HDF5 in L-Galaxies

RHYS POULTON

Supervisor: Prof Peter Thomas

University of Sussex

Overview

What is HDF5

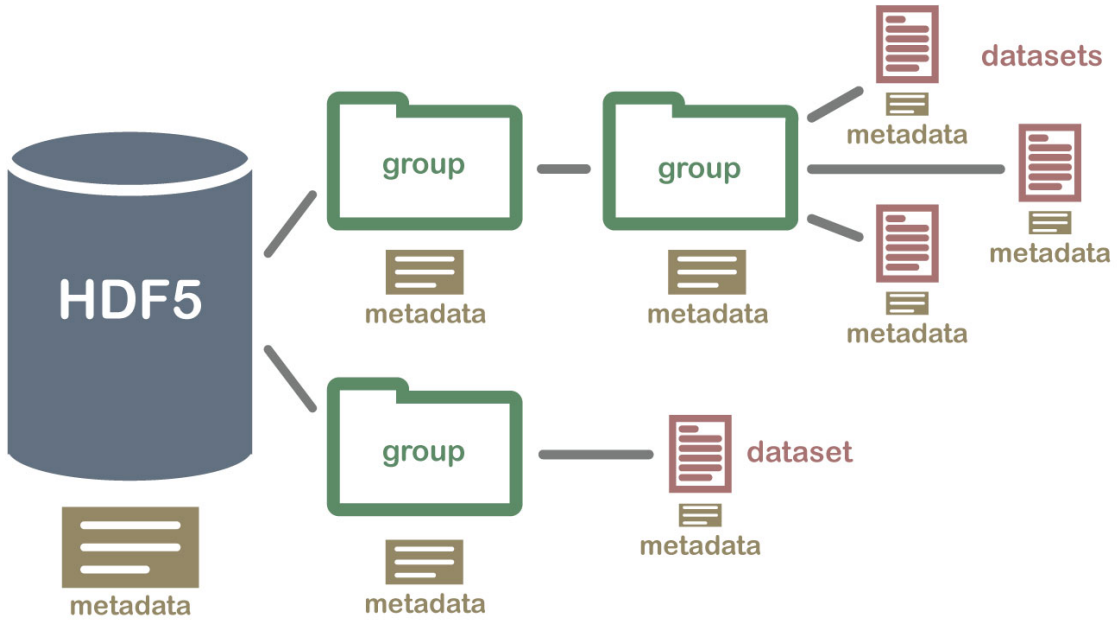
What can it do

What it looks like

What are the benefits/ current problems of using it.

What is HDF5

New type of data model and file format for storing and managing numerical data.



What can it do?

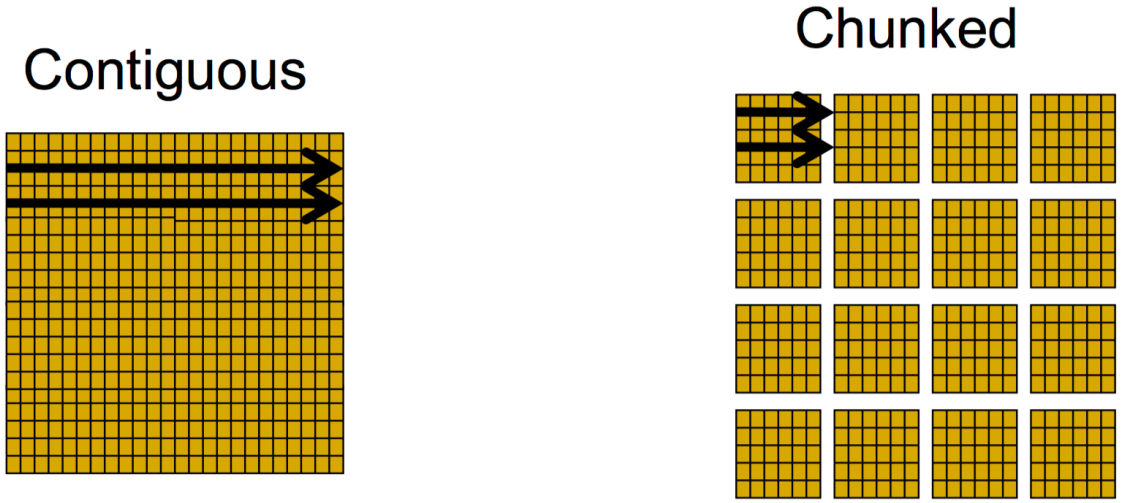
- Simple but versatile data model
- Unlimited size, extensibility and portability
- Unlimited variety of datatypes
- Flexible and efficient I/O
- Subsetting

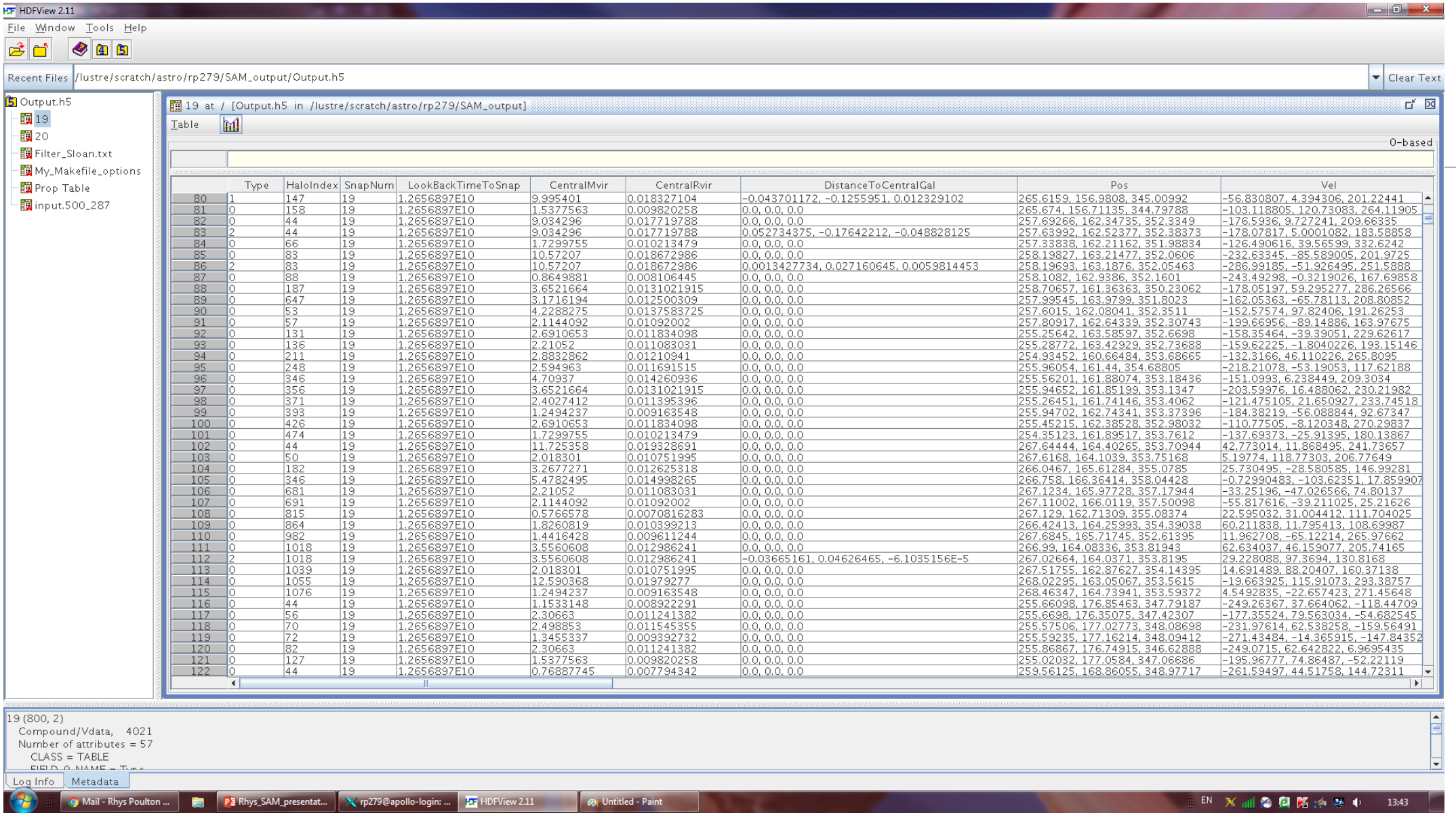
HDF5 chunking

- Data is stored in chunks of predefined size
- Two-dimensional instance may be referred to

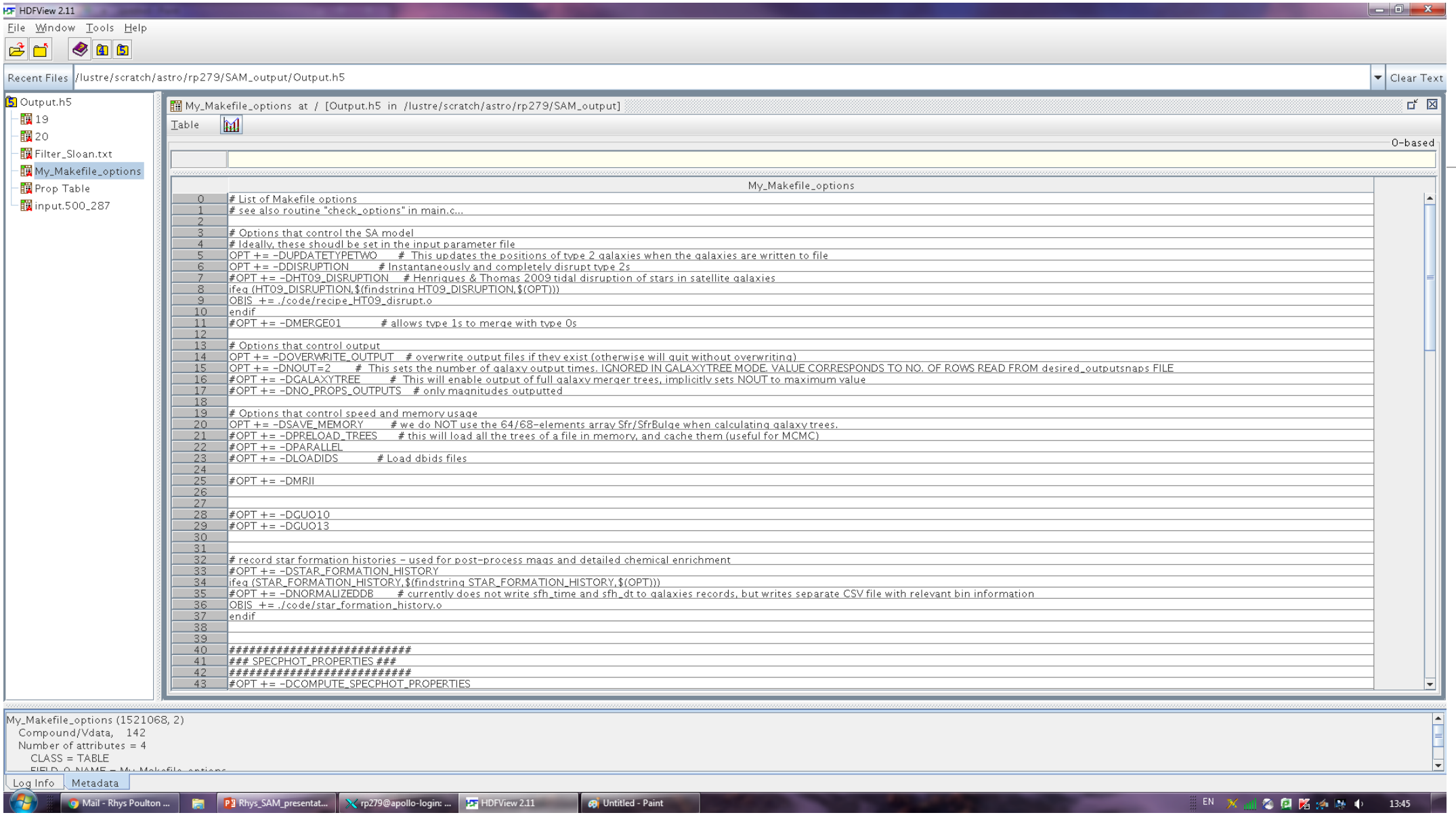
as data tiling

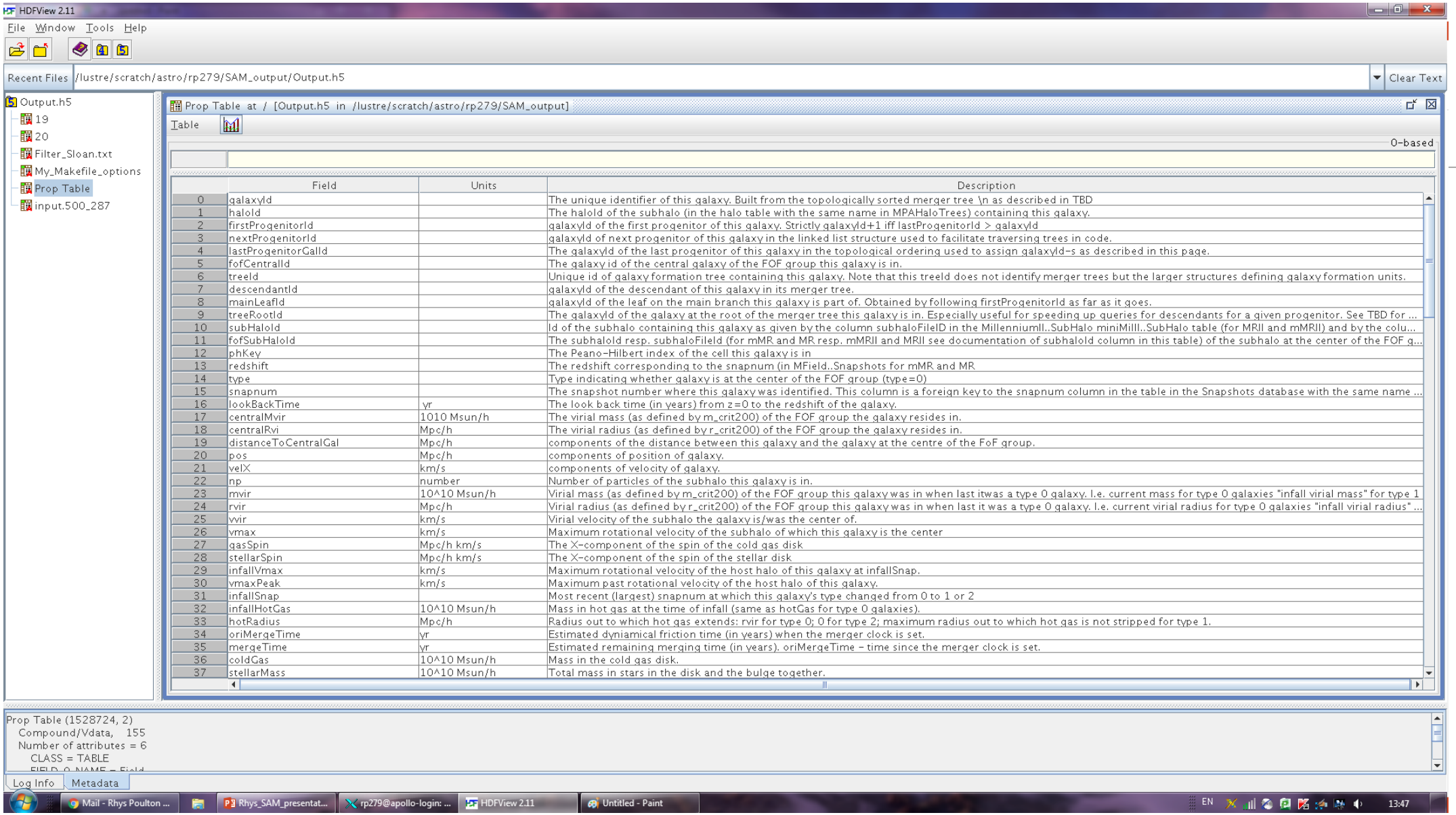
- HDF5 library always writes/reads the whole chunk



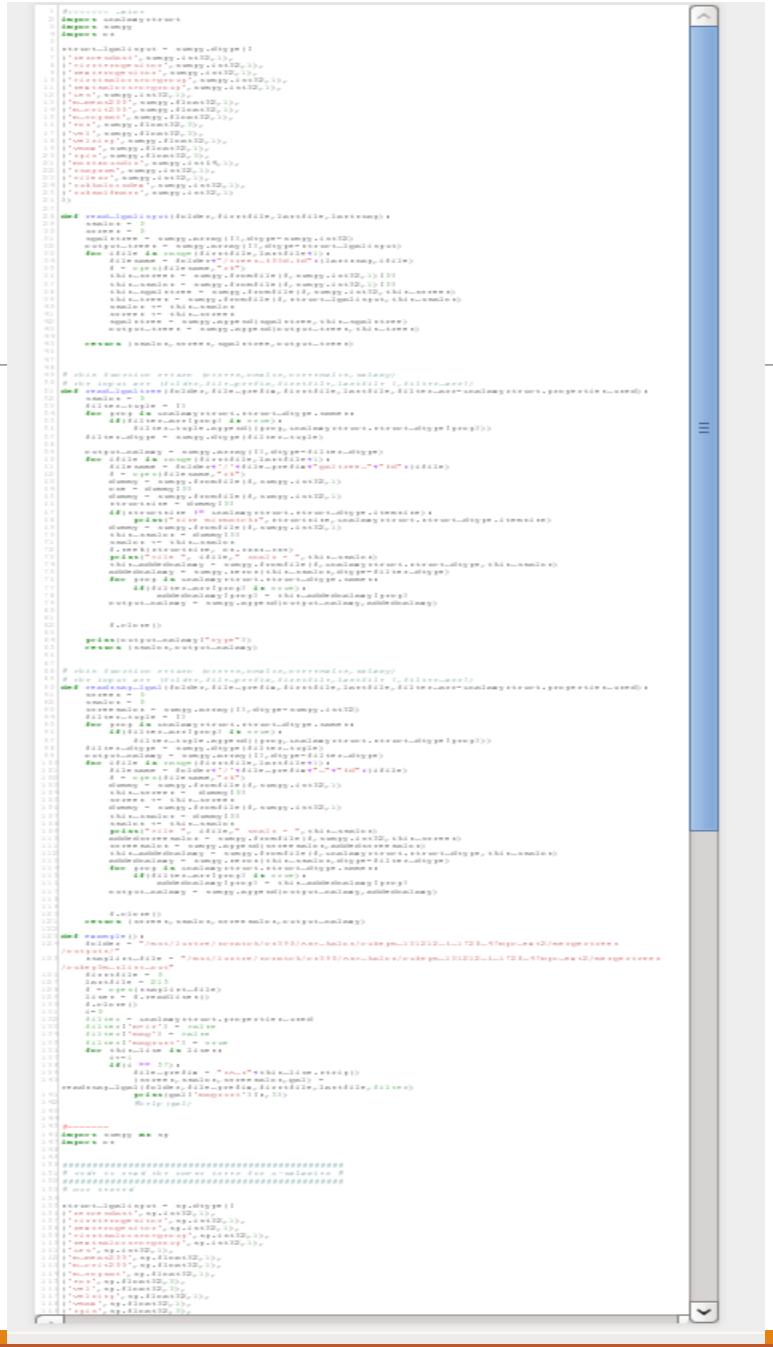


	Type	HaloIndex	SnapNum	LookBackTimeToSnap	CentralMvir	CentralRvir	DistanceToCentralGal	Pos	Vel
80	1	147	19	1.2656897E10	9.995401	0.018327104	-0.043701172, -0.1255951, 0.012329102	265.6159, 156.9808, 345.00992	-56.830807, 4.394306, 201.22441
81	0	158	19	1.2656897E10	1.5377563	0.009820258	0.0, 0.0, 0.0	265.674, 156.71135, 344.79788	-103.118805, 120.73083, 264.11905
82	0	44	19	1.2656897E10	9.034296	0.017719788	0.0, 0.0, 0.0	257.69266, 162.34735, 352.3349	-176.5936, 9.727241, 209.66335
83	2	44	19	1.2656897E10	9.034296	0.017719788	0.052734375, -0.17642212, -0.048828125	257.63992, 162.52377, 352.38373	-178.07817, 5.0001082, 183.58858
84	0	66	19	1.2656897E10	1.7299755	0.010213479	0.0, 0.0, 0.0	257.33838, 162.21162, 351.98834	-126.490616, 39.56599, 332.6242
85	0	83	19	1.2656897E10	10.57207	0.018672986	0.0, 0.0, 0.0	258.19827, 163.21477, 352.0606	-232.63345, -85.589005, 201.9725
86	2	83	19	1.2656897E10	10.57207	0.018672986	0.0013427734, 0.027160645, 0.0059814453	258.19693, 163.1876, 352.05463	-286.99185, -51.926495, 251.5888
87	0	88	19	1.2656897E10	0.8649881	0.008106445	0.0, 0.0, 0.0	258.1082, 162.9386, 352.1601	-243.49298, -0.3219026, 167.69858
88	0	187	19	1.2656897E10	3.6521664	0.0131021915	0.0, 0.0, 0.0	258.70657, 161.36363, 350.23062	-178.05197, 59.295277, 286.26566
89	0	647	19	1.2656897E10	3.1716194	0.012500309	0.0, 0.0, 0.0	257.99545, 163.9799, 351.8023	-162.05363, -65.78113, 208.80852
90	0	53	19	1.2656897E10	4.2288275	0.0137583725	0.0, 0.0, 0.0	257.6015, 162.08041, 352.3511	-152.57574, 97.82406, 191.26253
91	0	57	19	1.2656897E10	2.1144092	0.01092002	0.0, 0.0, 0.0	257.80917, 162.64339, 352.30743	-199.66956, -89.14886, 163.97675
92	0	131	19	1.2656897E10	2.6910653	0.011834098	0.0, 0.0, 0.0	255.25642, 163.58597, 352.6698	-158.35464, -39.39051, 229.62617
93	0	136	19	1.2656897E10	2.21052	0.011083031	0.0, 0.0, 0.0	255.28772, 163.42929, 352.73688	-159.62225, -1.8040226, 193.15146
94	0	211	19	1.2656897E10	2.8832862	0.01210941	0.0, 0.0, 0.0	254.93452, 160.66484, 353.68665	-132.3166, 46.110226, 265.8095
95	0	248	19	1.2656897E10	2.594963	0.011691515	0.0, 0.0, 0.0	255.96054, 161.44, 354.68805	-218.21078, -53.19053, 117.62188
96	0	346	19	1.2656897E10	4.70937	0.014260936	0.0, 0.0, 0.0	255.56201, 161.88074, 353.18436	-151.0993, 6.238449, 209.3034
97	0	356	19	1.2656897E10	3.6521664	0.0131021915	0.0, 0.0, 0.0	255.94652, 161.85199, 353.1347	-203.59976, 16.488062, 230.21982
98	0	371	19	1.2656897E10	2.4027412	0.011395396	0.0, 0.0, 0.0	255.26451, 161.74146, 353.4062	-121.475105, 21.650927, 233.74518
99	0	393	19	1.2656897E10	1.2494237	0.009163548	0.0, 0.0, 0.0	255.94702, 162.74341, 353.37396	-184.38219, -56.088844, 92.67347
100	0	426	19	1.2656897E10	2.6910653	0.011834098	0.0, 0.0, 0.0	255.45215, 162.38528, 352.98032	-110.77505, -8.120348, 270.29837
101	0	474	19	1.2656897E10	1.7299755	0.010213479	0.0, 0.0, 0.0	254.35123, 161.89517, 353.7612	-137.69373, -25.91395, 180.13867
102	0	44	19	1.2656897E10	11.725358	0.019328691	0.0, 0.0, 0.0	267.64444, 164.40265, 353.70944	42.773014, 11.868495, 241.73657
103	0	50	19	1.2656897E10	2.018301	0.010751995	0.0, 0.0, 0.0	267.6168, 164.1039, 353.75168	5.19774, 118.77303, 206.77649
104	0	182	19	1.2656897E10	3.2677271	0.012625318	0.0, 0.0, 0.0	266.0467, 165.61284, 355.0785	25.730495, -28.580585, 146.99281
105	0	346	19	1.2656897E10	5.4782495	0.014998265	0.0, 0.0, 0.0	266.758, 166.36414, 358.04428	-0.72990483, -103.62351, 17.859907
106	0	681	19	1.2656897E10	2.21052	0.011083031	0.0, 0.0, 0.0	267.1234, 165.97728, 357.17944	-33.25196, -47.026566, 74.80137
107	0	691	19	1.2656897E10	2.1144092	0.01092002	0.0, 0.0, 0.0	267.11002, 166.0119, 357.50098	-55.817616, -39.211025, 25.21626
108	0	815	19	1.2656897E10	0.5766578	0.0070816283	0.0, 0.0, 0.0	267.129, 162.71309, 355.08374	22.595032, 31.004412, 111.704025
109	0	864	19	1.2656897E10	1.8260819	0.010399213	0.0, 0.0, 0.0	266.42413, 164.25993, 354.39038	60.211838, 11.795413, 108.69987
110	0	982	19	1.2656897E10	1.4416428	0.009611244	0.0, 0.0, 0.0	267.6845, 165.71745, 352.61395	11.962708, -65.12214, 265.97662
111	0	1018	19	1.2656897E10	3.5560608	0.012986241	0.0, 0.0, 0.0	266.99, 164.08336, 353.81943	62.634037, 46.159077, 205.74165
112	2	1018	19	1.2656897E10	3.5560608	0.012986241	-0.03665161, 0.04626465, -6.1035156E-5	267.02664, 164.0371, 353.8195	29.228088, 97.3694, 130.8168
113	0	1039	19	1.2656897E10	2.018301	0.010751995	0.0, 0.0, 0.0	267.51755, 162.87627, 354.14395	14.691489, 88.20407, 160.37138
114	0	1055	19	1.2656897E10	12.590368	0.01979277	0.0, 0.0, 0.0	268.02295, 163.05067, 353.5615	-19.663925, 115.91073, 293.38757
115	0	1076	19	1.2656897E10	1.2494237	0.009163548	0.0, 0.0, 0.0	268.46347, 164.73941, 353.59372	4.5492835, -22.657423, 271.45648
116	0	44	19	1.2656897E10	1.1533148	0.008922291	0.0, 0.0, 0.0	255.66098, 176.85463, 347.79187	-249.26367, 37.664062, -118.44709
117	0	56	19	1.2656897E10	2.30663	0.011241382	0.0, 0.0, 0.0	255.6698, 176.35075, 347.42307	-177.35524, 79.563034, -54.682545
118	0	70	19	1.2656897E10	2.498853	0.011545355	0.0, 0.0, 0.0	255.57506, 177.02773, 348.08698	-231.97614, 62.538258, -159.56491
119	0	72	19	1.2656897E10	1.3455337	0.009392732	0.0, 0.0, 0.0	255.59235, 177.16214, 348.09412	-271.43484, -14.365915, -147.84352
120	0	82	19	1.2656897E10	2.30663	0.011241382	0.0, 0.0, 0.0	255.86867, 176.74915, 346.62888	-249.0715, 62.642822, 6.9695435
121	0	127	19	1.2656897E10	1.5377563	0.009820258	0.0, 0.0, 0.0	255.02032, 177.0584, 347.06686	-195.96777, 74.86487, -52.22119
122	0	44	19	1.2656897E10	0.76887745	0.007794342	0.0, 0.0, 0.0	259.56125, 168.86055, 348.97717	-261.59497, 44.51758, 144.72311


HDFView 2.11
File Window Tools Help
Recent Files /lustre/scratch/astro/rp279/SAM_output/Output.h5
Clear Text
Output.h5
19
20
Filter_Sloan.txt
My_Makefile_options
Prop Table
Input.500_287
My_Makefile_options at / [Output.h5 in /lustre/scratch/astro/rp279/SAM_output]
Table
0-based
My_Makefile_options
0 # List of Makefile options
1 # see also routine "check_options" in main.c...
2
3 # Options that control the SA model
4 # Ideally, these shoudl be set in the input parameter file
5 OPT += -DUPDATETYPETWO # This updates the positions of type 2 galaxies when the galaxies are written to file
6 OPT += -DDISRUPTION # Instantaneously and completely disrupt type 2s
7 #OPT += -DHT09_DISRUPTION # Henriques & Thomas 2009 tidal disruption of stars in satellite galaxies
8 ifeq (HT09_DISRUPTION,$(findstring HT09_DISRUPTION,$(OPT)))
9 OBJS += ./code/recipe_HT09_disrupt.o
10 endif
11 #OPT += -DMERGE01 # allows type 1s to merge with type 0s
12
13 # Options that control output
14 OPT += -DOVERWRITE_OUTPUT # overwrite output files if they exist (otherwise will quit without overwriting)
15 OPT += -DNOUT=2 # This sets the number of galaxy output times. IGNORED IN GALAXYTREE MODE. VALUE CORRESPONDS TO NO. OF ROWS READ FROM desired_outputsnaps FILE
16 #OPT += -DGALAXYTREE # This will enable output of full galaxy merger trees, implicitly sets NOUT to maximum value
17 #OPT += -DNO_PROPS_OUTPUTS # only magnitudes outputted
18
19 # Options that control speed and memory usage
20 OPT += -DSAVE_MEMORY # we do NOT use the 64/68-elements array Sfr/SfrBulge when calculating galaxy trees.
21 #OPT += -DPRELOAD_TREES # this will load all the trees of a file in memory, and cache them (useful for MCMC)
22 #OPT += -DPARALLEL
23 #OPT += -DLOADIDS # Load dbids files
24
25 #OPT += -DMRII
26
27
28 #OPT += -DGUO10
29 #OPT += -DGUO13
30
31
32 # record star formation histories - used for post-process mags and detailed chemical enrichment
33 #OPT += -DSTAR_FORMATION_HISTORY
34 ifeq (STAR_FORMATION_HISTORY,$(findstring STAR_FORMATION_HISTORY,$(OPT)))
35 #OPT += -DNORMALIZEDDB # currently does not write sfh_time and sfh_dt to galaxies records, but writes separate CSV file with relevant bin information
36 OBJS += ./code/star_formation_history.o
37 endif
38
39
40 ############################
41 ### SPECPHOT_PROPERTIES ###
42 ############################
43 #OPT += -DCOMPUTE_SPECPHOT_PROPERTIES
My_Makefile_options (1521068, 2)
Compound/Vdata, 142
Number of attributes = 4
CLASS = TABLE
Log Info
Metadata
Mail - Rhys Poulton ...
Rhys_SAM_presentat...
rp279@apollo-login: ...
HDFView 2.11
Untitled - Paint
EN
13:45



	Field	Units	Description
0	galaxyId		The unique identifier of this galaxy. Built from the topologically sorted merger tree \n as described in TBD
1	haloId		The haloId of the subhalo (in the halo table with the same name in MPAHaloTrees) containing this galaxy.
2	firstProgenitorId		galaxyId of the first progenitor of this galaxy. Strictly galaxyId+1 iff lastProgenitorId > galaxyId
3	nextProgenitorId		galaxyId of next progenitor of this galaxy in the linked list structure used to facilitate traversing trees in code.
4	lastProgenitorGalId		The galaxyId of the last progenitor of this galaxy in the topological ordering used to assign galaxyId-s as described in this page.
5	fofCentralId		The galaxy id of the central galaxy of the FOF group this galaxy is in.
6	treeId		Unique id of galaxy formation tree containing this galaxy. Note that this treeId does not identify merger trees but the larger structures defining galaxy formation units.
7	descendantId		galaxyId of the descendant of this galaxy in its merger tree.
8	mainLeafId		galaxyId of the leaf on the main branch this galaxy is part of. Obtained by following firstProgenitorId as far as it goes.
9	treeRootId		The galaxyId of the galaxy at the root of the merger tree this galaxy is in. Especially useful for speeding up queries for descendants for a given progenitor. See TBD for ...
10	subHaloId		Id of the subhalo containing this galaxy as given by the column subhaloFileID in the MillenniumII..SubHalo miniMillI..SubHalo table (for MRII and mMRII) and by the colu...
11	fofSubHaloId		The subhaloId resp. subhaloFileId (for mMR and MR resp. mMRII and MRII see documentation of subhaloId column in this table) of the subhalo at the center of the FOF g...
12	phKey		The Peano-Hilbert index of the cell this galaxy is in
13	redshift		The redshift corresponding to the snapnum (in MField..Snapshots for mMR and MR
14	type		Type indicating whether galaxy is at the center of the FOF group (type=0)
15	snapnum		The snapshot number where this galaxy was identified. This column is a foreign key to the snapnum column in the table in the Snapshots database with the same name ...
16	lookBackTime	yr	The look back time (in years) from z=0 to the redshift of the galaxy.
17	centralMvir	1010 Msun/h	The virial mass (as defined by m_crit200) of the FOF group the galaxy resides in.
18	centralRvi	Mpc/h	The virial radius (as defined by r_crit200) of the FOF group the galaxy resides in.
19	distanceToCentralGal	Mpc/h	components of the distance between this galaxy and the galaxy at the centre of the FoF group.
20	pos	Mpc/h	components of position of galaxy.
21	velX	km/s	components of velocity of galaxy.
22	np	number	Number of particles of the subhalo this galaxy is in.
23	mvir	10^10 Msun/h	Virial mass (as defined by m_crit200) of the FOF group this galaxy was in when last itwas a type 0 galaxy. I.e. current mass for type 0 galaxies "infall virial mass" for type 1
24	rvir	Mpc/h	Virial radius (as defined by r_crit200) of the FOF group this galaxy was in when last it was a type 0 galaxy. I.e. current virial radius for type 0 galaxies "infall virial radius" ...
25	vvir	km/s	Virial velocity of the subhalo the galaxy is/was the center of.
26	vmax	km/s	Maximum rotational velocity of the subhalo of which this galaxy is the center
27	gasSpin	Mpc/h km/s	The X-component of the spin of the cold gas disk
28	stellarSpin	Mpc/h km/s	The X-component of the spin of the stellar disk
29	infallVmax	km/s	Maximum rotational velocity of the host halo of this galaxy at infallSnap.
30	vmaxPeak	km/s	Maximum past rotational velocity of the host halo of this galaxy.
31	infallSnap		Most recent (largest) snapnum at which this galaxy's type changed from 0 to 1 or 2
32	infallHotGas	10^10 Msun/h	Mass in hot gas at the time of infall (same as hotGas for type 0 galaxies).
33	hotRadius	Mpc/h	Radius out to which hot gas extends: rvir for type 0; 0 for type 2; maximum radius out to which hot gas is not stripped for type 1.
34	oriMergeTime	yr	Estimated dynamical friction time (in years) when the merger clock is set.
35	mergeTime	yr	Estimated remaining merging time (in years). oriMergeTime - time since the merger clock is set.
36	coldGas	10^10 Msun/h	Mass in the cold gas disk.
37	stellarMass	10^10 Msun/h	Total mass in stars in the disk and the bulge together.



Henriques et al. 2015

```python
import h5py

snap='20'

with h5py.File('../output/Output.h5','r') as f:

    #Get all the data from the snapshot
    snapdata=f[snap]

    #Get array of the Galaxy Type, Mass, disk mass and bulge mass
    Type=snapdata['Type']
    Mass=snapdata['Mass']
    Dmass=snapdata['DiskMass']
    Bmass=snapdata['BulgeMass']
```

What are the main benefits

Reduced file size (40%)

Easier to read out data

Faster read in (2x)

Data stored in a table format

Can store all the data needed in one file

What are the current problems

L-Galaxies execution significantly slower

-Source: High level API

Solution:

-Work with the Low level API

Conclusion

Main advantages of HDF5:

- Faster to read in data
- Easy to extract data
- Can contain multiple files within a single HDF5 file
- Reduce the file size up to 40%

Current disadvantages

- L-Galaxies execution much slower.