ANF2016 Architecture des services nomades de la PLM

Sandrine Layrisse

13 octobre 2016

mathrice

Sommaire

Progression

mathrice

Définition des services nomades de la PLM

- PLM = Plateforme en ligne pour les mathématiques
- services numériques accessibles via internet
  - d'organisation et de production personnelle,
  - de communication,
  - et de travail collaboratif

  Elle fédère aussi des services d'accès à l'information scientifique, à des services de calcul, et des services à destination des laboratoires.

- offerts à une communauté constituée de la communauté mathématique :
  - de l'ESR français,
  - et aussi internationale (UMI, GDRI)

mathrice

Pourquoi une PLM aujourd'hui ?

- Raison **historique**, début 2004, issu du réseau de métier MATHRICE ont été mis en place :
  - le partage de jetons logiciels Maple, Mathematica et Matlab à l'échelle nationale
  - l'annuaire national de la communauté
  - et une maquette de services (basés sur un annuaire LDAP pour la gestion des comptes utilisateur) comprenant :
    - un proxy d'accès aux revues électroniques,
    - un serveur de messagerie,
    - une solution de VPN

mathrice

Pourquoi une PLM aujourd'hui ?

- Raisons **spécifiques** au chercheur :
  - mobile (conf, séminaires, ...)
  - en math, politique de recrutement qui induit une grande mobilité au cours de leur carrière
  - utilisateur des services d'internet (messageries, dropbox, doodle, skype, ... de FAI privés et autres grands "monstres" de l'internet !)
- Raison liée à l'**offre** :
  - à l'époque pas d'ENT, certains laboratoires proposent quelques services utilisables à distance mais pas tous (RH)
  - aujourd'hui, à part les offres des très puissants de l'internet,

mathrice

Le projet aujourd'hui

Intégration dans le projet **"portail math"** présenté comme une panoplie de services numériques à destination des mathématiciens

- Porteur du projet : INSMI (CNRS)
- Partenaires : Mathdoc , Mathrice (CNRS), RNBM (CNRS)
- Hébergeur : Mathrice (CNRS)
- Guichet d'accès simplifié :
  - à la documentation scientifique et aux services associés, en accès libre ou contrôlé, avec des fonctions de recherche évoluées ;
  - aux services facilitant le travail nomade collaboratif ;
  - aux informations institutionnelles et professionnelles.

mathrice

Pourquoi les chercheurs utilisent toujours la PLM

Environ 3000 chercheurs au sein de 90 structures de recherche utilisent régulièrement la PLM.

Atout : possibilité d'**inviter des collaborateurs**

- y compris des mathématiciens étrangers
- y compris des chercheurs d'autres communautés

mathrice

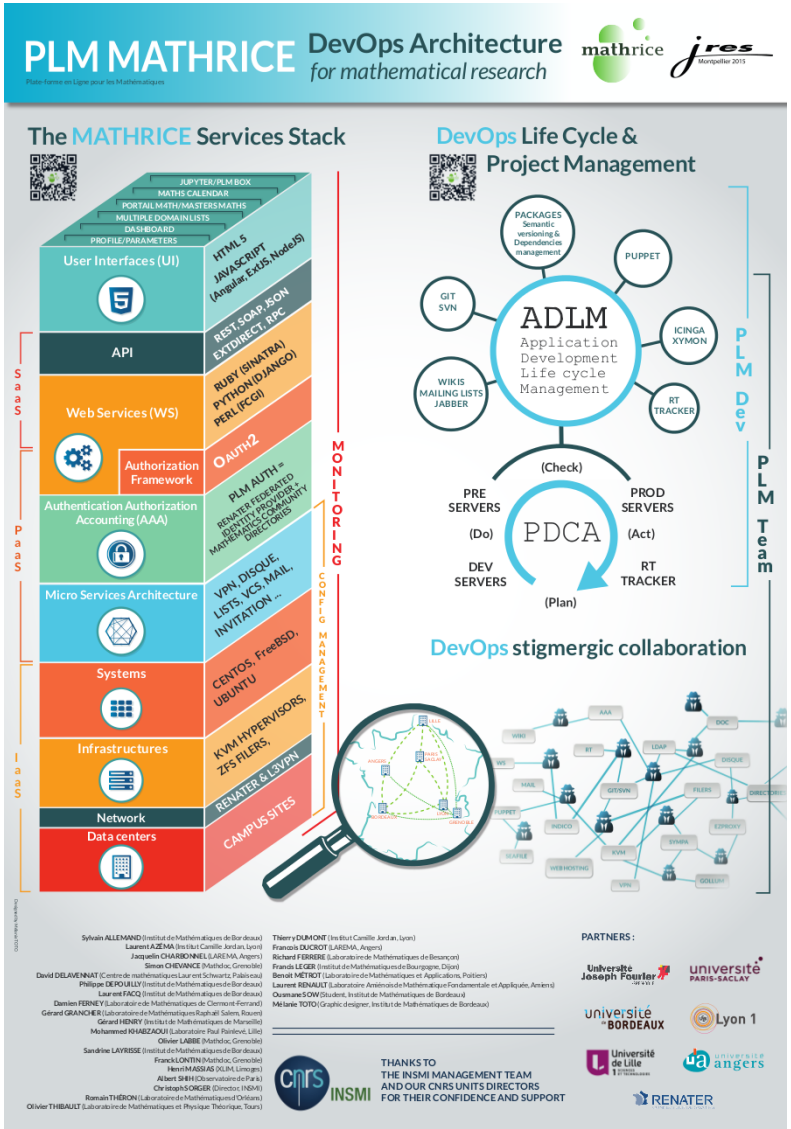
Représentation de l'architecture complète

- La pile des différentes ressources techniques et technologiques mises en œuvre

Progression

mathrice

La pile



Poster présenté aux JRES de 2015

IAAS

PAAS

Chaque service est **isolé** dans une machine virtuelle et décrit dans un outil de **gestion de configurations** auquel est associé un mécanisme d'historisation.

L'objectif de ce choix est de fournir des environnements de travail **reproductibles** afin de tendre vers une architecture de micro-services.

mathrice

Progression

mathrice

Architecture micro-services

Définition : une architecture microservices est un système de services **communicant entre eux**

Principe : isoler chaque partie métier de l'application pour la transformer en un service indépendant exposant sa propre interface (par exemple : communication à travers le protocole HTTP via une API REST, SOAP, JSON-RPC)

Intérêt des micro-services

Intérêt d'implémenter une architecture de micro-services :

- code de chaque service **limité** à ses besoins, gain en lisibilité et facilité de maintenance.
- **indépendance** du service : évolution sans dépendance, pas d'impact sur les autres services, liberté de choix du langage qui correspond le mieux à l'équipe dédiée à son développement ou au besoin précis du service (exemple : ruby pour vpn, javascript pour interface web, ...)
- de même pour les **bases de données**, choix qui correspond le mieux à l'usage (sqlite, MariadB, ...)

Limitations :

- implémentation de l'application, de l'API, du protocole

mathrice

Progression

mathrice

Exemple d'intégration d'application : PLMlatex, PLMbox, PLMwebconf, ...

PLMlatex = sharelatex intégré à la PLM

- implémentation de l'**authentification** par un stagiaire (OAuth2)
- utilisation du middleware **Passport** : permet de configurer plusieurs types d'authentification (ldap, local, Oauth)
- quelques **customisations** (modèle Mathrice, aide, mail d'invitation personnalisé, gestion des droits ...)

mathrice

Progression

mathrice

SAAS

L'architecture logicielle repose sur le développement de **web-services**, l'utilisation d'**API standards** et sur la production d'**interfaces utilisateur** intégrées et modulaires.

L'ensemble des applicatifs déployés sur la PLM sont instrumentés par des web-services (soit présents d'origine, soit développés spécifiquement), ce qui permet de les rendre tous interopérables. Les technologies de frameworks les plus adaptées sont utilisées : ruby Sinatra, python Django ou encore perl FCGI.

Les **protocoles d'échanges et de manipulation de données** utilisés au niveau des API sont aussi bien REST, SOAP, Ext.Direct ou JSON-RPC.

La **gestion de session** est assurée par Couchbase et l'**authentification** par le trio CAS+Shibboleth+OAuth2.

Enfin, la **présentation utilisateur** est développée en HTML5 et/ou javascript, notamment avec les frameworks MVC/MVVM Sencha, Angular ou NodeJS.

Le challenge

L'identification et l'authentification des utilisateurs = un challenge et le pilier de l'architecture des services nomades de la PLM !! Pourquoi ?

Mettre en place un système d'identification et d'authentification unique, fiable, et simple :

- ok, on est ASR, on sait faire : LDAP,Kerberos, ...
- et on peut s'appuyer sur la fédération d'identités RENATER

mathrice

Le challenge suite

Mais dans notre contexte, il faut ... :

- n'autoriser que les membres de la communauté mathématique (accords revues, limitation des budgets de fonctionnement) : comment les différencier parmi toutes les personnes d'un même établissement fournies par les fédérations d'identités ;
- limiter les créations de comptes inutiles (on en a tous déjà trop !!)
- reconnaître les utilisateurs de la PLM au travers de leurs multiples identités numériques (Etablissements : Université, INRIA, CNRS, compte Mathrice) ;
- tout en continuant à bénéficier d'une authentification unique (SSO) maintenant "exigée".

mathrice

Les clés de la solution

L'annuaire emath = annuaire des mathématicien(ne)s, membres des laboratoires de recherche (UMR, FR, UMI, EA ...), des départements d'enseignement supérieur, des sociétés savantes ...

- initié en mars 2001
- fonction : référencer tous les types de personnels (chercheurs, enseignants-chercheurs, doctorants, personnels administratifs et techniques ...).
- environ 10 000 personnes au sein d'une centaine d'entités
- objectif : pouvoir trouver les coordonnées (nom, prénom, téléphone, adresse électronique et URL de la page Web personnelle) de tout(e) mathématicien(ne) appartenant à une entité française

Donc pour n'autoriser que les membres de la communauté math : vérifier que l'identité qui se connecte au service est présente dans cet annuaire.

mathrice

Croisement des informations de connexion

Maintenant pour autoriser la personne reconnue membre de la communauté à se connecter aux services, il faut vérifier son authentification.

2 façons :

- connexion avec un **compte Mathrice** (mais pas toujours indispensable) :
  - login inconnu ou mot de passe erroné **ECHEC**
  - login/mot de passe vérifiés **OK**
- connexion via le **système d'authentification** fourni par RENATER (pour les services web) :
  - informations utilisateur retournés par l'établissement/authentification réussie **????**
  - + croisement avec les informations de l'annuaire emath **OK**

La convergence d'identités

Principe de croisement des informations "F-Id/Annuaire emath" : attributs **mail mailAlternateAddress**

- Pour les **services web**, pas besoin de gérer des comptes utilisateurs. On ne fait que s'appuyer sur la fédération d'identités RENATER et identifier une population par rapport à l'annuaire emath.
- Pour les **services CLI** (par opposition aux services web), besoin de comptes locaux.

Or, **association du compte Mathrice** à une personne de l'annuaire emath = **activation de la convergence**

- récupération appartenance à un laboratoire
- affectation automatique des comptes au groupe (branche de l'annuaire) du laboratoire

D'où l'importance de gérer cet annuaire correctement

mathrice

Rôle crucial du correspondant annuaire

- **interlocuteur** de Mathrice pour la mise en place et le suivi de la participation de son entité à cet annuaire
- **interlocuteur** des personnes référencées de son entité
- **responsable** de la mise à jour et de la mise à disposition des informations de sa structure (URL)
  - s'assurer que l'adresse courriel renvoyée par le fournisseur d'identité soit présente dans l'attribut mail ou mailAlternateAddress du fichier ldif qu'il met à disposition de l'annuaire

Convergence d'identité : l'intérêt de l'attribut mailAlternateAddress (qui peut être multivalué) est de spécifier les adresses courriel renvoyées par les fournisseurs d'identité susceptibles d'être utilisés par les membres de l'unité.

mathrice

L'annuaire techniquement

- géré par Mathrice : suivi quotidien effectué par 2 personnes, un "correspondant annuaire" désigné par chaque entitée référencée garant du contenu
- technologie d'annuaire LDAP
- construit à partir d'un moissonnage automatique auprès des entités référencées : un robot récolte chaque nuit, auprès de chaque entité, les informations la concernant.
- Format d'échange : le fichier à fournir est un fichier texte en codage UTF-8 au format LDIF, contenant les informations nominatives

mathrice

Authentification décentralisée

Authentification PLM différente du modèle d'un ENT

- services **indépendants** et accessibles directement sur leur URL (PLMbox, PLMwebconf, PLMlatex, ...)
- tableau de bord de pilotage des services également indépendants
- services **déployés** sur 4 sites différents

Solution : OAuth2, puis OpenID Connect

Principe : réutiliser les briques OAuth2 ou OpenID Connect qui existe sur le marché (souvent déjà intégrés au service) et les faire fonctionner sur l'authentification de la PLM "AuthPLM"

mathrice

Gestion des droits

Retour de la fédération d'identité d'un email enregistré dans emath

- appartenance à un labo
- applications autorisées avec droits labo (revues, cf présentation Damien)
- utilisation des applications où identifiant = email (plmbox, plmlatex, plmwebconf)

Si compte PLM associé :

- autorisation sur toutes les applications
- accès aux services avancés (mail, svn)
- activation de nouveaux services (disque, vpn)

A ajuster selon le niveau de **besoin du chercheur**

mathrice

Progression

mathrice

Gestion de l'accès nomade aux services de type CLI (svn, git, vpn, ...)

Gestion des données utilisateur, paramètres, états

- **Stockage des informations** utiles associées à chaque utilisateur
  - clés ssh
  - configuration VPN
  - favoris
- **Tableau de bord** de paramétrage des services
  - boutons ON/OFF des tuiles
  - gestion des droits sur les dépôts git, svn
  - sélection des favoris
- **Interface d'administration** déléguée aux correspondants Mathrice
  - gestion des comptes utilisateurs
  - activation de services pour l'unité : mise en place d'un vpn, délégation de déclaration DNS
- urbanisation de la gestion de tous ces services à travers un seul et même portail

Gérer ses services

Bouton ON/OFF des tuiles (exemple création/désactivation service disque, création boîte mail, sagemath)

Fonctionnement :

- 1 service = 1 groupe dans l'annuaire LDAP de Mathrice = 1 liste de login
- 1 nouveau login dans le groupe = création des ressources sur le micro-service distant
- cas communication synchrone avec le micro-service : git, svn
- cas communication asynchrone avec le micro-service : mail, disque, ?vpn?

A venir : vue des services favoris

mathrice

Progression

Conclusion

- architecture qui intègre les dernières technologies
- beaucoup d'investissement sur la résolution de l'authentification et la gestion des droits
- implémentation ambitieuse de l'offre de services Mathrice
- projet passionnant
- s'il n'y a pas la PLM (un seul point d'entrée), quelles sont les offres ? (des réponses demain)

mathrice

Merci de votre attention

Pour en savoir plus sur le projet : https://plm.wiki.math.cnrs.fr

Des questions ?