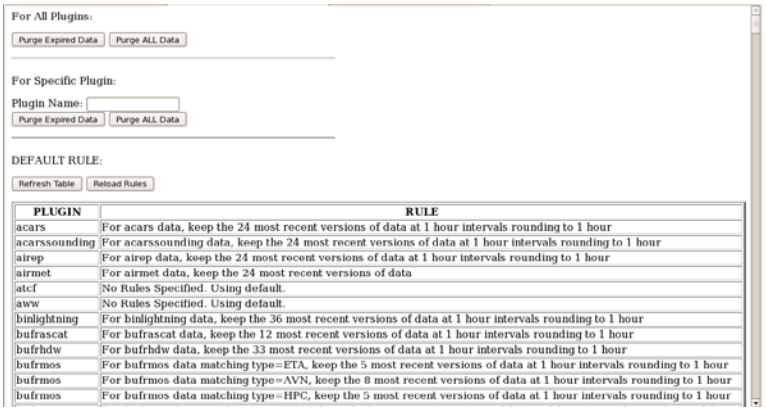
AWIPS-2 Reference Sheet

EDEX Purging: PurgeRules and PathKeys Files

Updated for AWIPS Build OB13.5

Introduction

EDEX has implemented a data purging system for processed (*i.e.,* HDF5) data files that is configured in a way that is analogous to AWIPS-1. EDEX's purging uses plugin-specific PurgeRules.XML files that tell the purging system how to work. PurgeRules files are named according to <plugin>PurgeRules.xml. The purging can be based on either retention-time or a number of versions. PurgeRules files are located either in common_static/base/purge (baseline) or common_static/site/<your site>/purge/ (site overrides). There is a defaultPurgeRules.xml which is used if no purgeRules.xml file exists for a given plugin. The defaultPurgeRules.xml file essentially keeps one day's worth of files. These files are editable using the Localization Perspective, and are found in the D2D section.



PLUGIN	RULE
acars	For acars data, keep the 24 most recent versions of data at 1 hour intervals rounding to 1 hour
acarssounding	For acarssounding data, keep the 24 most recent versions of data at 1 hour intervals rounding to 1 hour
airep	For airep data, keep the 24 most recent versions of data at 1 hour intervals rounding to 1 hour
airmet	For airmet data, keep the 24 most recent versions of data
atcf	No Rules Specified. Using default.
aww	No Rules Specified. Using default.
binlightning	For binlightning data, keep the 36 most recent versions of data at 1 hour intervals rounding to 1 hour
bufrascat	For bufrascat data, keep the 12 most recent versions of data at 1 hour intervals rounding to 1 hour
bufrhdw	For bufrhdw data, keep the 33 most recent versions of data at 1 hour intervals rounding to 1 hour
bufrmos	For bufrmos data matching type=ETA, keep the 5 most recent versions of data at 1 hour intervals rounding to 1 hour
bufrmos	For bufrmos data matching type=AVN, keep the 8 most recent versions of data at 1 hour intervals rounding to 1 hour
bufrmos	For bufrmos data matching type=HPC, keep the 5 most recent versions of data at 1 hour intervals rounding to 1 hour

Data Purge web page at http://dx3:8080/uEngineWeb/dataPurge/dataPurge.html

PathKeys files are used to construct purge rules for specific subsets of data. Without a key defined, a purge rule operates on all data for a given plugin. With a key, metadata fields can be used to refine the data for the purge, like a particular model, for example. PathKeys files **are not** editable using the Localization Perspective, but site override versions can be created manually.

The AWIPS-2 purge runs hourly on the half hour and can also be run manually using a web page at http://dx3:8080/uEngineWeb/dataPurge/dataPurge.html. From this web page, all the current purge rules can be viewed as well.

PurgeRules.xml

The purging system is analogous to AWIPS-1, and the fields in the purgeRules.xml files correspond to the fields in the AWIPS-1 pipe-delimited purgeInfo.txt files as described in Chapter 11 of the AWIPS-1 System Manager's Manual.

AWIPS-1 SMM Chapter 11 relevant material:

Understanding individual purge table entries

The **purgeInfo.txt** file is where purging information is entered for data keys. This file has **#include** lines for radar, satellite, Redbook, grids, and local keys. Each line in the file normally supplies the purging information for one data key, but sometimes one line can refer to ranges of keys. Entries are formatted as follows:

```
//
// key | dir | r,w,c,i,l | scourPer | ver,period,delta,round | ...
```

• Where: [only the ver,period, delta, round section is relevant for AWIPS-2]

- **ver** - Number of versions to keep; defaults to zero, which means do not version purge.
- **period** - Max period between the current time and oldest time stamp of file to keep; defaults to zero which means do not time purge. A leading tilde (~) on the period means calculate from the latest file time stamp instead of the current time.
- **delta** - A file with a time stamp separated by less than this amount from the next newest file will not be kept; defaults to zero which means do not consider time separation. If there is a leading equal (=), only keep files an exact multiple of this delta time, if there is a leading tilde (~), only keep the one file closest to an exact multiple of this delta time.
- **round** - Round times by this amount before deciding whether to purge; defaults to zero, which means "do not round." The rounding time interacts with the **delta**, but not the **period**. If it has a leading plus sign (+), add the time instead of rounding by it. If consecutive files round to the same time, if one is kept, they will all be kept. There must be at least four vertical bars (delimiting 5 fields) fields for a valid entry.

There can be any number of **ver, period, delta or round** groups.

Time lengths are encoded as *dd-hh:mm:ss*, where **dd** is days, **hh** is hours, **mm** is minutes and **ss** is seconds. Valid examples include:

- **2-** for 2 days
- **:30** for 30 minutes
- **:4:30** for 4.5 minutes
- **4:** for 4 hours
- **1-18** for 1.75 days

For each **ver, period, delta or round** group, a file must pass all of the tests to be considered for retention. If there are multiple groups, a file must only pass the tests for one group to be retained.

Example:

<table>
<tr><th>AWIPS-1
METAR purging</th><th>AWIPS-2
Localization Perspective: D2D▸Purge Rules ▸obsPurgeRules.xml
/awips2/edex/data/utility/common_static/base/purge/obsPurgeRules.xml</th></tr>
<tr><td>Using the above information and looking at the default entry for METAR purging:
<code>82 | | | 2- | 34 | 38,,=3:00 | 42,,=6:00 | 50,,=24:00,+12:00</code>
[In this example, there are] four purge groups. For a file to be kept, it must be OK with at least one of the purge groups. Within a group, it must pass all of the specified tests.
− The first group represents a straight version purge to 34 versions.
− The second group means keep up to 38 files that are time-stamped exactly at 3-hour intervals; this means 4 additional files (12 hours in this case) beyond the 34 kept by the first rule.
− The third group means keep up to 42 files time-stamped exactly at 6-hour intervals (4 more files, or 1 day’s worth)
− The last group means keep up to 50 versions stamped exactly at 24-hour intervals, at 12Z (8 more days). Because the largest versions parameter is 50, the purging will reduce the set of kept files to no more than 50 when a purge operation is complete.</td><td>
<pre>
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<purgeRuleSet xmlns:ns2="group">
    <rule>
        <id>
            <pluginName>obs</pluginName>
            <key></key>
        </id>
        <versionsToKeep>34</versionsToKeep>
        <delta>=00-01:00:00</delta>
        <round>00-01:00:00</round>

    </rule>
    <rule>
        <id>
            <pluginName>obs</pluginName>
            <key></key>
        </id>
        <versionsToKeep>38</versionsToKeep>
        <delta>=00-03:00:00</delta>
        <round>00-01:00:00</round>
    </rule>
    <rule>
        <id>
            <pluginName>obs</pluginName>
            <key></key>
        </id>
        <versionsToKeep>42</versionsToKeep>
        <delta>=00-06:00:00</delta>
        <round>00-01:00:00</round>
    </rule>
    <rule>
        <id>
            <pluginName>obs</pluginName>
            <key></key>
        </id>
        <versionsToKeep>50</versionsToKeep>
        <delta>=01-00:00:00</delta>
        <round>+00-12:00:00</round>
    </rule>
</purgeRuleSet>
</pre>
The above file was replaced in OB13.4.1 with the file below. The above file is shown here to illustrate how the AWIPS-1 purge rules were transferred to AWIPS-2. The updated file reduced the number of retained metar HDF files.
<pre>
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<purgeRuleSet>
    <defaultRule>
        <versionsToKeep>34</versionsToKeep>
        <delta>=00-01:00:00</delta>
        <round>00-01:00:00</round>
    </defaultRule>
    <defaultRule>
        <versionsToKeep>15</versionsToKeep>
        <delta>=00-03:00:00</delta>
        <round>00-01:00:00</round>
    </defaultRule>
    <defaultRule>
        <versionsToKeep>11</versionsToKeep>
        <delta>=00-06:00:00</delta>
        <round>00-01:00:00</round>
    </defaultRule>
    <defaultRule>
        <versionsToKeep>10</versionsToKeep>
        <delta>=01-00:00:00</delta>
        <round>+00-12:00:00</round>
    </defaultRule>
</purgeRuleSet>
</pre>
</td></tr>
</table>

PathKeys.xml

Plugin-specific PathKeys.xml files exist in …./common_static/base/path. They define keys that can be used in PurgeRules.xml files to make a purge rule specific to a subset of data. In the example below, the key info.datasetId must be defined in gridPathKeys.xml for the purge system to use it to specify different purge rules for different models. The key is a metadata field. In this case, it identifies a common ID for each model.

Localization Perspective: D2D▸Purge Rules ▸gridPurgeRules.xml
/awips2/edex/data/utility/common_static/base/purge/gridPurgeRules.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<purgeRuleSet>
    <key>info.datasetId</key>
    <!-- Default rule for data not covered by any rules,
         2 versions up to 7 days-->
    <defaultRule>
        <versionsToKeep>2</versionsToKeep>
        <period>07-00:00:00</period>
    </defaultRule>
    <!-- Purge rule for the NAM80 (ETA) model -->
    <rule>
        <keyValue>ETA</keyValue>
        <versionsToKeep>2</versionsToKeep>
    </rule>
    <!-- Purge rule for the RUC80 (RUC) model -->
    <rule>
        <keyValue>RUC</keyValue>
        <versionsToKeep>8</versionsToKeep>
    </rule>
    <!-- Purge rule for the GFS180 (AVN) model -->
    <rule>
        <keyValue>AVN</keyValue>
        <versionsToKeep>2</versionsToKeep>
    </rule>
    <!-- Purge rule for the LAPS model -->
    <rule>
        <keyValue>LAPS</keyValue>
        <versionsToKeep>30</versionsToKeep>
    </rule>
    <!-- Purge rule for the NGM80 (NGM) model -->
    <rule>
        <keyValue>NGM</keyValue>
        <versionsToKeep>2</versionsToKeep>
    </rule>
    <!-- Purge rule for the gfsLR (MRF) model -->
    <rule>
        <keyValue>MRF</keyValue>
        <versionsToKeep>2</versionsToKeep>
    </rule>
    <!-- Purge rule for the MSAS model -->
    <rule>
```

/awips2/edex/data/utility/common_static/base/path/gridPathKeys.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<pathKeySet>
        <pathKey>
                <key>info.datasetId</key>
                <order>0</order>
        </pathKey>
        <pathKey>
        <key>info.level.masterLevel.name</key>
        <order>0</order>
    </pathKey>
</pathKeySet>
```

Example for Satellite Data

It's possible to define multiple keys for a given purge rule. The satellite data purge rules file is an example because it keys off both the sector and the observing element of the satellite.

D2D▸Purge Rules ▸satellitePurgeRules.xml /awips2/edex/data/utilitycommon_static/ base/purge/satellitePurgeRules.xml (excerpt)	/awips2/edex/data/utility/common_static/ base/path/satellitePathKeys.xml

```
<?xml version="1.0" encoding="UTF-8" standalone="yes"?>
<purgeRuleSet>
    <key>sectorID</key>
    <key>physicalElement</key>
    <defaultRule>
        <versionsToKeep>24</versionsToKeep>
    </defaultRule>
    <rule>
        <keyValue>Alaska National</keyValue>
        <keyValue>Imager 11 micron IR</keyValue>
        <versionsToKeep>40</versionsToKeep>
    </rule>
    <rule>
        <keyValue>Alaska National</keyValue>
        <keyValue>Imager 6.7-6.5 micron IR (WV)</keyValue>
        <versionsToKeep>48</versionsToKeep>
    </rule>
    <rule>
        <keyValue>Alaska National</keyValue>
        <keyValue>Imager Visible</keyValue>
        <versionsToKeep>36</versionsToKeep>
    </rule>
    <rule>
        <keyValue>Alaska National</keyValue>
        <keyValue>Percent of Normal TPW</keyValue>
        <versionsToKeep>40</versionsToKeep>
    </rule>
    <rule>
        <keyValue>Alaska National</keyValue>
        <keyValue>Rain fall rate</keyValue>
        <versionsToKeep>48</versionsToKeep>
    </rule>
    <rule>
        <keyValue>Alaska National</keyValue>
        <keyValue>Sounder Based Derived Precipitable Water (PW)</keyValue>
        <versionsToKeep>48</versionsToKeep>
    </rule>
    <rule>
        <keyValue>Alaska Regional</keyValue>
        <keyValue>Imager 11 micron IR</keyValue>
        <versionsToKeep>96</versionsToKeep>
    </rule>
    <rule>
        <keyValue>Alaska Regional</keyValue>
        <keyValue>Imager 12 micron IR</keyValue>
```

```
<pathKeySet>
        <pathKey>
                <key>sectorID</key>
                <order>0</order>
        </pathKey>

        <pathKey>
                <key>physicalElement</key>
                <order>0</order>
        </pathKey>
</pathKeySet>
```

Example:

Consider this scenario. You work at a WFO where you have a local model that is in grib format. It successfully ingests into AWIPS-2, but you wish to set the data purging for this model so that the past 5 runs are kept by the system, except for the analysis and 3-hour forecast, for which you wish to keep 10 runs. Do to this, you would create and/or modify site versions of the PurgeRules.xml and PathKeys.xml files. Because the data are in grib format, the particular files to modify would be:

- edex_static/site/PAH/purge/gridPurgeRules.xml
    - add two rules: one for the 5 files and one for the 10 files
- common_static/site/PAH/path/gridPathKeys.xml
    - add a key based on the forecast hour