

# Setting up the Oil & Gas Governance Demo

October 2017: v1.4

Send feedback to: [neil.williams@us.ibm.com](mailto:neil.williams@us.ibm.com)

## 1. Provision and prep the IIS PoT image

This demo was created using both the IIS v11.5 PoT image and CAP. Instructions are verbose to help those new to IIS.

Information on using CAP is available at: <https://ibm.ent.box.com/s/98yno9z7kgakwxnbpap30o7unqspigs>

Tips for using CAP are in sections 3 and 4.

Copy and extract the demopackage folder to your IIS client and server images. Tips for copying the demopackage folder to the IIS images are in sections 3 and 4.

## 2. Choose full import or manual asset creation

Full IMAM, DS/QS (dstage1 project), and IGC export files are available. The IMAM export will overwrite existing assets where there's a conflict. **Follow the instructions in this section if you're ok with possibly overwriting your existing assets, otherwise skip to the section 3 for manual instructions.** Alternatively take a snapshot or backup of your image before proceeding with the imports in this section.

Export files are in the demopackage\full\_istool\_exports directory.

Before performing the imports below, do the following:

1. Create the oil\_db database and copy the PDF files as documented in section 3.
2. Disable workflow in IGC as documented in section 6.

Use istool to run the following import commands. Adjust file path as necessary [note quotation marks may not copy/paste well]. A tip on launching istool from CAP1 is in section 6.

```
import -dom is-server.ibm.com:9445 -username isadmin -password inf0server -archive  
"C:\demopackage\full_istool_exports\oilandgasall.isx" -all
```

```
import -dom is-server.ibm.com:9445 -username isadmin -password inf0server -archive  
"C:\demopackage\full_istool_exports\oilandgasdstage1.isx" -ds 'is-server.ibm.com/dstage1'
```

```
glossary import -dom is-server.ibm.com:9445 -username isadmin -password inf0server -  
filename "C:\demopackage\full_istool_exports\oilandgasigc.xml"
```

Post-import:

1. The IGC Data Connection Mapping was not imported. Instructions for creating the mapping are in section 8 below. The mapping is required to show lineage.
2. Data stewards were apparently created but not added to any assets by the import. You may want to add them to assets for your demo, the Permit term at the very least. Valid stewards are Pete Hornberger and Jack Donaghy.
3. The "Permit Number Exists" policy was imported but the import did not assign associated data rules. See section 13 to assign the rules to the policy.
4. Proceed to sections 15 and 16 for setting up Data Connect and Watson Analytics.
5. Demo setup is now complete. Refer to the Demo Script file for instructions on presenting the demo.

For reference, here are the export commands used to create the export files:

```
export -dom is-server.ibm.com:9445 -username isadmin -password inf0server -archive  
"C:\oilandgasall.isx" -all
```

```
export -dom is-server.ibm.com:9445 -username isadmin -password inf0server -archive  
"C:\oilandgasdstage1.isx" -ds 'is-server.ibm.com/dstage1/*/*.*'
```

```
glossary export -dom is-server.ibm.com:9445 -username isadmin -password inf0server -  
filename "C:\oilandgasigc.xml" -allcategories -allpoliciesrules -includeassignedassets -  
includestewardship -includelabeledassets -includeassetcollections -includeTermHistory
```

### 3. Prepare the IIS Server

#### **Create and import the database**

Use the IIS server image (e.g. CAP2) to import the database. The dbexport folder was created with the “db2move OIL\_DW export” command and contains everything necessary to create the source database (OIL\_DW) in DB2.

Log in to the IIS server as db2inst1 and uncompress then copy the demopackage folder.

Hints for the CAP2 image:

- You will need to use the capadmin/C@Pdemo4U! user for WinSCP, Putty, and Filezilla.
- Use WinSCP or Filezilla (available on CAP1 if you don’t have it on your laptop) to upload the folder to /tmp on CAP2. If using Filezilla, use the file manager feature to set protocol to sftp (quickconnect defaults to regular ftp).
- Use putty (available on CAP1) to log in as capadmin, then use `sudo su - db2inst1 -s /bin/bash` to switch to the db2inst1 user.
- Make a copy of the dbexport folder as user db2inst1 to avoid issues with permissions on the files, e.g. “`cp -R dbexport mydbexport`”.

Move to the dbexport directory in the demopackage folder and execute the create\_oildb.sh script using: `./create_oil_dw.sh`

The command should only take a couple minutes to run and should output messages telling how many rows were inserted (your prompt may initially appear to be frozen but it’s just silently creating the database and doing the import). To verify the database was created and loaded correctly, run the following from the command line. You should see a count of 3716 rows:

```
db2 connect to OIL_DW
```

```
db2 "select count(*) from WELLS_DATA"
```

If you need to manually create and load the tables (e.g. using a non-DB2 database) the table DDL and csv files are in the dbfiles folder in DB2 syntax/format. You will need to create the database (OIL\_DW) itself manually.

If you encounter any access permission issues with the database/tables, try assigning the permissions in the grant\_db2\_permissions.txt file in the dbfiles folder but this should not be necessary.

#### **Copy PDF files to the WAS directory**

Log in to the IIS server as root and copy blm\_gold\_book.pdf and eia\_pdr\_full.pdf from the demopackage folder to this directory: `/opt/IBM/WebSphere/AppServer/profiles/InfoSphere/installedApps/is-serverNode01Cell/igc-web.ear/igc-web.war/`

On the CAP2 image, you will need to use sudo to copy the file as root, e.g.:

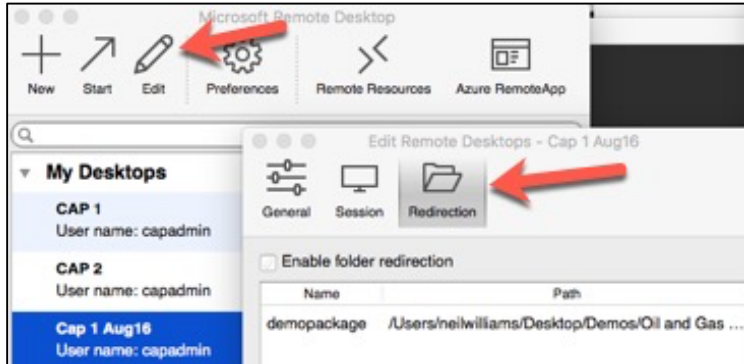
```
sudo cp *.pdf /opt/IBM/WebSphere/AppServer/profiles/InfoSphere/installedApps/is-  
serverNode01Cell/igc-web.ear/igc-web.war/
```

## 4. Login information

I used isadmin/inf0server for all IIS logins, except for the database connection in IMAM where I used db2inst1/inf0server.

Hints for CAP users:

- Use RDP to connect to CAP1
  - Mac users can download “Microsoft RDP” from the App Store
  - Mac users: put the RDP port after the IP address in RDP setup, e.g. 169.45.105.204:3333
  - Mac users: Copy/paste from host to RDP session isn’t supported. You can map a local folder to your CAP1 image via RDP folder redirection instead.



- OS username/password should be in your CAP confirmation email, e.g. capadmin / C@Pdemo4U!

## 5. Import the metadata via IMAM

Once your database is ready to go, import and share the metadata from the OIL\_DW database via IMAM. Use user db2inst1/inf0server in the connection parameters. The host is IS-SERVER.IBM.COM, the database server is DB2, and the database name is OIL\_DW.

### Create New Import Area

Name the import area, select a metadata interchange server, and select a bridge or connector.

\* Import area name:

Import area description:

\* Metadata interchange server:

Select a Bridge or Connector

- IBM Informix bridge
- IBM InfoSphere CSV
- IBM InfoSphere DB2 Connector**
- IBM InfoSphere Master Data Management

**Import Help**

**About this connector**

The IBM DB2 connector imports including database schemas and versions of IBM DB2.

For more information on the IBM metadata.

### New Data Connection

\* Name:

Description:

\* Database/Location:

Enter credentials if they are required for the connection:

User name:

### Create New Import Area

Enter values for identity parameters.

▼ Identity Parameters

Host system name:

IS-SERVER.IBM.COM

Database name:

OIL\_DW

DBMS server instance name:

### Create New Import Area

- ✓ Import area 'Oil and Gas demo' created.
- ✓ Staging area import complete.

Click OK to view the results.

For the initial import you will need to **Preview** the import then **Share to Repository**. Note the number of assets in my import below as a sanity check.

Import Areas > Oil and Gas demo

Close

Overview Staged Imports Shared Imports

Oil and Gas demo 001

Analyze Preview Share to Repository Reimport... More Actions

▼ Summary

Description:

initial import [Edit Description](#)

Imported:

2017-07-20 at 16:33:30 by isadmin [\(Download Import Log\)](#)

Analyzed:

2017-07-20 at 16:35:57 by isadmin

Previewed:

Not previewed

Shared:

Not shared

▼ Statistics

Assets	Total	Unique	Duplicate Sets	Invalid Identities
All	54	54	0	0
Candidate key	1	1	0	0
Database	1	1	0	0
Database column	43	43	0	0

Imported Assets

▼ Data connection

oil\_and\_gas

▼ Host

IS-SERVER.IBM.COM

Import Areas > Oil and Gas demo > View Share Preview: Oil and Gas demo 001

Share to Repository Refresh Close

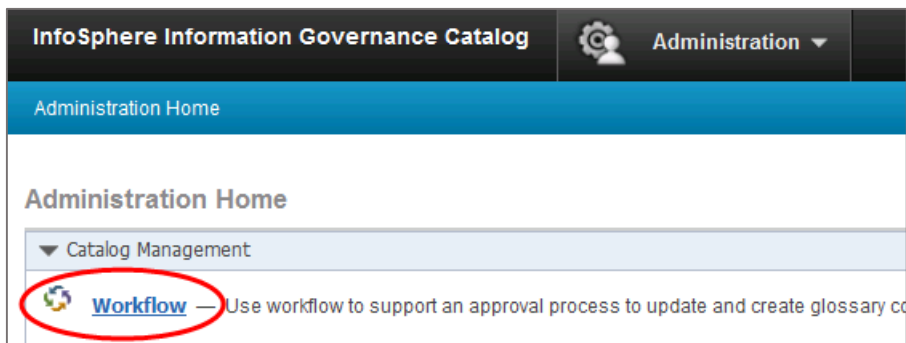
Share preview created 2017-07-20 at 16:39:29 by isadmin. Refresh old previews before st

Statistics

Asset Types	Total	Created	Merged	Deleted
All	54	53	1	0
Candidate key	1	1	0	0

## 6. IStool IGC Import

Before importing, I recommend turning off workflow. If workflow is enabled, assets will be imported in draft state.

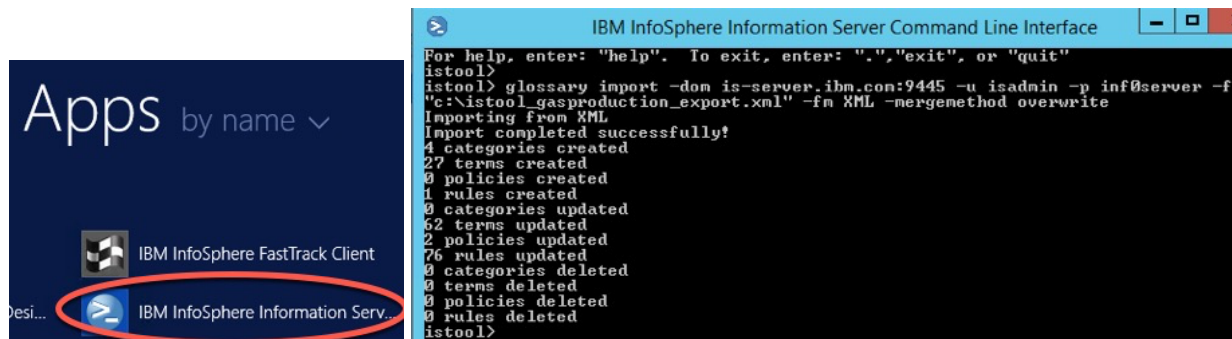


A full istool IGC export is in the file `istool_gasproduction_export.xml`. Copy the file to `c:\` on your client (e.g. CAP1). For CAP1 I mapped my local `c:\` drive to the image and copied the entire (uncompressed) demopackage folder.

Import `istool_gasproduction_export.xml` with istool and the command:

```
glossary import -dom is-server.ibm.com:9445 -u isadmin -p inf0server -f
"c:\istool_gasproduction_export.xml" -fm XML -mergemethod overwrite
```

CAP users: launch an istool command window from the programs list.



If you have any issues with the istool import, there is an alternate import file called `business-glossary-xml-export-nov-15.xml` which you can import via IGC under Administration -> Tools -> Import.

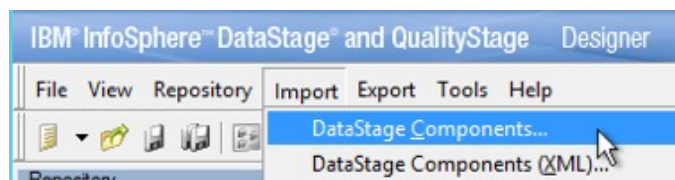
## 7. IGC Collection

The import did not create my collection, so create one for your demo with a name like "Well Assets". Assign some of the demo terms and columns to the collection.

## 8. Create lineage in DataStage

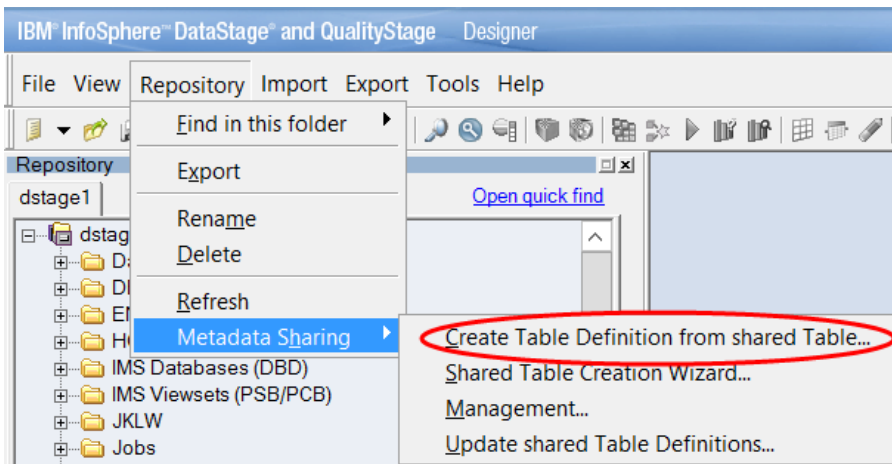
In DataStage you can import my project or create your own.

My DS project export is in the `dsexport` folder and can be imported directly from DataStage using the Import -> DataStage Components menu, then browsing to the `dsproject.dsx` file.

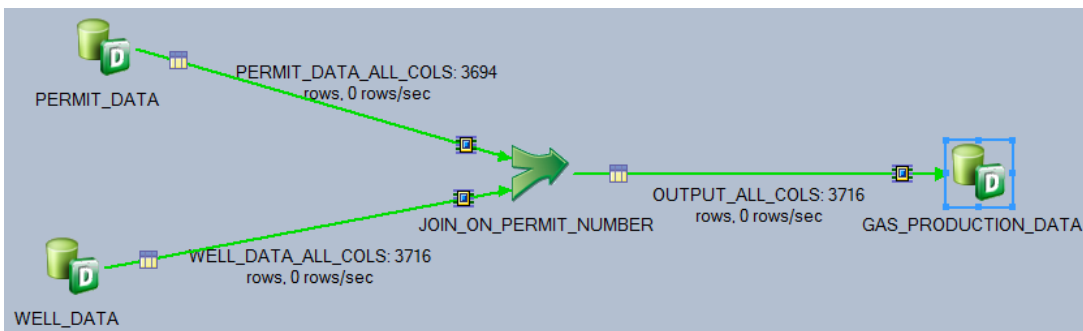


Once the import is complete you should have a project called `OIL_DW`.

To create your own project, you will need to import the shared table metadata and create/run a job.



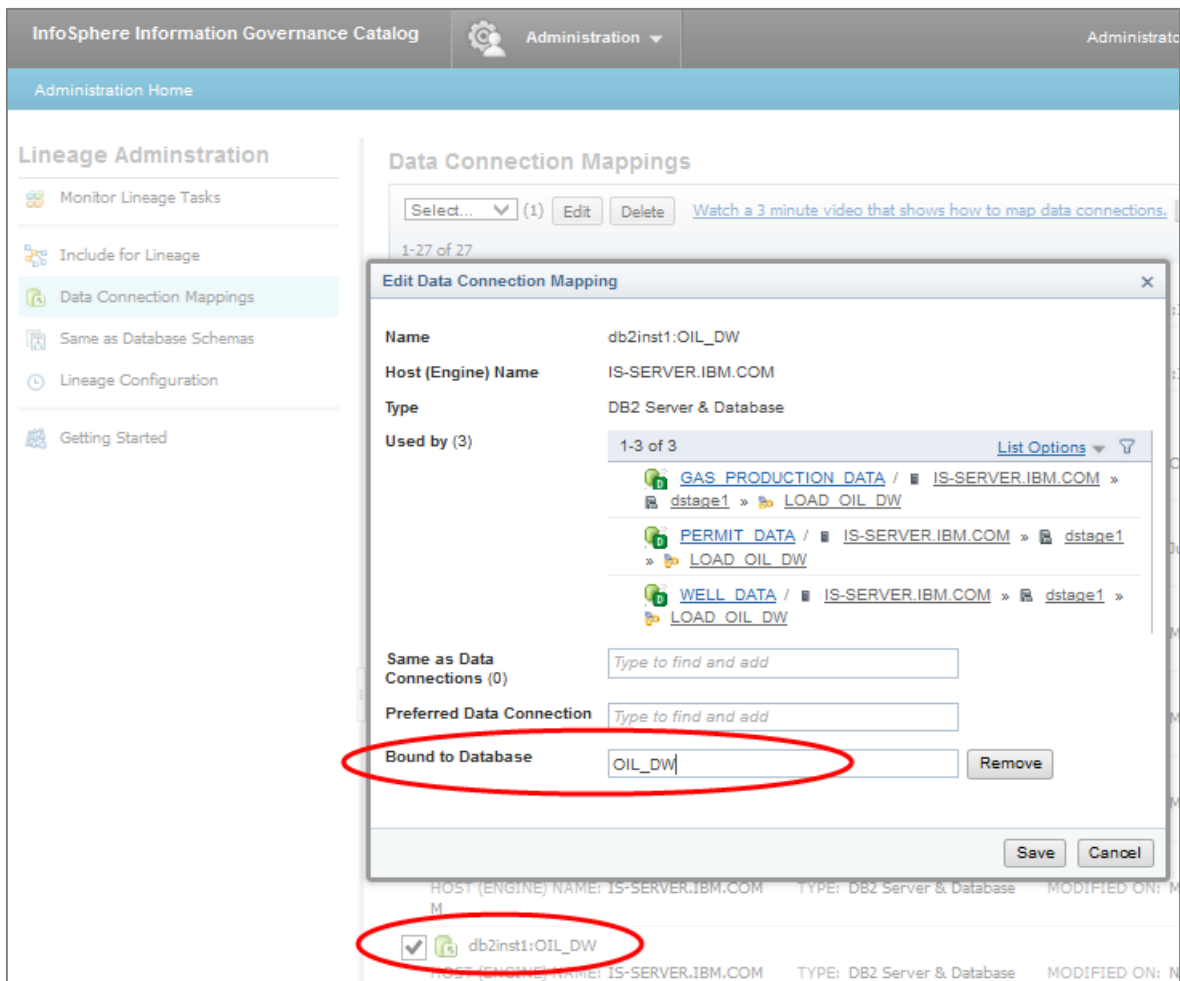
Create and run a DS job using your shared table definitions. I created a simple job joining `PERMIT_DATA` and `WELL_DATA` on `PERMIT_NUMBER` and writing the output to the `GAS_PRODUCTION_DATA` table. I used obvious names for the stages so they would make sense for the business analysts during the demo.



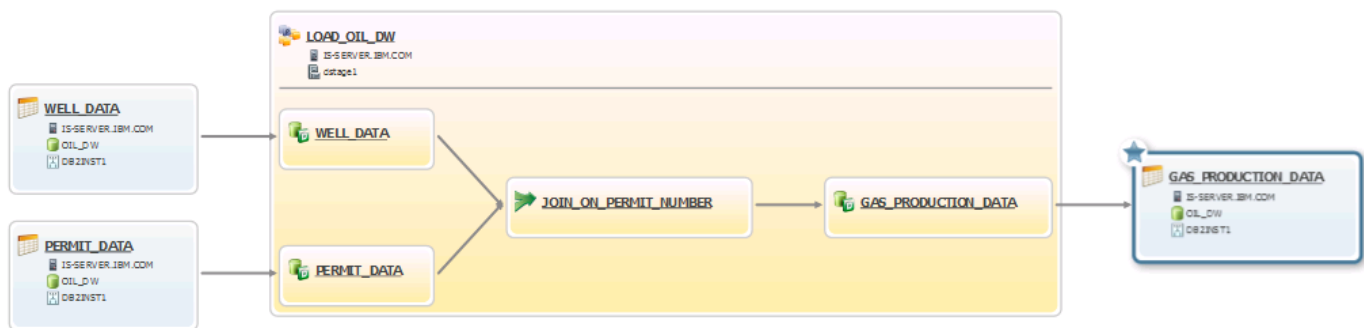
## 9. Map lineage in IGC

In order for lineage from your DS job to show in IGC you need to link the `OIL_DW` database in IGC to the `OIL_DW` connection (the connection won't show up in IGC until the DataStage job is run successfully or the project is imported). To do this in IGC go to: Administration -> Lineage Administration -> Data Connection Mappings. Edit the `db2inst1:OIL_DW` connection and enter `OIL_DW` in the "Bound to Database" field.





After doing the mapping, run lineage for the GAS\_PRODUCTION\_DATA table (you can find it assigned to the *Field* term) and you should have the lineage pictured below.



## 10. IGC PDF/document links in terms


Copy the blm\_gold\_book.pdf and eia\_pdr\_full.pdf files to your client desktop and give them simple names, e.g. “BLM Gold Book”.

Check the *Permit* term in IGC to see if you need to link the PDF’s you previously copied to the WAS directory on the IIS server. You should have copied them to this directory in an earlier step:

/opt/IBM/WebSphere/AppServer/profiles/InfoSphere/installedApps/is-serverNode01Cell/igc-web.ear/igc-web.war




The links in the terms should have been imported but if they didn't or you need to edit them, use the format below and link to any term you feel appropriate for the demo. I linked to *field*, *permit*, and *well*.


Link	<input type="text"/>
Link to more information on the Term	<input type="text" value="[https://is-server.ibm.com:9445/ibm/iis/igc/eia_dpr_full.pdf US Energy Information Administration Drilling Productivity Report 2016]"/>
Mainframe Data Source	<input type="text"/>

 **Permit**


The permit to drill gives an operator approval to begin the process of drilling a well.

**Long Description** The permit to drill gives an operator approval to begin the process of drilling a well. Before a Permit is granted, there are several approvals (including environmental compliance) that must be in place. There are several different types of permits. A new well involves [... Show More](#)

**Parent Category**  [Gas Production](#) »  [Field](#) »  [Permit](#)

**Stewards (1)**  [Data Analyst Pete Hornberger](#)

**Status** Accepted

**Governed by Rules (1)**  [Permit Number Exists](#)

**General Information**

**Extract data using Bluemix Data Connect** <https://console.ng.bluemix.net/>

**Link to more information on the Term** [BLM Gold Book 4th Edition](#)

If your browser doesn't display the pdf, you may need to edit your browser settings to show PDF files in-line.




## 11. Create custom data class in IGC




My custom class did not come through in the import, so you should create it manually. Create a valid-values class in IGC called *Fracking Fluid* using the values from the TRT\_FLUID\_TYPE column in the WELL\_DATA table. The values are FF\_TYPE1D, FF\_TYPE2D, FLUD, FLUD\_VAR3F, GEL, SLKW, U, X-LINKGEL.

Assign the Fracking Fluid data class to the TRT\_FLUID\_TYPE column in both WELL\_DATA and GAS\_PRODUCTION\_DATA tables. Also assign it to a few terms that will be easy to find in your demo, such as Well and Fracking Fluid. Be sure to check *Enabled: True* when defining your class.

See the image below for the links to define a custom class.

**n Governance Catalog**  
Information Assets ▾ Labels Queries Collections ▾ Data Integration ▾  
 [Browse](#)

**Browse Hierarchies**  
 Implemented Data Resources  
 Logical Data Models  
 Physical Data Models

**Manage**  
 Create Extension Mapping Document...  
 **Create Data Class...**  
 Import Extension Mapping Documents...

**Create Data Class**  
**NAME \***  
  
**PARENT DATA CLASS**  
  
**SHORT DESCRIPTION**  
  
**TYPE \***  
  
**CLASS CODE \***



Data Class Details

Edit
Delete
Add to Collection...

\*!

Fracking Fluid

Type of fluid used in hydraulic fracturing (aka "fracking")

Long Description

Data class encompassing all types of hydraulic fracturing fluid used. Must be a member of the "valid values" list below.

Type

Valid Values

Stewards (1)

Data Analyst Pete Hornberger

Assigned to Terms (4)

Fracking Fluid / Gas Production » Field » Well

Hydraulic fracturing fluid / Gas Production » Field » Well

Hydraulic Treatment Fluid Type (Fracking Fluid) / Gas Production » Field » Well

Well / Gas Production » Field » Well

Contains Data Classes

Data Classifications (5)

Selected Data Classifications (2)

TRT\_FLUID\_TYPE / IS-SERVER.IBM.COM » OIL\_DW » DB2INST1 » WELL\_DATA

TRT\_FLUID\_TYPE / IS-SERVER.IBM.COM » OIL\_DW » DB2INST1 » GAS\_PRODUCTION\_DATA

Detected Data Classifications (3)

TRT\_FLUID\_TYPE

TRT\_FLUID\_TYPE

TRT\_FLUID\_TYPE

Definition

Valid Values

FF\_TYPE1D

FLUD

FRAC

GEL

## 12. Import/create datarule in IA

Copy the file `ia_oil_dw_export.isx` to your client image (e.g. CAP1) in the folder `c:\IBM\InformationServer\Clients\`

Use an istool command window to import with the command below. **Note that the quotes at the end of the command are two single quotes, not one double-quote.** Copy and paste if in doubt.

```
istool import -dom is-server.ibm.com:9445 -u isadmin -p inf0server -ar
ia_oil_dw_export.isx -ia ''
```

If you get an authorization error when running column analysis, granting the DB2 database and table permissions in the `grant_db2_permissions.txt` file in the `demopackage/dbfiles` folder may fix the issue.

Run column analyses on the `OIL_DW` tables to create data quality scores and column classification for your “Fracking Fluid” custom class. Verify that the `trt_fluid_type` column is correctly assigned to the `Fracking Fluid` data class.

WELL\_DATA

View Analysis Summary

View Details

View the frequency distribution, data classes, properties, domain and completeness information, and formats for the column.

Select View:

TRT\_FLUID\_TYPE

Overview Frequency Distribution Data Class Properties Domain & Completeness Format

View the properties of Data Classes by Value. A data class categorizes the column. You can select a new data class if you class.

Inferred: Fracking Fluid

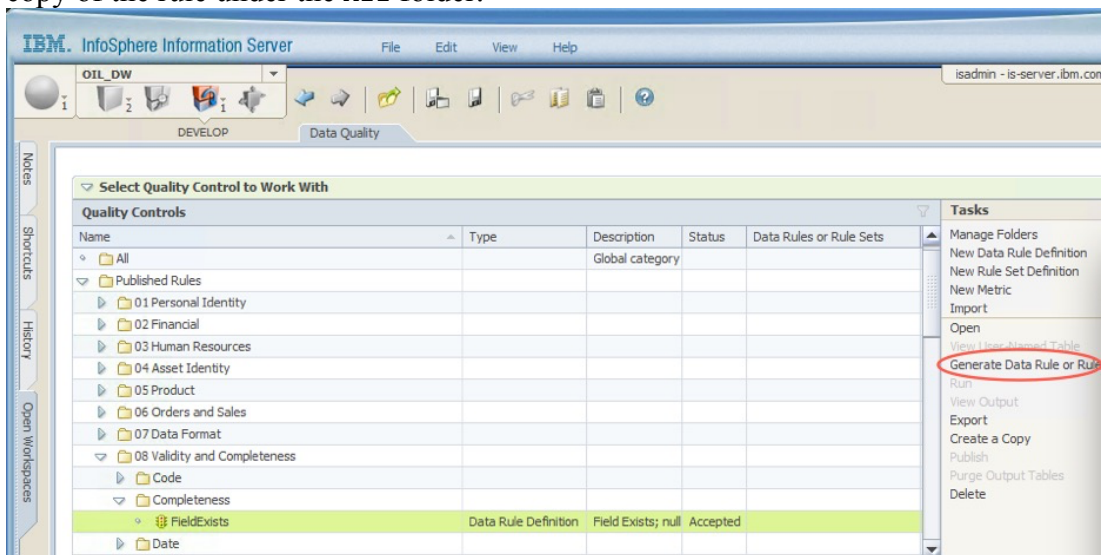
Related Terms: Hydraulic Treatment Fluid Type (Fracking Fluid)

Selected: \* Fracking Fluid

Personal Information: Yes

Manual instructions for creating the project and rules are below if you have any issues with the import.

1. Create a project call OIL\_DW using the data source OIL\_DW on host IS-SERVER.IBM.COM
2. Create two identical data rules for the PERMIT\_NUMBER column in the WELL\_DATA and PERMIT\_DATA tables using the FieldExists global rule definition (found under published rules > 08 validity and completeness > completeness). When you click “Generate Data Rule...” with FieldExists highlighted, you be prompted to create a copy. Click ok, then you will see your copy of the rule under the All folder.



3. Click your copy of the FieldExists rule under the All folder, then click Generate Data Rule...
4. Give your data rule a useful name the business users will understand during your demo, like PermitNumberExistsPermitData. Add a data steward and owner, and set the status to Accepted. Under the Bindings and Output tab, select the PERMIT\_NUMBER column of the

PERMIT\_DATA table as the binding for the rule.

The screenshot shows the 'Bindings And Output' tab. On the left, the 'Select View' menu has 'Bindings' selected. The 'Rule Logic Variables' table is as follows:

Name	Rule Definition	Data Type	Design Binding	Implemented Binding
field1	FieldExists	Any		PERMIT_DATA.PERMIT_NUMBER

Below the table is a text box containing 'Field1 EXISTS' and a 'Discover Bindings' button. On the right, the 'Implemented Data Resources' panel shows a tree structure. The 'PERMIT\_NUMBER' resource is highlighted under the 'OIL\_DW' folder. At the bottom of this panel, the 'Set as Bindings' button is circled in red.

5. Validate, Save and Close your rule. Repeat the above steps to create an identical rule on the WELL\_DATA table. Run your rules when finished, and view the output, which should show 100% met.
6. Run column analyses on the OIL\_DW tables to create data quality scores and column classification for your “Fracking Fluid” custom class. Verify that the trt\_fluid\_type column is correctly assigned to the Fracking Fluid data class.

WELL_DATA: (1 of 17 columns)							
	Name	Sequence	Record	Definition	Cardinality Percent	Data Class	
	TRT_FLUID_TYPE	12	3716		0.21528525	Inferred	Fracking Fluid

### 13. Assign the data rules to a governance rule in IGC

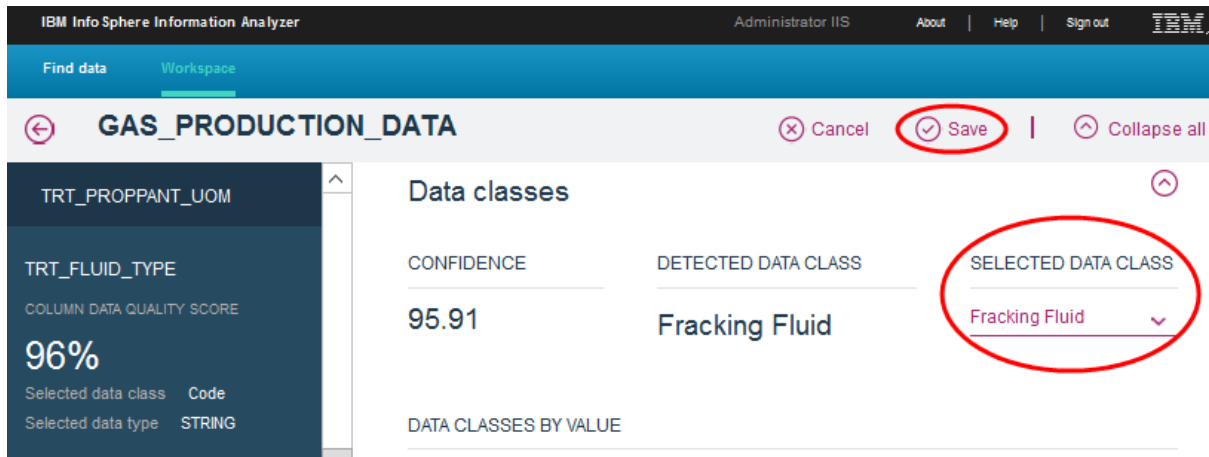
Once your data rules are ready, link them to the governance rule “Permit Number Exists” in IGC. The governance rule should have been imported; you can find it under the Permit, Field, and Well terms. If you don’t have it, then create it and bind to the terms: Field, Permit, Permit Number, and Well. Also bind to technical assets such as the PERMIT\_NUMBER column in the WELL\_DATA and PERMIT\_DATA tables, along with the tables themselves.

The screenshot shows the 'Permit Number Exists – Edit Information Governance R...' dialog box. The 'Implemented By' section is expanded, showing a list of assets. The 'PermitNumberExistsWellTable' and 'PermitNumberExists' assets are selected. The 'Implemented by Assets' section shows a list of assets with checkboxes. The 'PermitNumberExistsWellTable' and 'PermitNumberExists' assets are selected. The 'Implemented by External Assets' section is empty.

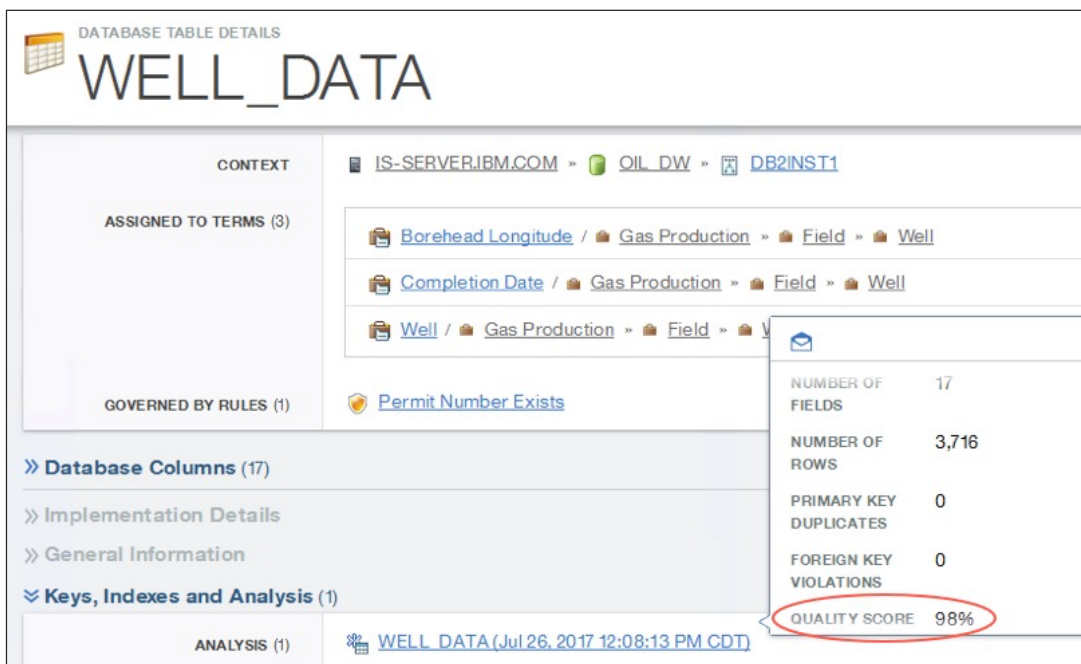
### 14. Analyze and verify tables in IA thin client

In the IA thin client open the OIL\_DW workspace (or whatever you named your project in IA) and analyze/publish all the tables. Be sure to analyze and publish the GAS\_PRODUCTION\_DATA table.

You may need to manually set the custom “Fracking Fluid” data class in IA. Save and analyze/publish after your changes.



Once you’ve analyzed/published the tables in the IA thin client, the analysis results should be available in IGC:



## 15. Configure Bluemix Data Connect

If you don’t have a Bluemix or Watson Analytics account, you will need:

1. An IBM ID (*NOT* your W3 ID)
  - a. If you do not have an IBM ID, create one at the IBM ID registration webpage: <https://www.ibm.com/account/profile/us?page=reg>
2. Your IBM ID to be linked to your W3 ID
  - a. *This step may no longer be necessary*
  - b. If having problems with signing on, you can link them at the Central Sign On webpage: <https://w3-03.sso.ibm.com/tools/cso/index.jsp>
3. An IBM Bluemix account (<https://console.ng.bluemix.net/>)
  - a. If you have not used Bluemix, first attempt to LOG IN with your IBM ID (*NOT* your W3 ID)
  - b. If unsuccessful, then you must SIGN UP
4. An IBM Watson Analytics account (<https://watson.analytics.ibmcloud.com/>)

- a. Again, attempt to LOG IN with your IBM ID
- b. Otherwise, you must TRY IT FOR FREE

If you're new to Data Connect, I recommend working through Rick Buglio's lab at: <https://w3-connections.ibm.com/files/app#/file/e3bcbcff-2f1e-40ea-aec9-6a14f8cd97a2>

Also see Rick's Connections page: <https://w3-connections.ibm.com/profiles/html/profileView.do?userid=164578c0-5775-102b-8e72-f4f7902a5c09>

For the demo, I show Bluemix Data Connect for filtering and joining the data. You could configure a Data Connect connection directly to your db2 database in the IIS image. Instead, I just loaded the `PERMIT_DATA` and `WELL_DATA` tables in dashDB and used that for the Data Connect part of the demo. Use the files in the `demopackage/dbfiles` folder to create and load your dashDB database if you want to do the same.

Once you have dashDB and/or Data Connect configured, join `WELL_DATA` and `PERMIT_DATA` and output the result to Watson Analytics. Change the column names in Data Connect to more understandable English terms to facilitate the Watson part of the demo. A description of the columns is in the 2<sup>nd</sup> tab of the spreadsheet "Anonymous Gas Production Data v2.xlsx" (spreadsheet and data courtesy of Shad Griffin).

Create an identical job using the "Data Flow" interface of Data Connect to show customers the alternate interface.

## 16. Verify Watson Analytics

Log on to Watson Analytics and verify that your dataset works. Your dataset will be in the *Personal* folder.

For the Watson demo, I start off by typing in the question "first 12 months gas by frac fluid type" and choose the "Packed Bubble" chart.

