
Subject: Re: reading data-files from archive?

Posted by [Maarten\[1\]](#) on Fri, 12 Oct 2007 11:11:22 GMT

[View Forum Message](#) <> [Reply to Message](#)

On Oct 12, 10:03 am, Sven Utcke <utcke+n...@informatik.uni-hamburg.de> wrote:

- > is IDL capable of reading data out of archive-files? What I mean by
- > that, given a somfile.zip or somefile.arc, could I somehow access the
- > /some/path/to/image.tiff stored inside the somfile.zip? If not, would
- > it be easy to add this functionality?

While the file routines will deal with individual gzipped files on the fly, I'm not aware of the ability to use zip, arc or tar archives on the fly. Of course you can call the system tools (spawn & friends) to unzip/untar the data in a temporary location, but I can see why you'd want to avoid that.

- > Background: we're doing tomography here, and expect to create some PB
- > of data each year, starting late 2008/early 2009. Unfortunately the
- > only system able to deal with that much data (dCache) is only well
- > suited for files between 100MB and 2GB, i.e. too big for a single
- > slice, and too small for a stack of slices. Using archives would be
- > one way around this limitation...

I think you'll be better off using scientific data formats, in particular hdf. The files can be up to 2 GB *hdf4) or way larger (hdf5). Note that the formats are very different: hdf4 has data types for image data, etc., while hdf5 basically stores arrays. In my opinion hdf4 was over-engineered. For storing multiple items under potentially the same name (but a different location within the file), you're better off with hdf5.

It is easy to add metadata to the data-sets in hdf, although I would advise to duplicate at least some of the metadata in a database for easier retrieval later on.

Maarten
