ALANIS AUDIT REPORT GENERATION TOOL RELEASE 1.0

PREFACE

DISCLAIMERS

Altran products are intended for commercial uses. Without the appropriate network design engineering, they must not be sold, licensed or otherwise distributed for use in any hazardous environments requiring fail-safe performance, such as in the operation of nuclear facilities, aircraft navigation or communication systems, air traffic control, direct life-support machines, or weapons systems, in which the failure of products could lead directly to death, personal injury or severe physical or environmental damage. The customer hereby agrees that the use, sale, license or other distribution of the products for any such application without the prior written consent of Altran, shall be at the customers sole risk. The customer hereby agrees to defend and hold Altran harmless from any claims for loss, cost, damage, expense or liability that may arise out of or in connection with the use, sale, license or other distribution of the products in such applications.

This document may contain information regarding the use and installation of non-Aricent products. Please note that this information is provided as a courtesy to assist you. While Aricent tries to ensure that this information accurately reflects information provided by the supplier, please refer to the materials provided with any non-Altran product and contact the supplier for information provided about non-Altran products.

However, this does not constitute a representation or warranty. The warranties provided for Altran products, if any, are set forth in contractual documentation entered by Aricent and its customers.

This document was originally written in English. If there is any conflict or inconsistency between the English version and any other version of a document, the English version shall prevail.

Ed	Date	Status	Comments	Author	Approval
1			Creation of documents		

TABLE OF CONTENTS

1.INTRODUCTION

1.1 SCOPE

The scope of this specification is to develop Alanis Audit Report Generation Tool within Altran and Nokia. All the sections in this document describe the Beta with regards to Alanis Audit Report Generation Tool release 1.1 only. The ultimate aim of this tool is to extract the xml files from tarfile(.tgz) and convert into pdf file with reference to the 1.2 template document.

1.2 Referenced documents

Title	Description
alanis_report_annexe_Poland_GSMR_R5.3_GSMR_2019-02-20_081100_annexe	
oracle_alanis_report_Italy_WIND_R5.0_Live_2015-08-06_151400	

2. MAINTENANCE WINDOW DETAILS, ASSUMPTIONS, CONSTRAINTS & DEPENDENCIES, PREREQUISITES

2.1 Maintenance Window Details

Contact Name	Role	Telephone Cell / Office	Email	Supervisor
Vijayaraj G	Software Engineer	9994193068	vijayaraj.g@aricent.com	Vungarala.Srinivas@aricent.com

2.2 ASSUMPTIONS

- As servers and services are grabbed from "/result/servers" file and "/result/services" file respectively. It is assumed that the input file having same folder hierarchy to perform the extraction of required data from XML's inside it.
- Use of a latest web browser can give exact UI (user interface) as the web design done using recent version of Php.

2.3 CONSTRAINTS & DEPENDENCIES

- There is a non-printable ascii character TAB presents in some xml files. This is removed via code and then processed for extraction.
- There are some xml tags are opened but not closed. It is closed via code and then processed for extracting the data's inside it. i.e. the </comment> end tag is missing in some xml files.

2.4 PREREQUISITES

- Server's username and password should be available. Hence login credentials shared via mail based on approval

3. SYSTEM REQUIREMENTS

3.1 HARDWARE REQUIREMENTS

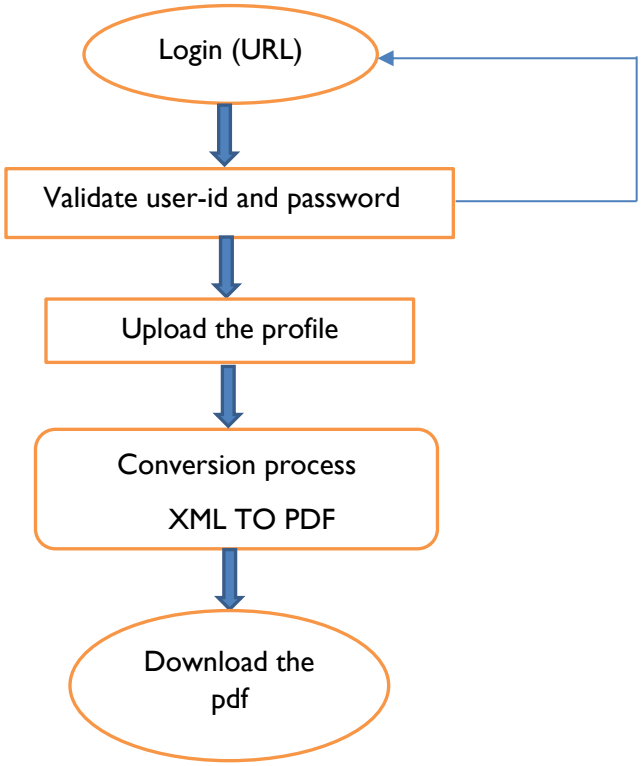
Processor : Intel dual core 64 bit

3.2 SOFTWARE REQUIREMENTS

Operating System : Windows 7

Language Used : Python 3.7, Php 7.4, Apache 2.4

3.3 DESIGN ARCHITECURE



4. DESIGN AND IMPLEMENTATION

4.1 DESIGN MODULES

The execution of the activities would be as follows:

- Install and configure python, apache and php
- Login validation using server-side scripting (i.e. using php)
- Software description

4.2 MODULES DESCRIPTION

4.2.1 INSTALL AND CONFIGURE APACHE, PYTHON AND PHP

- **Installation of Python**
  Python is an open-source software it can be downloaded freely
  The procedures are as follows

  - Open a browser window and navigate to the Download page for Windows at python.org.
  - Underneath the heading at the top that says **Python Releases for Windows**, click on the link for the **Latest Python 3 Release - Python 3.x.x**. (As of this writing, the latest is Python 3.7.3.)
  - Scroll to the bottom and select either **Windows x86-64 executable installer** for 64-bit or **Windows x86 executable installer** for 32-bit. (See below.)
  - Download the installer and run it by double click on the downloaded installer
  - Tick the check box of Add Python 3.X.X to PATH to ensure the interpreter will be placed in your execution path

- **Installation of Apache**
  Apache is also an open-source software, so it can be downloaded freely
  The procedures are as follows
  - Install the Apache web server on your PC. Go to http://www.apachelounge.com/download/. Scroll down the page until you find the download for the 'Apache 2.4 win64 binaries' and download. You need to be careful that the module dll in PHP matches the version of Apache you install. Apache won't load otherwise.
  - Unzip the file into C:\. You should end up with a directory 'Apache24' (or whatever the latest version is).
  - Find Start > All programs > Accessories > Command Prompt..... BUT, right click, and select 'Run as administrator'.
  - Enter the following commands
    ```
    cd \Apache24\bin
    httpd -k install
    httpd -k start
    ```
  - you may well get a warning about the server name. Don't worry about it. Don't close this window, you will need it again in a minute.
  - To test it worked type 'http://localhost' into your browser. You should get a screen up with a text “It works” to the effect that Apache is installed and working.

  Configuration of apache to python CGI
  - open the httpd.conf apache configuration file located in the apache install directory in the conf directory. Search the httpd.conf file for the line
  - line no 60: Listen: Allows you to bind Apache to specific IP addresses and/or ports, instead of the default.
  - For this application I have set to “Listen 10.203.28.245:8056”, here the instead of IP address localhost can also be given and the port number can be any number from 0 to 65535 except the reserved ports, avoid first 1024 port numbers because their considered as system ports
  - DocumentRoot: The directory out of which you will serve your documents. By default, all requests are taken from this directory, but symbolic links and aliases may be used to point to other locations.

- Line no 254: DocumentRoot "D:/Alanis_Tool_Reports/xxxxxxx" this document root is an example you can set any location as you wish.

- Set the same path as the DocumentRoot to line no 255: <Directory "D:/Alanis_Tool_Reports/generated_pdfs">

  Change the configuration of the directory according to your requirements

- ScriptAlias: This controls which directories contain server scripts. ScriptAliases are essentially the same as Aliases, except that documents in the target directory are treated as applications and run by the server when requested rather than as documents sent to the client.
- Set line no 371 ScriptAlias /cgi-bin/ "c:/Apache24/cgi-bin/"
- Set line no 387 as below

```
<Directory "c:/Apache24/cgi-bin">
   AllowOverride all
   Options ExecCGI
   Require all granted
</Directory>
```

- Uncomment line no 435 by removing the # in front of the line, and add a .py to the end of the line. The new line should look like this:
- AddHandler cgi-script .cgi .py
- Save the httpd.conf file
- **Note:** the first line of your python program must tell where the python interpreter is located for example #!c:/python37/python.exe This line needs changed to match the location of your Python installation

Restart Apache

- Now, the apache web server needs to be restarted. Find Start > All programs > Accessories > Command Prompt..... BUT, right click, and select 'Run as administrator'.
- Enter the following commands

```
cd \Apache24\bin
httpd -k stop
httpd -k restart
```

Run a test Python page

- Here is an example assuming Python is installed in the C:\Python37 location

  #!/Python37/python print "Content-type: text/html" print print "<html><head>" print "" print "</head><body>" print "Hello." print "</body></html>"

  Note the line **#!/Python37/python**. This line needs changed to match the location of your Python installation.

  Save this file as test.py to your htdocs folder under your apache installation directory or wherever you set the script alias i.e., cgi-bin. Open your web browser and type in your apache host (and :port if the port is something other than 80) followed by test.py, for example

  http://localhost/test.py or IP address followed by port number according to your config.

➢ **Installation of PHP**

PHP is also an open-source software it can be downloaded freely
The procedures are as follows

- Now install the PHP scripting language on your PC. Go to http://www.php.net/download. In the current stable release section click on link for Windows 7.2.x binaries and source. Scroll down to the newest 'Zip' for **VC14 x86 Thread Safe** (non thread safe doesn't have the Apache dll. the VC version may be different, and you may need x64 version) PHP (again, the newest versions of PHP didn't have this but it shouldn't matter) and download.
- Open the zip file and extract to C:\php7.4

Configuration of apache to PHP

- You now need to edit Apache's httpd.conf file again. In the file explorer navigate to C:\Apache24\conf\httpd.conf. Open it in Notepad. At the end of this file (or wherever you like if you want to be more organized) add the following lines:

    LoadModule php7_module "C:/PHP/php7apache2_4.dll"

    AddHandler application/x-httpd-php .php

    PHPIniDir C:/php7.4

- The version of the module file matters (2_4 in this case). It MUST match the Apache version installed.
- In the same file. Search for the line starting **DirectoryIndex**. Change it as follows

    DirectoryIndex index.php index.html

- Now, navigate to C:\PHP7.4, and rename php.ini-development to php.ini. Edit this file, find the following lines and modify them as follows,
  You need to specify the extensions required for php. Find the 'Dynamic Extensions' section and change the following lines (uncomment and add the correct path):

    extension=c:/php/ext/php_curl.dll

    extension=c:/php/ext/php_gd2.dll

    extension=c:/php/ext/php_intl.dll

    extension=c:/php/ext/php_mbstring.dll

    extension=c:/php/ext/php_mysqli.dll

    extension=c:/php/ext/php_openssl.dll

    extension=c:/php/ext/php_soap.dll

    extension=c:/php/ext/php_xmlrpc.dll

- Set line no 757 as extension_dir = "c:\Apache24\php7.4\ext"
- (these are a minimum. You may need others - e.g. LDAP - for specific functions) ...and save.
- Back in the 'cmd' window for Apache, you need to restart it to load your changes...

    httpd -k restart

Run a test PHP page

- Navigate to C:\Apache24\htdocs or wherever you set the script alias i.e., cgi-bin and create a file called 'test.php'.

  In this file enter the single line...

    <?php phpinfo();

    ?>

  And then, in your browser, navigate to http://localhost/test.php or according to your config. You should see a screen with masses of information and the PHP logo at the top. Check a few lines down for 'Loaded Configuration File' and make sure it says c:\php\php.ini.

4.2.2 LOGIN VALIDATION USING SERVER-SIDE SCRIPTING

- The tool can be accessed through URL (Uniform Resource Locator) for that to protect unauthorized access to the site a login page has been implemented. So, there will be backend database either a file or a relational database like MySQL, mangoDB etc.,
- Based on approval only authorized persons will be added in the database and thereby a unique username and password will be given to the authorized person.
- Here we used PHP a widely known scripting language for web development.
- when the user entered the login credentials by submitting the login form the php script will validate the username and password provided by the user is correct or incorrect through comparing it with the database (text file). If correct, then the user will be redirected to the profile upload files else again redirected to login page for entering the username and password

4.2.3 SOFTWARE DESCRIPTION

- **Required modules for converting tarfile to pdf file**

  It requires to import below modules necessarily:

  1. reportlab -3.5.28
  2. PyPDF2 1.26.0
  3. lxml 4.4.1

- **Description of modules**

  - Reportlab

    ReportLab is the time-proven, ultra-robust open-source engine for creating complex, data-driven PDF documents and custom vector graphics. It's free, open-source, and written in Python.

    - The reportlab handling submodules are:
      - reportlab.pdfgen.canvas
      - reportlab.platypus
      - reportlab.lib.styles
      - reportlab.lib.pagesizes
      - reportlab.lib.enums
      - reportlab.lib.colors
      - reportlab.lib.units
    - PyPDF2

      PyPDF2 is a pure-python PDF library capable of splitting, merging together, cropping, and transforming the pages of PDF files. It can also add custom data, viewing options, and passwords to PDF files. It can retrieve text and metadata from PDFs as well as merge entire files together.

    - lxml

      lxml is a Pythonic, mature binding for the libxml2 and libxslt libraries. It provides safe and convenient access to these libraries using the ElementTree API.

      - The lxml handling submodules are:
        - lxml etree
        - lxml objectify

- **Default modules**

Apart from the two above externally downloaded modules the below listed default modules are also used. They are,

- xml:
  - The xml handling submodules are:
    - ❖ xml.etree.ElementTree
- shutil
- tarfile
- re (regular expression)
- glob
- os
- The os handling submodules are:
  - ❖ os.path
- datetime
- cgi(common gateway interface)
- csv (comma separated values)
- sys
- io
- cgitb

- **Required Classes & methods**

Class name	Method name	Functionality
Pagebreak		Move on to the next page in the document. This works by consuming all remaining space in the frame
	xml.get	It used to get the data from xml file
	append	add the data into table at last or list at last
Paragraph		Data should be in paragraph format
	xml.find	finds the *first* child with a particular tag
Tablestyle		It describes style of table
	xml.description	It describes get the data from particular tag name
	readlines	It returns a list containing each line in the file as a list item
	startswith	returns True if a string starts with the specified prefix(string). If not, it returns False.
	endswith	returns True if a string ends with the specified suffix. If not, it returns False
	restorestate	restore the graphics state to the matching saved state
	Savestate	Save the current graphics state to be restored later by restoreState
Image		At the present time images as flowables are always centered horozontally in the frame. We allow for two kinds of lazyness to allow for many images in a document which could lead to file handle starvation
	build	build(...) must be called for most reasonable uses since it builds a document using the page template.
Spacer		A spacer just takes up space and doesn't draw anything - it guarantees a gap between objects
	setStyle	This is the only way to get tables to appear in a nicely formatted way
Table		Tables are created by passing the constructor an optional sequence of column widths, an optional sequence of row heights, and the data in row order
	save	it represents to save the document
	fromstring	Parses an XML section from a string constant
		Parses an XML section into an element tree. *source* is a filename or file object containing XML data. *parser* is an

	parse	optional parser instance. If not given, the standard XMLParser parser is used. Returns an ElementTree instance
	xml.findall	finds only elements with a tag which are direct children of the current element
	extractall	This method will extract all the contents of the any file to the current working directory
	glob	Return a possibly-empty list of path names that match *pathname*, which must be a string containing a path specification
	rmtree	function to delete all the contents of a directory i.e. It accepts 3 arguments ignore_errors, onerror and path. path argument should be a path of the directory to be deleted
	split	returns a list of strings after breaking the given string by the specified separator
	rsplit	returns a list of strings after breaking the given string from right side by the specified separator
	iterchildren	the returned elements can be reversed with the 'reversed' keyword and restricted to find only elements with a specific tag
	join	It provides a flexible way to concatenate string. It concatenates each element of an iterable (such as list, string and tuple) to the string and returns the concatenated string
PdfFileWriter		This class supports writing PDF files out, given pages produced by another class
PdfFileReader		Initializes a PdfFileReader object. This operation can take some time, as the PDF stream's cross-reference tables are read into memory.
	enable	function control whether the report is displayed in the browser and whether the report is logged to a file for later analysis
BytesIO		A stream implementation using an in-memory bytes buffer. It inherits BufferedIOBase.The buffer is discarded when the close() method is called.Theoptionalargument *initial_bytes* isa bytes-like object that contains initial data. It provides or overrides these methods in addition to those from BufferedIOBase and IOBase
	extractText	Locate all text drawing commands, in the order they are provided in the content stream, and extract the text. This works well for some PDF files, but poorly for others, depending on the generator used. This will be refined in the future. Do not rely on the order of text coming out of this function, as it will change if this function is made more sophisticated
	getPageNum	Retrieve page number of a given PageObject
	getPage	Retrieves a page by number from this PDF file
	addPage	Adds a new page to the document
FieldStorage		The optional argument keep_blank_values is a flag indicating whether blank values in URL encoded form data should be treated as blank strings. The default is false, which means that blank values are ignored as if they were not included.
	getValue	just returns the entire contents of the stream regardless of current position
	reader	Read the data from a csv file
	strip	The strip() removes characters from both left and right based on the argument (a string specifying the set of characters to be removed)

	Strftime	returns a string representing date and time using date, time or datetime object.
	listDir	Returns the list of files inside the directory
	Open	It will open the file

5. RESULTS AND DISCUSSIONS

5.1 Login URL Page

- open google chrome and type the following URL

  http://10.203.28.245:8052/code_files/login_page.php

- Enter the registered user_id and password
- Click “Login” button
- Once you log into the page then you will get the profile upload page

5.2 Profile Upload Page

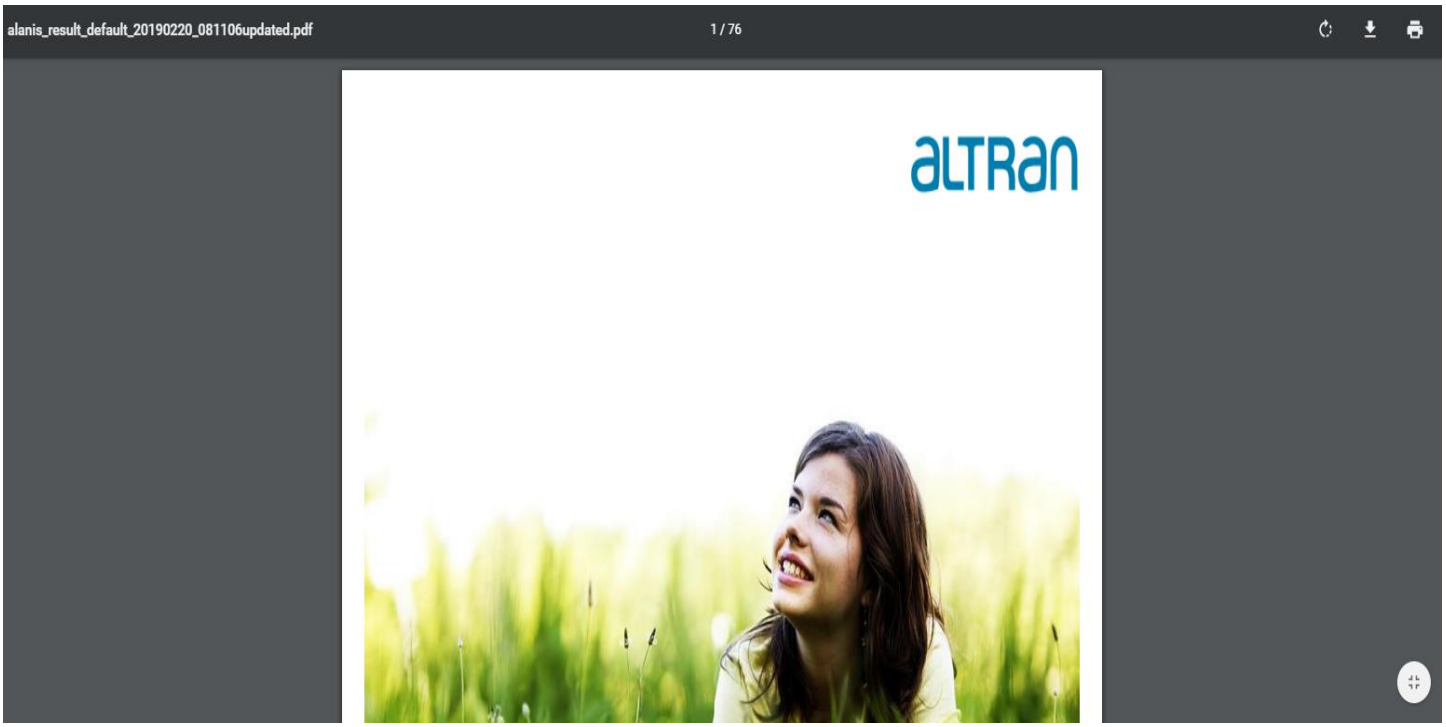
- Click “choose profile’ for upload file and select .tgz file only
- After selecting the file, click “generate” button

5.3 Download page

- Once the file upload done successfully then in the backend PDF generation processing will take place
- Then it will display a page with download button.
- By clicking on that button, the output pdf can be downloaded.

5.4 Saving the PDF

- By clicking the download icon, it can be saved in local storage from the server



6. SUMMARY

In this work developed a tool for automating alanis audit report generation. The work involves using a input file which must be of .tgz extension and same file structure to generate a pdf file. The conversion will not work if the input profile consists improper tags and characters present in the xml files. The older versions of OSP version profiles may not work as it may consists different file structure and data in it.