Tidal Enterprise Orchestrator: Using Web activities to Scrape Web Pages

Document ID: 113232

Contents

Introduction

This document provides information about how you can use Tidal Enterprise Orchestrator (TEO) to scrape web pages.

Prerequisites

Requirements

You should have knowledge of General Web technologies and TEO Process Authoring.

Components Used

The information in this document is based on Tidal Enterprise Orchestrator 2.1 or later.

The information in this document was created from the devices in a specific lab environment. All of the devices used in this document started with a cleared (default) configuration. If your network is live, make sure that you understand the potential impact of any command.

Conventions

Refer to Cisco Technical Tips Conventions for more information on document conventions.

Ways to web scrape

There are a few ways to web scrape a webpage. You can use either the Web HTTP Request activity or VBScript.

VBScript

An end user can use VBScript and the Execute Windows Script activity to scrape a website. You need to use the MSXML2.ServerXMLHTTP object. For more information, refer to www.microsoft.com or search for Web Scraping VBScript.

Using a TEO Activity

Complete these steps:

1. Drag and drop the Web HTTP Request activity to your process.
2. For 2.1.X, on the HTTP Request page, input the URL. Then, set the Method to **Get** and the Content Type to **text/html; charset=utf–8**.
3. For 2.2, on the HTTP Request, input the URL and leave the method as **Get**. Change the content type to **text/html;charset=utf–8** on the HTTP Headers tab. You can also change the output type on the Output Format tab.
4. If the webpage requires an integrated authentication, you can override the credentials tab to include this.

Related Information

- **Technical Support & Documentation – Cisco Systems**