Compréhension du XML d’un job « PJB »

Obtenu par extraction des jobs d’une plateforme Datastage par l’outil istool.

# Obtenir de la matière ISX

L’outil se nomme istool. Il est disponible tant dans le monde Windows que dans le monde Unix. C’est d’ailleurs le seul outil pour ce genre de travail dont nous disposons sous Unix. C’est également le seul outil avec lequel on peut ne sauvegarder qu’une branche de l’arborescence des jobs du serveur, voire d’un ensemble de fichiers (à éprouver)

## Méthode

[eta2458@slzudisten05 cli]$ ./istool.sh export -archive /tmp/istool\_test.isx -u ETA2458 -p Evelyne@6969 -ds 'SLZUDISTEN05.yres.ytech/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/\*.\*'

## Résultats :

Beginning Export of DataStage

[1/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_ACQ\_Acquisition\_Xml\_E26\_2.pjb

[2/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_BSN\_UNVRS\_DE\_BSN\_025\_1.pjb

[3/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_FAM\_OFR\_025\_1.pjb

[4/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_FAM\_OFR\_FAM\_OFR\_025\_1.pjb

[5/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_FAM\_PROD\_025\_1.pjb

[6/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_FAM\_PROD\_FAM\_PROD\_025\_1.pjb

[7/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_GRP\_CHOIX\_025\_1.pjb

[8/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_LIEN\_ACTE\_OFR\_BSN\_024\_1.pjb

[9/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_LIEN\_OFR\_PROD\_025\_1.pjb

[10/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_MISE\_EN\_MRCH\_024\_1.pjb

[11/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_OFR\_FAM\_OFR\_025\_1.pjb

[12/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_OFR\_MRCH\_025\_1.pjb

[13/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_OFR\_SEGM\_MRCH\_CIAL\_025\_1.pjb

[14/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_OPTN\_CHOIX\_SUR\_OPTN\_025\_1.pjb

[15/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_OPTN\_GRP\_CHOIX\_025\_1.pjb

[16/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_OPTN\_OFR\_025\_1.pjb

[17/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_OPTN\_PROD\_CIAL\_025\_1.pjb

[18/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_PROD\_CIAL\_FAM\_PROD\_025\_1.pjb

[19/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_PROD\_CIAL\_GRP\_CHOIX\_025\_1.pjb

[20/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_PROD\_CIAL\_MRCH\_025\_1.pjb

[21/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_PROD\_CIAL\_PROD\_CIAL\_025\_1.pjb

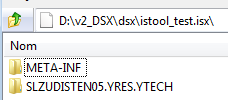
[22/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_PROD\_PRODTR\_025\_1.pjb

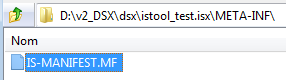
[23/23] SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/CDO/Jx\_S\_DTA\_WRK\_Enrichissement\_PROD\_PRODTR\_PROD\_PRODTR\_025\_1.pjb

Elapsed time: 00:00:26

Exported 23 assets

## Aperçu du contenu du fichier





Ce manifeste contient, semble-t-il, la liste des jobs et leur « emplacement ».

<?xml version="1.0" encoding="UTF-8"?>

<com.ibm.iis.istools.ai.core:Manifest xmi:version="2.0" xmlns:xmi="http://www.omg.org/XMI" xmlns:com.ibm.iis.istools.ai.core="http://com.ibm.iis.istools.ai.core/manifest" creationDate="2017-08-25T09:51:40.728+0200" createdBy="ETA2458" chunksCount="1">

<entries name="Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_COMN\_026\_1" path="SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ADRESSE/Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_COMN\_026\_1.pjb" createdBy="ETA2458" createdOn="2017-08-25T09:51:40.729+0200" description="" product="DataStage">

<id key="uri" value="istp://DataStageParallelJob@SLZUDISTMS04.yres.ytech:9080/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ADRESSE/Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_COMN\_026\_1?server=SLZUDISTEN05.YRES.YTECH"/>

<additionalInfo key="path" value="SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ADRESSE/Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_COMN\_026\_1.pjb"/>

<additionalInfo key="version" value="850"/>

<additionalInfo key="includedesigns" value="true"/>

<additionalInfo key="includeexecutable" value="false"/>

<additionalInfo key="modificationtimestamp" value="1484064552132"/>

</entries>

<entries name="Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_E27\_2" path="SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ADRESSE/Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_E27\_2.pjb" createdBy="ETA2458" createdOn="2017-08-25T09:51:40.897+0200" description="" product="DataStage">

<id key="uri" value="istp://DataStageParallelJob@SLZUDISTMS04.yres.ytech:9080/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ADRESSE/Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_E27\_2?server=SLZUDISTEN05.YRES.YTECH"/>

<additionalInfo key="path" value="SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ADRESSE/Jx\_S\_DTA\_WRK\_Enrichissement\_ADR\_E27\_2.pjb"/>

<additionalInfo key="version" value="850"/>

<additionalInfo key="includedesigns" value="true"/>

<additionalInfo key="includeexecutable" value="false"/>

<additionalInfo key="modificationtimestamp" value="1484064564485"/>

</entries>

…/…

</entries>

<entries name="Jx\_S\_DTA\_WRK\_Enrichissement\_PST\_FNCT\_RESPONS\_EDS\_025\_1\_save" path="SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ZZZ-Perso/VEG/Jx\_S\_DTA\_WRK\_Enrichissement\_PST\_FNCT\_RESPONS\_EDS\_025\_1\_save.pjb" createdBy="ETA2458" createdOn="2017-08-25T09:52:55.609+0200" description="" product="DataStage">

<id key="uri" value="istp://DataStageParallelJob@SLZUDISTMS04.yres.ytech:9080/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ZZZ-Perso/VEG/Jx\_S\_DTA\_WRK\_Enrichissement\_PST\_FNCT\_RESPONS\_EDS\_025\_1\_save?server=SLZUDISTEN05.YRES.YTECH"/>

<additionalInfo key="path" value="SLZUDISTEN05.YRES.YTECH/ZUDA0\_DTWH\_01/DTWH\_01/0010-QAG/02-ToDelivery/ZZZ-Perso/VEG/Jx\_S\_DTA\_WRK\_Enrichissement\_PST\_FNCT\_RESPONS\_EDS\_025\_1\_save.pjb"/>

<additionalInfo key="version" value="850"/>

<additionalInfo key="includedesigns" value="true"/>

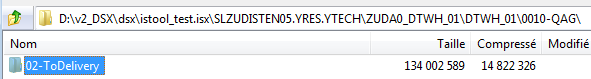
<additionalInfo key="includeexecutable" value="false"/>

<additionalInfo key="modificationtimestamp" value="1501158991079"/>

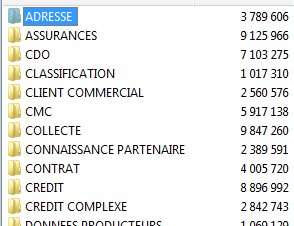
</entries>

</com.ibm.iis.istools.ai.core:Manifest>

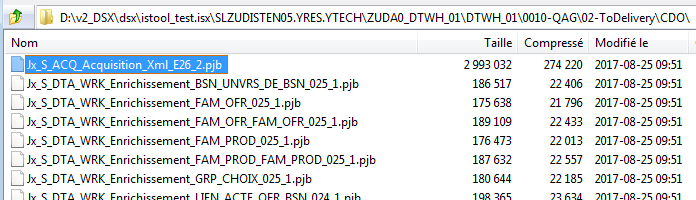
La « branche » 02-ToDelivery a été sauvegardée. Le fichier archive contient donc une suite de répertoires vers ce nœud de la branche.



Puis là, les répertoires contenus dans ce nœud.



Voici un extrait du répertoire CDO



Les fichiers .pjb sont des fichiers XML, un par job

Il peut exister d’autres extensions de nom de fichier, .pjb est pour les jobs parallèles.

## Un fichier .pjb : Le XML Datastage ISX

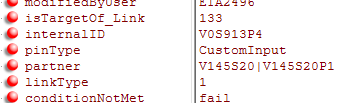
TODO : Décrire ce que l’on cherche

Nous sommes certains que les jobs tels que fabriqués jusqu’aujourd’hui ne contiennent qu’UN SEUL stage de type « Transformer »

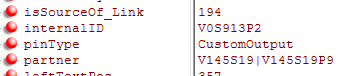
### En amont, le stage CP\_READ





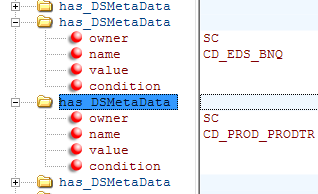




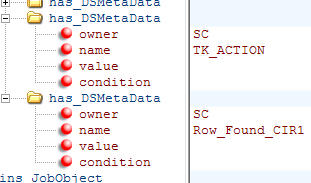




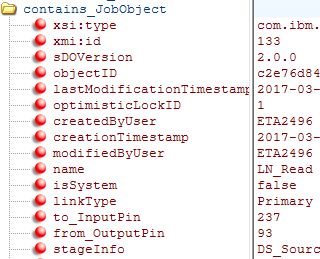
Il existe d’autres has\_DSMetaData, avec un autre owner

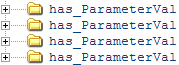


…/…

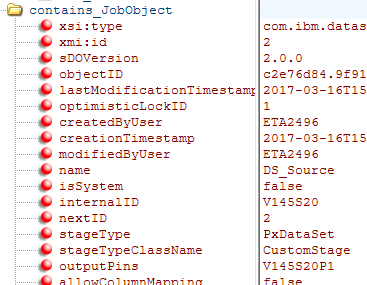


### En amont, le lien LN\_Read



Contient 

### En amont, le stage DS\_SOURCE



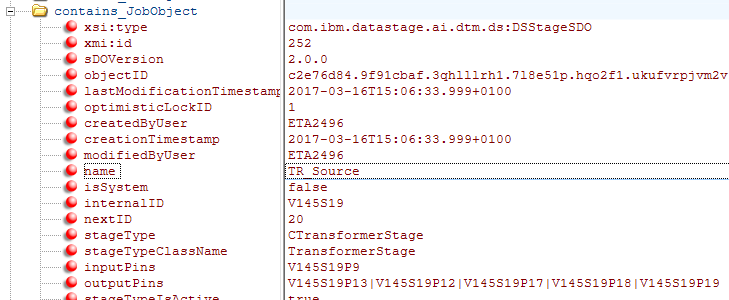




Pas de has\_InputPin ! on est donc à l’entrée du job.

### Le « Transformer »

Il y est décrit de la façon suivante :

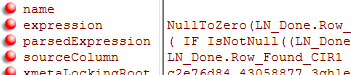










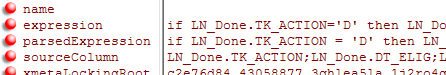






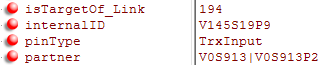




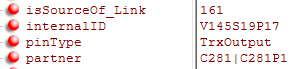


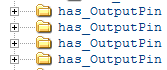
…/…





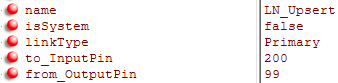


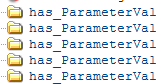


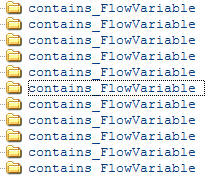


### En aval, le lien LN\_Upsert



















### En aval, le stage CP\_Upsert





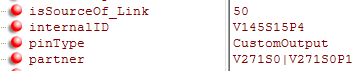












### En aval, le lien LN\_Write





### En aval, le stage DB\_Target









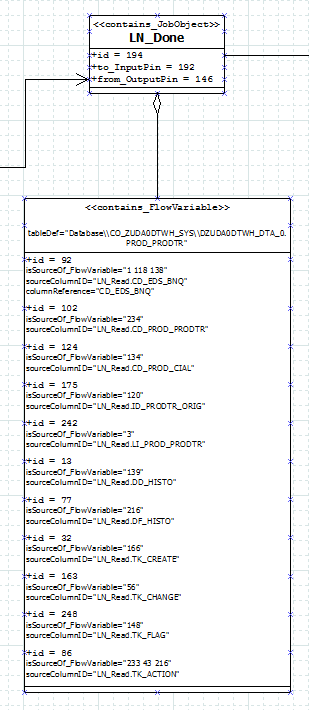


Pas de has\_OutputPin ! on est donc à la fin du job.

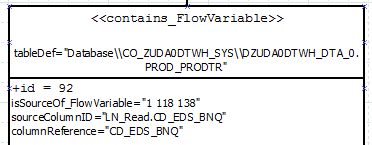
### Synthèse de cet examen

Les informations susceptibles de nous intéresser sont :

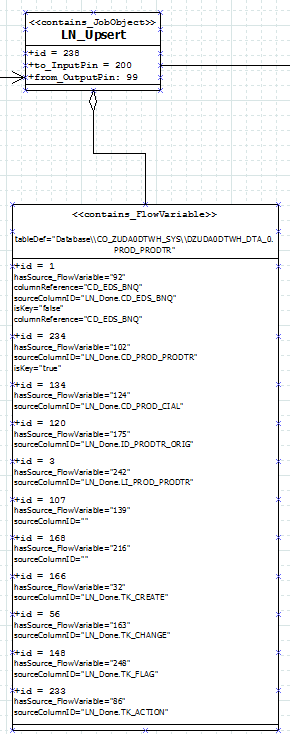
Pour les champs en entrée ; dans le noeud « LN\_Done »



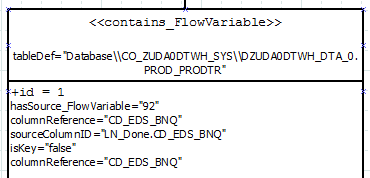
Qui contient lui-même des nœuds « contains\_FlowVariable » dont un des attributs est nommé « isSourceOf\_FlowVariable »



Pour les champs en sortie ; dans le noeud « LN\_Upsert »



Qui contient lui-même des nœuds « contains\_FlowVariable » dont un des attributs est nommé « hasSource\_FlowVariable »



## A noter !

Ici un utilisateur standard est utilisé pour extraire les données, avec un mot de passe périssable. Il conviendra donc qu’un utilisateur système soit défini, avec un mot de passe permanent.

<http://www-01.ibm.com/support/docview.wss?uid=swg21437906>

<https://www.ibm.com/support/knowledgecenter/en/SSZJPZ_11.5.0/com.ibm.swg.im.iis.iisinfsv.assetint.doc/topics/depasset.html>

<https://exldmbd.wordpress.com/2015/10/03/lautomatisation-des-processus-dimportexport-datastage/>

# Traitement de la matière obtenue

## Quel utilitaire, dans les machines de type SL, permet de désarchiver ce fichier .isx ?

[eta2458@slzudisten05 ~]$ unzip --help

UnZip 6.00 of 20 April 2009, by Info-ZIP. Maintained by C. Spieler. Send

bug reports using http://www.info-zip.org/zip-bug.html; see README for details.

Usage: unzip [-Z] [-opts[modifiers]] file[.zip] [list] [-x xlist] [-d exdir]

Default action is to extract files in list, except those in xlist, to exdir;

file[.zip] may be a wildcard. -Z => ZipInfo mode ("unzip -Z" for usage).

-p extract files to pipe, no messages -l list files (short format)

-f freshen existing files, create none -t test compressed archive data

-u update files, create if necessary -z display archive comment only

-v list verbosely/show version info -T timestamp archive to latest

-x exclude files that follow (in xlist) -d extract files into exdir

modifiers:

-n never overwrite existing files -q quiet mode (-qq => quieter)

-o overwrite files WITHOUT prompting -a auto-convert any text files

-j junk paths (do not make directories) -aa treat ALL files as text

-U use escapes for all non-ASCII Unicode -UU ignore any Unicode fields

-C match filenames case-insensitively -L make (some) names lowercase

-X restore UID/GID info -V retain VMS version numbers

-K keep setuid/setgid/tacky permissions -M pipe through "more" pager

See "unzip -hh" or unzip.txt for more help. Examples:

unzip data1 -x joe => extract all files except joe from zipfile data1.zip

unzip -p foo | more => send contents of foo.zip via pipe into program more

unzip -fo foo ReadMe => quietly replace existing ReadMe if archive file newer

[eta2458@slzudisten05 ~]$

L’utilitaire réalise le désarchivage directement là où il est appelé. Il convient donc de lui désigner un répertoire, qu’il créera si nécessaire.



## Créer la liste des fichiers de type « Parallel jobs» ainsi obtenus, sans les répertoires (\*.pjb)





486 fichiers .pjb

## Remplacer les caractères non-imprimables, présents dans les fichiers…

On trouve, malencontreusement, des entités représentant des caractères non imprimables dans des parties résultant de saisie. En l’occurrence :

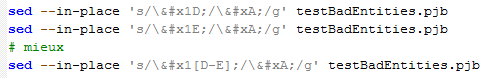
* **&#x1D;**
* **&#x1E;**

Nous allons les transformer en entités de retour à la ligne **&#A;**

Par exemple, un fichier testBadEntities.txt contient

azerty&#x1D;

&#x1E;qsdfgh



azerty&#xA;

&#xA;qsdfgh

## … récursivement

Quelle est la situation ? Combien de fichiers, et lesquels, contiennent des entités indésirables ?



-c count

-r récursif

(il n’est pas possible de filtrer les fichiers avec grep, tous les fichiers passeront par grep.)

! le –exec de find interprète le caractère « & » et ne l’échappe pas par « \& »

La commande finale devient donc :

find . -type f -name "\*.pjb" -exec sed --in-place 's/#x1[D-E];/#xA;/g' "{}" \;