Self-service data onboarding with Pentaho Metadata Injection
Setup

Setup

Website

1. Download onboard.zip.
2. Extract to [Pentaho Install]/server/data-integration-server/tomcat/webapps/. This will create an onboard folder under webapps.
3. Edit onboard.jsp and update password to the Pentaho PostgreSQL database (set when installing Pentaho). E.g. **String password = "mypassword";**
4. Edit PostgreSQL URL port number, if not the default. You can check the ports at [Pentaho Install]\installatino-summary.txt. E.g. **String url = "jdbc:postgresql://localhost:5432/postgres";**
5. Edit upload path. E.g. **String filePath = "C:/Pentaho/server/data-integration-server/tomcat/webapps/onboard/data/";**
6. Go to **http://localhost:9080/onboard/index.jsp?customer=Acme** to verify the form appears. Note, the host and IP should match the Pentaho Data Integration server.
7. Edit onboard/js/onboard.js to use correct url host, port, and path to onboard/PDI/t_file2hdfs.ktr. Note: **%2f** is forward slash "/". For example:
   url = http://localhost:9080/pentaho-di/kettle/executeTrans/?trans=C:%2fPentaho%2fserver%2fdata-integration-server%2ftomcat%2fwebapps%2fonboard%2fPDI%2ft_file2hdfs.ktr
8. Edit username and password in onboard/js/onboard.js. For example:
   'Authorization': "Basic " + btoa("admin:password")

Metadata Database

The metadata database stores customer metadata. The steps below will create the table and load some sample data located in CSV files. The CSV files can be edited.

1. Open Pentaho Data Integration client (Spoon)
2. Files for creating the metadata database are under onboard/setup/
3. Open j_create_metadata_db.kjb
4. Click on View tab
5. Expand Jobs folder, j_create_metadata_db, Data connections
6. Right-click on PentahoPGSQL and click Share (if not already)
7. Right-click on PentahoPGSQL and click Edit
8. Update password (same as above)

9. Update port, if not the default. You can check the ports at [Pentaho Install]\installatino-summary.txt
10. Click OK to save
11. Go to Action menu and click Run
12. Click Launch
13. Verify that the job completed successfully
    a. Success job entry should have a green check mark.
    b. Open k_browse_metadata.ktr
    c. Go to Action menu and click Run
    d. Click Launch
    e. Click on the Browse Metadata step
    f. Click Preview Data tab and verify results

Configure Transformations

Set data path:

- Open onboard/t_file2hdfs.ktr
- Right-click on the canvas and choose Properties
- Click Parameters tab
- Update FILE_DIR Default Value to your onboard/data path
- Update Calculate paths step with hdfs_path

Testing

1. Go to http://localhost:9080/onboard/index.jsp?customer=Acme
2. Enter a different customer name in the URL (please no spaces or special characters)
3. Fill in form. Sample files are in onboard/sample_data/
4. Click GO.
5. Verify the file you uploaded appears in onboard\[file_type]_[customer_name].csv
6. Run the k_browse_metadata.ktr (see steps above) and verify the form data appears.
7. Verify data appears in target. Check hue /metadata_injection/[file_type]/

Additional To Do's:

- Create a sample file for uploading employee data. The key things to show are:
    o Source field names differ from target
    o Source field order differs from target
    o Source file has extra fields that are not in the target