



SURVEILLANCE DU CLOUD PAR LE CLOUD

Lahoucine BENLAHMR

Référent UPMC : Gérard Nowak

7 Septembre 2012

Responsable Wallix : Fabien Boucher

Master Sciences et Technologies - mention Informatique -
Spécialité «Systèmes et Applications Répartis» (SAR)
Parcours Systèmes Répartis et Middleware (SRM)

WALLIX ?

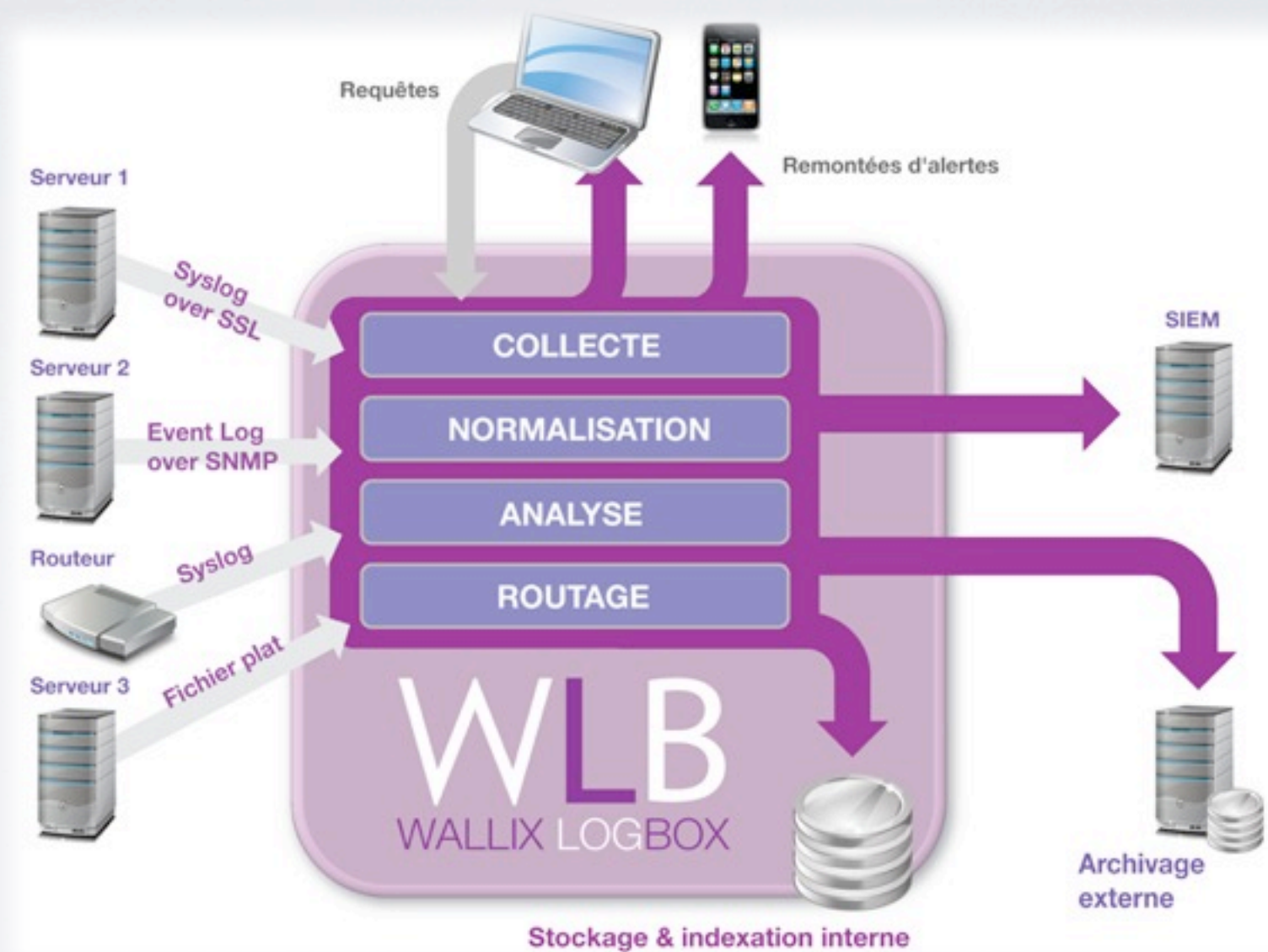
- Editeur de solutions de sécurité informatique
- Fondée en 2003
- Effectif: 35 salariés
- Distribution des solutions à travers un réseau de partenaires: France, au Benelux, en Suisse, au Royaume-Uni, en Afrique du Nord et au Moyen-Orient



WALLIX ?




LA LOGBOX



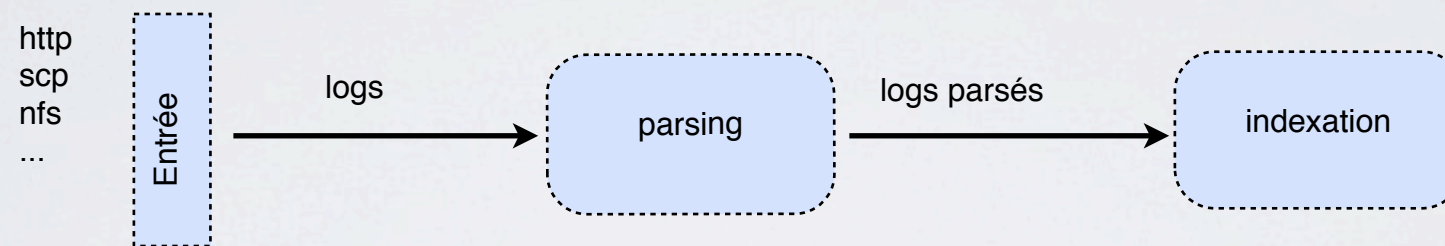
CONTEXTE DU PROJET RÉSILIENCE

- Risques du Cloud
- Projet initié par l'état Français
- Fournir un Cloud Français Open Source et résilient
- durée : 3 ans
- Acteurs:
 - **Nexedi**, Morpho, Astrium GEO Information, le CEA, Alcatel-Lucent Bell Labs France.
 - l'Institut Télécom, l'Université Paris 13, l'INRIA.
 - 4 PME du logiciel libre : Alixen, Alter Way, **Wallix** et Xwiki.

OBJECTIFS DU STAGE

- Réalisation d'un collecteur de logs pour le Cloud:
 - **Adapté** au système SlapOS
 - Entièrement distribué
 - Tolérant aux fautes
-  Nouvelle architecture pour La LogBox

OBJECTIFS DU STAGE



- Système de stockage distribué
- Système d'indexation distribué
- Traitements distribués des données collectées

SLAPOS

- Système distribué pour le Cloud
- Automatisation du déploiement d'application. (Buildout)
- Une vision «tout est processus»
- Un isolement par les droits utilisateurs
- Utilisation des adresses de type IPV6

ETUDE DES COMPOSANTS

SYSTÈME DE FICHIERS DISTRIBUÉ



Point de défaillance



Espace Noyau



Non compatible avec le système SlapOS

MONGODB

- SGBD NoSQL
- Orienté documents
- Popularité grandissante
- Utilisé par SourceForge et GitHub

```
{  
  // object id of the chunk in the _chunks collection  
  "_id" : <unspecified>,  
  // _id of the corresponding files collection entry  
  "files_id" : <unspecified>,  
  // chunks are numbered in order, starting with 0  
  "n" : chunk_number,  
  // the chunk's payload as a BSON binary type  
  "data" : data_binary,  
}
```



GRIDFS

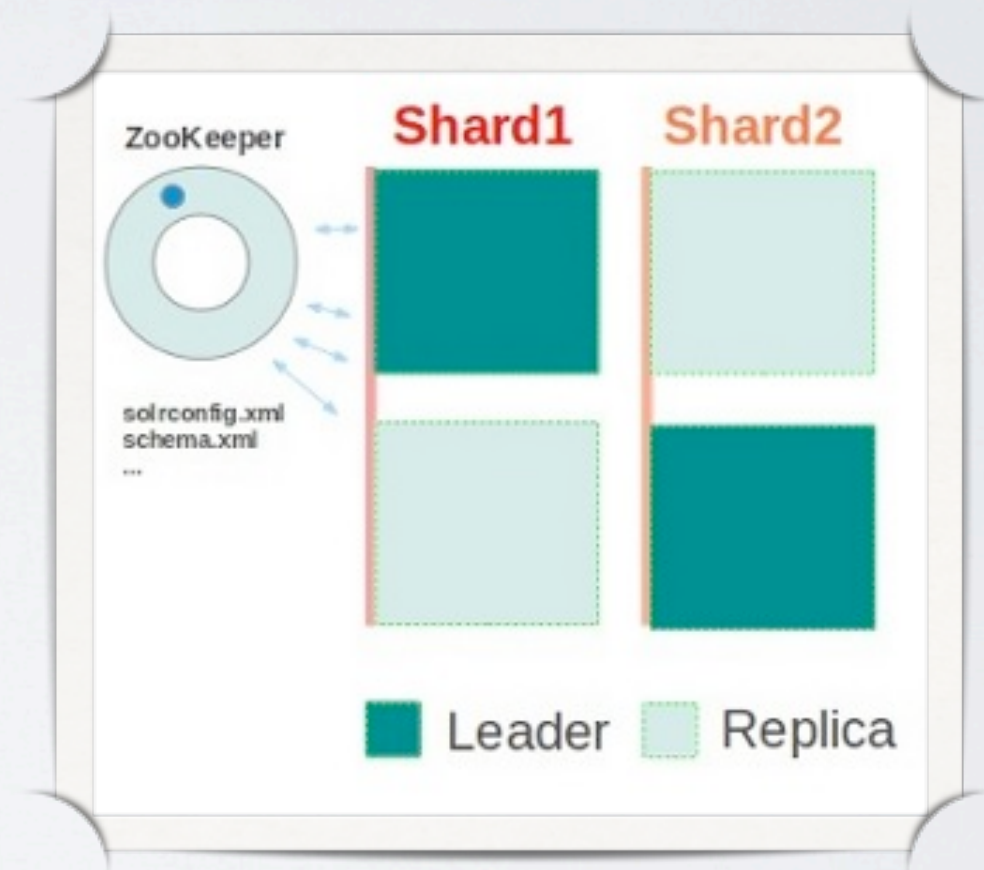
- Système de fichiers distribué
- Repose sur MongoDB
- Stocke les fichiers en morceaux
- Utilisation de clients dédiés pour manipuler les fichiers

SOLR

- Moteur de recherche
- Etend le principe de Lucen :
 - Facilite l'administration des index
 - Filtres de recherche
 - Manipulation des résultats

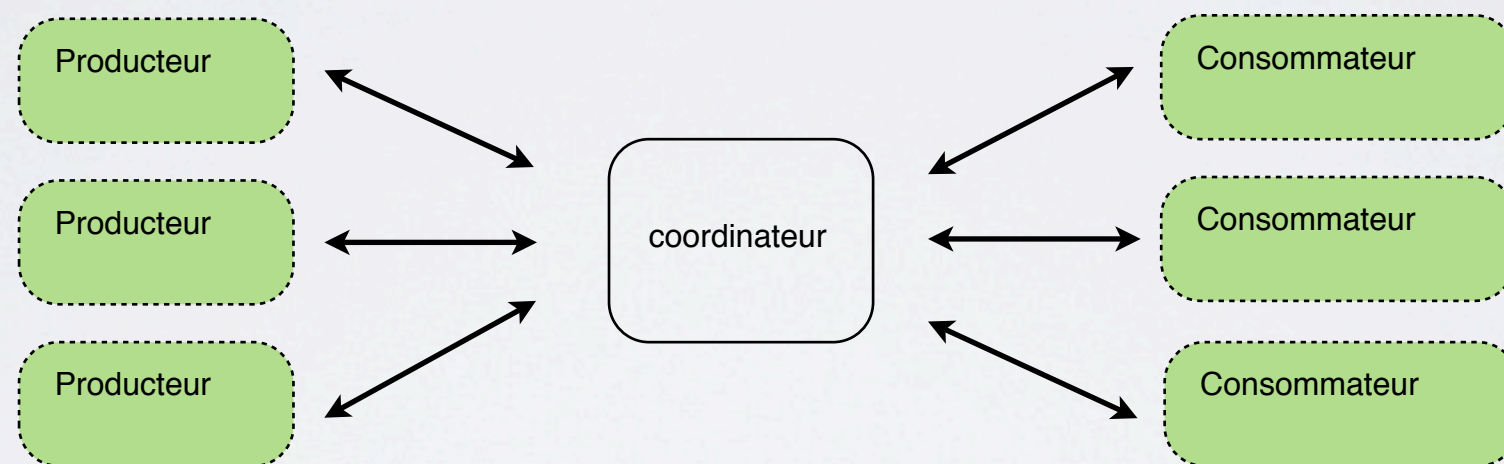
SOLRCLOUD

- Fonctionnalités de Solr
- Tolérance aux fautes
- Traitements distribués
- Indexation distribuée



RÉALISATION DE LA SOLUTION

TRAITEMENTS DISTRIBUÉS DE DONNÉES



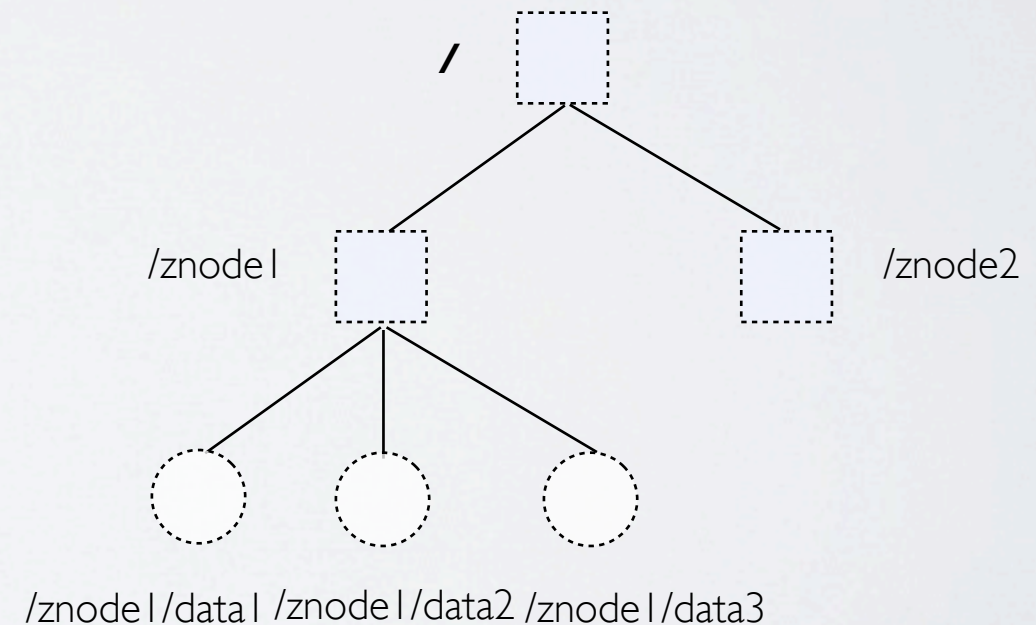
Modèle producteurs / consommateurs

PROGRAMMATION DISTRIBUÉE

- Réseau très souvent instable
- Besoins communs des applications réparties:
 - Techniques de localisation (Annuaire, découvertes, ...)
 - Horloges globales
 - Identifiants uniques
 - Espace de stockage (disque, mémoire, ...)

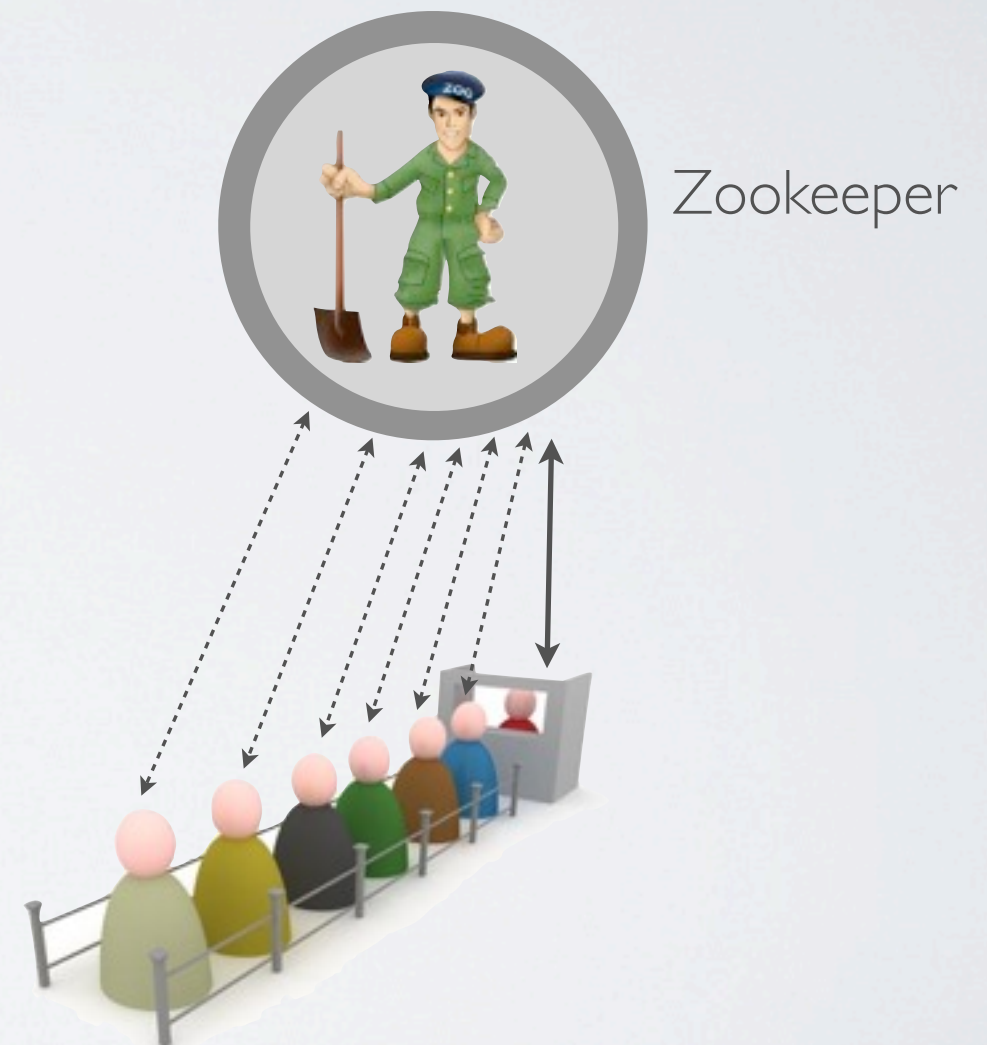
ZOOKEEPER

- 🔗 Service de coordination d'applications réparties
- 🔗 Permet d'implémenter des protocoles complexes:
 - 🔗 lock distribué
 - 🔗 élection de leader
 - 🔗 barrière de synchronisation
 - 🔗 ...
- 🔗 Passage à l'échelle
- 🔗 Service fiable

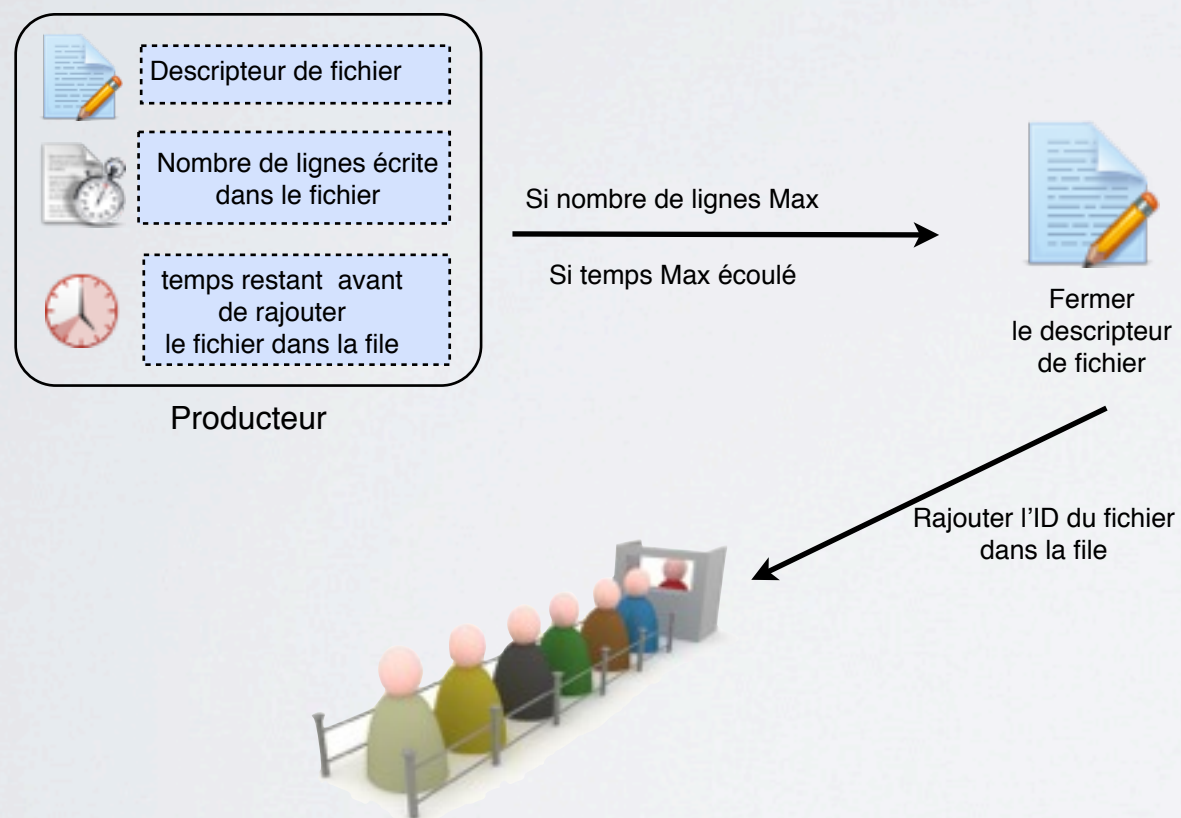


FILE DISTRIBUÉE FIABLE

- File coordonnée par Zookeeper
- Consommation fiable
- Traitement distribué

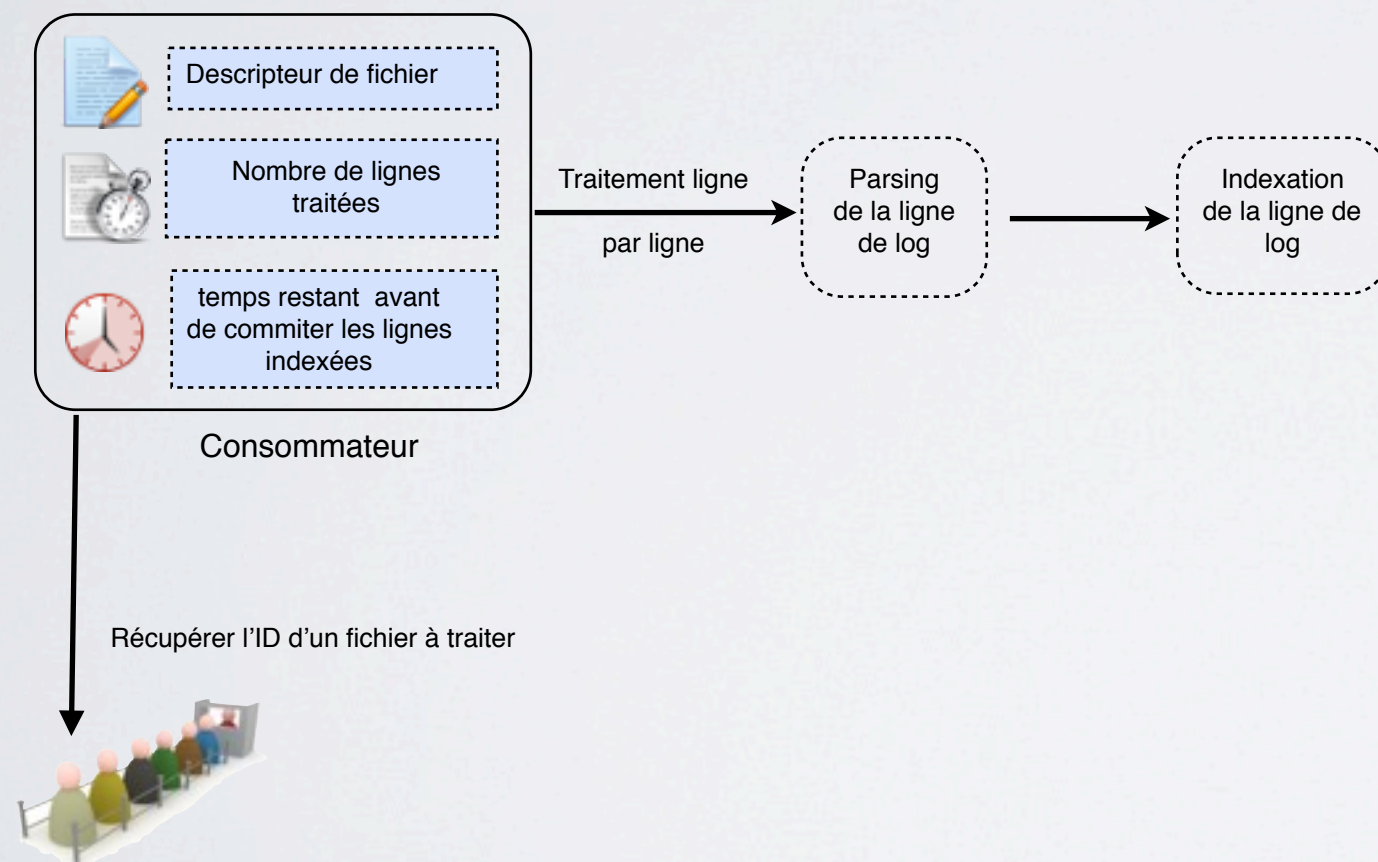


PRODUCTEUR



- Module d'entrée https
- Vérification de l'identité de la source
- Stockage des lignes de logs collectés
- Déclaration du fichier collecté dans la file
- Gestion de plusieurs sources de collecte

CONSOMMATEUR



Parsing de logs:

• PyLogsParser

Indexation de logs

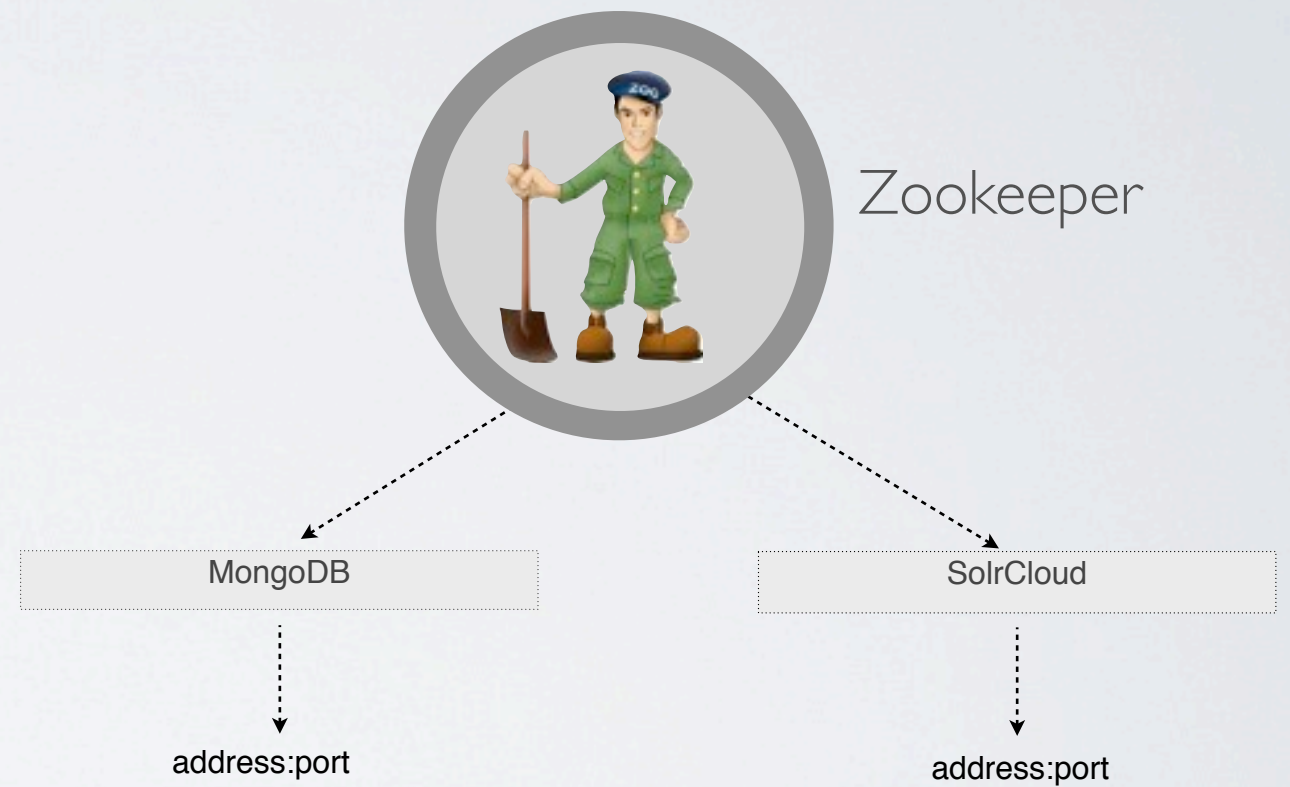
• mySolr

Consommation fiable:

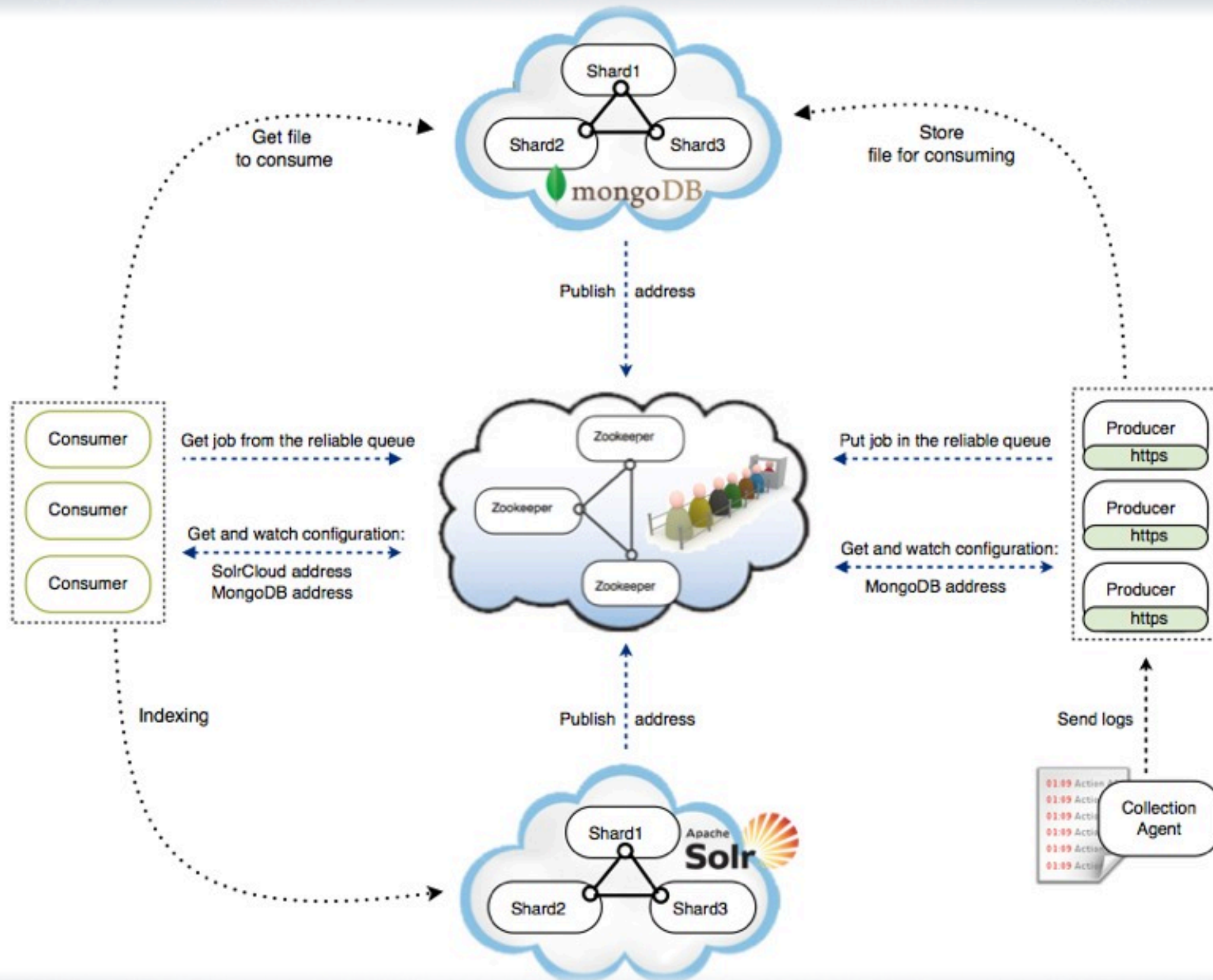
• S'assurer que chaque fichier est entièrement consommé

CONFIGURATION DYNAMIQUE

- Nature dynamique du Cloud
- Instabilité du réseau
- Besoin de re-configuration



ARCHITECTURE GLOBALE



CONCLUSION

- 🔗 Mise en pratique des connaissances acquises.
- 🔗 Participation à un projet OpenSource d'envergure, qui est le projet Résilience.
- 🔗 Monter en compétences dans les technologies du Cloud:
 - 🔗 Apache SolrCloud
 - 🔗 MongoDB
 - 🔗 Apache Zookeeper
 - 🔗 GridFs ...
- 🔗 Participation aux projets OpenSource:
 - 🔗 GlusterFS de Red Hat
 - 🔗 SlapOS
- 🔗 L'aventure continue !!!
 - 🔗 Embauche après le stage pour un contrat CDI (Contrat à Durée Indéterminée)

A conceptual image featuring a hand in a light blue shirt sleeve reaching up towards a large, stylized blue cloud with a white outline. The word "QUESTIONS?" is written in white capital letters across the center of the cloud. The background is a dark blue gradient.

QUESTIONS?

BIBLIOGRAPHIE

- K. Chodorow. Scaling MongoDB. O'Reilly Media, 2011.
- MongoDB, page d'accueil. <http://www.mongodb.org>
- Jeff Rush, Eggs and Buildout Deployment in Python - March 13, 2008.
- Thom White, hadoop the definitive guide. publié par O'REILLY
- Patrick Hunt and Mahadev Konar Yahoo! Grid, Flavio P. Junqueira and Benjamin Reed Yahoo! Research. ZooKeeper : Wait-free coordination for Internet-scale system.
- <http://www.slapos.org>
- <http://www.gluster.org/>
- <http://ceph.com/>
- <http://hadoop.apache.org/hdfs/>
- <http://lucene.apache.org/solr/>