Iowa DOT FME Best Practices Guide

Public Version 1.0 | 06/2019

SANITIZED PUBLIC VERSION

Contents

Guide created by David Runneals, GIS Developer at HNTB, for the Iowa DOT using a combination of existing knowledge and workflows of FME at the Iowa DOT, FME UC presentations, FME knowledge forum posts by the greater FME community, and other FME documentation. For additional tips and tricks on improving performance and other best practices, check out the following links: Performance Tuning FME, Layouting Your Workspace, Best Practice Validation Project, FME Desktop Training Best Practices, Code Smell and FME Golf, and Advanced Tricks when using Bookmarks.

Notes

FME does NOT currently support SAML authentication with Portal reader/writer, however it is in the works for FME 2020. This means that you will NOT be able to use your Azure AD account account to access ArcGIS Portal. A suggested workaround is to write to a file geodatabase and then upload that.

For this public version, I have highlighted and set generalized specific references to the following (which you can replace with your own):

- FME Production server url: yourprodfmeserver.gov
- FME Test server url: yourtestfmeserver.gov
- Generic UNC paths: \\yourfmeserver\
- Generic FME user account that runs your FME server: yourfmeserveruser
- Additional generalized references do exist and all of them are highlighted in magenta

General Tips

- Use the newest version of FME possible EXCEPT when publishing to FME server when you should use the version that matches FME server.

  By using the latest version of FME, some previous issues may have been resolved and performance will likely increase due to evolving transformers and modifications to the software.

- Look at the FME Desktop Training Best Practices (which has been partially included in this document)
- Determine if you require either an all or nothing approach (if a record fails, than don't write any data at all) or if invalid features can be filtered out.
- Implement validation checks for error trapping, so workspaces won't fail. Log features (if logs will be reviewed regularly) or send email notification alerts to the data owner, so that they can fix it.
- Keep both test and production writers in the workbench, but just disable the ones that aren't needed before uploading to server.
- If your workbench file is excessively large, you may have to copy over everything to a new file to reduce the size and may resolve issues with corrupted workspaces.

Workspace Organization

- Keep it Simple Silly! Keep it clean and understandable!

  You should be able to show your workspace to a co-worker and have them understand what is occurring. Reduce "Code Smell". **GOAL: Have a co-worker be able to understand and troubleshoot an issue in the workspace in less than an hour.**

- Use bookmarks to group transformers together that distinctly define different sections, components, or sub-tasks

Using bookmarks allows you to collapse complicated sub-tasks into a smaller footprint. A recommended best practice is having between 5-10 transformers per bookmark. Don't over use bookmarks!

Label bookmarks with a short and sweet description (name) that describes what is happening within the bookmark.

Appropriately color code your bookmark (see the color coding sub-section below)

Bookmarks can be SUPER useful when used appropriately. See some advanced tricks when using bookmarks here and bookmark documentation here.

- Rename transformers to give a 1-2 word description of what you're doing.

   Convert a transformer with a name like "AttributeManager_6" to "AM_DescriptionHere" or "DescriptionHere". Don't forget to use CamelCase.

- Add annotations!

   Use annotations broadly to describe functionality where it happens or attach it to a transformer to describe in detail what is happening. A best practice is having one annotation for every 3-5 transformers.

   Appropriately color code your annotation (see the color coding sub-section below)

- Use junctions, tunnels, and hidden connections where appropriate.

   Use junctions, tunnels, and hidden connections to help keep your workspace clean and understandable. More information on junctions, tunnels, and hidden connections can be found here. Best practices in using these can be found here. If it makes it more complicated to understand, don't use them! Make sure to label appropriately when using them!

- Ensure that ALL metadata in the workspace parameters completed!

   See the sub-section below on required metadata. See the workspaces parameters documentation for additional guidance.

- Use consistent object layout

   Object layout is a personal preference but keep it consistent! More information on object layout can be found here.

Bookmarks/Annotations Color Coding

- Doesn't work/Needs improvement/To-Do (red)
- In progress (yellow)
- Completed & Working (green)
- Notes (orange)
- Summary Annotation (blue)

REQUIRED Metadata

- Workspace Name: (NOTE this is NOT the file name, but the name in the workspace parameters.) Use the standard naming conventions.

  WorkspaceName_Dev(01-99)(A-Z) when developing or testing workspaces. For example, if you were updating the production workspace DBtoAGOL_Prod04 and this was your first update it would be named DBtoAGOL_Dev04A)

  WorkspaceName_Prod(01-99) for workspaces that have been finalized and meet or exceed all documented requirements and best practices.

- Overview: Describe what the workspace does. Where does the data come from and where does it go to? What is it for/why is it needed? Make a note if the python version requires a specific version to run.
- Help: Include a point of contact for someone that supports the workspace and tips to help with common issues (if any occur).
- History: When finalizing a workspace for production use by others, you MUST include your name, date, and a detailed description of what you changed since the last time the workspace was production quality. This is also useful to keep track of minor changes throughout the development lifecycle, so you know what all you've done!

Readers

General Tips

- Use the newest readers

  Ensure that your readers match the latest version of FME (or match the version that is on FME server if uploading there). The great developers over at SAFE constantly refine and update readers and using the latest ones ensures that you get the best performance out of your workspace.

- Use as few readers as possible

  If you have multiple tables that are coming from the same database with the same user, use 1 reader with multiple feature types. If you are optimizing existing workspaces, pick a reader and import the tables to that reader and remove the other ones. This helps make updating workspaces easier and makes it cleaner! Note that you will still need multiple readers if you are using transformers such as the FeatureMerger where you set suppliers first.

- Disable feature types that aren't being used

  When feature types are left enabled on workspaces and not required, it creates clutter on your canvas and reduces performance because the feature type data is still read in.

- Use Database/Web Connections where possible

  When you use a connection instead of embedding the connection parameters, you can easily create and update a single connection in the future for all your workspaces instead of having to embed the parameters to each location that requires it. See the section below on Web/Database Connections for best practices.

  If reading from a database, and you use OS authentication, ensure that yourfmeserveruser users have access to the database prior to uploading your workspace to FME server.

- Expose only attributes that you need!

  If looking to speed up existing workspaces, the easiest and quickest thing to do is to define attributes on your readers. This article gets into detail and compares methods of filtering data, but an easy way to think of it is that each of your attributes is a weight and your feature is a horse. The more weight (attributes) you put on your horse (feature) the slower it (workspace) will run.
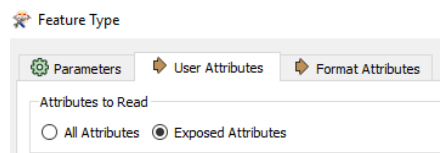
- Expose only features that you need!

Similar to the previous point, the less records you read in, the quicker your workspace will run! When possible, use the WHERE or SELECT parameters to define features prior to reading them in. Also use the bounding box search if you only need a specific geographic area.

- When developing, set the "Max Features to Read" parameter for each of your readers.

    When developing and testing, a good idea is to set the Max Features to Read parameter. This is super helpful when using large datasets, but you only want to verify your process is working with a smaller subset before running them all through. **Ensure that you REMOVE this prior to publishing to FME server!**

- Enable Feature Caching on FME Desktop

    When developing and testing, a good idea is to enable feature caching. It helps speed up your reading and writing by writing each transformers output to a table and allows you to easily visualize the data in inspector. It also speeds up the workspace by only running from the point in your workspace where changes have been made, so you don't have to constantly re-run the first half of your workspace that you didn't change.

- Attempt to have all value-added data (i.e. route ID, reference post, etc) in the source dataset.

    Having value-added data such as route ID or reference post in the source dataset saves a LOT of time and improves accuracy of the data. A great example of this is a workspace that spent 20+ minutes getting route ID and reference posts, when it could be added to the original dataset, and only had to be ran once per feature. Think smarter, not harder.

ArcGIS Server Feature Service

When trying to connect to authenticated REST services, you will need to set the authentication to "NTLM" and then for username enter yourorganizationaddomain\USERNAME

ArcGIS Portal

FME 2019 does not support SAML authentication. Only built-in user accounts are supported.

Attribute names in portal feature services are ALL lowercase!

Oracle Spatial

Pre-filter SDO Geometry

When using an oracle spatial reader for reading geometry, ensure that you include the following in the **SELECT** statement area (which can be found by going to the *Navigator > Feature Type > Parameters > Format* OR opening up a feature type by clicking the gear), which will ensure that your job will always run successful, even if there is no geometry or if there is bad geometry in the tables. Replace the highlighted portion with your schema and table name.

Lines
SELECT * FROM SCHEMA.TABLENAME LINETABLE WHERE SDO_GEOM.VALIDATE_GEOMETRY_WITH_CONTEXT (geometry, 0.005) = 'TRUE' AND GEOMETRY IS NOT NULL AND (LINETABLE.GEOMETRY.SDO_GTYPE = 2002 OR LINETABLE.GEOMETRY.SDO_GTYPE = 2006)

SELECT * FROM <mark>SCHEMA.TABLENAME</mark> POLYGONTABLE WHERE SDO_GEOM.VALIDATE_GEOMETRY_WITH_CONTEXT (geometry, 0.005) = 'TRUE' AND GEOMETRY IS NOT NULL AND (POLYGONTABLE.GEOMETRY.SDO_GTYPE = 2003 OR POLYGONTABLE.GEOMETRY.SDO_GTYPE = 2007)

Points

SELECT * FROM <mark>SCHEMA.TABLENAME</mark> POINTTABLE WHERE SDO_GEOM.VALIDATE_GEOMETRY_WITH_CONTEXT (geometry, 0.005) = 'TRUE' AND GEOMETRY IS NOT NULL AND POINTTABLE.GEOMETRY.SDO_GTYPE = 2001

Geodatabase SDE

Filter feature classes

Set the WHERE Clause (which can be found by going to the *Navigator > Feature Type > Parameters > Format* OR opening a feature type by clicking the gear OR for an entire reader can be set *Navigator > Parameters*)

- Filter out RAMS geometry

  Use the current RAMS network geometry with the following WHERE statement:

  EFFECTIVE_END_DATE IS NULL

- Filter out invalid geometry

  Line geometry (where SHAPE.LEN is the name of the shape length field):

  <mark>SHAPE.LEN</mark> IS NOT NULL

  Other geometry (where SHAPE is the name of the shape field):

  <mark>SHAPE</mark> IS NOT NULL

Update

When updating your SDE Readers, ensure that your WHERE clauses are set correctly. Updating broke some of them.

Writers

General Tips

- Use Database/Web Connections where possible

  When you use a connection instead of embedding the connection parameters, you can easily create and update a single connection in the future for all your workspaces instead of having to embed the parameters to each location that requires it. See the section below on Web/Database Connections for best practices.

  If reading from a database, and you use OS authentication, ensure that yourfmeserveruser users have access to the database prior to uploading your workspace to FME server.

- Keep test writers in production workspaces.

  When publishing workspaces to production, simply disable the test writer in the navigator panel. This helps during future testing.

Geodatabase SDE

Transaction Type should always be set to "None" when writing to PGIS or TGIS.

Truncate for ETL

It is highly advised to issue a truncate command before writing. This helps truncate tables faster than FME and ensures that your table will only have the most current data. If your FME workspace fails, the table will be empty because it was truncated and can't roll back the data, which is good in some cases and bad in others. There is also the potential that your table will remain empty while the FME job is writing to it, so for real-time applications this may not be the best option. The statement below should be put in the **SQL To Run Before Write** (which can be found in: *Navigator > Geodatabase_SDE Writer > Parameters > Advanced*). Replace the highlighted portion with your schema and table name.

FME_SQL_DELIMITER ;
TRUNCATE TABLE SCHEMA.TABLENAME;

ArcGIS Online Feature Service

If writing to ArcGIS Online more than once (scheduled job) or a lot of data, consider writing to a file geodatabase and using this python script to append the data (much faster than the applyEdits, which the writer uses).

ArcGIS Portal

FME 2019 does not support SAML authentication for Portal. Only built-in user accounts are supported.

Attribute names in portal feature services are ALL lowercase!

Transformers

General Tips

- Reduce duplicated and similar transformer types to improve performance and make your workspace layout cleaner.

  Attempt to combine attribute transformers like AttributeCreator, AttributeRemover, AttributeRenamer, etc into a single AttributeManager transformer.

  Duplicating transformers is a bad practice as it requires multiple transformers to be edited individually, which is harder to maintain.

  If you find yourself using an AttributeFilter with duplicated AttributeManager transformers, Try using the AttributeValueMapper or SchemaMapper transformer to allow scaling with a minimum number of edits.

- Use the least amount of custom python scripting as possible.

  This not only reduces dependencies and other issues when switching between python versions, but also makes it easier for others to understand your workspace. If you do have to use python scripting, use comments in the code!

- Use as few transformers as possible.

  This makes workspaces more efficient, easier to understand, easier to update, and your canvas less cluttered.

- When creating attributes in your workspace, prefix them with _

  This helps the attribute move up to the top in inspector. It also helps it to stand out, so you know that it wasn't part of any of the original datasets.

Attribute Manager

- Try to condense multiple attribute managers into a single attribute manager. This is especially important when using @Expression() in multiple transformers, to cut down on TCL calls and compiling.

Date/Time

To parse out the hour to add:
@Substring(@DateTimeNow(local),24,1)

To add time (Replaces DateTimeCalculator):
@DateTimeAdd(@Left(@Value(timestamp),14),PT@Substring(@DateTimeNow(local),24,1)H)

To calculate difference (Replaces DateDifferenceCalculator):
- **Both times are required to be in same format**

@DateTimeDiff(@Value(CURRENT_DT),@Value(PHOTO_FILEDATE),Minutes)

Regular Expression

(.[^<]*)

This will find anything between the characters to the left and right of it. Use with a string searcher.

Web/Database Connections

Using database and web connections where possible make it simpler for managing parameters when reading/writing to a database or an online service. Following the naming conventions below is strongly suggested when publishing to FME server. Note that if you use OS authentication for database access, yourfmeserveruser must be granted access to the database when publishing to FME server.

ArcGIS Online

UserName@AGOL (ie: IowaDOT_GIS@AGOL)

Workspace Parameters

Metadata

Notice how this is mentioned TWICE? It's **VERY** important!!!

- Workspace Name: (NOTE this is NOT the file name, but the name in the workspace parameters.) Use the standard naming conventions.
    - WorkspaceName_Dev(01-99)(A-Z) when developing or testing workspaces. For example, if you were updating the production workspace DBtoAGOL_Prod04 and this was your first update it would be named DBtoAGOL_Dev04A.
    - WorkspaceName_Prod(01-99) for workspaces that have been finalized and meet or exceed all documented requirements and best practices.
- Overview: Describe what the workspace does. Where does the data come from and where does it go to? What is it for/why is it needed? How often is the source data updated? How often does it need to run? Make a note if the python version requires a specific version to run.
- Help: Include a point of contact for someone that supports the workspace and tips to help with common issues (if any occur).
- History: When finalizing a workspace for production use by others, you MUST include your name, date, and a detailed description of what you changed since the last time the workspace was production quality. This is also useful to keep track of minor changes throughout the development life cycle so you know what all you've done!

Scripting

Use **Python 3.7+** by default for future support. If you are writing to ArcGIS Online via the python script, use the **ArcGIS Pro 2.x 3.6.** Some transformers may still require 2.7, so you may have to downgrade python. Make sure to note this in your workspace.

FME Server Publishing Standards Checklist

FME Workspace General Requirements

- o User's FME version must match that of the FME server to which they are publishing to. Temporary exceptions may be made to accommodate bug workarounds.
- o FME Workspace Parameters are filled in and complete.
    - o Name
    - o Category – Choose one from the FME default dropdown.
    - o Overview – Define information about what the workspace does, along with other requirements.
    - o Help – Who should be contacted if questions arise regarding this workspace.
    - o History – Ensure you keep track of changes made and when they were made, especially when making changes to workspaces in production.
- o Workspace is properly documented with appropriate number of annotations and bookmarks that logically group parts of your process together.
- o **Ensure that debugging is DISABLED when uploading workspaces to FME server (unless absolutely necessary!**

Readers/Writers

- o There are no references to files on local drives. All files and connections must be referenced via UNC path to the specified shared directory on FME Server that matches the user's repository folder.
- o Use Web and Database Connections where needed and use the standard naming conventions to reduce duplicate connections. All connections will be managed by the GIS Team. If you require one that isn't already created, contact the GIS Team to help get one created.
- o When uploading workspaces to FME server, ensure that the test writer is disabled in the navigator panel. (*Tools > FME Options > Translation > Log Message Filter*)

Transformers

- o If custom transformers are required, work with the GIS Team to get them uploaded to the shared transformers folder so others can use them. Never upload any versions of SkyFire to FME server, as that is already managed by the GIS Team.
- o There should be no inspectors or loggers enabled when uploading to server. This slows down the job.

Production Publishing Process

DESKTOP

- a. Create workspace in FME Desktop and run to ensure that it works correctly.
- b. Ensure debugging is disabled.

FME TEST Server

- a. Verify publishing standards have been met and files are referenced using \\yourtestfmeserver\ UNC paths.
- b. Publish to FME Test

c. Run workspace on FME Test. Verify output is correct.
d. Once you have a working version, commit the version. **If you are updating a workspace, don't commit unless it actually works!** In the commit notes, document what you changed.
e. Continue to the process for publishing to FME production server.

FME PRODUCTION Server

a. Verify files are referenced using \\yourprodfmeserver\ UNC paths.
b. Have a FME-using colleague review your workspace.
c. Publish to FME Production.
   a. Ensure that you commit your FME Job to version control and add a date and other descriptive text about what you modified (or "Initial Commit" if new). This helps provide a history of changes that have been made to roll back to older versions if necessary.
d. Run workspace on FME Production. Verify output is correct.
e. Once you have a working version, commit the version. **If you are updating a workspace, don't commit unless it actually works!** In the commit notes, document what you changed.
f. Contact the GIS Team to schedule your job (if needed). If your job takes longer than 10 minutes to run or is to be scheduled to run more than once a week, contact the GIS Team to help optimize your workspace to conserve server resources.

FME Server

Security

\\yourfmeserver\[REPOSITORYFOLDER]\ - All FME users will have read/write access to repository folders that they have access to. This folder will allow users to keep data on the server and access it via unc path.

\\yourfmeserver\[REPOSITORYFOLDER]\SDE - All FME users will have read access to repository folders that they have access to. This folder will consist of SDE connection files.

Permissions will be granted via the "Advanced Sharing" functionality of Windows. Security will be locked down via the security tab.

Version Control

Version control is set up on FME server and should be leveraged whenever updates are made to jobs, so they can be rolled back if needed. Versions will automatically be committed once per day on Production.

Notes on Upgrading Workspaces to latest version

Read notes above to ensure that you include best practices in your workspace.

- All readers and writers should be updated to the latest version to take advantage of the latest functionality and performance enhancements.
- Transformers should also be updated when documented performance enhancements exist or when possible in other circumstances.
- WHEN UPDATING GEODATABASE_SDE Writers, ensure that the table handling is set correctly. Previously it was "Truncate", but it gets reset to "Create If Needed".
- Ensure that your select statements/where clauses get set correctly (if they were set).

Upgrading FME Server

The following procedures should be followed when upgrading FME Server. This process will be used to replicate the production environment into a test environment and then replicate the test environment back to production to ensure proper testing is completed.

Upgrading Test

Notes:

- If any issues are encountered through this process, ensure that they are DOCUMENTED and contact SAFE Software for mitigation or to report a bug.

Prior to Upgrading Test (a week before upgrade):

a) Notify all users with an account on FME server and others who utilize FME server jobs, that an upgrade is planned. Include outage date/time.

Prior to Upgrading Test (on day of upgrade):

a) Ensure that a copy is made of the repositories folder (C:\ProgramData\Safe Software\FME Server\repositories).
b) Run the Bulk Commit for Version Control workspace and ensure that the local history is pushed remotely to GitHub.
c) Create a backup of Test. Note that this backup only includes the components listed here.

Following Upgrading Test (on day of upgrade):

a) Restore the Test Backup. (This will restore all the jobs that are on test)
b) Configure License
c) Replicate Production
    a. Create a backup from Production. Note that this backup only includes the components listed here.
    b. Restore the Production Backup to Test. Overwrite existing jobs (because production is more accurate than test)
    c. Disable all schedules

> > d. Update jobs that need to have references updated
> d) Configure Version Control
> > a. [Create a new personal access token](#) on GitHub
> > b. Enable and configure Version Control.
> > c. Fetch from Remote
> e) Configure the Config files appropriately
> > a. [Configure FME Download Service to use https](#): in the fmeEngineConfig.txt file, add the ResultPrefix parameter under the section SUB_SECTION FILE_DOWNLOAD_SERVICE, on the line that starts SUCCESS_RESPONSE, and before the ResultRootDir parameter: ...|ResultPrefix=https://yourtestfmeserver.gov|ResultRootDir=...
> f) Configure System Email
> > a. SMTP Server: yoursmtpserver
> > b. Server Port: 25
> > c. Connection Security: None
> > d. Email From: fmetestserver@yourdomain.gov
> g) Upload Files
> > a. [Custom Coordinates System Definition File](#)
> > b. [Custom Transformers](#)
> h) Testing
> > a. Request a Test License for # of engines that match Production
> > b. Enable schedules for jobs that run more than once an hour on test and disable them on production.
> > c. Run jobs manually that run less than once an hour.
> > d. Test to ensure that jobs run properly, produce expected results, and engines restart correctly.
> > e. Document any issues experienced during testing and report to FME.
> > f. Once completed with testing, disable schedules on test, and re-enable schedules on production.

Upgrading Production

Prior to Upgrading Production (a week before upgrade):

a) Notify all users with an account on FME server and others who utilize FME server jobs, that an upgrade is planned. Include outage date/time.

Prior to Upgrading Production (on day of upgrade):

a) Ensure that a copy is made of the repositories folder (C:\ProgramData\Safe Software\FME Server\repositories).
b) Run the Bulk Commit for Version Control workspace and ensure that the local history is pushed remotely to GitHub.
c) [Create a backup from Production](#). Note that this backup only includes the [components listed here](#).

Following Upgrading Production (on day of upgrade):

a) [Restore the Production Backup](#)

b) Configure License
c) Configure Version Control
   a) [Create a new personal access token](#) on GitHub
   b) Enable and configure Version Control.
   c) Fetch from Remote
d) Configure the Config files appropriately
   a) [Configure FME Download Service to use https](#). In the fmeEngineConfig.txt file, add the ResultPrefix parameter under the section SUB_SECTION FILE_DOWNLOAD_SERVICE, on the line that starts SUCCESS_RESPONSE, and before the ResultRootDir parameter:
   ...|ResultPrefix=https://yourprodfmeserver.gov|ResultRootDir=...
e) Configure System Email
   a) SMTP Server: yoursmtpserver
   b) Server Port: 25
   c) Connection Security: None
   d) Email From: fmeserver@yourdomain.gov
f) Upload Files
   a) [Custom Coordinates System Definition File](#)
   b) [Custom Transformers](#)