



The HDF Group



Groups and Links



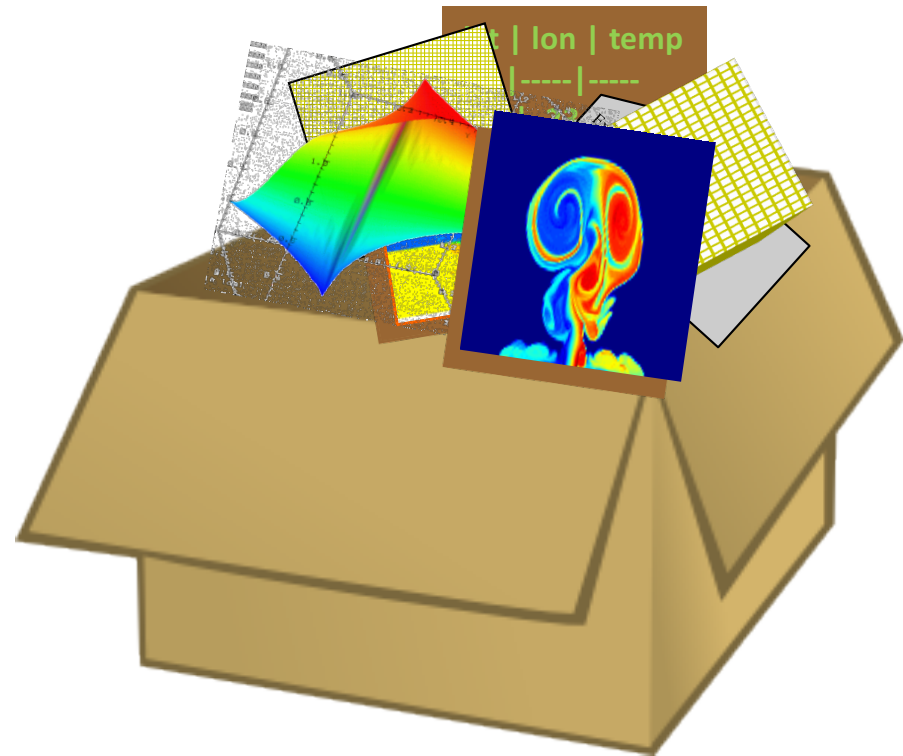
Goals

- Introduce to Groups and Links



HDF5 File

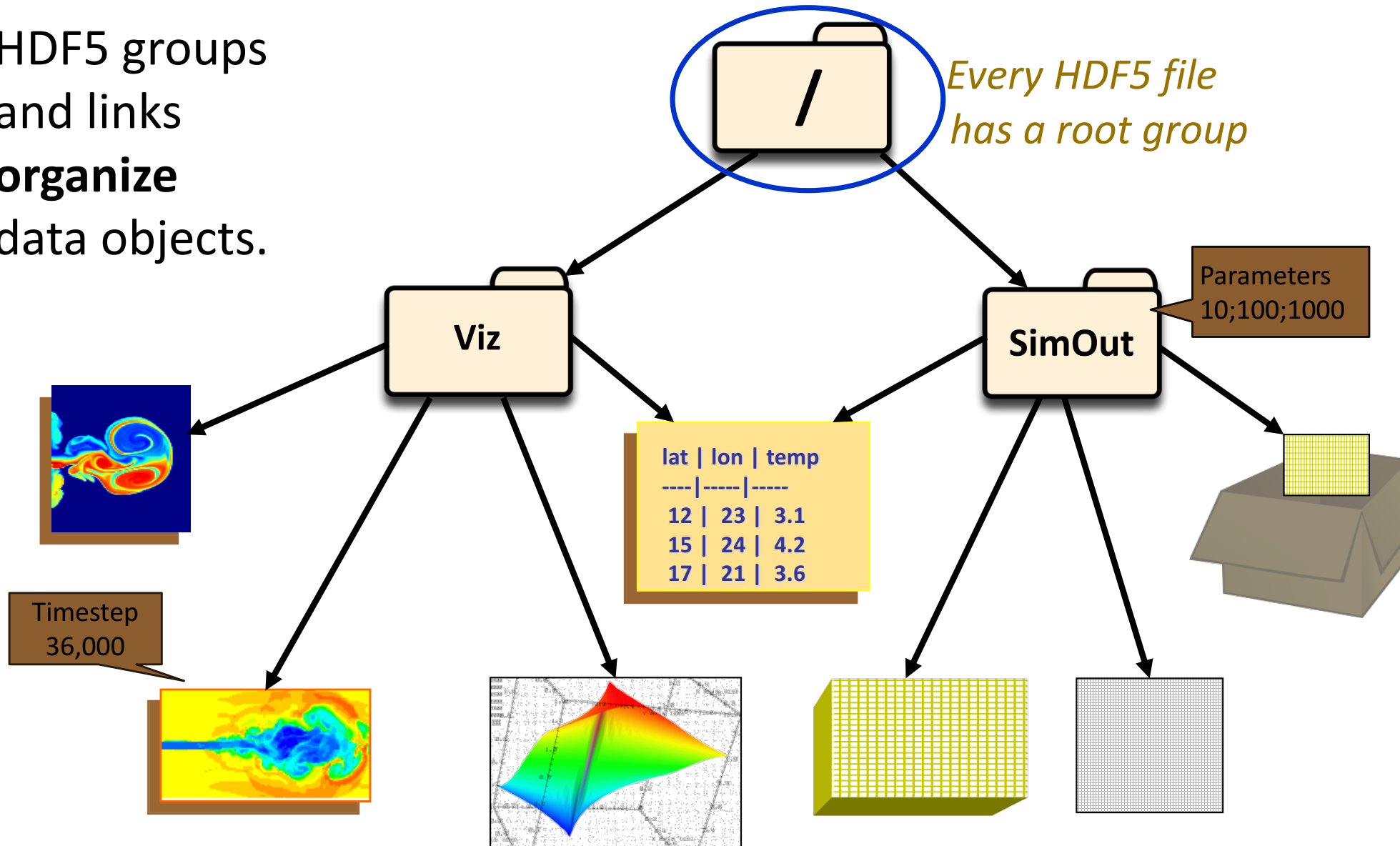
An HDF5 file is a **container** that holds data objects.





HDF5 File

HDF5 groups and links **organize** data objects.





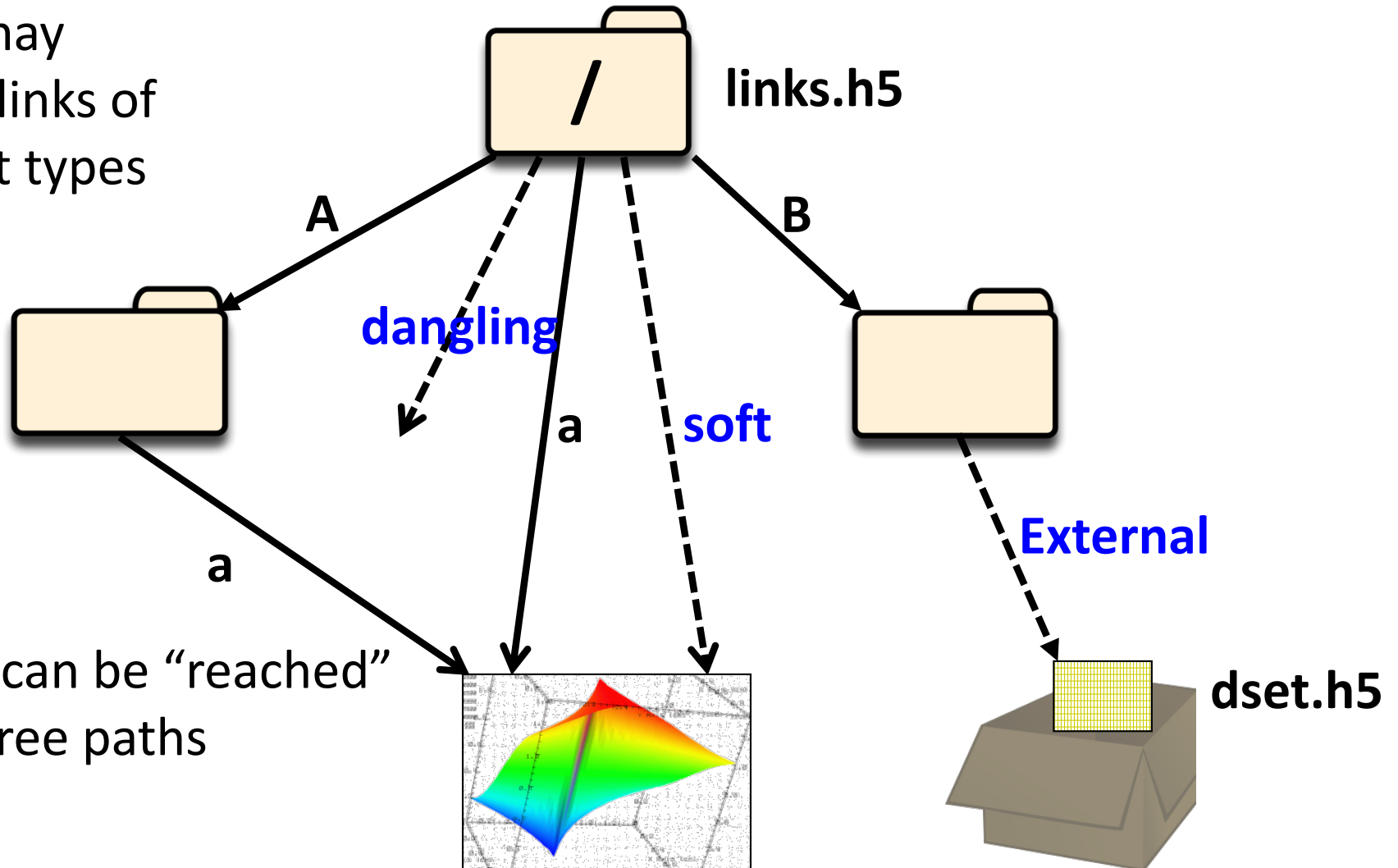
GROUPS AND LINKS



Groups and Links

- Groups are containers for links (graph edges)
- Links were added in 1.8.0
- Warning: Many APIs in H5G interface are obsolete - use H5L interfaces to discover and manipulate file structure

Group may contain links of different types



Dataset can be “reached” using three paths

/A/a

/a

/soft

Dataset is in a **different** file
/B/External



Link Names

- Name
 - Example: “A”, “B”, “a”, “dangling”, “soft”
 - Unique within a group; “/” are not allowed in names
- Best practices for interoperability with netCDF4
 - Names should be composed of letter, digits, and underscores
 - Names should begin with a letter
 - Use UTF-8 encoding by setting link creation property with `h5set_char_encoding_f`



Link Types

- Internally a link is represented by (name, value) pair; value defines link's type
- Type
 - Hard Link
 - Value is object's address in a file
 - Created automatically when object is created
 - Can be added to point to existing object
 - Soft Link
 - Value is a string , for example, “/A/a”, but it can be anything
 - Use to create aliases



Links Types

- Type
 - External Link
 - Value is a pair of strings , for example, (“dset.h5”, “dset”)
 - Use to access data in other HDF5 files
 - See `h5lcreate_external_f`
 - C only
 - For caching of files opened via external links see `H5Pset_elink_file_cache_size`
 - For location external files see `H5Pset_efile_prefix`



Links Properties

- Links Properties
 - ASCII or UTF-8 encoding for names
 - Create intermediate groups
 - Saves programming effort

- Example

```
CALL h5pcreate_f (H5P_LINK_CREATE_F,...)
```

```
CALL h5pset_create_intermediate_group_f ( lcpl_id, 1,... )
```

```
CALL h5gcreate (fid, "A/B", lcpl_id, ...)
```

- Group “A” will be created if it doesn’t exist



Operations on Links

- See H5L interface in Reference Manual
- Create
- Delete
 - Doesn't delete the object; use h5repack to release space
- Copy
- Iterate
- Check if exists (High-Level Library API)



Exercise

- In directory Demo-1 modify file.h5 using functions listed in H5L

https://support.hdfgroup.org/HDF5/doc/RM/RM_H5L.html

- Add group /C/D using example on slide 11
- Create a soft link /F that points to group D
- Copy file.h5 to t.h5
- Create external link in the Root group / to point to dataset /A in file t.h5



Groups Properties

- Creation properties
 - Type of links storage
 - Compact
 - Used with a few members (default under 8)
 - Dense (default behavior)
 - Used with many (>16) members (default)
 - Tunable size for a local heap
 - Save space by providing estimate for size of the storage required for links names
 - Can be compressed
 - Many links with similar names (XXX-abc, XXX-d, XXX-efgh, etc.)
 - Requires more time to compress/uncompress data



Example

- In directory EX-4 review and run
h5ex_g_compact.f90
h5ex_g_phase.f90



Groups Properties

- Creation properties
 - Links may have creation order tracked and indexed
 - Indexing by name (default)
 - A, B, a, dangling, soft
 - Indexing by creation order (has to be enabled)
 - A, B, a, soft, dangling



Example

- In directory EX-4 review and run
`h5ex_g_corder.f90`



Discovering HDF5 file's structure

- HDF5 provides C and Fortran 2003 APIs for recursive and non-recursive iterations over the groups and attributes
 - H5Ovisit and H5Literate (H5Giterate)



Example (EX-4)

`h5ex_g_visit_F03.f90` visits all objects in a file

```
$ h5dump -n h5ex_g_visit.h5
```

```
HDF5 "h5ex_g_visit.h5" {  
  FILE_CONTENTS {  
    group      /  
    group      /group1  
    dataset    /group1/dset1  
    group      /group1/group3  
    dataset    /group1/group3/dset2 -> /group1/dset1  
    group      /group1/group3/group4  
    group      /group1/group3/group4/group1  
    group      /group1/group3/group4/group1/group5 -> /group1  
    group      /group1/group3/group4/group2  
    group      /group2 -> /group1/group3  
  }  
}
```



Example (EX-4)

h5ex_g_iterate_F03.f90 iterates through links in a group; compare with h5dump output

```
$ h5dump -n h5ex_g_iterate.h5
```

```
HDF5 "h5ex_g_iterate.h5" {
```

```
FILE_CONTENTS {
```

```
group      /
```

```
dataset    /DS1
```

```
datatype   /DT1
```

```
group      /G1
```

```
dataset    /G1/DS2
```

```
dataset    /L1 -> /G1/DS2
```

```
}
```



Example (EX-4)

h5ex_g_travers_F03.f90 traverses the file

```
$ h5dump -n h5ex_g_traverse.h5
```

```
HDF5 "h5ex_g_traverse.h5" {
```

```
FILE_CONTENTS {
```

```
group      /
```

```
group      /group1
```

```
dataset    /group1/dset1
```

```
group      /group1/group3
```

```
dataset    /group1/group3/dset2 -> /group1/dset1
```

```
group      /group1/group3/group4
```

```
group      /group1/group3/group4/group1
```

```
group      /group1/group3/group4/group1/group5 -> /group1
```

```
group      /group1/group3/group4/group2
```

```
group      /group2 -> /group1/group3
```

```
}
```



Checking a path in HDF5

- HDF5 1.8.8 provides HL C and Fortran 2003 APIs for checking if paths exists
 - H5LTvalid_path (h5ltvalid_path_f)
 - Example: Is there an object with a path /A/B/C/d ?
 - TRUE if there is a path, FALSE otherwise



Hints

- Use latest file format (see `H5Pset_libver_bound` function in RM)
 - Save space when creating a lot of groups in a file
 - Save time when accessing many objects (>1000)
- Caution: Tools built with the HDF5 versions prior to 1.8.0 will not work on the files created with this property

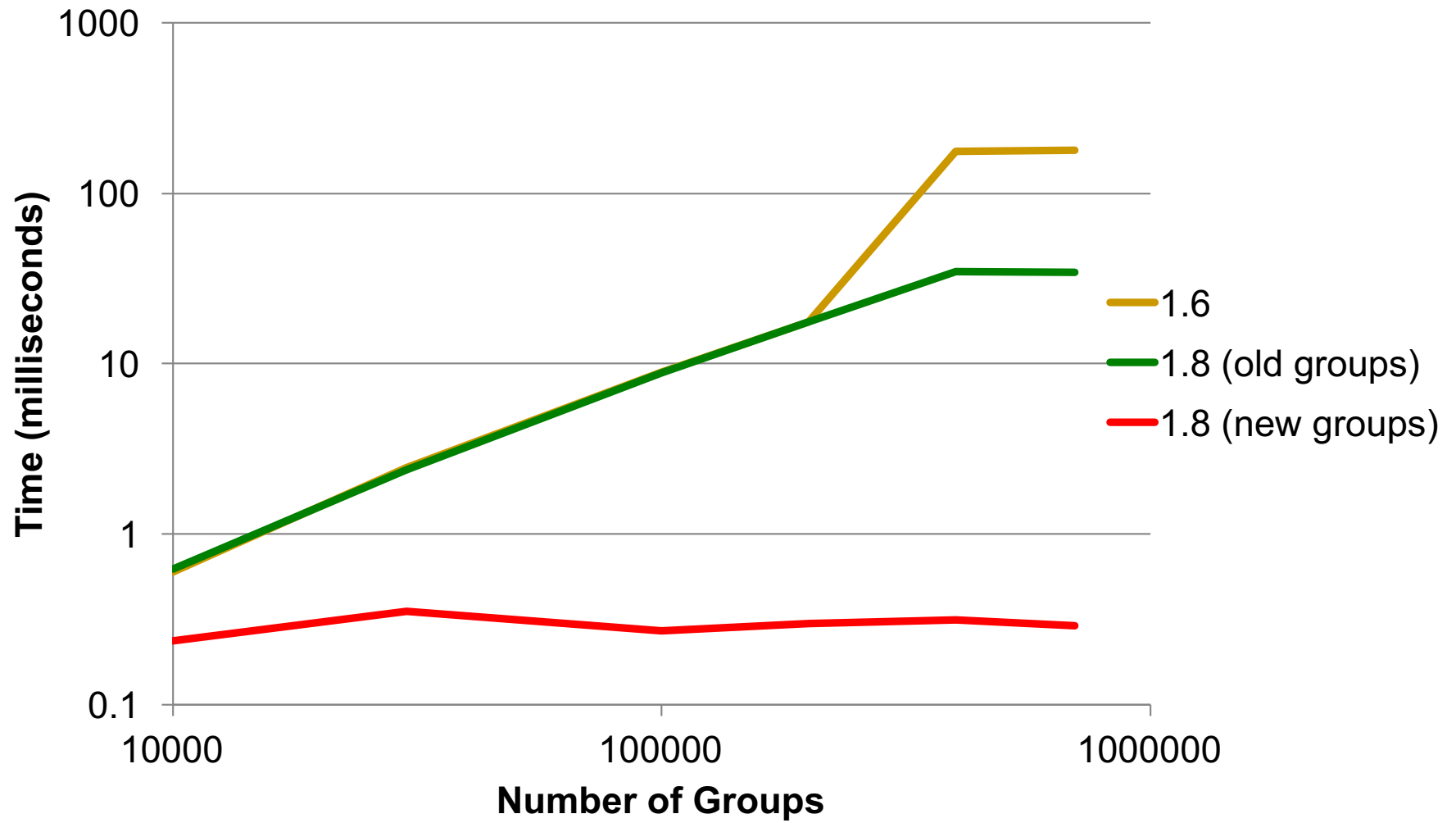


Informal Benchmark

- Create a file and a group in a file
- Create up to 10^6 groups with one dataset in each group
- Compare files sizes and performance of HDF5 1.8.1 using the latest group format with the performance of HDF5 1.8.1 (default, old format) and 1.6.7
- Note: Default 1.8.1 and 1.6.7 became very slow after 700000 groups

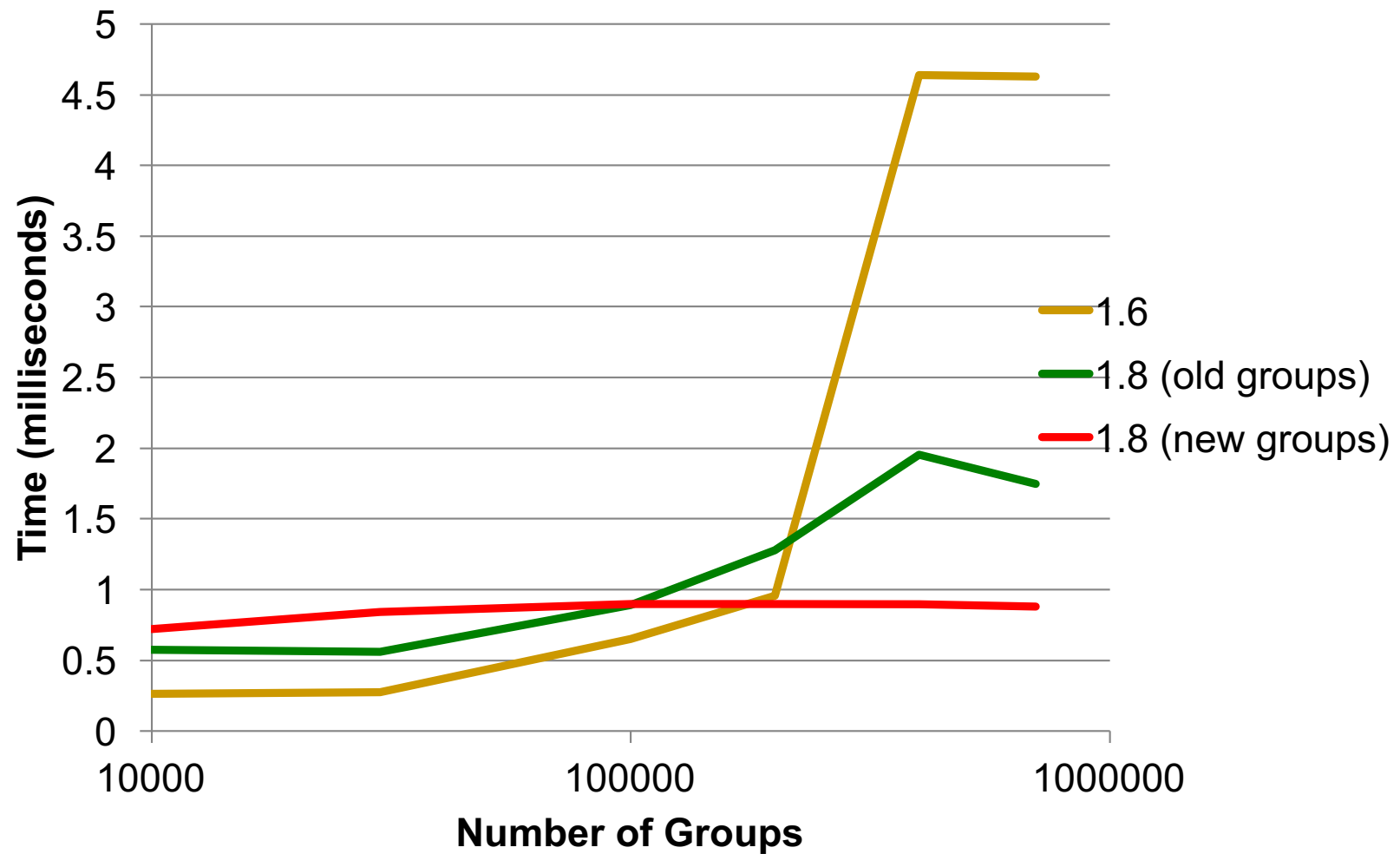


Time to Open and Read a Dataset



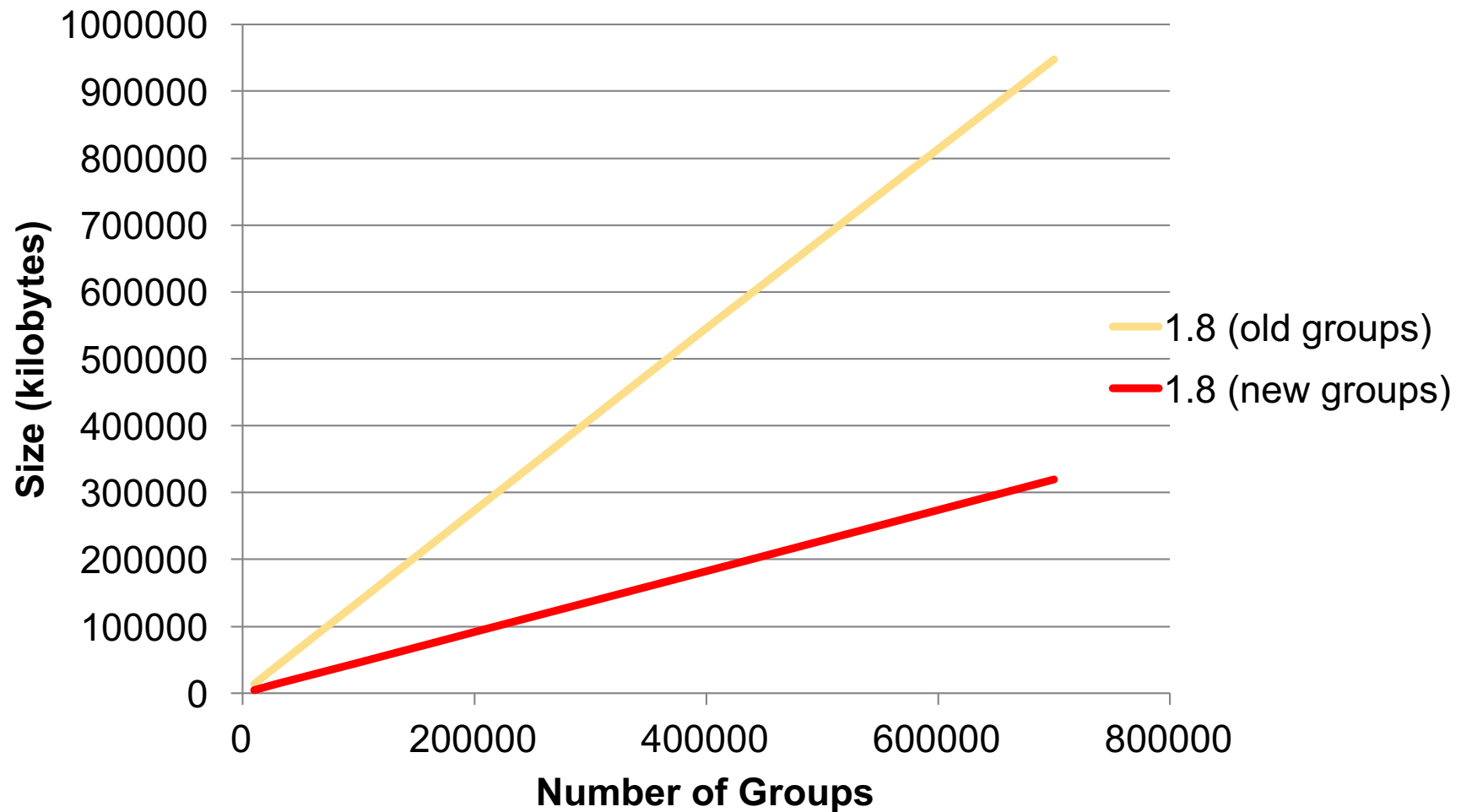


Time to Close the File





File Size





Thank You!

Questions?