, IaaS, partage d'expertise, calcul scientifique

1 Introduction

France Grilles vise à fédérer des moyens de calcul et de stockage distribués, situés sur les sites des différents organismes qui participent au GIS (groupement d'Intérêt Scientifique) ou qui sont fournis par les communautés d'utilisateurs. Les missions de France Grilles, initialement focalisées sur les grilles, s'étendent depuis 2012 au domaine du « *Cloud Computing* ». Apparu comme une évolution naturelle des infrastructures de calcul distribué, il permet, par une grande souplesse, l'extension de leur domaine d'application. Il s'agit d'une formidable opportunité car le « *Cloud Computing* » intègre les avancées technologiques nécessaires pour mieux répondre aux besoins variés des communautés de recherche, notamment en termes de flexibilité d'utilisation des ressources et d'extension ou de réduction à la demande.

La communauté France Grilles a donc développé, en complément de l'infrastructure de grilles, toujours sur le principe de partage et de mutualisation d'expertise, FG-Cloud, une infrastructure nationale de type IaaS, construite à partir d'une *fédération de Clouds IaaS*, rassemblant des ressources distribuées (calcul, stockage, réseau et humaines) fournies librement par les différents sites participants en leur imposant le minimum de contraintes.

Une première partie présentera la stratégie de France Grilles dans le domaine du *Cloud* (construction et opération d'un *Cloud* académique destiné aux applications scientifiques et à la recherche) et ses objectifs (mettre à disposition du monde académique une infrastructure basée sur les technologies *Cloud*, pour répondre aux besoins des communautés scientifiques, s'intégrer dans une infrastructure internationale de manière à servir les collaborations scientifiques dont le cadre dépasse celui de la France).

2 FG-Cloud

2.1 Stratégie

France Grilles vise à la construction et l'opération d'une infrastructure de *Cloud* académique destinée à répondre aux besoins des applications scientifiques. L'accès aux ressources doit être ouvert aux communautés d'utilisateurs historiques de France Grilles comme aux nouvelles communautés. Cette *fédération de Cloud* ne se positionne pas en concurrente d'autres acteurs du domaine, qu'ils soient publics ou privés. Elle doit pouvoir s'intégrer dans une infrastructure internationale de manière à servir les communautés dont le cadre dépasse celui de la France. France Grilles tient à ce titre le rôle de représentant pour la France au sein de ces infrastructures ou projets.

2.2 Objectifs

Les objectifs de FG-Cloud sont les suivants :

- développer une communauté d'experts des technologies *Cloud* ;
- valoriser leur expérience par le biais de publications et d'interventions lors de meetings et conférences ;
- mutualiser les expériences de chacun afin de faciliter l'adoption des technologies par de nouveaux sites et utilisateurs en diffusant la connaissance et les compétences au travers de formations et documentation publique ;
- créer une infrastructure pouvant être utilisée par la communauté scientifique pour effectuer des travaux de recherche ;
- être le partenaire de choix entre les communautés Françaises et internationales pour toutes les problématiques liées à l'utilisation de *Clouds* dans des contextes de recherche scientifique.

2.3 Organisation

Le projet est piloté par un groupe technique qui se réunit deux fois par mois. Le point est fait sur l'évolution des sites, le suivi des utilisateurs et les développements en cours. L'expertise en terme d'administration et d'opération de *Clouds* est partagée entre les membres du groupe FG-Cloud, afin de mutualiser au mieux savoir, expériences et bonnes pratiques. Les éventuels développements sont aussi mutualisés et chacun peut participer à la conception et à l'élaboration d'outils additionnels pour opérer ces *Clouds*. Les échanges de ce groupe se font sur une liste de discussion.

Les compétences de ce groupe sont proposées aux administrateurs souhaitant déployer une infrastructure et aux utilisateurs pour qu'ils puissent exploiter de la manière la plus efficace possible les ressources disponibles.

3 Infrastructure du Cloud fédéré

L'infrastructure nationale est de type IaaS. Il s'agit d'une fédération de ressources de calcul et de stockage distribuées sur un ensemble de sites français. Elle est agnostique quant aux technologies. Les ressources, peuvent être utilisées de manière homogène et transparente, l'infrastructure est extensible à de nouvelles plateformes.

3.1 Fournisseurs de ressources

Les sites de *Cloud* participant à la fédération France Grilles mettent à disposition de la communauté d'utilisateurs une partie de leur ressources sous la forme d'un *Cloud IaaS* sur lequel ces derniers peuvent instancier des machines virtuelles qui vont servir à effectuer les calculs scientifiques nécessaires pour les projets de recherche scientifique.

La construction de la fédération est progressive. La fédération a débuté en 2012 avec la mise en place de trois sites pilotes, le CC-IN2P3 à Villeurbanne, le LAL à Orsay et CloudMIP déployée à l'IRIT à Toulouse. Depuis, la fédération s'est étendue successivement à plusieurs autres sites, l'IPHC à Strasbourg, le LUPM à Montpellier, le LPSC à Grenoble et l'Université Lille 1. Ce travail est réalisé par le groupe d'administrateurs des sites participants appuyé par des experts du domaine. Il est actuellement coordonné par le directeur technique de France Grilles.

Différents niveaux de fonctionnement ont été définis : test, pré-production et production permettant l'entrée progressive des sites dans la fédération et permettant d'assurer finalement le meilleur niveau possible de qualité de production aux scientifiques utilisateurs. Enfin l'intégration dans le *Cloud fédéré EGI* est l'objectif de la plupart des sites. Le site de l'IPHC est le premier à avoir réalisé cette étape.

La fédération comprend actuellement 8 sites répartis en trois catégories, comme le montre la Figure 1 :

- Les sites en production sont :
 - Centre de Calcul de l'IN2P3 – Villeurbanne
 - Laboratoire de l'Accélérateur Linéaire – Orsay
 - Institut Pluridisciplinaire Hubert Curien – Strasbourg
 - Université de Lille 1 – Lille
 - Institut de Recherche en Informatique de Toulouse – Toulouse
 - Laboratoire Univers et Particules de Montpellier – Montpellier
- Les sites en préproduction sont :
 - Laboratoire de Physique Subatomique et Cosmologie – Grenoble
 - Laboratoire Leprince-Ringuet – Palaiseau
- Un site est en cours d'installation :
 - Laboratoire de Physique Corpusculaire – Clermont-Ferrand

Sans imposer de contraintes strictes aux sites participant, les technologies utilisées dans la fédération de *Cloud*, ont en commun des caractéristiques d'ouverture (code, formats, communautés d'utilisateurs), de versatilité en terme de couches de virtualisation possibles, de souplesse et d'adaptabilité au matériel disponible. Chaque site est autonome dans ses choix de technologies.

La **Figure 1** détaille les *middleware* de *Cloud* utilisés dans l'infrastructure fédérée.

Les déploiements et maintenances sont automatisés grâce aux outils **Quattor** ou **Puppet**. Ainsi, la configuration et les versions des logiciels utilisés sont entièrement maîtrisées, assurant une stabilité logicielle des plateformes.

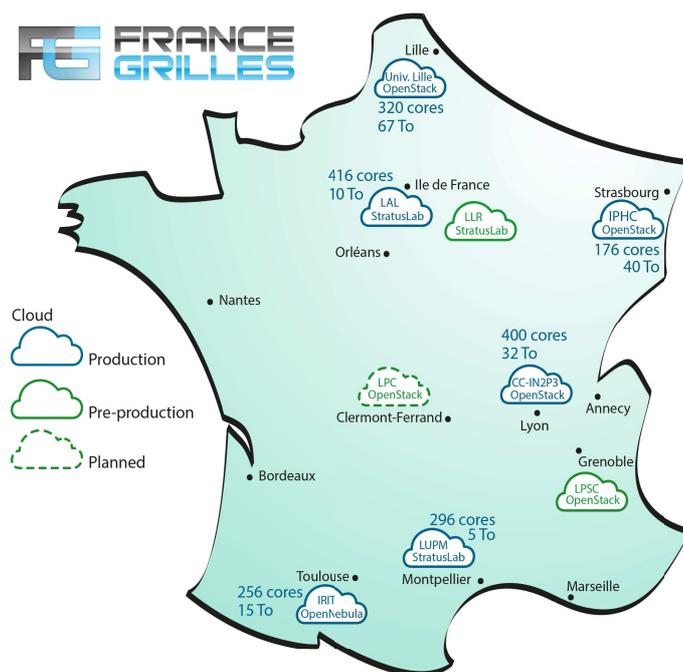


Figure 1 - Carte de France des sites

3.2 Ressources disponibles

Cette fédération d'infrastructures dispose globalement de :

- 1500 CPU ;
- 10 TB RAM ;
- 100 TB de stockage.

De nouvelles ressources sont régulièrement ajoutées pour faire face à la demande croissante des utilisateurs des différentes communautés scientifiques.

3.3 Intégration de nouveaux fournisseurs de ressources

Les sites intéressés pour rejoindre la fédération sont accueillis par le groupe d'experts et aidés tout au long de leur processus de mise en place du *Cloud*.

Des réunions régulières de suivi sont organisées afin de faciliter le transfert de connaissances et d'expertise vers les nouveaux administrateurs.

L'intégration se déroule en trois étapes :

- installation / test – dans un premier temps, un suivi de l'installation matérielle et logicielle est réalisée, des conseils et bonnes pratiques sont fournis ;
- préproduction – une fois l'installation réalisée, la nouvelle infrastructure de *Cloud* est ouverte aux premiers utilisateurs. Une documentation est rédigée pour que les utilisateurs puissent utiliser cette nouvelle ressource ;
- production : finalement, le site est intégré à la fédération et peut accueillir des utilisateurs de différents domaines. Les ressources sont étendues au fur et à mesure des besoins ressentis et des sources de financements disponibles.

3.4 Certification EGI

Après la mise en production au sein de *France Grilles* d'un site, le passage à l'échelle internationale est l'intégration du site dans la *fédération EGI*. Ceci représente pour le site la reconnaissance de la qualité de production par la communauté *EGI* et permettra l'accès et l'utilisation de ses ressources par des chercheurs du monde entier. Ceci est particulièrement important pour servir les communautés de recherche internationales.

Cette certification nécessite plusieurs étapes, tel que le support de l'authentification par certificat, le déploiement automatisé des images virtuelles ou encore la centralisation de l'*accounting*. Pour que le site franchisse facilement les différentes étapes, les experts du groupe FG-Cloud proposent leur aide à la mise en place des services et l'accompagnement technique tout au long de ce processus de certification.

3.5 Technologies employées

- *OpenStack* : la suite logicielle (open source) *middleware* de *Cloud IaaS* leader du domaine, que l'on retrouve dans bon nombre d'offres commerciales (*Rackspace*, *HP Helion*, *OVH*, etc.) ;
- *StratusLab* : une autre suite logicielle de gestion de *Cloud IaaS*, ainsi qu'un *marketplace* sur lequel trouver des images virtuelles toutes prêtes pour les *Clouds* ;
- *SlipStream* : un logiciel de déploiement d'applications sur clusters virtuels (*PaaS*) et d'ordonnancement de VM sur *Clouds* hybrides ;
- *KVM* : une solution de virtualisation pour architecture matérielle x86 (*Intel*, *AMD*) permettant d'avoir très peu de différence de performance par rapport à une exécution native ;
- *Libvirt* : une interface logicielle d'abstraction des différentes technologies de virtualisation ou de conteneurs disponibles ;
- *OpenVSwitch* : Un *switch* réseau virtuel multicouche programmable et extensible, supportant de nombreux protocoles et standards. Distribuable sur de multiples serveurs physiques pour supporter la montée en charge ;
- *Linux* : les distributions *Scientific Linux 6.x* et *CentOS 7* sont utilisées comme système d'exploitation stable sur lequel baser cette infrastructure. Elles ont été choisies pour leur degré de support et de stabilité.

4 Utilisation de l'infrastructure

4.1 Services disponibles

4.1.1 Instanciation de clusters à la demande

Les ressources disponibles (CPU, RAM, stockage, réseau) peuvent l'être sous forme de simples machines virtuelles indépendantes ou bien d'un réseau de machines virtuelles qui fonctionnent de concert et peuvent avoir chacune des caractéristiques différentes, par exemple une machine virtuelle (VM) pour héberger une base de données de résultats, une pour présenter les données sous la forme d'un site web accessible depuis internet, et quelques VM pour effectuer les calculs... Tout ceci peut être réalisé de manière dynamique en fonction des besoins ponctuels.

4.1.2 Accompagnement

Des formations sont organisées pour l'utilisation des piles *Cloud* (Stratuslab, Openstack). Une formation plus complète pour les utilisations avancées du *Cloud* sera organisée au printemps 2016. Enfin, un support personnalisé est proposé aux utilisateurs.

4.1.3 Documentation

Une documentation collaborative est disponible, et couvre les sujets connexes au *Cloud* et de son usage ainsi que des bonnes pratiques d'utilisation.

4.1.4 Développements

L'équipe France Grilles effectue des développements spécifiques pour améliorer la qualité du service rendu par le *Cloud* fédéré, par exemple la création d'outils de gestion des images virtuelles.

Afin de rendre l'accès aux ressources plus facile pour les utilisateurs, certains services additionnels ont été mis en place en plus des APIs natives et clients en ligne de commande pour instancier des machines virtuelles sur l'infrastructure.

- SlipStream – un logiciel d'instanciation et d'ordonnement de clusters complexes sur des *Clouds* hybrides
- VM-DIRAC – une solution pour présenter un environnement de calcul scientifique homogène quel que soit le *Cloud* utilisé
- FG-iRODS – un système de stockage objet virtualisé et distribué déjà disponible aux utilisateurs de façon indépendante est accessible également à travers FG-Cloud.

4.1.5 Accès au Cloud EGI

Si les besoins de(s) utilisateur(s) sont très importants, ils pourront être accompagnés pour le passage à l'échelle supérieure, et la mise en production de leurs calculs scientifiques sur la *fédération de Cloud EGI* qui dispose de ressources beaucoup plus importantes mises à disposition de la recherche et de l'innovation.

4.2 Accès au Cloud France Grilles

Les infrastructures France Grilles sont accessibles à l'ensemble des partenaires du GIS pour les travaux liés à leurs recherches. Elles sont également accessibles aux organismes ou entreprises dans le cadre de projets communs avec ces partenaires. Plus largement, les chercheurs ou ingénieurs travaillant en France peuvent y avoir accès pour répondre à leurs besoins de ressources de calcul, stockage. Ainsi, de nombreux utilisateurs ont déjà utilisé les ressources disponibles dans cette fédération.

Pour utiliser les ressources de France Grilles, il faut [prendre contact avec l'équipe](#). Il faut ensuite renseigner un questionnaire assez simple pour évaluer les besoins et évaluer l'adéquation entre ces besoins et les services qui peuvent être offerts.

Il n'y a pas de contrepartie financière à l'utilisation des ressources du *Cloud fédéré* : les infrastructures ont été financées depuis plusieurs années et sont encore financées par différentes sources (participation des organismes, subventions régionales, projets, Infrastructure de Recherche France Grilles, etc...). Les personnels sont financés par les organismes participants ou dans le cadre de projets.

Cependant, si les besoins de l'équipe utilisatrice ou de la communauté deviennent importants, une participation sous forme de ressources humaines ou d'achat d'équipement pourra être envisagée.

Par contre, pour être en mesure de justifier la pertinence scientifique de son infrastructure et de ses services, France Grilles demande en retour de l'utilisation, de participer à la vie de la communauté des utilisateurs, de présenter les travaux réalisés lors des rencontres scientifiques, et de référencer les publications réalisées grâce à France Grilles dans la collection de publications hébergées sur HAL.

4.3 Cloud Challenge

France Grilles a organisé un test grandeur nature d'utilisation massive de l'infrastructure par un utilisateur expert : le *Cloud Challenge*. Après sélection d'un projet parmi les différents candidats, le test a été effectué courant 2015.

Il s'agissait d'une simulation Monte Carlo quantique massivement parallèles pour la chimie. Ce projet implique une équipe du Laboratoire de Chimie et Physique Quantiques de l'université Paul Sabatier de Toulouse et une équipe de l'IRIT.

Les résultats de cette expérience ont été présentés lors des journées développement logiciel de l'enseignement supérieur et de la recherche, en juillet 2015, et des résultats plus aboutis lors des Journées de rencontres scientifiques des utilisateurs de calcul intensif, de *Cloud* et de stockage, en novembre 2015.

Ce test a permis de vérifier à grande échelle la validité de l'utilisation de FG-Cloud pour effectuer des calculs scientifiques et nous a permis de tirer des enseignements sur la pertinence de cette solution. Les différences de performance par rapport à des solutions sans couche de virtualisation, dues aux particularités des *Clouds* actuellement disponibles ne sont pas rédhibitoires et la souplesse d'utilisation ainsi que la disponibilité permettent d'en compenser les effets. Cette variabilité de l'infrastructure est à prendre en compte lors des études avant-projet afin de choisir les sites d'exécution avec pertinence. Les CPUs disponibles dans les machines virtuelles peuvent présenter des différences significatives quant aux codes de calcul utilisés par les projets scientifiques, et présenter des fonctionnalités suboptimales pour l'exécution de ceux-ci.

4.4 Exemples d'utilisation

Il y a encore assez peu de publications scientifiques de projets réalisés avec des traitements de données ou des calculs faits sur l'infrastructure de *Cloud* France Grilles. Cependant, on peut citer quelques exemples :

- Retour d'expérience en Astrophysique : utilisation du *Cloud IaaS* pour le traitement de données des missions spatiales [1]. Une étude de faisabilité et de validation de choix technologiques sur l'adéquation de l'utilisation d'un *Cloud* de calcul pour la recherche en astrophysique pour les missions spatiales *Euclid* (étude de l'énergie noire) et *eLISA* (recherche d'ondes gravitationnelles).
- Formation SGBD [2] au *CEFE* : Utilisation de machines virtuelles pour simuler un environnement de test et de recherche pour une formation aux logiciels de gestion de bases de données. Le *Cloud* permet d'instancier aisément des SGBD de différents types pour animer les travaux pratiques.

5 Conclusion

La phase de production étant atteinte pour un certain nombre de sites, l'accueil de nouveaux utilisateurs ainsi que l'extension de la fédération à de nouveaux sites et l'intégration progressive des sites actuels dans la fédération EGI sont les défis majeurs sur la feuille de route de France Grilles. L'accompagnement de certains sites en cours de migration fait aussi partie des tâches qui nous occupent. La constante nécessité de rester vigilant du point de vue sécurité est une de nos grandes préoccupations. Des travaux de recherche et développement sont effectués dans les domaines de la supervision mutualisée des sites de la fédération, ou de l'*accounting* qui nous permettront de mieux connaître la consommation dynamique des ressources.

Bibliographie

- [1] Cécile Cavet. Retour d'expérience en Astrophysique : utilisation du *Cloud IaaS* pour le traitement de données des missions spatiales. *École informatique IN2P3 2014 : Maîtriser le Cloud*, Jul 2014, Lyon, France. <hal-01132587>
- [2] Marie-Claude Quidoz. Initiation aux bases de données. UE PRO Terrain – Master SVT M2 http://rdbd.cnrs.fr/IMG/pdf/initiation_bdd.pdf?112/a90a83f0d943cd359b375b2758c3773ff75b0dac