Extract Transfer Load (ETL) Administration Guide

For Aviation Management

San Francisco International Airport

July 2018

QUALITY

At Woolpert, quality is the cornerstone of our business. We invite your comments and suggestions for improving this document.

TRADEMARKS

All brand names and product names are trademarks or registered trademarks of their respective companies.

NOTICE OF PROPRIETARY INFORMATION

© 2018, Woolpert, Inc., Dayton, Ohio.

All rights reserved to Woolpert. This document was designed, prepared, and submitted by Woolpert to be used only by the recipient.

None of this material is permitted to be reproduced in any way or distributed to anyone other than the authorized representatives of the recipient.

Table of Contents

1 Executive Sign-Off 3
2 Revision History and Change Control 4
3 Purpose, Software, Process 5
3.1 Reference Documents 5
3.2 Acronyms 5
3.3 Software Versions 5
3.4 ETL Process 6
4 Data & Scripts 7
4.1 AVM CAD Files 7
4.1.1 ArcGIS for AutoCAD 7
4.2 AVM Geodatabase 8
4.3 PROD Geodatabase 9
4.4 FME Scripts 9
4.4.1 Floorplan 10
4.4.2 Grid 12
4.4.3 Public Architecture 15
4.4.4 Topology 18
4.4.5 AVM to PROD 21
4.4.6 Workspace Runner 23
5 Running ETL Tool 24
6 Production Web Services 25
7 Updating ETL Administration Guide 26
Appendix A: AVM Building Level List 28
Appendix B: AVM CAD Files 29
Appendix C: AVM Geodatabase Schema 30
Appendix D: PROD Geodatabase Schema 31
Appendix E: Topology Errors 32

List of Figures

Figure 1: ETL Diagram 6
Figure 2: CAD Folder Structure 7
Figure 3: AVM Geodatabase 9
Figure 4: PROD Geodatabase 9

List of Tables

Table 1: Revision History 4
Table 2: Software Versions 5
Table 3: CAD Layer Name 15
Table 4: Web Services 25

1 Revision History

Table 1: Revision History			
Name	**Date**	**Reason for Change**	**Version**
Your name here	Date here	Initial draft	1.0

2 Purpose, Software, Process

AVM maintains SFO interior floor plan drawings as AutoCAD DWGs with standard AutoCAD A-AREA and A-DOOR layers that include object data. IIM and AVM have agreed to store the areas and doors in the airport's GIS, using the ArcGIS for AutoCAD extension ("ArcGIS extension") to edit those specific layers. The Extract Transfer Load (ETL) tool automates the process of converting CAD/GIS data from AVM to IIM utilizing Safe Software's FME along with providing geometry checks and fixes to maintain correct, current, and accurate A-Area data for SFO.

2.1 SFO ETL Tool Manager

Hanson Guy Michael
GIS Analyst | Planning, Design and Construction
San Francisco International Airport | P.O. Box 8097 | San Francisco, CA 94128
Tel 650-821-5590 | flysfo.com

2.2 Reference Documents

SFO CAD Standard
SFO GIS Standard
Others?

2.3 Acronyms

GIS	-	Geographic Information System
CAD	-	Computer Aided Design
AGIS	-	Airport GIS
ETL	-	Extract Transfer Load
ESRI	-	Environmental Systems Research Institute
FME	-	Feature Manipulation Engine
SFO	-	San Francisco International Airport
AVM	-	Aviation Management
IIM	-	Interior Information Management
GDB	-	Geodatabase
PROD	-	Production
Others		

2.4 Software Versions

The ETL tool for SFO utilizes the software listed below to process CAD files into the AVM & Production databases for managing Building Doors and Areas.

Table 2: Software Versions

Company	Software	Version
AutoDesk	AutoCAD Map 3d	2017

AutoDesk	AutoCAD Civil 3d	2017
ESRI	ArcGIS for AutoCAD	Build 370
ESRI	ArcGIS for Desktop	10.5
ESRI	ArcGIS Server	10.5
SAFE	FME Server	2017
Oracle	Oracle Database	12g

2.5 ETL Process

The ETL tool process takes AVM CAD Files that are constantly being edited and updated, and converts the data to a Production Geodatabase for other SFO personnel and the public to utilize. Figure 1 below describes the process of managing and converting the data. It lists the referenced servers utilized in the process.

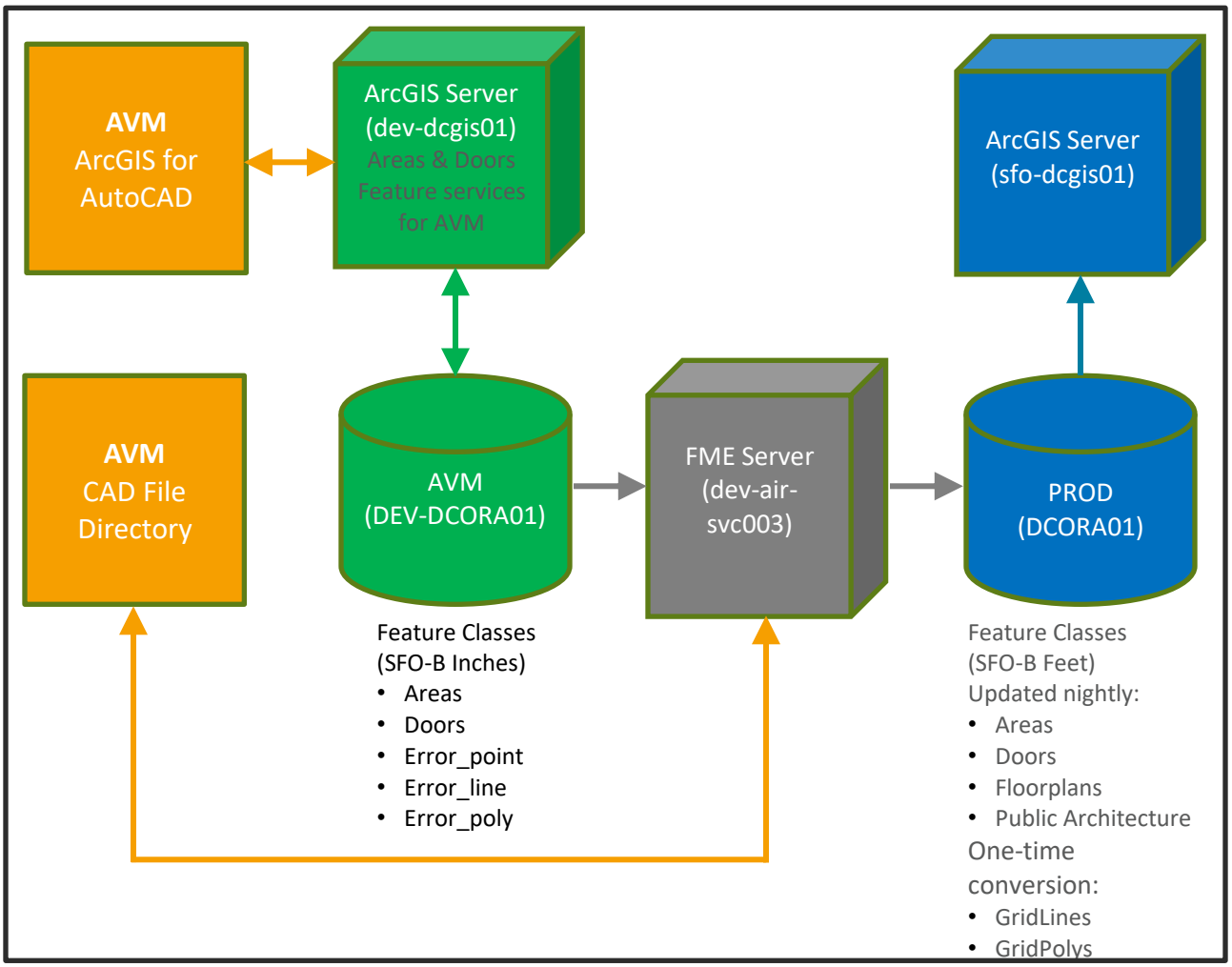


Figure 1: ETL Diagram

3 Data & Scripts

This section explains the type of data, where it is located, and how it is used in the ETL tool for managing interior floorplan data. It briefly describes how the scripts and data interact in the ETL tool.

3.1 AVM CAD Files

The AVM CAD Files is where all interior floorplan data is managed and updated at the airport. The AVM group are the only personnel authorized to edit the AVM CAD Files and they are stored here:

<Enter SFO internal path>

The CAD Files are in the following folder hierarchy based on building name and level under the before mentioned path. For a detailed list of buildings and levels see Appendix A. The CAD Files also contain Object Data that is used to identify attributes for A_Areas and A_Doors.

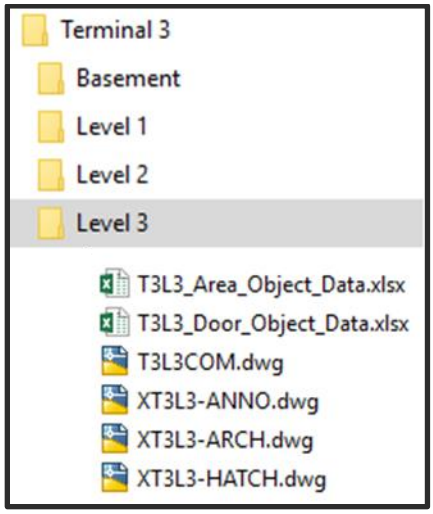


Figure 2: CAD Folder Structure

The AVM CAD Files follow the SFO CAD Standards and only utilize the layers listed in Appendix B. In editing the CAD Files, AVM also must utilize the ArcGIS for AutoCAD tool to edit the feature services for areas and doors for the ETL Tool to work correctly and maintain updated data.

3.1.1 ArcGIS for AutoCAD

This section explains the process of connecting and editing the Area and Door feature classes necessary for the ETL tool.

<Not sure the below are described in another document if so just reference document>

- Finding and connecting to Area and Door Feature Services from the AVM ArcGIS Server.
    - Step 1

<describe and screenshot>

- Step 2

<describe and screenshot>

- How to add new Area Feature using template
  - Step 1

<describe and screenshot>

  - Step 2
- How to add new Door Feature using template
  - Step 1

<describe and screenshot>

  - Step 2
- How to edit Area Feature
  - Step 1

<describe and screenshot>

  - Step 2
- How to edit Door Feature
  - Step 1

<describe and screenshot>

  - Step 2
- How to update Attributes for Doors and Areas (Required)
  - Step 1

<describe and screenshot>

  - Step 2

3.2 AVM Geodatabase

The AVM Geodatabase is where the Area and Door feature classes are stored for editing by AVM and conversion by the ETL tool to the Production Database. In addition to the A_Area and A_Door features

there are three Error files (Point, Line, Polygon). These features collect topology errors from the ETL tool when it is run, however all the errors might not be actual errors in the real world. The schema for the AVM database is in Appendix C. The current AVM Geodatabase is stored on DEV-DCORA01 and the features are accessed through ArcGIS Server dev-dcgis01. Access to this geodatabase and server are managed through the ETL tool manager. Here is a view of the features stored in the AVM Database, which utilizes SFO's custom coordinate system in inches.

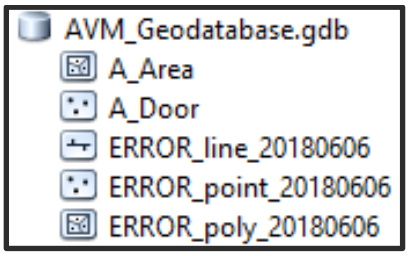


Figure 3: AVM Geodatabase

3.3 PROD Geodatabase

The PROD database is where all the final AVM data that has been updated and cleaned is stored for consumption by other SFO personnel and the public. The PROD Geodatabase is stored on DCORA01. The PROD Geodatabase not only receives the A_Area and A_Door features from the AVM geodatabase but also data from CAD Files AVM maintains. AVM CAD File Layers are abstracted and transferred to the Architectural_Data, Floorplans, Grid_Bubble, and Grid_Line feature classes. The schema for the PROD database features utilized by the ETL tool is in Appendix D. Here is a view of the features stored in the PROD Database, which utilizes SFO's custom coordinate system in feet.

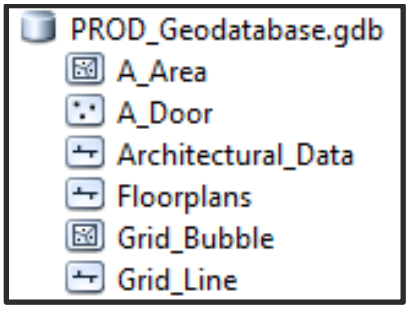


Figure 4: PROD Geodatabase

3.4 FME Scripts

The FME scripts are the brains of the ETL process. These scripts automate the process to extract, transfer, validate, clean, and load the interior floorplan data into the AVM and PROD Geodatabases mentioned before. The scripts check and clean topology along with automatically populating attributes for features from the CAD files. These FME scripts are located in the following file path:

<Enter SFO internal path>

All the scripts have sections that describe what the script is doing

3.4.1 Floorplan

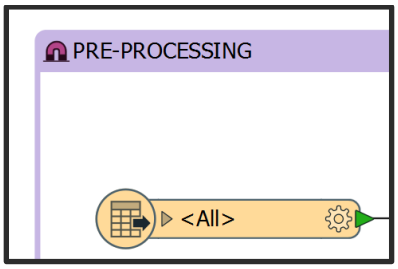
The Floorplan script is used to extract floorplans from the AVM CAD Files for each building and level and convert it to a single feature in the PROD Geodatabase to give a simple overview of the floorplan.

A. **Input** – The input for Floorplan script is multiple CAD Files where the AVM CAD Folder is located followed by ***ARCH.dwg:

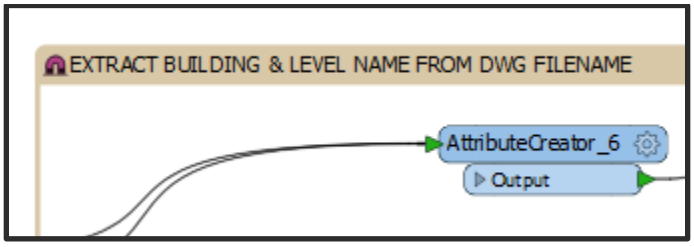
\\<FolderPath>***ARCH.dwg

B. **Transformers** (listed in processing order)

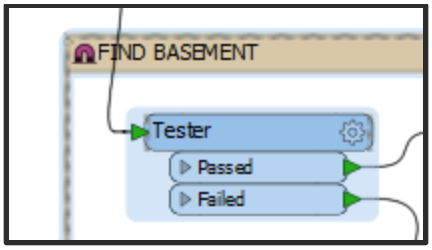
i. Pre-Processing Transformer performs selects only layers needed for the floorplan, validates the geometry of those features and then simplifies and cleans the data for better display.



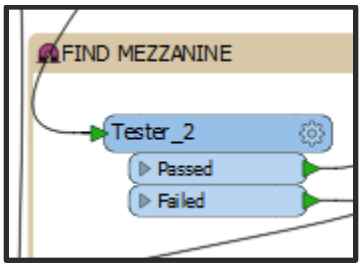
ii. Extract Building & Level Name from DWG Filename Transformer reads the filename that CAD Features are coming from and adds them as an attribute to the feature.



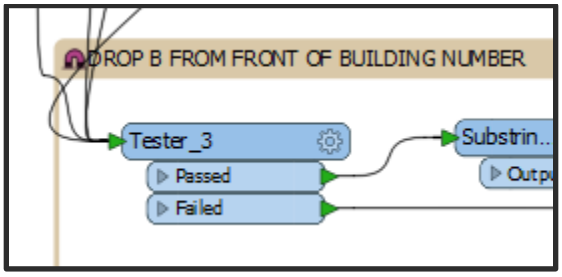
iii. Find Basement Transformer identifies basement level features in order to populate level name and ordinal value correctly (Other Levels have a number value).



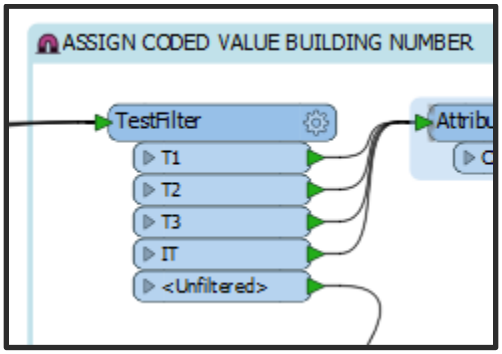
iv. Find Mezzanine Transformer identifies mezzanine level features in order to populate level name and ordinal value correctly (Other Levels have a number value).



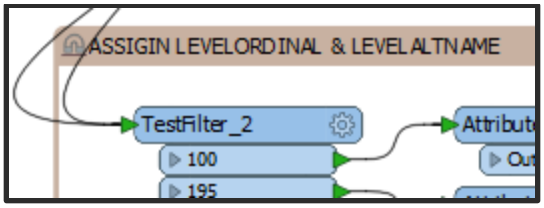
v. Drop B From Front of Building Number Transformer removes the leading B from the Building number populated earlier from the file name.



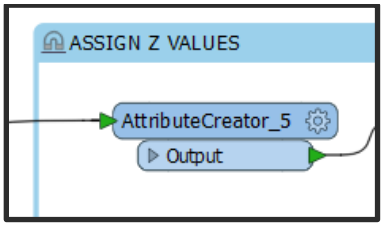
vi. Assign Coded Value Building Number Transformer converts Terminal Building Names there coded building number



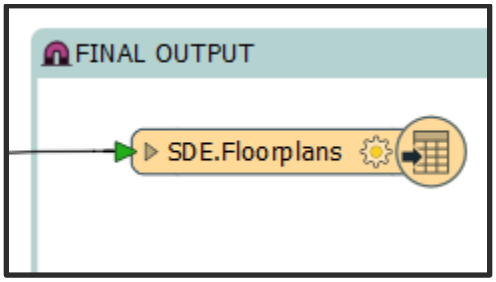
vii. Assign LevelOrdinal & LevelALTName Transformer populates the Level Ordinal and Level Alternate name based on pre-determined values.



viii. Assign Z Values Transformer converts the ordinal value as the geometry Z value.



C. **Output** - The output for the Floorplan script is the SDE.Floorplans feature in the PROD Geodatabase.



3.4.2 Grid

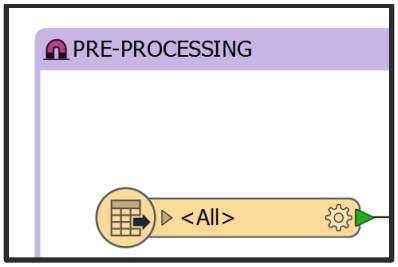
The Grid script is used to extract the grid lines and grid bubble from the AVM CAD Files for each building and level and attribute it with its grid annotation.

A. **Input** – The input for Grid script is multiple CAD Files where the AVM CAD Folder is located followed by ***ARCH.dwg:

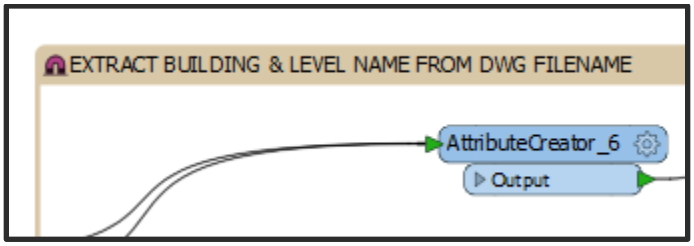
\\<FolderPath>***ARCH.dwg

B. **Transformers** (listed in processing order)

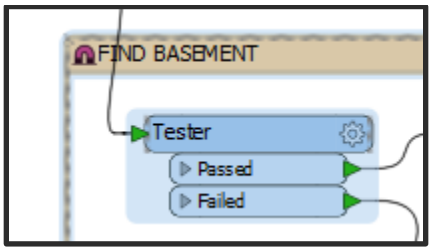
i. Pre-Processing Transformer performs selects only layers needed for the grid features, validates the geometry of those features and then simplifies and cleans the data for better display.



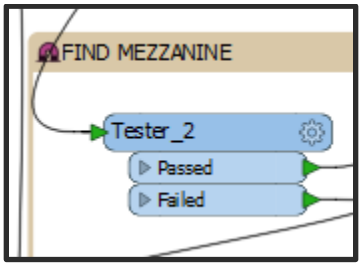
ii. Extract Building & Level Name from DWG Filename Transformer reads the filename that CAD Features are coming from and adds them as an attribute to the feature.



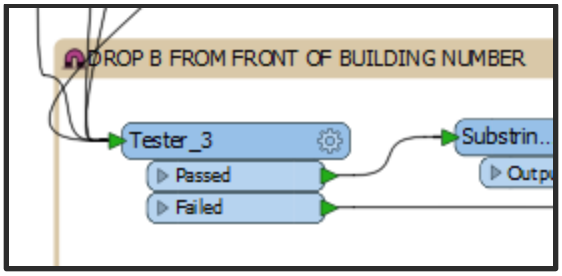
iii. Find Basement Transformer identifies basement level features in order to populate level name and ordinal value correctly (Other Levels have a number value).



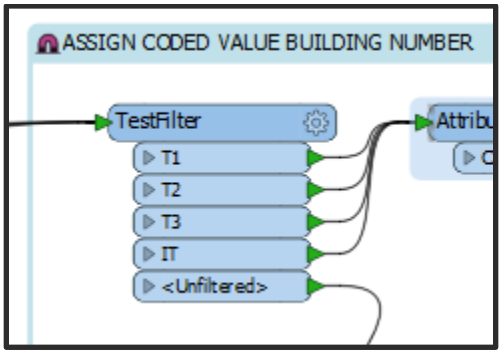
iv. Find Mezzanine Transformer identifies mezzanine level features in order to populate level name and ordinal value correctly (Other Levels have a number value).



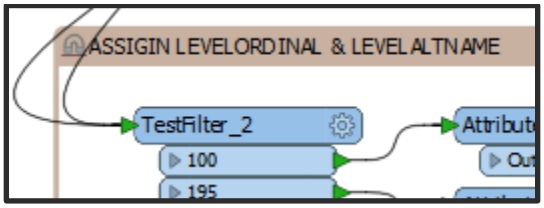
v. Drop B From Front of Building Number Transformer removes the leading B from the Building number populated earlier from the file name.



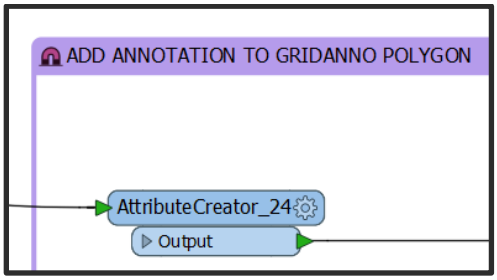
vi. Assign Coded Value Building Number Transformer converts Terminal Building Names there coded building number.



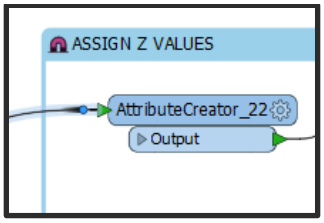
vii. Assign LevelOrdinal & LevelALTName Transformer populates the Level Ordinal and Level Alternate name based on pre-determined values.



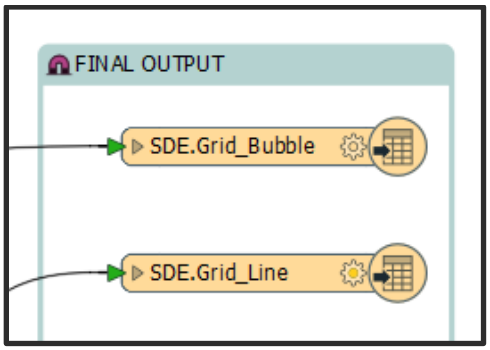
viii. Add Annotation to GridAnno Polygon Transformer creates polygons from Bubbles with annotation and populates them with the annotation grid value.



ix. Assign Z Values Transformer converts the ordinal value as the geometry Z value.



D. **Output** - The output for the grid script is the SDE.Grid_Bubble and SDE.Grid_Line features in the PROD Geodatabase.



3.4.3 Public Architecture

The Public Architecture script is used to extract floorplan linework for public areas into a single feature class for public use. s from the AVM CAD Files for each building and level and convert it to a single feature in the PROD Geodatabase. The following is a a list of CAD Layers that are for public use.

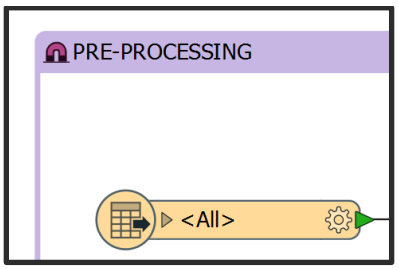
Table 3: CAD Layer Name		
A-DOOR-FRMG	E-FIRE-ALRM	A-HRAL
A-DOOR-PANL	M-EQPM	S-BEAM
A-DOOR-SWNG	M-ESCA	S-BRCG
A-FLOR-OTLN	M-EVTR	A-HRAL
A-GLAZ	P-EQPM	I-FURN
A-WALL-EXTR	S-COLS	S-STRS-TRED
A-WALL-INTR	S-STRS-RAIL	

A. **Input** – The input for Public Architecture script is multiple CAD Files where the AVM CAD Folder is located followed by ***ARCH.dwg:

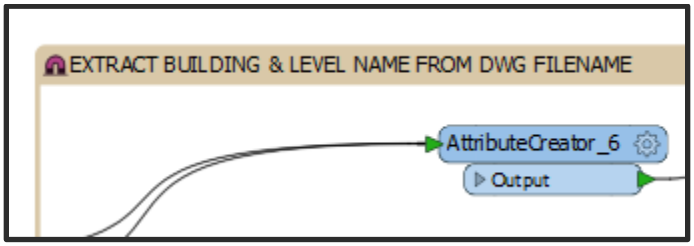
\\<FolderPath>***ARCH.dwg

B. **Transformers** (listed in processing order)

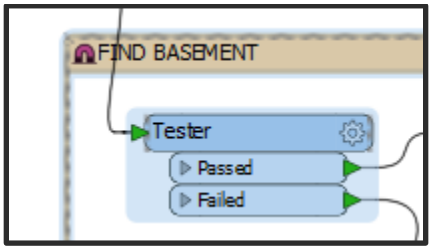
i. Pre-Processing Transformer performs selects only layers needed for the Public Architecture feature, validates the geometry of those features and then simplifies and cleans the data for better display.



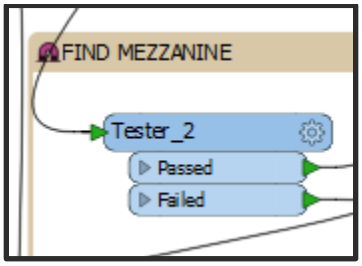
ii. Extract Building & Level Name from DWG Filename Transformer reads the filename that CAD Features are coming from and adds them as an attribute to the feature. This section also populates Feature Type attribute based on CAD Layer Name.



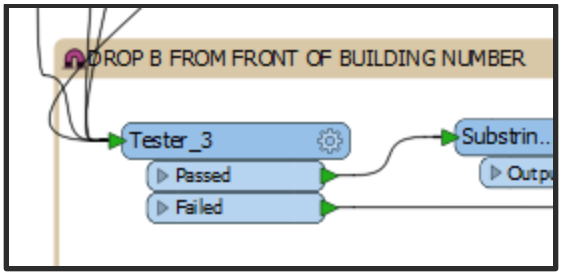
iii. Find Basement Transformer identifies basement level features in order to populate level name and ordinal value correctly (Other Levels have a number value).



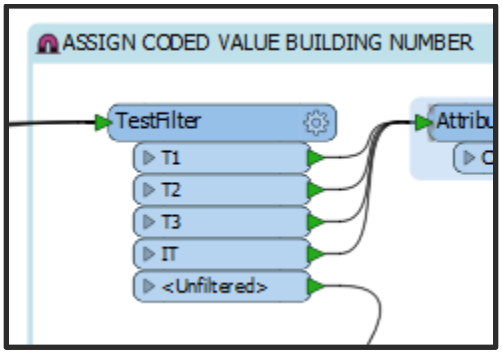
iv. Find Mezzanine Transformer identifies mezzanine level features in order to populate level name and ordinal value correctly (Other Levels have a number value).



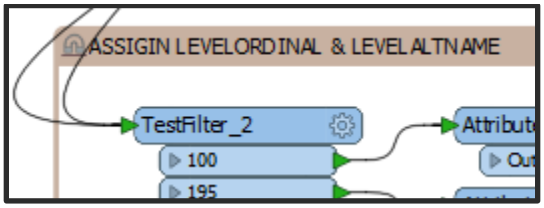
v. Drop B From Front of Building Number Transformer removes the leading B from the Building number populated earlier from the file name.



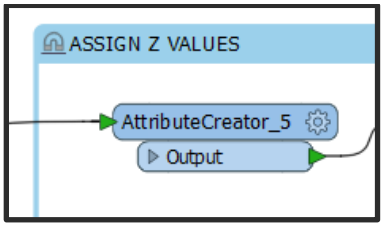
vi. Assign Coded Value Building Number Transformer converts Terminal Building Names there coded building number



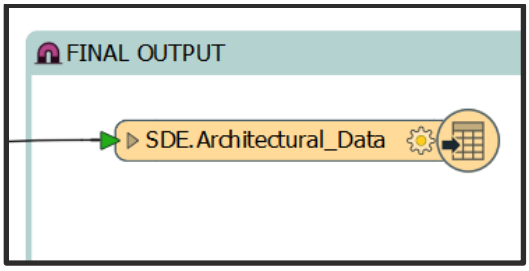
vii. Assign LevelOrdinal & LevelALTName Transformer populates the Level Ordinal and Level Alternate name based on pre-determined values.



viii. Assign Z Values Transformer converts the ordinal value as the geometry Z value.



C. **Output** - The output for the Floorplan script is the SDE.Architectural_Data feature in the PROD Geodatabase.



3.4.4 Topology

The Topology script is used to validate geometry, automatically repair some geometry issues and flag issues found in the AVM A_Door and A_Area feature classes located in SFO AVM SDE.

The A_Door feature class will be checked for the following:

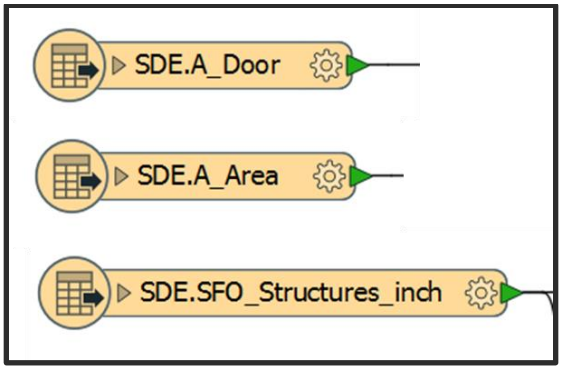
- Spatial Duplicates using attributes and geometry (Builiding, LevelName, x, y)
- Duplicate Door Numbers
- Incorrect Building Number and Incorrect Level Name
- Door falls inside Building Footprint

The A_Area feature class will be checked for the following:

- Null or Corrupt Geometry
- Self-Intersections
- Spatial Duplicates using attributes and geometry (Builiding, LevelName, x, y)
- Duplicate Space ID
- Incorrect Building Number, Level Name, Lease Type,
- Incorrect Building Number and Incorrect Level Name
- Overlaps and Gaps
- Door falls inside Building Footprint

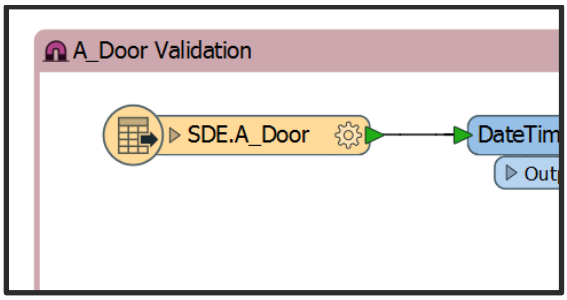
The script cleans and simplifies geometry where possible to limit the amount of errors created by the topology checks to be handled manually.

A. **Input** – The input for Topology script is the AVM A_Area, A_Door, and Structures (Building Footprint) feature classes located on DEV-DCORA01.

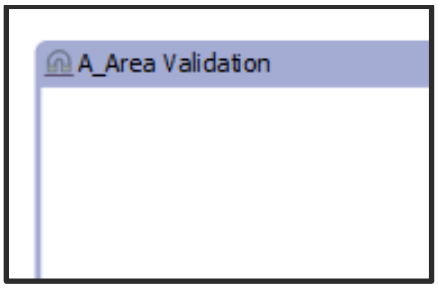


B. **Transformers** (listed in processing order)

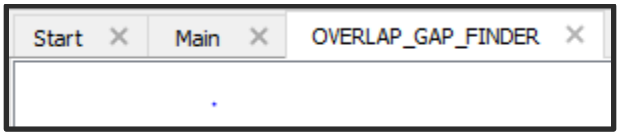
i. A_Door Validation Transformer performs the topology checks mentioned previously for the A_Door feature class except for the building footprint check.



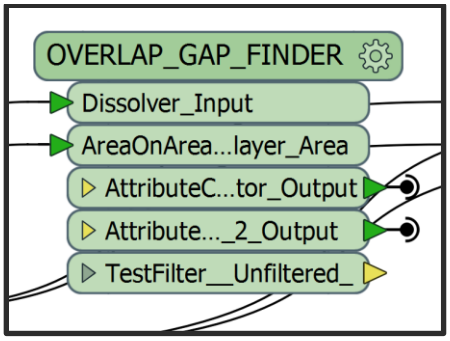
ii. A_Area Validation Transformer performs the topology checks mentioned previously for the A_Area feature class except for the building footprint check.



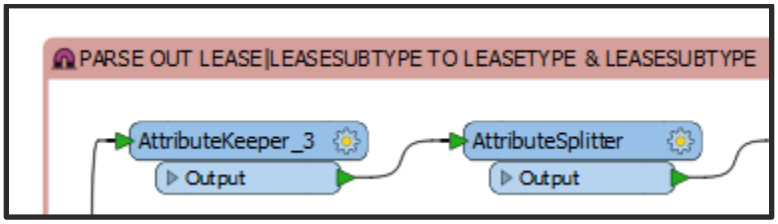
iii. **Embedded Transformer** is a transformer that is a custom built transformer that is used in the main model and located on the tab next to the main tab and shows the name of the custom transformer (Overlap_Gap_Finder).



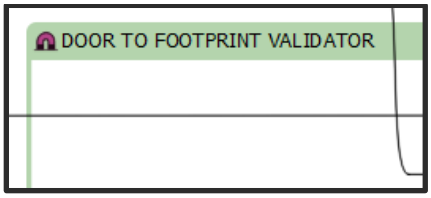
iv. Overlap_Gap_Finder Transformer is located within the A_Area validation transformer and performs the gaps and overlaps topology check.



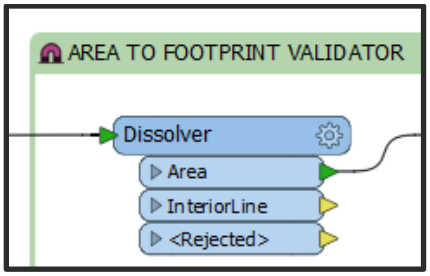
v. Parse Out Lease|LeaseSubType to LeaseType & LeaseSubType Transformer parses out the A_Area LeaseSubtype Field into two representative fields that are automatically populated, those fields are LeaseType and LeaseSubType.



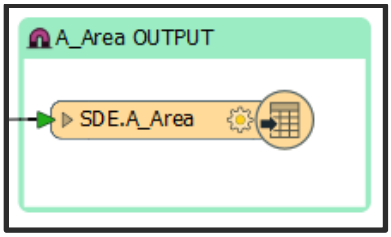
vi. Door to Footprint Validator Transformer checks to see that all A_Doors are in or on the boundary of the Building Footprints based on building name.

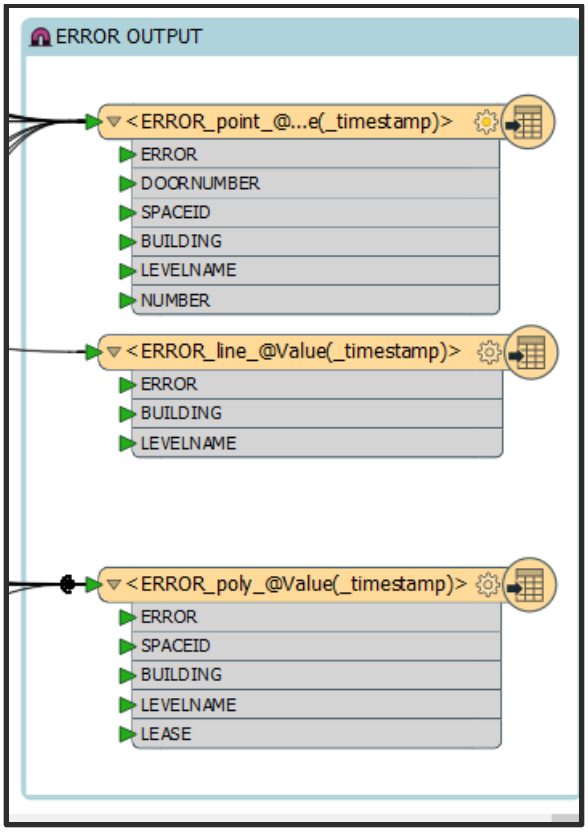


vii. Area to Footprint Validator Transformer checks to see that all A_Areas are in or on the boundary of the Building Footprints based on building name.



C. **Output** - The output for the Topology script is the SDE.A_Areas <does this go back to AVM database and overwrite A_Area?> and Error features. The Error features are Points, Lines, and Polygons with attributes to identify features and errors. The errors that can be generated and troubleshooted are in Appendix E.

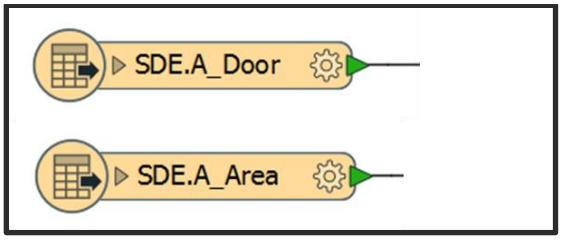




3.4.5 AVM to PROD

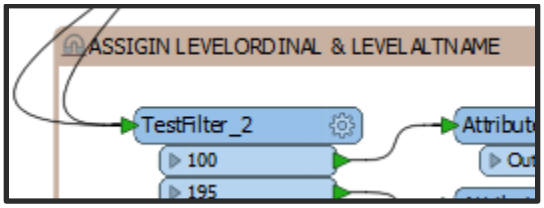
The AVM to PROD script is used to transfer the A_Door and A_Area feature class from the AVM database to the PROD database and to automatically populate attributes needed in the PROD database environment.

A. **Input** – The input for AVM to PROD script is the AVM A_Area and A_Door feature classes located on DEV-DCORA01.

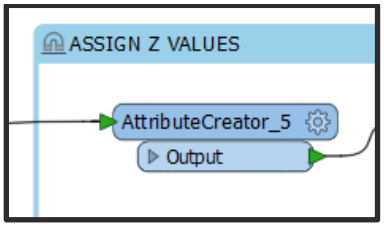


B. **Transformers** (listed in processing order)

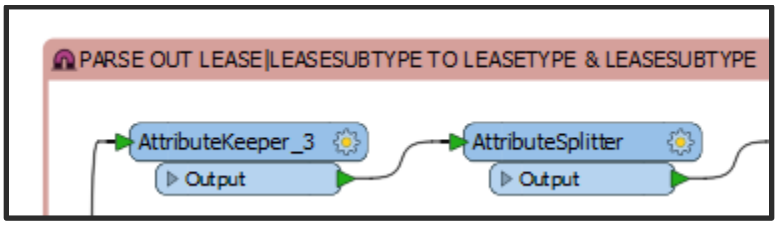
i. Assign LevelOrdinal & LevelALTName Transformer populates the Level Ordinal and Level Alternate name based on pre-determined values.



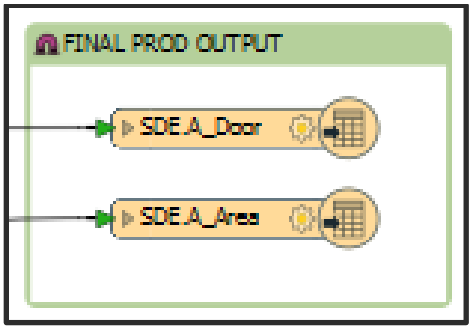
ii. Assign Z Values Transformer converts the ordinal value as the geometry Z value and also populates the BASE field and Height field for A_Areas.



iii. Parse Out Lease|LeaseSubType to LeaseType & LeaseSubType Transformer parses out the A_Area LeaseSubtype Field into two representative fields that are automatically populated, those fields are LeaseType and LeaseSubType.



C. **Output** - The output for the AVM to PROD script is the SDE.A_Areas and SDE.A_Doors in the PROD Geodatabase located on server DCORA01.



3.4.6 Workspace Runner

<May not be needed with Server workflow for running tools>

4 Running ETL Tool

<Steps and Process to run ETL tool. Along with reviewing errors and pushing to production. Will have to document based on SFO setup>

5 Production Web Services

The ArcGIS map templates and web services are described in this section. The following is a list of maps used to create the web services along with their path location and required settings.

Table 4: Web Services

WebServiceName	MapTemplate	PathLocation	RequiredSetting
Floorplans	Floorplans.mxd	// sfo-dcgis01/GISdata/Maps	MapScale : 1000
Grids	Grids.mxd	// sfo-dcgis01/GISdata/Maps	Grid Bubble Labeled
AreasDoors	AreasDoors.mxd	// sfo-dcgis01/GISdata/Maps	Feature Service

- Finding and connecting to Area and Door Feature Services from the AVM ArcGIS Server.
    - Step 1

      <describe and screenshot>
    - Step 2

      <describe and screenshot>
- How to add new Area Feature using template
    - Step 1

      <describe and screenshot>
    - Step 2

6 Updating ETL Administration Guide

The ETL tool will require updates to account for future and legacy data, software updates, and to adhere to evolving project requirements. Any and all changes to the FME Script, CAD template, and GIS Schema will need to be accounted for to keep the translation up to date and functioning properly.

- FME and programs that it uses to produce data are always changing and therefore the script and project data sources will always need to be looked at.
- Data Files are not always created the same way and the FME tool will not capture the true nature of the data.
- FME is a tool to speed up human processing and is not perfect, therefore data should always be reviewed and if necessary changes to the tool made.
- In conjunction with the FME script being updated the CAD template needs to be updated along with the GIS schema as necessary.
- In conjunction with the FME script, GIS schema. And CAD template SFO will need to account for the updates and changes within this document.

It is the responsibility not only of the ETL tool manager but also the personnel utilizing the CAD Files, Geodatabases, and scripts to help ensure that all changes and updates are known and documented.

Appendix A: AVM Building Level List

Appendix B: AVM CAD Files

Appendix C: AVM Geodatabase Schema

Appendix D: PROD Geodatabase Schema

Appendix E: Topology Errors

Error Shapefile	Feature	Error	Troubleshoot
Error_Point	A_Door	Bad Value for Building	
Error_Point	A_Door	Bad Value for Level Name	
Error_Point	A_Door	Duplicate Door Location	
Error_Point	A_Door	Duplicate Door Number	
Error_Point	A_Door	No Door to Footprint Match	
Error_Point	A_Door	Door to Footprint Building Number Mismatch	
Error_Point	A_Area	Geometry Validator Failed	
Error_Point	A_Area	Geometry Validator Issue Locations	
Error_Point	A_Area	No Area to Footprint Match	
Error_Point	A_Area	Area to Footprint Building Number Mismatch	
Error_Line	A_Area	Geometry Validator Failed	
Error_Poly	A_Area	Geometry Validator Invalid Parts	
Error_Poly	A_Area	Duplicate Space ID	
Error_Poly	A_Area	Gap	
Error_Poly	A_Area	Overlap	