UNIX
Data Integration Robot
Implementation Guide

Version 1.0

Document Change History

Date	Version	Description	Author
August 6, 2020	1.0	Initial documentation	Content and Configuration, Galvanize

Contents

1 Overview and Context

This guide describes how to implement the UNIX Data Integration Robot. It assumes that you are familiar with ACL Robotics technologies and terminology.

Data integration robots are designed to support pre-built analysis robots currently in development, for release later in 2020. They also serve as a starting point for building custom analysis robots.

The interaction between data integration robots and analysis robots is shown below. The process is to implement the data integration robot and share the resulting tables so they can be used by any analysis robot. The target analysis robot links to the shared tables required to perform the analysis. Resulting exceptions can be exported to HighBond (Results) from within the analysis robot. Data integration robots do not export raw source data to HighBond (Results).
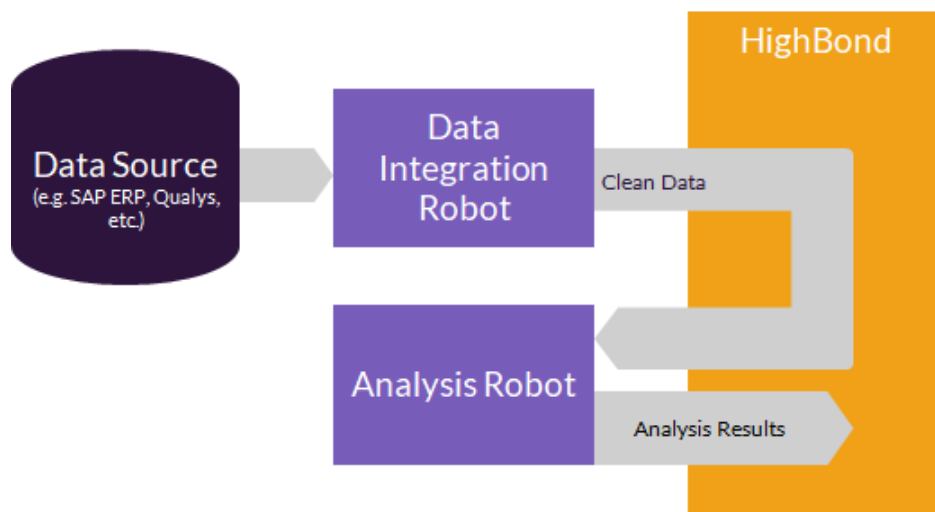
2 Before you Start

Please review the pre-requisites below to confirm your readiness and collect all the appropriate information for the configuration process.

2.1 System Requirements Checklist for Data Integration Robots

Item	Notes	Status
On-premise robot agent 14.2 or higher installed on the server	• Confirm Unicode or non-Unicode version of the agent	
ACL Analytics Version 14.2 or higher installed*	• * A local installation of ACL Analytics can be useful for troubleshooting purposes or for developing custom scripts. If running exclusively as a robot, a local installation may not be required. • Ensure same encoding as the robot agent (Unicode, non-Unicode) • Data integration robots may utilize functionality not available in earlier versions	
Data integration robot is available in the customer organization in HighBond.	• The robot has been successfully deployed into the customer organization via toolkits.highbond.com	

2.2 Source System Access Checklist for Data Integration Robots

Item	Notes	Status
A mechanism for providing UNIX extracts to the Input / Output tab has been established.	• Unlike other data integration robots, the UNIX data integration robot relies on extracts of specific UNIX files, to be uploaded to the Input / Output tab • Extracts can be from a single system, or multiple systems (see expected naming conventions below)	

3   Implementation Steps

Pre-built robots are delivered into the customer's HighBond organization via toolkits.highbond.com. The process creates the robot in development mode, and provisions the required Default Import Configuration File in the Input / Output tab. To complete the configuration of the robot, follow the steps below.

3.1   Open the Robot in HighBond

1)   Log into HighBond Launchpad (www.highbond.com) and browse to the data integration robot.
2)   If the robot opens in production mode, toggle the switch to Development mode, and open the Input / Output tab.

3.2   Review the Default Configurations

The robot is delivered with default configurations - stored in the Default Import Configuration File - that are intended to apply to most customers. These configurations include the list of import scripts to run, or filters applicable to certain tables such as document types or date ranges. A customer may require changes to these default settings if they are not applicable to their environment.

The provided default configurations can be reviewed in two locations:

1)   Refer to the Reference section of this Implementation Guide, or
2)   Download a copy of the Default Import Configuration File from the Input / Output tab.

Do not make any changes to the Default Import Configuration File or overwrite the provisioned file. Do not add, delete, or modify the content or names of any worksheets. Doing so may cause unexpected behavior.

When reviewing the default configurations, take note of any items that require modification, including:

1)   Import scripts that need to be disabled.
  a)   Examples are tables that are not being used by any pre-built or custom analysis robots for the time being, tables to which the user has not been given access, or tables which require customization to align with the customer's needs (refer to section Custom Scripts).
2)   Parameter input that requires different values than those provided with the default configuration.

3.3 Configure the User Import Configuration File (optional)

The User Import Configuration File template is available from an implementation consultant. It is not provisioned by toolkits.highbond.com.

Configuration of the User Import Configuration File is only necessary if changes to the default settings are required. If the default settings are acceptable and no changes are required, leave the User Import Configuration File as-is, with its original worksheet names and field names intact.

1) To make changes, follow these steps as applicable:

    a) In the User_Config_Imports worksheet, disable specific import scripts from running, as follows:

        i) Enter the affected import script name(s) in column Import_Name exactly as they appear in the Default Import Configuration File.
        ii) Enter one script per row and mark column Disable_Import with X (required).

    b) In the User_Config_Params worksheet, apply customized parameter values as follows:

        i) To ensure correct configuration of the new parameter value, it is recommended to download a copy of the Default Import Configuration File and copy the affected parameters to the User Import Configuration File to act as a template for the new input.
        ii) Adjust the copied parameter values as needed, paying attention to the correct formatting as described in the Parameter Description field and in the documentation below. Do not change the variable names.

2) Save any changes made to the User Import Configuration File, if applicable.

Note: Default parameters not overwritten via the User Import Configuration File continue to apply.

3.4 Upload required files to the robot's Input / Output tab

1) Upload the User Import Configuration File to the Input / Output tab, whether it has been changed or not.
2) If used, upload any custom scripts or external source files to the Input / Output tab. For more information, refer to section Custom Scripts.

3.5   Set up the robot task

1)   In Robots, switch to the Tasks tab, and click Create task. Configure the task as follows:

    a)   Assign a meaningful name to the task.
    b)   In the next screen, set the task input parameters.

        i)   As a first test run, it may be helpful to select the Sample Data mode parameter and familiarize a new user with the various concepts of the robot and its output, using small sets of fictitious data. For more information on this option, refer to section Sample Data Mode.

        ii)   Rerun the task on actual data to confirm that a connection to the data connector can be established. (Note: UNIX Data Integration robot does not utilize a data connector.)
            (1)   If the task was initially created in Sample Data mode, edit the task to set the Sample Data mode parameter to False, or leave blank.
            (2)   To test the connection quickly, set the following parameter to import only the first 500 records to True. This option is available only in data integration robots that connect to data connectors. Data integration robots that import external source files do not have this option.
            (3)   Populate the task parameters with the user's credentials and other connectivity settings as prompted.
            (4)   Run the task.

        iii)   Once a successful connection has been confirmed, edit the task to the turn the 500-record limitation off to prevent future imports from receiving limited data.

2)   After confirming that a given data integration task runs successfully, share all source tables (S_*) and the robot's Import Error Log table via the Input / Output tab so that they can later be linked by analysis robots.

3.6   Activate the robot in Production Mode

1)   After confirming the successful implementation in Development mode, activate the robot in Production mode.

    a)   All external files in the Input / Output tab need to be manually uploaded to Production mode. This includes:

        i)   Default Import Configuration File (download from Development mode, do not make changes).
        ii)   User Import Configuration File (download from Development mode).
        iii)   Any custom scripts or external source files, if applicable.

    b)   Once the robot has run in Production mode, the resulting source tables need to be manually shared in the Input / Output tab for use by analysis robots also running in Production mode.

4 Sample Data Mode

Sample Data mode can be used temporarily to create source tables populated with fictitious data. The sample data resides in the Default Import Configuration File. This mode provides a method to showcase a robot's components and concepts without the need to connect to an actual data source.

Most task parameters such as user account information can be skipped in Sample Data mode, except for any hashed input such as passwords. Enter a random string of text to be able to save the task. An actual password is not required in Sample Data mode. Parameter input from the Default and User Import Configuration Files is not used in this mode. The 500-record limitation, if selected, does not apply.

5 Custom scripts

Data integration robots import a variety of tables to support analysis robots. A customer may need to import additional tables to support their own custom analytics, or they may need to import an existing table with modifications such as additional filters or fields. Both examples require the use of custom scripts.

The original data integration robot **should not** be modified internally to incorporate new scripts or modify existing scripts. Doing so jeopardizes the future ability to deploy official updates seamlessly into the customer's organization. Custom scripts provide a mechanism for these types of changes, as described in the following scenarios.

5.1 Importing an additional table

1) Create a separate custom script for every additional table.

2) To ensure compatibility with standard data integration and analysis robots processes, source tables must be named as follows, using SAP ERP as an example. Adapt the naming conventions to the data integration robot being built:

   a. Minimum table name: S_<application_system>_<table_name>. For example: S_SAP_LFA1
   b. If there are filters being applied, the table name can be expanded to include a reference to the filter. For example: S_SAP_LFA1_LAND1 if filtered on LAND1.
   c. The table name must not match an existing table name (see next scenario).
   d. The script must be saved as an external aclscript file, with the same name as the table. For example: S_SAP_LFA1_LAND1.aclscript.
   e. The custom script must be uploaded to the Input / Output tab in both Development mode and Production mode.

5.2 Customizing an existing table import

1) If a custom script replaces an original script, disable the original script via the User Import Configuration File.

2) Create a separate custom script for every customized table import.

3) To prevent conflicts with the original table import, the custom table name must be unique.
   a. For example, if the original table S_SAP_LFA1 is to be replaced by a custom script, the new table name could be S_SAP_LFA1_Custom.
   b. The script must be saved as an external aclscript file, with the same name as the table. For example: S_SAP_LFA1_Custom.aclscript.
   c. The custom script must be uploaded to the Input / Output tab in both Development mode and Production mode.

Existing variables from the Default Import Configuration File or User Import Configuration File are available to the custom script. The 500-record limitation is also available to the custom scripts. Copy the corresponding syntax from an original script.

Custom scripts can be developed in an ACL project to facilitate testing or as an external text file. An existing script can be used as a template.

Custom scripts run last, in alphabetical order on script name. They can use the tables created by the enabled original import scripts as input if needed. If a custom script is dependent on another custom script, control the order in which they run via the script names.

6 Reference

6.1   Default Parameters

The Default Import Configuration File provides the following parameter input for the import scripts:

Variable Name	Worksheet Name	Type	Applies to	Default Value
Not applicable				

6.2   Overwriting default parameters

If any of the provided default parameters do not apply or are incomplete, declare the required custom values in the User Import Configuration File. The format and naming conventions must match the Default Import Configuration File.

Variable Name	Worksheet Name	Type	Applies to	Default Value
Not applicable				

Note: All parameter values are ignored if the sample data approach has been selected.

6.3 S_UNIX_GROUP

General approach
- Unlike other data integration robots, the UNIX process does not connect to the production data via Data Connector but instead, requires extracts to text files.
  - Requires extracts from one or more UNIX systems to be uploaded to the Input / Output tab in Robots.
  - File naming convention:
    - S_<system_name>_etc_group.txt
    - In case of extracts from a single system, the naming convention S_etc_group.txt is sufficient.
    - The extract must include all available fields, delimited by a colon.

- Although the system name is optional for single source systems, the extract must always be prefixed with S_.
- As an example, if there are extracts from multiple systems, the naming conventions could be S_SystemA_etc_group.txt and S_SystemB_etc_group.txt.
- The data integration robot combines data from all group files into a single source table, S_UNIX_Group and keeps track of the source system name each record came from.

- Required fields:
  - Group_Name
  - Password
  - GID
  - Group_List

Import Filters
Not applicable

- Note: The 500-record limitation available in other data integration robots is not available in the UNIX data integration robot. The record limitation, intended for faster initial testing, is available only for use with data connectors.

Post Processing Script
- Applicable only to sample data mode, exports the sample data imported from Excel to text files, to match the structure of the production extracts. Then calls the standard utility scrips Import_UNIX_Extracts and Import_UNIX_Extracts_A01 to append the etc_group files from sample data source systems "SystemA" and "SystemB" into table S_UNIX_Group.
  - Calls the Rename_UNIX_Fields script to prefix the field names with the file name. This step allows the various source files (group, passwd, shadow) to be joined without duplicating common field names.

- Applicable to both sample data and production processes, determines the number of individual etc_group extracts expected based on a count of unique systems.
  - Calls error handling procedures if one or more files for a given system appear to be missing, see below.

Error Logging
- Reports if any of the source tables could not be created due to missing extracts for one or more systems.
- Reports which systems' files could not be located.

6.4 S_UNIX_PASSWD

General approach

- Unlike other data integration robots, the UNIX process does not connect to the production data via Data Connector but instead, requires extracts to text files.
    - Requires extracts from one or more UNIX systems to be uploaded to the Input / Output tab in Robots.
    - File naming convention:
        - S_<system_name>_etc_passwd.txt
        - In case of extracts from a single system, the naming convention S_etc_passwd.txt is sufficient.
        - The extract must include all available fields, delimited by a colon.

- Although the system name is optional for single source systems, the extract must always be prefixed with S_.
- As an example, if there are extracts from multiple systems, the naming conventions could be S_SystemA_etc_passwd.txt and S_SystemB_etc_passwd.txt.
- The data integration robot combines data from all passwd files into a single source table, S_UNIX_Passwd and keeps track of the source system name each record came from.

- Required fields:
    - User_Name
    - Password
    - UID
    - GID
    - GECOS
    - Home_Directory
    - Command_Shell

Import Filters
Not applicable

- Note: The 500-record limitation available in other data integration robots is not available in the UNIX data integration robot. The record limitation, intended for faster initial testing, is available only for use with data connectors.

Post Processing Script

- Applicable only to sample data mode, exports the sample data imported from Excel to text files, to match the structure of the production extracts. Then calls the standard utility scrips Import_UNIX_Extracts and Import_UNIX_Extracts_A01 to append the etc_passwd files from sample data source systems "SystemA" and "SystemB" into table S_UNIX_Passwd.
    - Calls the Rename_UNIX_Fields script to prefix the field names with the file name. This step allows the various source files (group, passwd, shadow) to be joined without duplicating common field names.

- Applicable to both sample data and production processes, determines the number of individual etc_passwd extracts expected based on a count of unique systems.
    - Calls error handling procedures if one or more files for a given system appear to be missing, see below.

Error Logging

- Reports if any of the source tables could not be created due to missing extracts for one or more systems.
- Reports which systems' files could not be located.

6.5 S_UNIX_SHADOW

General approach

- Unlike other data integration robots, the UNIX process does not connect to the production data via Data Connector but instead, requires extracts to text files.
    - o Requires extracts from one or more UNIX systems to be uploaded to the Input / Output tab in Robots.
    - o File naming convention:
        - ▪ S_<system_name>_etc_shadow.txt
        - ▪ In case of extracts from a single system, the naming convention S_etc_shadow.txt is sufficient.
        - ▪ The extract must include all available fields, delimited by a colon.

- Although the system name is optional for single source systems, the extract must always be prefixed with S_.
- As an example, if there are extracts from multiple systems, the naming conventions could be S_SystemA_etc_shadow.txt and S_SystemB_etc_shadow.txt.
- The data integration robot combines data from all shadow files into a single source table, S_UNIX_Shadow and keeps track of the source system name each record came from.

- Required fields:
    - o User_Name
    - o Password
    - o Last_Pwd_Change
    - o Minimum
    - o Maximum
    - o Warn
    - o Inactive
    - o Expire
    - o Reserved_Field

- File /etc/shadow is not used in UNIX implementation
    - o Some flavors of UNIX do not use /etc/shadow. In this case, disable the S_UNIX_Shadow script in the User Import Configuration File.
    - o A custom import script may need to be created to provide the required data from other sources.
    - o This may be especially relevant to the determination of a user's account status for use (active / inactive) in pre-built analysis robots.
        - ▪ Pre-built analysis robots determine the status based on the following characteristics:
            - A blank or future date in field Expire in file etc_shadow, or
            - A password not preceded by * or ! in file etc_shadow.
            - Since not all flavors of UNIX are configured in this manner, it may be necessary to create custom import scripts and / or analysis scripts to determine the status in another manner.

Import Filters
Not applicable

- Note: The 500-record limitation available in other data integration robots is not available in the UNIX data integration robot. The record limitation, intended for faster initial testing, is available only for use with data connectors.

Post Processing Script

- Applicable only to sample data mode, exports the sample data imported from Excel to text files, to match the structure of the production extracts. Then calls the standard utility scrips Import_UNIX_Extracts and Import_UNIX_Extracts_A01 to append the etc_shadow files from sample data source systems "SystemA" and "SystemB" into table S_UNIX_Shadow.
    - Calls the Rename_UNIX_Fields script to prefix the field names with the file name. This step allows the various source files (group, passwd, shadow) to be joined without duplicating common field names.

- Applicable to both sample data and production data, determines the number of individual etc_shadow extracts expected based on a count of unique systems.
    - Calls error handling procedures if one or more files for a given system appear to be missing, see below.

- Applicable to both sample data and production data, creates computed fields to c_etc_shadow_Expire and c_etc_shadow_last_Pwd_Change into a legible format, calculated as 1970/01/01 plus the number of days shown in the source fields.

Error Logging

- Reports if any of the source tables could not be created due to missing extracts for one or more systems.
- Reports which systems' files could not be located.

7 Frequently Asked Questions / Troubleshooting

7.1 The parameter values or disabled scripts in the User Import Configuration File are not being applied

There are multiple situations in which the user configurations are not applied.

1) Check the Import Error Log table (uploaded as part of the result files of the robot). The error log reports issues such as a missing or incorrectly configured User Import Configuration File. In these cases, the Default Import Configuration File is used.
   a. Ensure the file has been uploaded to the Input / Output tab in the mode the robot is running in (Development and / or Production).
   b. Ensure the worksheet names and field names match the expected conventions. To confirm the expected worksheet names and field names, refer to the Default Import Configuration File, worksheet Expected_User_Config_Fields.

2) An invalid code has been entered in field Disable_Import.
   a. Ensure that the codes entered match those specified in variable **v_main_disabled_codes** in the start script (_<source system>_Imports). Codes not specified in the variable do not disable the import script.

3) If the robot is running in Sample Data mode, the parameter configurations from the User Import Configuration File are not being used.
   a. The Import Error Log has a note stating the robot is running in sample data mode.

7.2 Error message "Enter the file name" causes the process to fail

This error message can be caused by a misspelled import script name in the Import_Name column in the User Import Configuration File.

7.3 The record count of imported tables appears to be limited to 500 records

1) Edit the task to make sure the 500-record limit has been turned off.
2) Review any scripts, including custom scripts, to make sure the 500-record limitation is not hard-coded, for example, when copying syntax from the log to create the script.

7.4   In the Input / Output tab, a disabled script is showing a record count

1) This scenario is possible when a robot has been run once before with the table enabled, and the table has subsequently been disabled.
    a. The record count and update date reflect the last time the table was populated.
    b. Any analytics still using this table are using the available data as of the last run date.
    c. To prevent this situation, click and delete the table from the Input / Output tab.
    d. Ensure that all analytics requiring this table in any analysis robots are disabled via the analysis robots' User Analytic Configuration Files.

7.5   What is "post processing"?

Post processing refers to data cleansing taking place within the data integration robot. The tasks performed by post processing scripts vary by data integration robot. Most commonly, there may be a difference between data types imported from the production source system and in sample data mode. Post processing scripts align the data types so that any dependent analytics have a consistent data set to work with.

Other post processing may involve renaming sample data tables imported from the Default Import Configuration File if the table name was truncated in Excel. To ensure the source tables have the expected table names regardless of whether they came from a data connector or the sample data process, post processing checks the names of these tables for truncation and applies the expected name.

Post processing is an optional task in a data integration robot, and its use is determined by the developer. If a post processing script is present, it runs last, after all imports are completed. Post processing scripts are limited to original scripts, they are not intended to be used with custom scripts. Ensure all custom scripts perform their own data cleansing as needed.

7.6 What is the difference between data cleansing and data preparation?

You may hear the terms "data cleansing" in the context of a data integration robot, and "data preparation" in the context of an analysis robot.

Data cleansing is generally defined as harmonization or other processing – such as creation of computed fields – that apply to a table for most or all use cases. For example, a transaction date field is being imported as character. Every conceivable use of this field would require a DATE data type. Therefore, converting the data from character to date is a suitable application of data cleansing to be performed in the data integration robot. By performing the task in data integration, the correct data type is available to all dependent analysis robots. Joining of header and line item tables (if imported separately) is another application of data cleansing. These types of tasks are performed in post processing scripts that apply to data either imported via the data connector or sample data.

Data preparation is generally defined as harmonization or other processing limited to a small number of specialized analytics. For example, to obtain a quarterly summarization of data, a computed field showing only the year and the month of the transaction date is required. This type of task, which supports only a small number of analytics, is considered data preparation and is performed only in the analysis robots that need this field.

The line between data cleansing and data preparation can be blurry. If a decision to include a given task in either data cleansing or data preparation is causing issues, please report the scenario via the standard support / reporting channels for evaluation.

7.7 The default parameter values, imported tables, or imported fields do not match most clients' needs

Please report these findings via the standard support / reporting channels for evaluation. If the case is compelling and affects most customers, the master default configuration and master list of imported scripts may be updated in a new release to eliminate the need for customization.

8 Running a data integration robot in ACL Analytics

Although data integration robots have been optimized for use as a robot, they can also be run in ACL Analytics.

1) If not previously tested, ensure that a connection to the source system can be established from within ACL Analytics.

2) Ensure that all expected files such as Default Import Configuration Files, User Import Configuration Files, and any custom scripts or external source files are in the same folder as the ACL project. Ensure all files are closed.

3) Within the correct version of the ACL project (Unicode, non-Unicode), open the start script, named _<source_system>_Imports, for example _SAP_Imports.

4) Review the analytic header for any changes that need to be made to the parameter prompts.
   a. Default values have been provided for some parameters such as sample data mode, the 500-records limitation, or some common connectivity settings. Update the default values as needed.
   b. Other parameters require custom input, such as user account names or customer specific connectivity settings specific to the source systems.

5) Run the start script to kick off the process.

6) If any of the source tables are required by an analysis robot, save the data integration robot to update the project with the latest working copy.
   a. In the analysis robot, copy the table layouts from the data integration robot.
        i. If the table layouts are not available for selection, it is likely that the data integration project was not saved.
   b. Copy the .fil files to the analysis robot project folder or manually link the copied tables to the source files, if needed.
   c. Repeat these steps if toggling between sample data mode and production mode as the layouts will differ, resulting in skewed data.