Talend Context variable standard and naming-convention

I. Purposes

Centralize parameters used in Talend projects
Rationalize / standardize contexts through naming convention and values used
Distinct three environments : Developement (DEV), Testing/Quality/UAT (UAT) and Production (PROD)
Initiate Talend development standard which should be completed along with the projects done

II. Reminder : what is a « Context »

- A context in Talend could be assimilate to a parameter which could be used in read/write by any component used in a Talend job
- In programming, it could be refer to a global or a constant variable
- In general, context are defined at Talend project level
- The context value could be define :

  a. Within Talend project (« hard coded »)
  b. External files or database through the option « implicit context load »
  c. In TAC, at deployed job level (job conductor)

According to the latest point and in order to standardize the usage, it is recommended to use the implicit context load from an external data source. It will be easier for Production team to change context value without using any Talend tools. The external data base solution will be kept (only a restricted population should be able to change value in the database).

III. Centralization

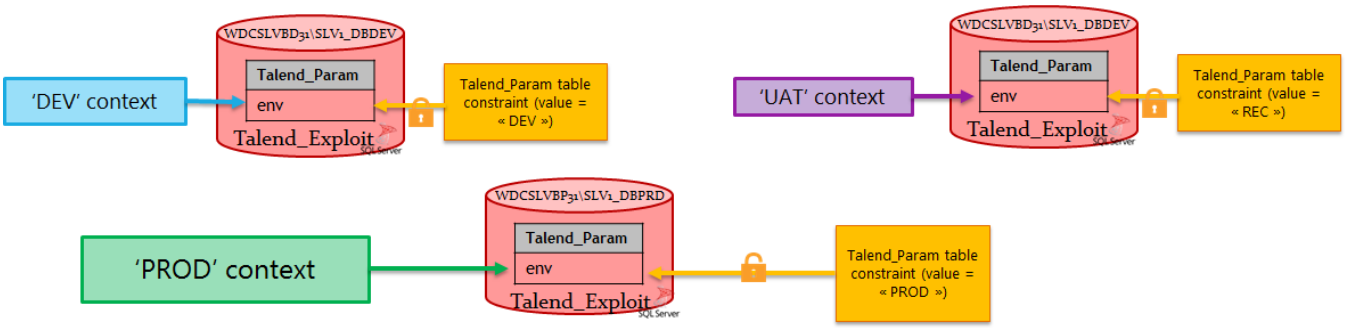
A. Target centralization storage

Environment	DEV	UAT	PROD
Server	WDCSLVBD31\SLV1_DBDEV	WDCSLVBD31\SLV1_DBDEV	WDCSLVBP31\SLV1_DBPRD
Database	TALEND_EXPLOIT	TALEND_EXPLOIT	TALEND_EXPLOIT
Table	TALEND_PARAM	TALEND_PARAM	TALEND_PARAM

B. Centralization TALEND_PARAM

- All context values will be stored in Talend_param table
- Talend_param table will be available on each server. It will contain all data by environment
- The table which could be used by a Production team will be stored on Talend_exploit database on WDCSLVBP31\SLV1_DBPRD server and should only contain prod values

env	key	value	LastModifiedOn	CreatedOn	Comment
Possible values : DEV / UAT or PROD	Context variable name	Value affected	Latest update date	Creation date	Comment



IV. Context

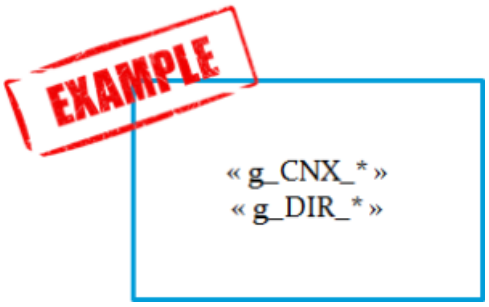
A. Context - Naming convention

- In general we will distinguish two kind of context, global context vs talend project local context
  1. Global context example
     - Source / target database connexions
     - Source / target systems and/or applications connexions (LDAP, WebService, API, etc.)
     - SMTP server
     - Workspace/target root folders used by Talend projects
  2. Local context example
     - Variable dedicated to a specific project (switch of year date, flags and so on)
     - Dedicated workspace folders (storage, rejects, logs and so on)
     - Dedicated Filenames (storage, rejects, logs and so on)
     - Dedicated e-mail distribution list

B. Global context – Naming convention

- Global context name should start by « g_* »
- The name should then follow by one of the following type :

Context type	Description
« CNX_* »	Related to system / application connexions
« DIR_* »	Related to path folders



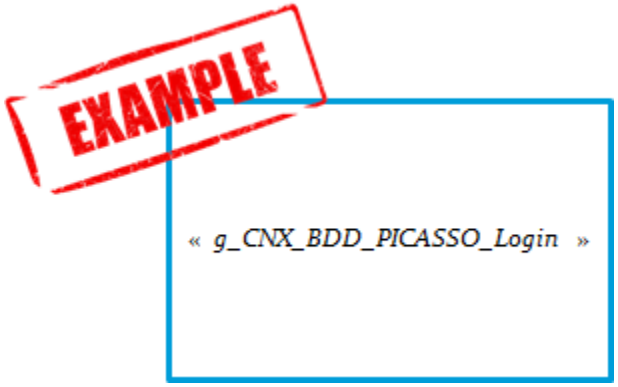
a. Global context – Naming convention « CNX »

SYNOPSIS : « g_CNX_<CONNECTIONTYPE>_<Application>_Key »

- Global context name should start by « g_CNX_* »
- Followed by one of the type below :

<CONNEXIONTYPE>	Description
« BDD »	For Database
« SMTP »	For email communication
« LDAP »	For Active Directory
« SF »	For SalesForce
« FTP »	For File Transfer Protocole
« API »	For Applicative API

- Followed by the System / application name (LABWARE, PICASSO etc.)
- At the end by the key name (Login, Server, etc.)



b. Global context – Naming convention « DIR »

SYNOPSIS : «g_DIR_<PROJECT>»

- Global context name should start by « g_DIR_* »
- Following by the Talend project name which produced the files

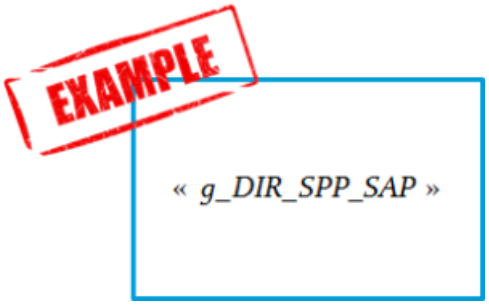
All Talend project directory will be stored as following :

« DiskLetter:/DATA/ » Which located on Talend servers, Depending the talend server we might find the folders DEV, UAT or PROD

env	key	Value*	LastModifiedOn
DEV	g_DIR_SPP_SAP	D:/DATA/DEV/SPP_SAP/	2017-01-20 14:00:00:000
UAT	g_DIR_SPP_SAP	F:/DATA/UAT/SPP_SAP/	2017-01-20 14:00:00:000
PROD	g_DIR_SPP_SAP	F:/DATA/PROD/SPP_SAP/	2017-01-20 14:00:00:000

* : the slash at the end of the value is important since it simplified the concatenation with a filename in Talend

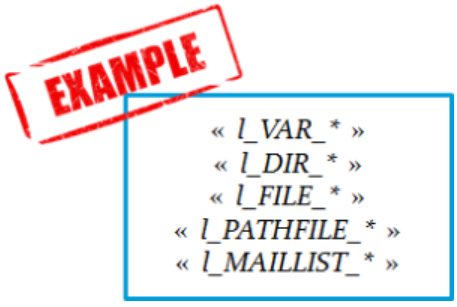
Suggestion : If possible set a network shared folder instead of storing the data directly on Talend servers



C. Local context – Naming-convention

- Local context name should start by « l_* »
- Followed by one of the context type:

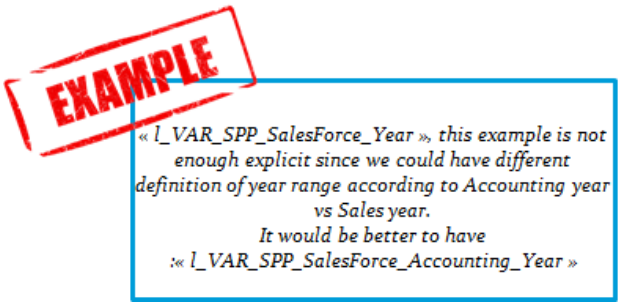
Context type	Description
« VAR_* »	Related to variable / parameter storage
« DIR_* »	Related to file system directory
« FILE_* »	Only filename, no path
« PATHFILE_* »	Related to absolute path filename. It is recommended to concatenate DIR_* and FILE_*
« MAILLIST_* »	Related to e-mail distribution list



a. Local Context – Naming-convention « VAR »

SYNOPSIS : « l_VAR_<PROJECT>_<Key> »

- Local variable/parameter context type should start by « l_VAR_* »
- Followed by the Talend project name
- Ended by an explicit description of the variable usage



b. Local context – Naming-convention « DIR »

SYNOPSIS : « l_DIR_<PROJECT>_<DirectoryType> »

- Local directory context type should start by « l_DIR_* »
- Followed by one of the directory type :

Context type	Description
« Input »	Related to input files
« Output »	Related to output files after ETL processing
« Tmp »	Related to working file, the content could be deleted at any time
« Rejects »	Related to rejects files
« Log »	Related to log file in order to provide some trace for debugging
« Done »	Should contain all files used after ETL processing
« Archive »	Archive files
« Script »	Should contain all script used by ETL processing

c. Local context – Naming-Convention « FILE »

SYNOPSIS : « l_FILE_<PROJECT>_<Key> »

- Local file context type should start by « l_FILE_* »
- Followed by Talend Project name

- At the end set an explicit description of the file content

d. Local context – Naming-convention « PATHFILE »

SYNOPSIS : « l_PATHFILE_<PROJECT>_<Key> »

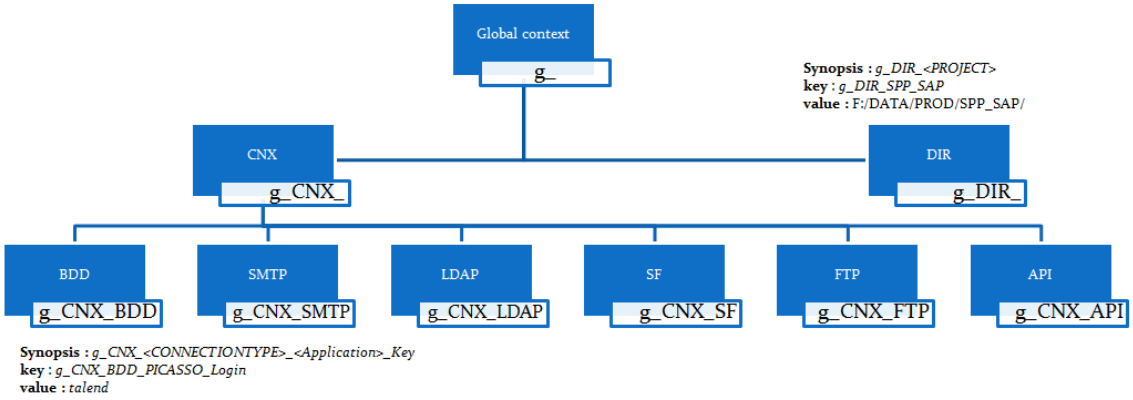
- Local path context type should start by « l_PATHFILE_* »
- Followed by the Talend project name
- At the end set an explicit description related to the file content

e. Local context – Naming-convention « MAILLIST »

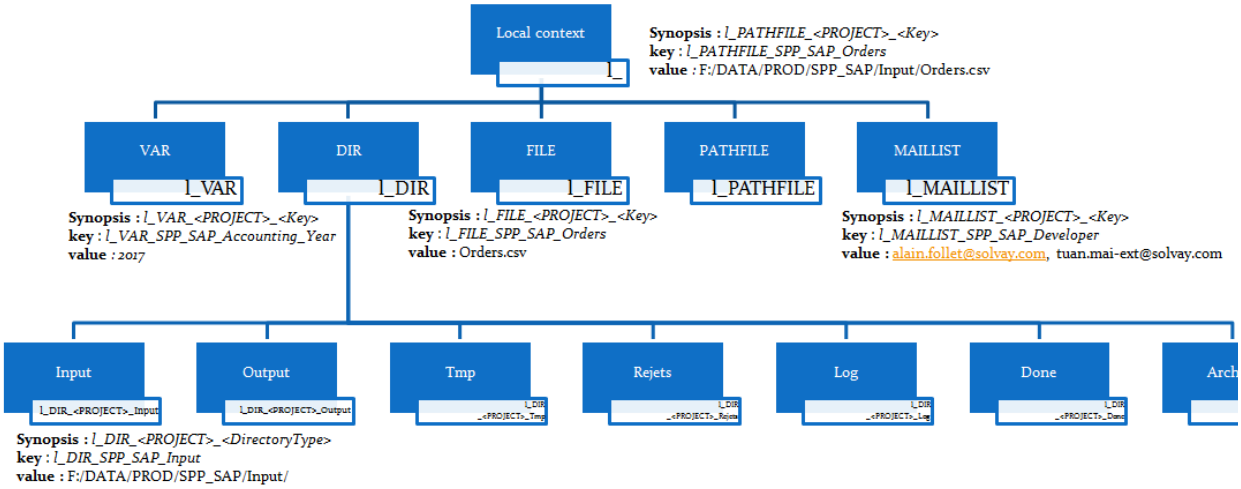
SYNOPSIS : « l_MAILLIST_<PROJECT>_<Key> »

- Local email distribution list context type should start by « l_MAILLIST_* »
- Followed by the Talend project name
- At the end set an explicit description related to the distribution list target population

V. Naming-convention synthesis – Global context



VI. Naming-convention synthesis – Local context



VII. Project folders suggestion

DiskLetter:/ or //serverName/ or alias //server_talend/

- DATA/
  - DEV
    - SPP_SAP
    - SPP_SALESFORCE
      - Input
      - Output
      - Tmp
      - Rejects
      - Log
      - Done
      - Archive
    - REC
  - PROD
    - SPP_SAP
    - SPP_SALESFORCE
      - Input
      - Output
      - Tmp
      - Rejects
      - Log
      - Done
      - Archive

VIII. Misc context vs globalMap – Naming-convention

- In order to simplify the development, some context variable could be used within a Talend project without being centralized in Talend_param table. For that case the naming-convention is to start the context name by : « v_* »
- For context variable only useful in one job and in one talend project could be stored in Talend globalMap component and the naming convention is to start the context name by : « m_* »