ATTACHMENT D - Installing Test Data

D.1 NOTE: The AWIPS II TO8 release is localized for KOAX, so the provided KOAX data sets are required for testing and training at this time. Localization for all sites is handled in a later TO release, so we have to work around the limited capability to verify the provided features.

D.2 This procedure describes installing and loading test data. A script is provided to copy data into the EDEX data ingest directories to process the test data. The user owning the EDEX services (awips) should perform the steps in this procedure.

D.2.1 OAX SBN Data

D.2.1.1 The gz'ed tarballs of OAX SBN data are provided for testing the system. See the AWIPS II TO8 Software Distribution Document to obtain the KOAX data. Each tarball contains an hour's worth of data. Together, the tarballs represent consecutive hours of data. Tarball names use time-stamps indicating the most recent time-stamp of any file in that data set (e.g., a tarball with the string "1339" means the tarball contains data collected from roughly 1239 - 1339z). The uncompressed data is about 50% larger than the tarball size.

D.2.2 Directions for use

D.2.2.1 As the user that owns the EDEX services (awips), find a directory to hold the data from the tarballs and be sure that the file system has sufficient space to hold the unpacked data. You do not need to unpack all the tarballs. Unpack as many as you want. The directory that you unpack the tarballs into must not contain "unrelated" subdirectories. That is, all subdirectories must be only those directories that are created after you have unpacked the tarballs. Say you have a directory "/data/source" with sufficient available disk space. Be sure that /data/source is empty to begin with.

D.2.2.2 Unpack the tarball 0303_1339.tgz into /data/source.

D.2.2.3 You will then have one subdir in /data/source: "0303_1339".

D.2.2.4 You may unpack the other tarballs into /data/source if you like - but it's not necessary. Be sure that you know the absolute path to the edex sbn file endpoints (ADE ingest file endpoints). Say you installed edex into /awips (root of the edex install). The absolute path to the sbn target dir would be "/awips/edex/opt/data/sbn".

D.2.2.5 Run the script "ingestData.sh" to read data from your data source into your target directory.

D.2.2.6 Using the above information, you would run (using default speed = 1):
ingestData.sh /data/source /awips/edex/opt/data/sbn

D.2.2.7 **Detailed discussion of usage:**

D.2.2.7.1 USAGE: ingestData.sh <root of data src dir> <root of target sbn dir> [speed]

D.2.2.7.1.1 First two arguments are mandatory and must be directory names.

D.2.2.7.1.2 "root of data source directory" is the absolute path of the directory that contains one or more subdirectories consisting of raw one-hour SBN data.

D.2.2.7.1.3 "root of target sbn directory" is the absolute path of the sbn directory in the edex runtime containing the sbn file endpoints (subdirs for sbn data ingest).

D.2.2.7.1.4 Third argument is optional and is the speed or rate of data ingest.

D.2.2.7.1.4.1 Note that the value for speed: 1 <= integer <= 10

D.2.2.7.1.4.2 If left blank, speed defaults to 1. Each integer given for speed represents the divisor to be used with the dividend 60 (60 seconds) to produce a value representing the number of seconds of real time to process one minute of raw source data. Reasons why one would want to accelerate rate of data ingest: performance testing, system loading, impatience, etc.

D.2.2.7.1.4.3 For example, if speed is 1, the script will take 60 seconds of real time to ingest/copy one minute of raw source data. This attempts to simulate the actual SBN ingest data rate. If speed is 2, the script will take 30 seconds of real time to process one minute of raw source data. And so on. At low speed settings, the ingest (copying) may not be uniform over each time interval (the copying will occur at the beginning of interval and may finish well ahead of the end of the time interval). In theory, if speed is set to 10, the script will take only 6 seconds to process/copy one minute of raw source data. There are practical limits to the speed setting which are imposed by system resources and the amount of time to simply traverse the algorithm in the script. On most machines, a true speed of 10 may not be achievable.

D.2.2.7.1.5 **EXAMPLE:**

```
ingestData.sh /data/source /awips/edex/opt/data/sbn 2
```

D.2.2.7.1.6 In this example, the script will copy raw data files from subdirs in /data/source to edex sbn file endpoints located in /awips/edex/opt/data/sbn with speed=2. A speed of 2 means that each minute of raw source data will be copied to the endpoints in 30 seconds (so data ingest will be double the rate of actual sbn ingest from the SBNCP).

D.2.2.7.2 Here is what /data/source should look like - `ls -l /data/source`

drwxr-xr-x 13 root root 4096 Mar 3 13:39 0303_1339

drwxr-xr-x 13 root root 4096 Mar 3 14:43 0303_1441

drwxr-xr-x 13 root root 4096 Mar 3 15:45 0303_1545

drwxr-xr-x 13 root root 4096 Mar 3 16:48 0303_1647

drwxr-xr-x 13 root root 4096 Mar 3 17:52 0303_1750

D.2.2.7.2.1 Each of the above directories 0303_1339, 0303_1441, etc contains about one hour's worth of raw sbn data. Each directory must have subdirs named 'sat', 'radar', etc which contain sbn data for that datatype. That is,

D.2.2.7.2.1.1 `ls -l /data/source/0303_1339`

drwxrwxrwx 2 root root 4096 Mar 3 13:30 airep/

drwxrwxrwx 2 root root 4096 Mar 3 13:39 binlightning/

drwxrwxrwx 2 root root 8192 Mar 3 13:35 bufrua/

drwxrwxrwx 2 root root 98304 Mar 3 13:36 grib1/

drwxrwxrwx 2 root root 4096 Mar 3 13:38 grib2/

drwxrwxrwx 2 root root 8192 Mar 3 13:38 metar/

drwxrwxrwx 2 root root 12288 Mar 3 13:35 pirep/

drwxrwxrwx 2 root root 65536 Mar 3 13:39 radar/

drwxrwxrwx 2 root root 4096 Mar 3 13:32 sat/

drwxrwxrwx 2 root root 28672 Mar 3 13:38 sfcobs/

drwxrwxrwx 2 root root 8192 Mar 3 13:39 taf/

D.2.2.7.3 Here is a listing of the target edex sbn directory -

D.2.2.7.3.1 `ls -l /awips/edex/opt/data/sbn`

drwxrwxr-x 2 awips fxalpha 4096 Feb 9 19:46 airep

drwxrwxr-x 2 awips fxalpha 4096 Feb 9 19:48 binlightning

drwxrwxr-x 2 awips fxalpha 8192 Feb 9 16:30 bufrua

drwxrwxr-x 2 awips fxalpha 4096 Feb 8 14:12 gfe

drwxrwxr-x 2 awips fxalpha 684032 Feb 9 17:55 grib

drwxrwxr-x 2 awips fxalpha 8192 Feb 9 19:48 metar

drwxrwxr-x 2 awips fxalpha 12288 Feb 9 19:46 pirep

drwxrwxr-x 2 awips fxalpha 69632 Feb 9 19:48 radar

drwxrwxr-x 2 awips fxalpha 4096 Feb 8 14:12 recco

drwxrwxr-x 2 awips fxalpha 16384 Feb 9 19:48 sat

drwxrwxr-x 2 awips fxalpha 110592 Feb 9 19:48 sfcobs

drwxrwxr-x 2 awips fxalpha 4096 Feb 8 14:12 shef

drwxrwxr-x 2 awips fxalpha 12288 Feb 9 19:48 taf

drwxrwxr-x 2 awips fxalpha 4096 Feb 8 14:12 test

drwxrwxr-x 2 awips fxalpha 4096 Feb 8 14:12 text

D.2.2.7.4 Note that you can stop the script at any time using Ctl-C if running interactively or kill (of the pid of the script process) if running in the background.

D.2.2.7.5 Log file is created in $HOME/.SBN.read.log

D.2.2.7.6 Scratch dir is created in $HOME/.SBN

D.2.2.7.7 **Assumptions:**

D.2.2.7.7.1 The user running this script has write access to the EDEX sbn file endpoints (target sbn dir) and the EDEX services are able to move the files out of the sbn dir and into processing. This will not be a problem if the user running this script is also the owner of the edex service processes. The local machine uses a 24-hour clock and user is running this script entirely between 00:00 - 23:59 on one calendar day. You should not need to set GMT on your local machine. The raw source data does not span multiple 24-hour days. That is, all the data to be read in falls in a single 24-hour day (0z to 24z). This is default. Each data dir contains no more than one hour's worth of data. This is default. System load will affect the rate of data reading. On a busy system or a system with "limited" resources, high rates of data reading may

not be possible. Thus, for "high" rates of data reading, the speed argument may be more theoretical than real. This script has been "tested" on RHEL4 and Fedora 6 and 7 linux distributions. It will probably work on most modern linux distributions.