

Creating a (distributed) MeasurementSet

Ger van Diepen
10 Februari 2011

Introduction

BBS works on distributed MeasurementSets. Normally they are created by the LOFAR online system, but for simulation purposes it must be possible to create them from scratch. Furthermore the MeqTrees system is often used to do simulations for which a new MeasurementSet is needed.

Makems creates a MeasurementSet by filling in all metadata, while the data columns are filled with zeroes. Optionally the columns like MODEL_DATA needed for the CASA/AIPS++ imager are created too in a way they are recognized by the imager.

The MeasurementSet can be partitioned. The parts can be created in a distributed way on a cluster, so it is directly suitable for BBS. The distribution is done using the MWCommon distribution framework, in particular the script `startdistproc`.

The VDS files describing a (partitioned) MeasurementSet are created too.

Running makems

A non-distributed (but optionally partitioned) MeasurementSet can be created using:

```
makems parsetfile
```

If no parsetfile argument is given, it defaults to `makems.cfg`. The parset file tells how the MeasurementSet has to be created. It is described in the next section.

A distributed MeasurementSet can be created using:

```
makemsdistr [-hfn hfn] [parsetfile [clusterdescfile  
[dir1 dir2, ...]]]
```

<code>-hfn</code>	can be used to give the name of the 'machinefile' to use. Usually the default name is fine.
<code>parsetfile</code>	name of the parset file.
<code>clusterdescfile</code>	if not given or empty, the <i>ClusterDescName</i> keyword in the parset file is used.
<code>dir1 ...</code>	can be used to tell where to create the distributed parts. It overrides the <i>Dirs</i> keyword in the parset file described below.

Using the `startdistproc` mechanism distributed makems processes will be started. Each process creates its MeasurementSet part and VDS file in the given node:directory. At the end, all VDS files are combined in a global VDS file.

`makems` can partition the MeasurementSet using the *NParts* keyword. In that case all parts are created in the same directory. Note that it also creates the VDS files and global VDS file (even if *NParts=1*).

The total number of rows in the MeasurementSet is the product of the number of times, fields, bands, and baselines. The rows are written in order of time, field, band, and baseline. Each part contains `nbands/nparts` bands and each band contains `nfreqs/nbands` channels.

ParameterSet File

Makems(distr) is controlled by a so-called parset file telling how the MeasurementSet has to be created.

A parset file is a text file in which parameters are given in 'key=value' format. A value can be a vector of values that needs to be enclosed in square brackets and separated by commas. A detailed description of the parset format can be found in the ParameterSet document.

The following keywords are used by makems(distr):

NParts	number of parts in which the MeasurementSet is to be split If 0, one part is created and no '_p' suffix is added to the MS name. For makemsdistr this keyword must always be defined.
NBands	number of subbands to use; must be divisible by NParts
NFrequencies	total number of frequencies; must be divisible by NBands
StartFreq	start frequency in Hz. If one value given, all bands are adjacent. Otherwise a vector of length NBands with start freq per band.
StepFreq	frequency step in Hz. If one value given, all bands have the same step. Otherwise a vector of length NBands with step freq per band.
StartTime	start time like 2000/08/03/13:22:30
StepTime	time step in seconds
NTimes	number of time steps
RightAscension	vector of J2000 right ascensions like 16:38:28.205274248 or 2.1rad
Declination	vector of J2000 declinations like 62.34.44.313606568 or 62deg RightAscension and Declination give the position of the beam center(s) in the FIELD subtable.
TileSizeFreq	data tile size in frequency (for advanced users); default no tiling.
TileSize	data tile size in kbytes (for advanced users); default 32 KBytes.
WriteAutoCorr	T = also write autocorrelations (default F)
WriteImagerColumns	T = write the columns needed for the CASA imager (default F)
AntennaTableName	name of the antenna table to use; it can be the ANTENNA subtable of an existing MeasurementSet. The antenna table must at least contain the POSITION column containing the antenna (station) positions in ITRF coordinates. Other columns (as defined in the MS ANTENNA definition) will be used if present. Otherwise they'll get a default (LOFAR related) value.
ClusterDescName	optional name of the clusterdesc file to be used by makems. For makemsdistr it can also be given on the command line.
Dirs	vector of directories (mountpoints) where to create the distributed MeasurementSet parts. It can be overridden by command line arguments in makemsdistr. The names used should match the mountpoints as used in the ClusterDesc file used. Thus if needed, a directory name must be prefixed with a node name to make it unique. The number of directories given does not need to match NParts. If less, the directory list is recycled.
MSName	the name of the MS to create; in principle this should only be the base name, although it can be prepended with one or more directories. The full name of each part is constructed by prepending it with the relevant entry from the Dirs. To make each part name unique, the part number is made part of the name, which is done by replacing %d with the part number. If no %d is given, _p%d is added to the name (unless NParts=0).

VDSPath	The path where to put the VDS files. In case of a distributed MeasurementSet, it must be on a file system accessible to all nodes. It defaults to the directory part of the MSName, which is usually fine for makems. For makemsdistr it needs to be filled in.
FlagColumn	If given, an integer flag column is made; in the future column FLAG will be mapped to it.
NFlagBits	Number of bits in FlagColumn.
MapFlagBits	If an integer flag column is made, tell if BitFlagsEngine<T> has to be used to map the flag bits to the Boolean FLAG column. Default is False meaning that an independent FLAG column is made.

Appendix A Examples

Create a simple non-distributed MeasurementSet with one band of 64 frequencies.

```
makems my.parset

NParts=1
NBands=1
NFrequencies=64
StartFreq=1170e6
StepFreq=100000
StartTime=2000/08/03/13:22:30
StepTime=10
NTimes=14
RightAscension=16:38:28.205274248
Declination=62.34.44.313606568
TileSizeFreq=8
TileSizeRest=10
WriteAutoCorr=T
AntennaTableName=~ /WSRT_ANTENNA
MSName=test.MS
VDSPath=.
```

Distributed MS consisting of 4 parts, each part containing 1 band of 16 frequencies. The MS parts are created on the given file systems as /lifs001/diepen/test.MS, etc. Note that MSName is a relative path name which gets appended to /lifs001, etc.

```
makemsdistr my.parset /app/clusterdesc.lioffen
                /lifs001 /lifs002 /lifs003 /lifs004

NParts=4
NBands=4
NFrequencies=64
StartFreq=30e6
StepFreq=10000
StartTime=2008/08/03/13:22:30
StepTime=10
NTimes=14
RightAscension=16:38:28.205274248
Declination=62.34.44.313606568
TileSizeFreq=8
TileSizeRest=10
AntennaTableName=~ /LOFAR_ANTENNA
MSName=diepen/test.MS
VDSPath=.
```