Machine Learning

R Extensibility Instructions

Contents

Initial Steps 1

Repeatable Steps (for each data mart) 1

Repeatable Steps (for each destination entity) 2

Sample Seed Scripts 2

This document instructs the user how to inject their designated R scripts into the Catalyst loaders. It begins by installing the required SSIS package and defining new system level attributes. It continues by injecting the new SSIS package into the designated loader step. It concludes by defining the necessary variables for each destination entity.

Initial Steps

1. Establish local folder on ETL server.
   1. Configure permissions to allow EDW loader account to read, write, and execute
2. Install ExternalScriptExtensibility ISPAC on ETL server
   1. Locate inside folder \\SSISDB\CatalsytExtensibility\
   2. Verify permission to allow EDW loader account to execute
   3. Configure ExternalScriptExecution.dtsx parameter called StagingDirectory with the local folder established in step 1
3. Seed new attribute names into EDWAdmin.CatalystAdmin.AttributeBASE
   * + RInterpreterPath
     + ExternalScriptType
     + ExternalScriptSourceEntity
     + ExternalRScript
4. Seed new system attribute value into EDWAdmin.CatalystAdminObjectAttributeBASE
   * + RInterpreterPath: “C:\Program Files\R\R-3.3.1\Rscript.exe”

Repeatable Steps (for each data mart)

1. Seed EDWAdmin.CatalystAdmin.ETLEngineConfigurationBASE
   * + ExtensionPointNM: “OnPostStageToProdLoad”
     + DatamartID: {data mart ID}
     + SSISPackagePathOrSPToExecuteTXT: “\\SSISDB\CatalystExtensibility\ExternalScriptExtensibility\ExternalScriptExecution.dtsx”
     + IsSSISPackageFLG: 1
     + ExtensionOwnerNM: “Health Catalyst”
     + RunInAsyncModeFLG: 1
     + Use32BitRuntimeFLG: 0
     + ExecutionOrderNBR: 1
     + ActiveFLG: 1
     + RequiredParametersTXT: “BatchID, TableID”
     + FailsBatchFLG: 1

Repeatable Steps (for each destination entity)

1. Seed new destination entity attribute values into EDWAdmin.CatalystAdmin.ObjectAttributeBASE
   * + ExternalScriptType: “R”
     + ExternalScriptSourceEntity: “SAM.Sample.SampleEntity”
     + ExternalRScript: {content of R script file with qualified single quotes}
2. Configure dependency based on need

Sample Seed Scripts

**EDWAdmin.CatalystAdmin.AttributeBASE**

IF NOT EXISTS (SELECT \* FROM CatalystAdmin.AttributeBASE WHERE AttributeNM = 'RInterpreterPath')

INSERT INTO CatalystAdmin.AttributeBASE (AttributeNM, AttributeDSC) VALUES ('RInterpreterPath','Local path to RScript.exe')

IF NOT EXISTS (SELECT \* FROM CatalystAdmin.AttributeBASE WHERE AttributeNM = 'ExternalScriptType')

INSERT INTO CatalystAdmin.AttributeBASE (AttributeNM, AttributeDSC) VALUES ('ExternalScriptType','Python or R')

IF NOT EXISTS (SELECT \* FROM CatalystAdmin.AttributeBASE WHERE AttributeNM = 'ExternalScriptSourceEntity')

INSERT INTO CatalystAdmin.AttributeBASE (AttributeNM, AttributeDSC) VALUES ('ExternalScriptSourceEntity','Source entity from which to calculate predictions')

IF NOT EXISTS (SELECT \* FROM CatalystAdmin.AttributeBASE WHERE AttributeNM = 'ExternalRScript')

INSERT INTO CatalystAdmin.AttributeBASE (AttributeNM, AttributeDSC) VALUES ('ExternalRScript','HCRTools script that contains HC functions')

**EDWAdmin.CatalystAdmin.ETLEngineConfigurationBASE**

INSERT INTO CatalystAdmin.ETLEngineConfigurationBASE (ExtensionPointNM, DatamartID, SSISPackagePathOrSPToExecuteTXT, IsSSISPackageFLG, ExtensionOwnerNM, RunInAsynchModeFLG, Use32BitRuntimeFLG, ExecutionOrderNBR, ActiveFLG, RequiredParametersTXT, FailsBatchFLG)

VALUES ('OnPostStageToProdLoad',{SAM datamart ID},'\\SSISDB\CatalystExtensibility\ExternalScriptExtensibility\ExternalScriptExecution.dtsx',1,'Health Catalyst',1,0,1,1,'BatchID, TableID',1)

**EDWAdmin.CatalystAdmin.ObjectAttributeBASE**

INSERT INTO CatalystAdmin.ObjectAttributeBASE (ObjectID,ObjectTypeCD,AttributeNM,AttributeTypeCD,AttributeValueTXT)

VALUES (0,'System','RInterpreterPath','string','C:\Program Files\R\R-3.3.1\bin\Rscript.exe')

INSERT INTO CatalystAdmin.ObjectAttributeBASE (ObjectID,ObjectTypeCD,AttributeNM,AttributeTypeCD,AttributeValueTXT)

VALUES ([tableid],'Table','ExternalScriptType','string','R')

INSERT INTO CatalystAdmin.ObjectAttributeBASE (ObjectID,ObjectTypeCD,AttributeNM,AttributeTypeCD,AttributeValueTXT)

VALUES ([tableid],'Table','ExternalScriptSourceEntity','string','SAM.Sample.SampleSAMSampleBindingBASE')

INSERT INTO CatalystAdmin.ObjectAttributeBASE (ObjectID,ObjectTypeCD,AttributeNM,AttributeTypeCD,AttributeValueLongTXT)

VALUES ([tableid],'Table','ExternalRScript','longstring','[script]')