



SYSTEME D'INFORMATION RESIF

Document de conception technique préliminaire

Transmission des données validées des nœuds A vers le nœud B

version 1.0

Historique des modifications

Date	Version	Description	Auteur
25/04/2012	0.1	Document initial	P.V0lcke
27/04/2012	0.2	Ajout descriptif traitement données validées	C.Pequegnat
27/04/2012	0.3	Prise en compte relecture C.Pequegnat	P.V0lcke
02/05/2012	0.4	Prise en compte relecture P.Barra	P.V0lcke
03/05/2012	0.5	Ajout des cas d'utilisation	P.V0lcke
24/05/2012	1.0	Prise en compte du document de revue	P.V0lcke

Sommaire

Historique des modifications.....	2
Sommaire.....	2
1 Introduction.....	3
1.1 Objectif.....	3
1.2 Références.....	3
2 Description générale.....	4
2.1 Protocole de transfert	4
2.2 Organisation du dispositif.....	4
2.3 Rapports d'intégration de données.....	6
2.4 Éléments de sécurité informatique.....	6
3 Cas d'utilisation.....	6
4 Pré-requis systèmes sur les nœuds A.....	9
5 Réponse aux exigences.....	10
5.1 Réponse aux exigences fonctionnelles.....	10
5.2 Réponse aux exigences non fonctionnelles.....	11

1 Introduction

1.1 Objectif

L'objectif de ce document est de décrire les principes techniques pour l'implémentation de la fonctionnalité de transmission des données validées entre "nœuds A" et "nœud B".

On entend par données validées l'ensemble des données produites et validées par les nœuds A à destination du nœud B (exemples non exhaustif : données sismologiques validées, métadonnées sismologiques validées).

Ce document décrit de façon préliminaire le dispositif. En seconde étape, des documents de spécification détaillées viendront compléter ce document initial.

Des documents spécifiques à chaque type de données transférées décriront les contraintes et mécanismes particuliers liées à ces données (exemples : taxonomie et formats des fichiers fournis par les nœuds A, description des mécanismes d'intégration au nœud B, description des rapports d'intégration fournis par le nœud B).

1.2 Références

Documentation de spécification logicielle - Transmission des données validées des nœuds A vers le nœud B

https://resif.dt.insu.cnrs.fr/fileviewer.php?file_id=207

SDS

<http://www.seiscomp3.org/wiki/doc/applications/slarchive>

SEED manual

http://www.iris.edu/manuals/SEED_chpt1.htm

Python programming language

<http://www.python.org/>

Rsync

<http://rsync.samba.org/>

2 Description générale

2.1 Protocole de transfert

Le dispositif de transfert s'appuie sur *rsync*, un protocole de transfert de données basé sur TCP. Le dispositif s'appuie en outre sur le logiciel éponyme *rsync*, fournissant une implémentation du protocole *rsync* (côtés client et serveur).

Le dispositif retient que les nœuds A transmettent leurs données en les "poussant" (*upload*) vers le nœud B. Ceci implique que les nœuds A soient *clients rsync* et le nœud B *serveur rsync*.

Il est noté que le principe de transfert décrit dans ce document prévoit les opérations d'ajout ou de remplacement de données par les nœuds A dans le nœud B. L'opération de suppression sera réalisée après contact auprès des opérateurs du nœud B, selon une procédure ne faisant pas partie de ce document.

2.2 Organisation du dispositif

Le dispositif se découpe en composants :

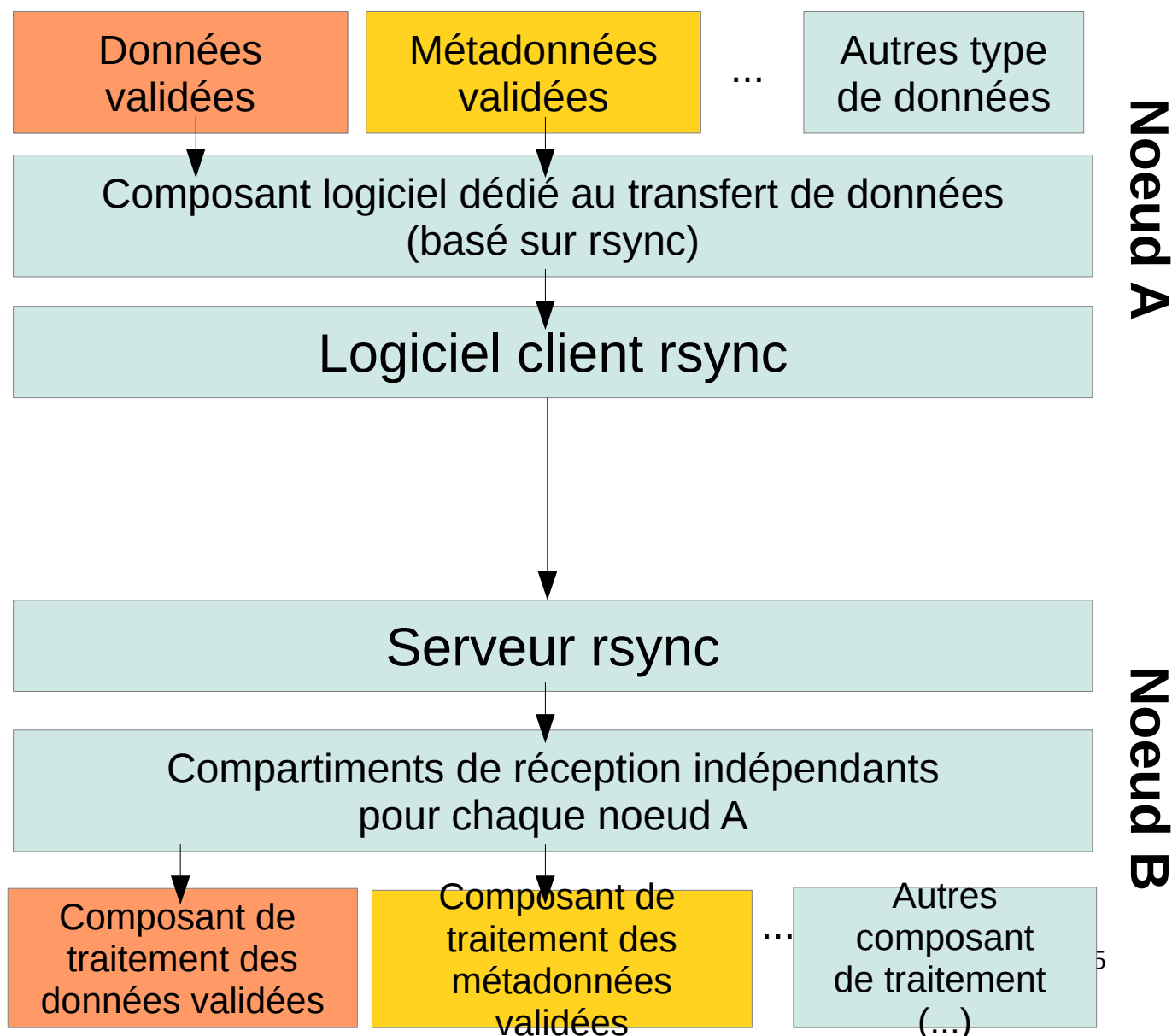
- Un **composant logiciel** déployé identiquement sur chaque nœud A :
 - ce composant est fourni et maintenu par le nœud B.
 - il fournit les fonctionnalités spécifiques au transfert des données RESIF selon les exigences fonctionnelles, puis repose sur *rsync* pour effectuer le transfert effectif.
 - il est paramétrable par les opérateurs des nœuds A.
 - il fonctionne en *ligne de commande*.
- Un ensemble de composants déployés au niveau du nœud B. Ces composants sont organisés en plusieurs niveaux, résumés ainsi :
 - un niveau *système*, qui assure aux nœuds A la disponibilité d'un serveur de réception et d'un espace réservé au transfert de données. Ce niveau système garantit notamment que ces espaces soient cloisonnés entre eux, afin d'éviter des débordements bloquants pour l'ensemble des nœuds A.
 - un second niveau *logiciel*, qui, pour chaque type de données transférées, va effectuer les opérations incombant au nœud B (exemple : vérification des données, intégration, génération des rapports destinés aux opérateurs des nœuds A, etc.). Pour chaque type de données, le fonctionnement détaillé du composant logiciel associé sera détaillé dans un document séparé.

Le schéma page suivante synthétise l'agencement du dispositif.

A *contrario* du système temps-réel basé sur un flux continu de données, le transfert de données validées fonctionnera par *lots*. Un lot correspond à un ensemble de données préparé *hors ligne* par un nœud A, puis transmis en une session unique (au sens du protocole *rsync*) au nœud B. Suite à une transmission réussie, le lot est clos (le nœud A ne peut plus y apporter de modifications, sauf à transmettre un nouveau lot complémentaire ou correctif), puis ajouté à la file de traitement du nœud B.

Enfin, le composant logiciel installé au nœud A est constitué de deux types de fichiers :

- un code exécutable, identique pour tous les nœuds A
- un ou plusieurs fichiers de configuration modifiables par les opérateurs des nœuds A. Ces fichiers de configuration sont dépendants des données transmises par un nœud A et déterminent notamment, pour chaque type de données, le répertoire source des données sur le nœud A et leur destination sur le nœud B. A ce titre, les paramètres de connexion (*login*, mot de passe, destination) sont fournis aux opérateurs du nœud A par les opérateurs du nœud B et renseignés dans ces fichiers de configuration.



2.3 Rapports d'intégration de données

Les outils d'intégration présents au nœud B (décrits dans des documents séparés) produisent un ensemble de rapports permettant aux opérateurs des nœuds A de suivre le cycle suivi par les données.

Dans un premier temps, ces rapports sont textuels.

Ces rapports sont rendus disponibles en téléchargement au niveau du nœud B pour les nœuds A. Les nœuds A utilisent alors le composant logiciel déjà utilisé pour le transfert de données pour rapatrier (*download*) les rapports les concernant.

2.4 Éléments de sécurité informatique

Les composants logiciels dédiés au transfert de données déployés au nœuds A et au nœud B vérifieront plusieurs principes fondamentaux de sécurité informatique. En particulier, seules certaines machines identifiées au sein des nœuds A seront autorisées à transférer des données au nœud B (voir section 4). De plus, les fichiers sensibles (ex. fichiers de configuration contenant des mots de passe) devront respecter certains critères de protection d'accès.

Le nœud B, afin de garantir le bon fonctionnement de ses systèmes et de se prémunir d'incidents causés par un afflux inhabituel de données, pourra imposer certaines limites de volume sur les lots transmis, ainsi que sur la fréquence des transferts. Ces limites pourront être contournées dans certains cas particuliers (ex : transfert initial d'un grand jeu de données), avec accord ponctuel des opérateurs du nœud B.

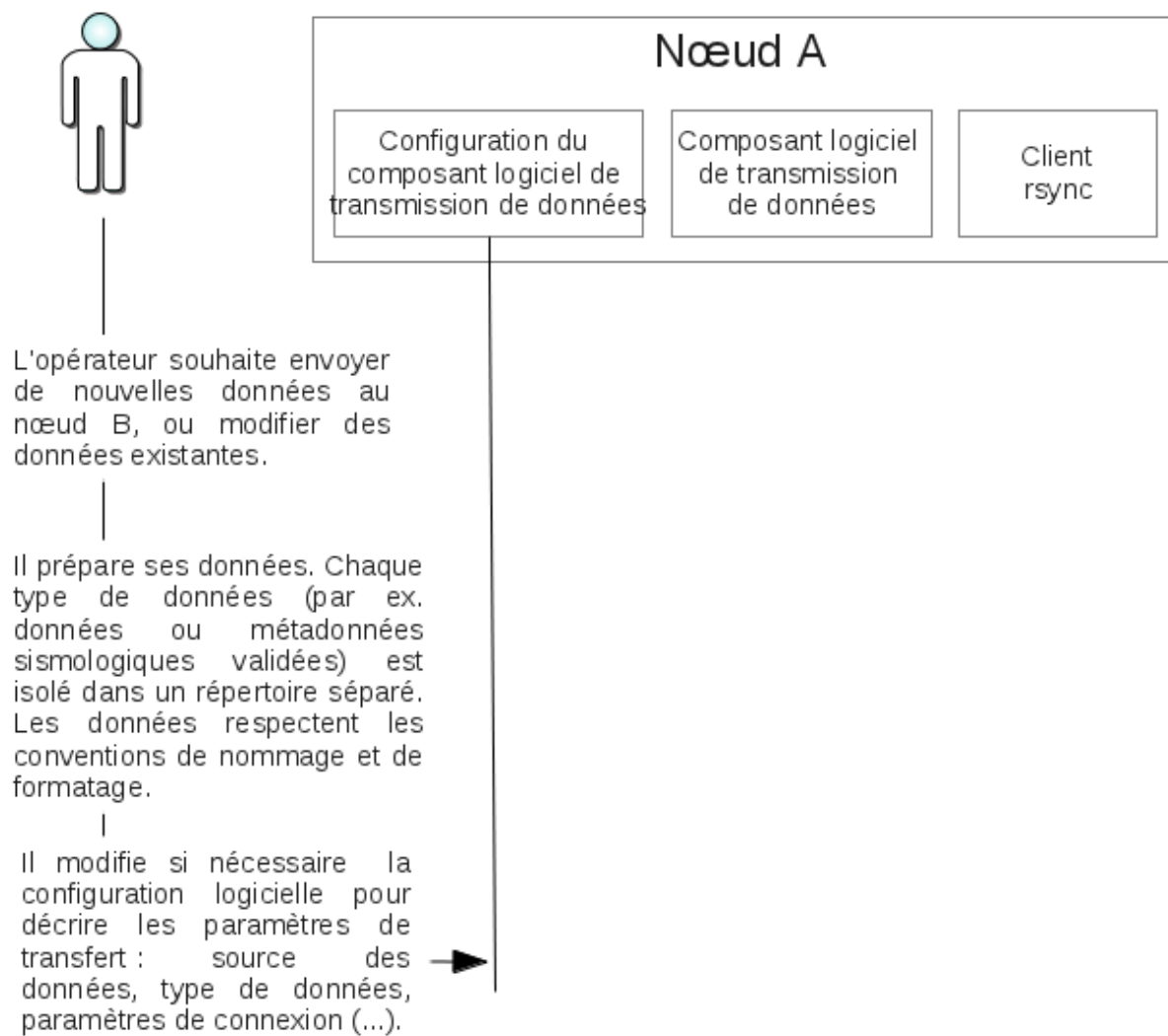
3 Cas d'utilisation

Les deux cas d'utilisation suivants illustrent les opérations de transferts de données du point de vue de l'opérateur d'un nœud A :

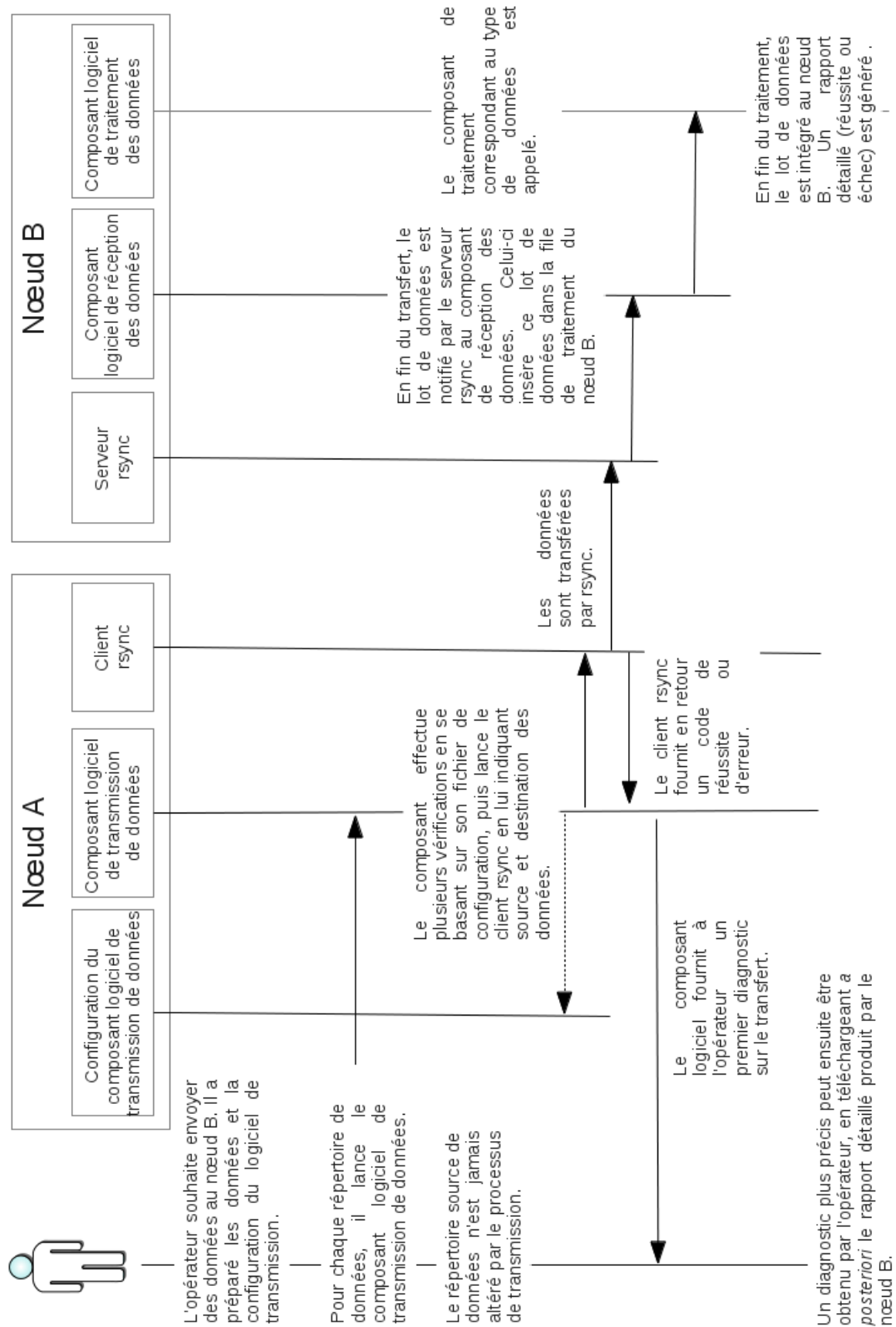
- configuration du composant logiciel de transmission de données
- lancement du composant logiciel de transmission de données

Sur le nœud A, les données à transmettre doivent être disposées dans un répertoire spécifique (distincts les uns des autres s'il y a plusieurs type de données à transmettre).

L'opérateur configure ensuite le composant logiciel de transmission de données, afin notamment de le faire pointer vers le bon répertoire. L'opérateur lance ensuite le composant. Les opérations d'ajout de nouvelles données et de remplacement de données existantes se réalisent de la même façon.



Configuration du logiciel de transmission de données



Lancement du logiciel de transmission de données

4 Pré-requis systèmes sur les nœuds A

L'implémentation du composant logiciel déployé sur chaque nœud A implique les pré-requis suivants :

- Pour chaque type de données transférées, chaque nœud A désigne au nœud B le nom DNS (*FQDN*) d'une machine unique depuis laquelle les dites données seront transmises (ainsi il est possible d'enregistrer une machine pour le transfert des données validées et une autre pour les métadonnées).
- Chaque machine ainsi désignée par les nœud A doit avoir accès au port *tcp/873* du serveur *rsync* RESIF-B (le nœud B informe les nœuds A du *FQDN* du serveur *rsync* afin que les éventuelles autorisations réseaux soient mises en place)
- Le composant logiciel déployé au nœud A sera validé pour un ensemble nécessairement limité de système d'exploitation (OS). Les opérateurs des nœuds A devront s'assurer que l'OS de la machine dédiée au transfert de données soit compatible avec les familles de systèmes suivants, dans leurs versions les plus à jour :
 - *Redhat Enterprise Linux* & assimilés, 32 ou 64 bits
 - *Debian Linux* & assimilés, 32 ou 64 bits
- Chaque machine désignée par un nœud A devra disposer des outils suivants :
 - un client *rsync* version 3.0.8 et supérieur, tel que fourni sur *rsync.samba.org*
 - un interpréteur *Python* 2.6.6 et au delà, hors branche 3.x, avec sa *standard library*

Afin de faciliter le déploiement et de réduire la complexité du composant logiciel déployé sur les nœuds A, la plupart des opérations de vérifications et de traitements des données sont implémentées au niveau du nœud B.

Ces pré-requis sont sujets à évolution pendant la durée de vie du composant logiciel. Ces évolutions seront proposées dans le cadre du COPIL.

5 Réponse aux exigences

Cette section reprend les exigences telles que décrites dans la section 3 du document "*Transmission des données validées des nœuds A vers le nœud B*". Pour chaque exigence, le type de réponse apportée est précisée.

5.1 Réponse aux exigences fonctionnelles

Exigence n°1 – Le composant logiciel travaille sur tout type de fichiers. La plupart des traitements spécifiques aux types de données étant appliqués au niveau du nœud B, le composant logiciel au nœud A est faiblement dépendant au type de données transmises.

Exigence n°2 – Le composant logiciel du nœud A travaille en prenant comme source de données des fichiers organisées dans un système de fichiers quelconque.

Exigence n°3 – voir rubrique 2.4

Exigence n°4 – Le composant logiciel est identique sur tous les nœuds A. Les étapes de mise à jour seront coordonnées par le nœud B avec les nœuds A. Un calendrier de mise en obsolescence des précédentes versions du composant logiciel sera proposé au COPIL.

Exigence n°5 – L'implémentation envisagée prévoit que le nœud B mette à disposition un serveur de réception et que les nœuds A "poussent" les données de leur propre initiative. Le composant logiciel pourra être utilisé soit en déclenchement manuel soit en tâche automatisée, au choix des opérateurs.

Exigence n°6 – L'implémentation envisagée prévoit que - une fois le composant logiciel lancé sur les données désignées - toutes les étapes nécessaires soient réalisées (dont le transfert effectif des données). Le composant logiciel peut-être lancé soit manuellement par un opérateur, soit automatiquement par programmation horaire au niveau des systèmes du nœud A.

Exigence n°7 – L'implémentation utilise le protocole *rsync*, largement répandu pour le transfert et synchronisation de données d'un point à un autre.

Exigence n°8 – La complétude et l'intégrité pourra être vérifiée au premier ordre grâce aux réponses du protocole *rsync*. Au second ordre, elles seront vérifiées dans les rapports générés *a posteriori* par le nœud B (même chose pour le processus d'intégration). Il sera proposé un mécanisme spécifique pour vérifier l'intégrité des fichiers transmis et intégrés, probablement basé sur des sommes de contrôle (*checksum*).

Exigence n°9 – voir section "Procédure d'intégration des données validées"

Exigence n°10 – voir section "Production de rapports pour les nœuds A"

Exigence n°11 – N/A

Exigence n°12 – voir section "Procédure d'intégration des données validées"

Exigence n°13 – voir section "Procédure d'intégration des données validées"

Exigence n°14 - voir section "Procédure d'intégration des données validées"

Exigence n°15 – N/A

Exigence n°16 – N/A

Exigence n°17 – le composant logiciel déployé au nœud A pourra être configuré pour envoyer différents type de données. Les traitements spécifiques s'effectuent au nœud B.

Exigence n°18 - voir section "Procédure d'intégration des données validées"

Exigence n°19 - voir section "Procédure d'intégration des données validées"

Exigence n°20 - voir section "Procédure d'intégration des données validées"

Exigence n°21 - voir section "Procédure d'intégration des données validées"

Exigence n°22 - voir section "Procédure d'intégration des données validées"

Exigence n°23 - voir section "Procédure d'intégration des données validées"

Exigence n°24 - le protocole de transfert n'est incompatible avec aucun système de sécurité classique (*firewall*, tunnel *SSL*, *VPN*, ...), tant que les systèmes sous-jacents sont configurés en accord.

5.2 Réponse aux exigences non fonctionnelles

Exigence n°1 – *rsync* est largement répandu et utilisé dans les milieux IT et enseignement/recherche. Il est stable et peu sujet aux failles de sécurité. Il est relativement peu coûteux à mettre en œuvre et maintenir.

Exigence n°2 – l'implémentation du composant logiciel nœud A s'appuie sur *Python*, langage portable et ouvert. Il sera proposée une documentation d'exploitation du composant logiciel aux opérateurs des nœuds A.