An HDF5 Schema for SKA Scale Image Cube Visualization

Angus Comrie, Adrianna Pińska, Rob Simmonds and Russ Taylor

Interactive Data Visualization

- Image cubes getting very large
  - MeerKAT: 16K x 16K per channel
  - SKA Phase 1: Up to 64K x 64K
  - ALMA: Lots of channels
- Remote DataViz (Cloud server, Web/Desktop client)
- “Interactive” speeds
  - Quick access to data
  - Accelerate commonly used workloads
- IDIA: Explore new approaches for implementation in new viewers

Hierarchical Data Format (HDF) 5

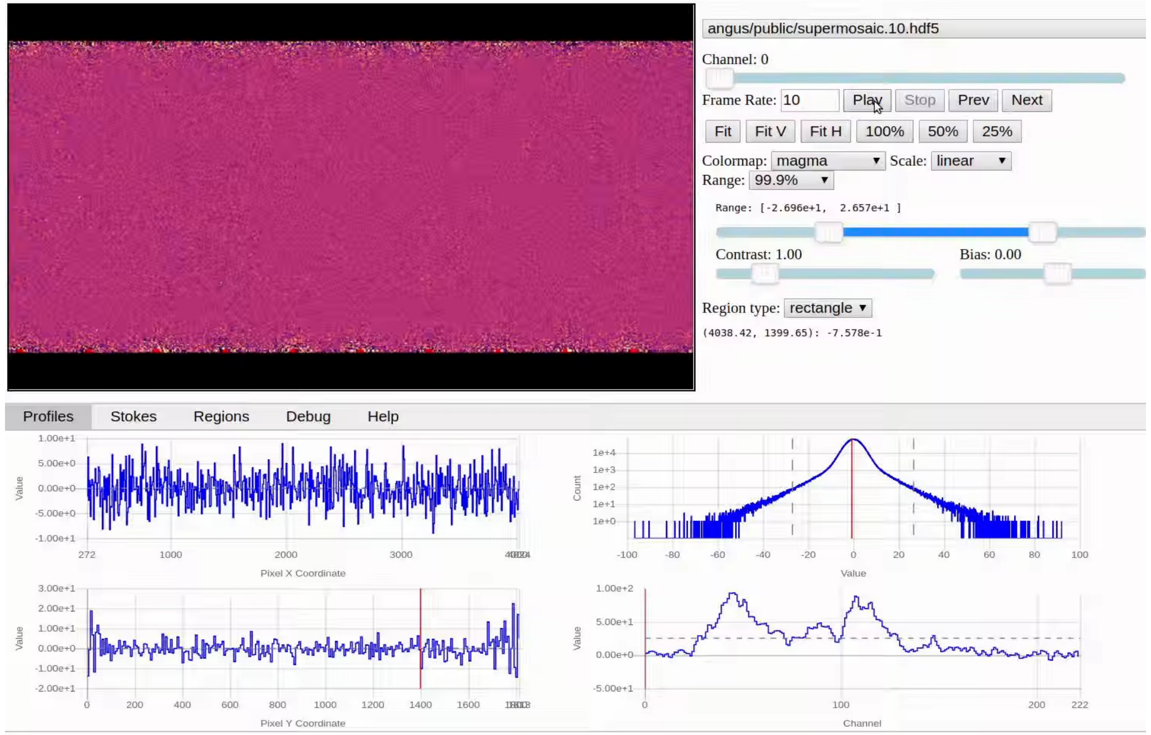
- Filesystem-in-a-file
- Easy to structure data product and derivatives
- Purpose-built schemas
- Existing FITS replacements:
  - LOFAR (Anderson et al. 2010)
  - HDFITS (Price et al. 2015)

Common Workloads

- Widely used: XY spatial, Z spectral
- “W” sometimes Stokes
- X contiguous on storage
- Different axes, different I/O (sequential / random)

Slices (2D subset)

- Plane-aligned (XY, YZ, XZ)
- Very common: XY slice (fixed Z)
- Animate through channels



angus/public/supermosaic.10.hdf5

Channel: 0

Frame Rate: 10

Colormap: magma Scale: linear

Range: 99.9%

```
Range: [-2.696e+1,  2.657e+1 ]
```

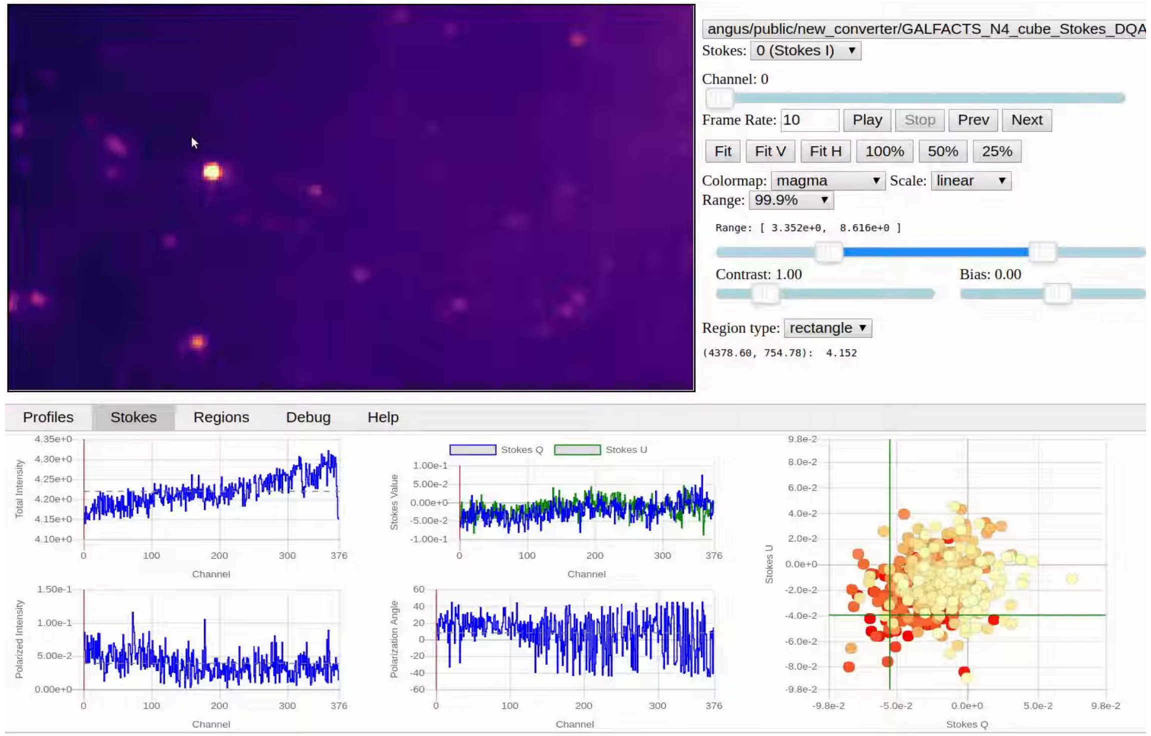
Contrast: 1.00 Bias: 0.00

Region type: rectangle

```
(4038.42, 1399.65): -7.578e-1
```

Profiles (1D subset)

- Common: Axis-aligned
- X, Y, Z profiles for given pixel
- Polarization: Stokes as well
- Z Axis: 1 single pixel read / channel



angus/public/new_converter/GALFACTS_N4_cube_Stokes_DQA

Stokes: 0 (Stokes I)

Channel: 0

Frame Rate: 10

Colormap: magma Scale: linear

Range: 99.9%

```
Range: [ 3.352e+0,  8.616e+0 ]
```

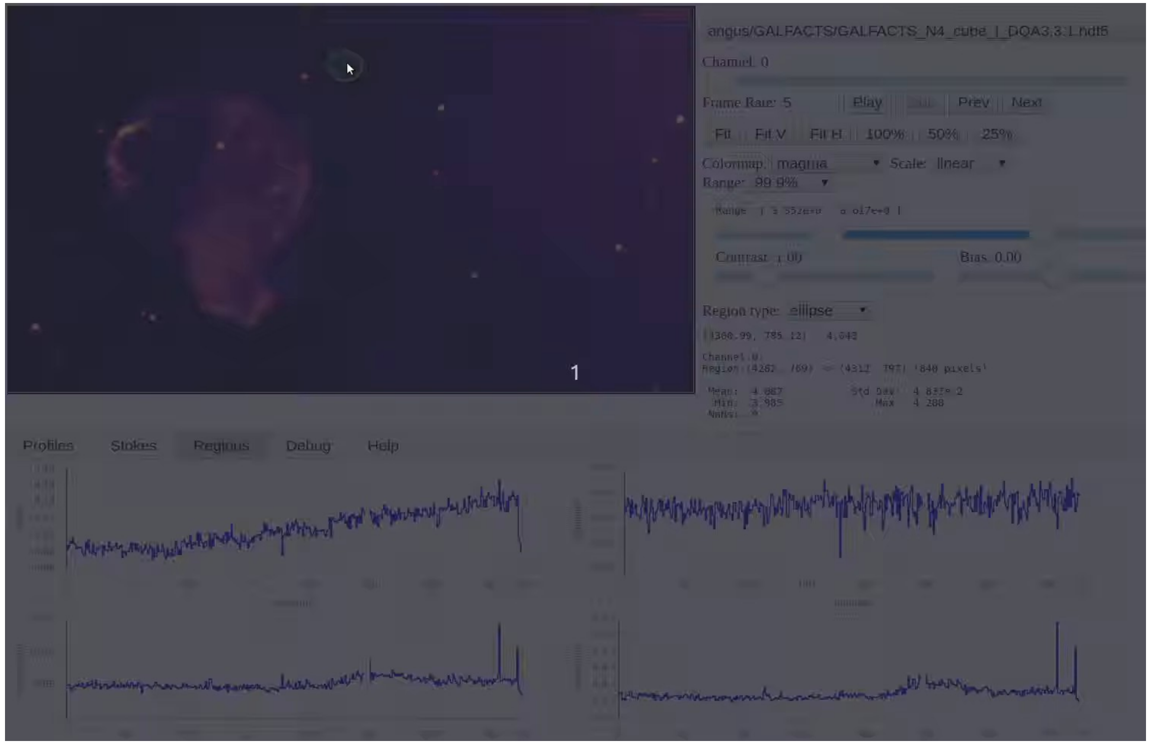
Contrast: 1.00 Bias: 0.00

Region type: rectangle

```
(4378.60, 754.78):  4.152
```

Regions (3D subset)

- Common: Small XY subset
- Z profiles for subset
- Still a lot of random reads!



angus/GALFACTS/GALFACTS_N4_cube_I_DQA3.3.1.hdf5

Channel: 0

Frame Rate: 5

Colormap: magma Scale: linear

Range: 99.9%

Contrast: 1.00 Bias: 0.00

Region type: ellipse

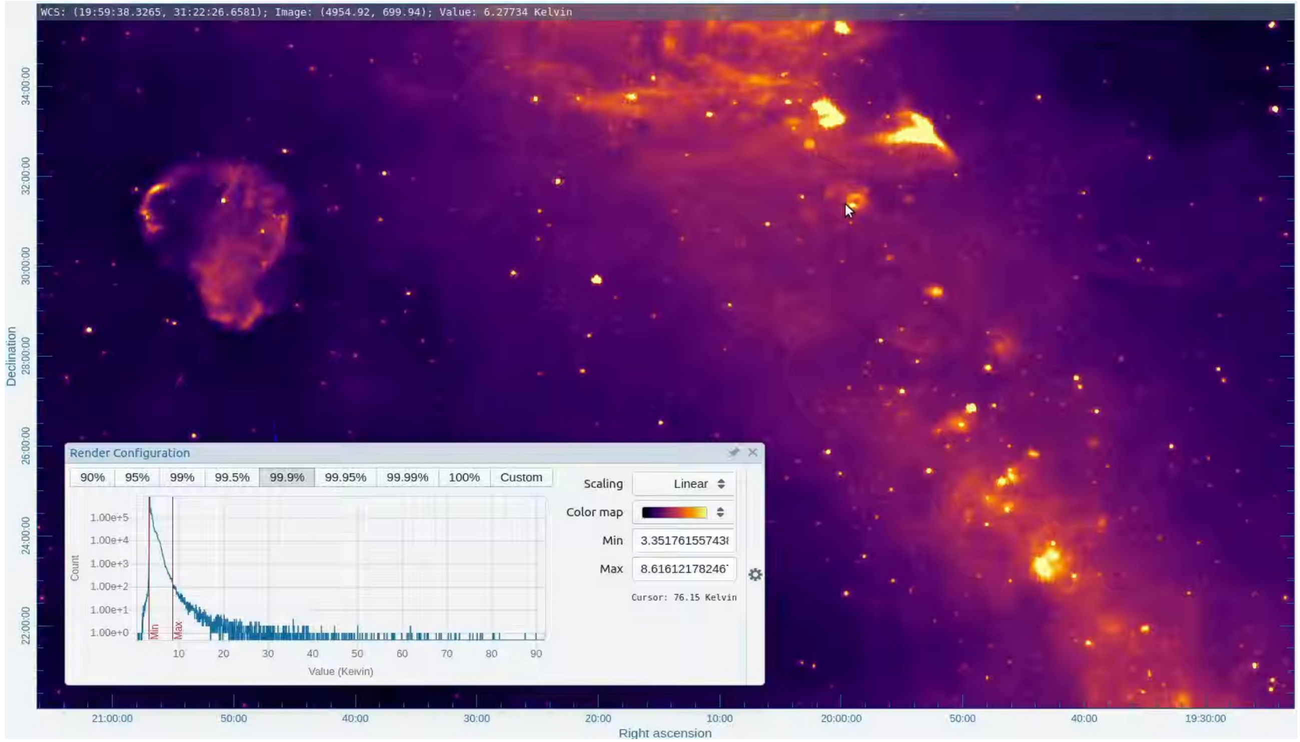
```
(1300.99, 785.12)   4.043

Channel:0:
Region:(4282, 769) -> (4312, 797) (840 pixels)

 Mean:  4.087          Std Dev:  4.877e-2
  Min:  3.985              Max:  4.288
 NaNs:  0
```

Histograms (Slice or Cube)

- Used to define color map bounds
- Can be used to approximate percentiles
- Global / Local bounding



WCS: (19:59:38.3265, 31:22:26.6581); Image: (4954.92, 699.94); Value: 6.27734 Kelvin

Render Configuration

Scaling: Linear

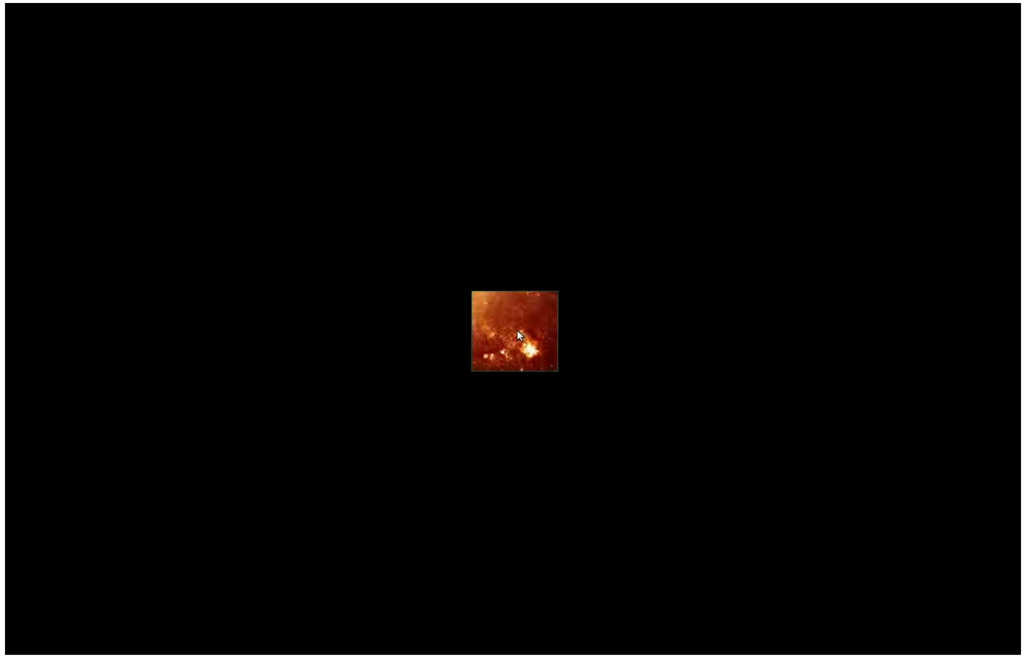
Min: 3.35176155743!

Max: 8.61612178246

Cursor: 76.15 Kelvin

Dynamic resolution

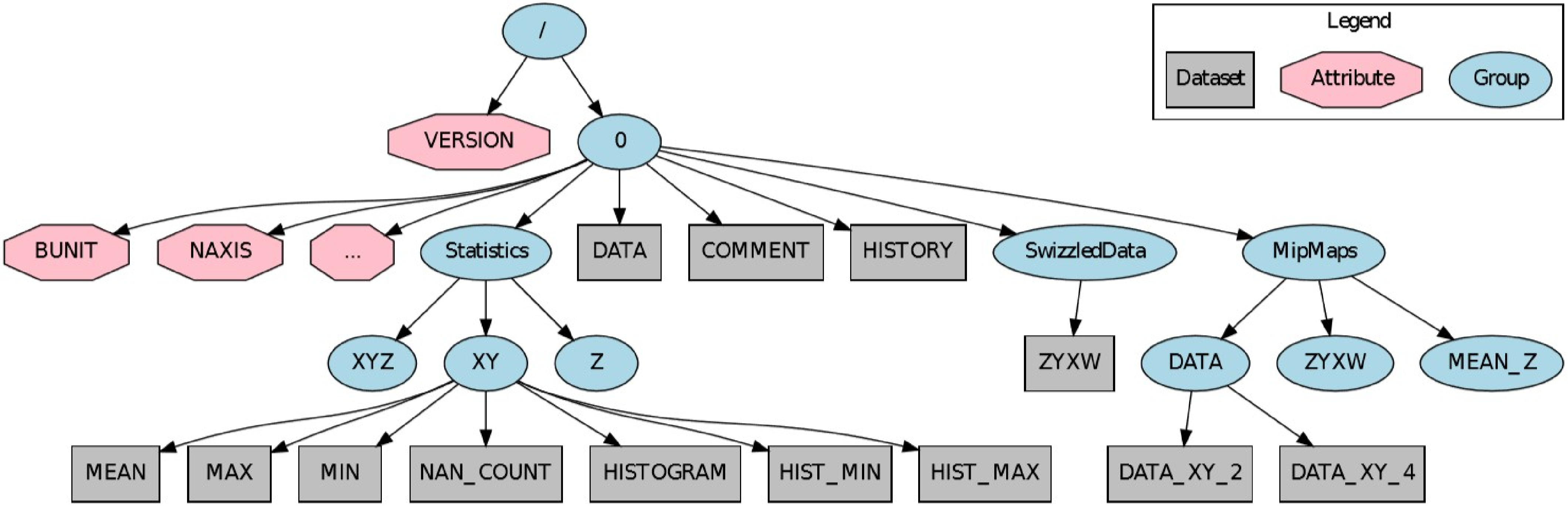
- Deliver appropriate image sizes to client
- Increase resolution as user zooms in
- Requires reading slices and down-sampling
- Tiled rendering



Schema Features

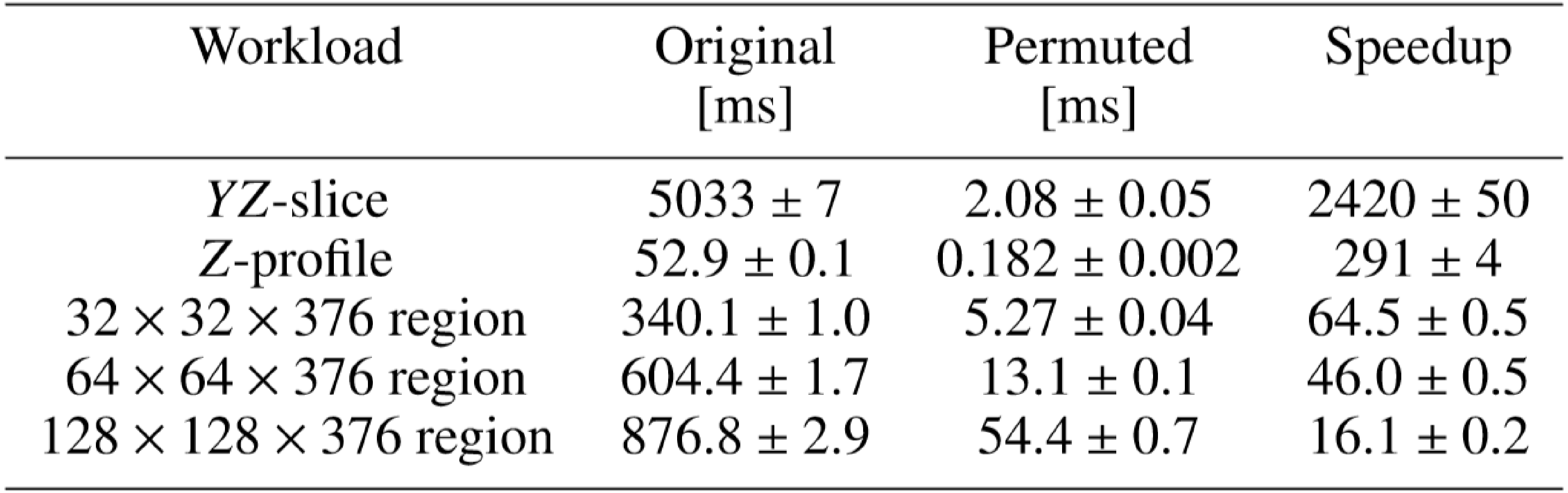
- FITS round-trip compatibility
- Permuted Datasets (XYZ -> ZYX)
    - Z-profile and region acceleration (Z is sequential)
    - Helps with YZ slices too
- Mip-mapped Datasets
    - Down-sampling in powers of 2
    - Helps with tiled delivery
- Slice and cube histograms
    - Very small data product

Schema Features



https://github.com/idia-astro/hdf5converters/wiki/HDF5-Image-Schema

Performance



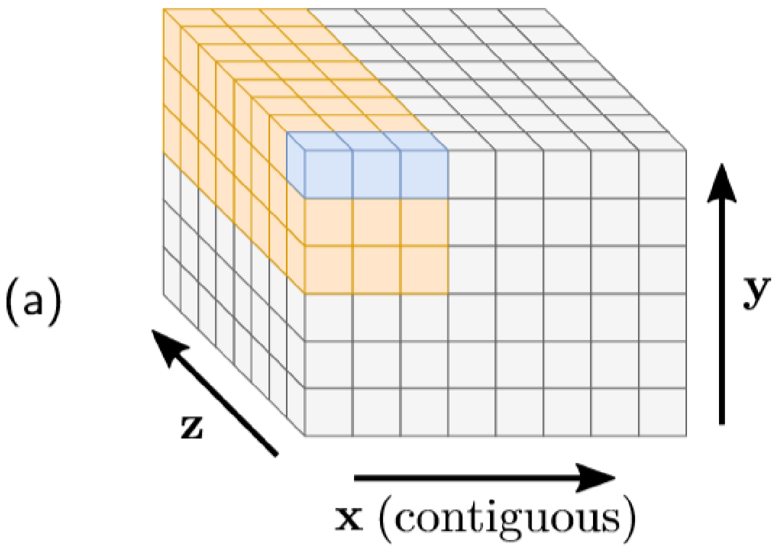
Workload	Original [ms]	Permuted [ms]	Speedup
YZ-slice	5033 ± 7	2.08 ± 0.05	2420 ± 50
Z-profile	52.9 ± 0.1	0.182 ± 0.002	291 ± 4
$32 \times 32 \times 376$ region	340.1 ± 1.0	5.27 ± 0.04	64.5 ± 0.5
$64 \times 64 \times 376$ region	604.4 ± 1.7	13.1 ± 0.1	46.0 ± 0.5
$128 \times 128 \times 376$ region	876.8 ± 2.9	54.4 ± 0.7	16.1 ± 0.2

(Tests performed on 5840 x 1074 x 376 image,
fast SSD)

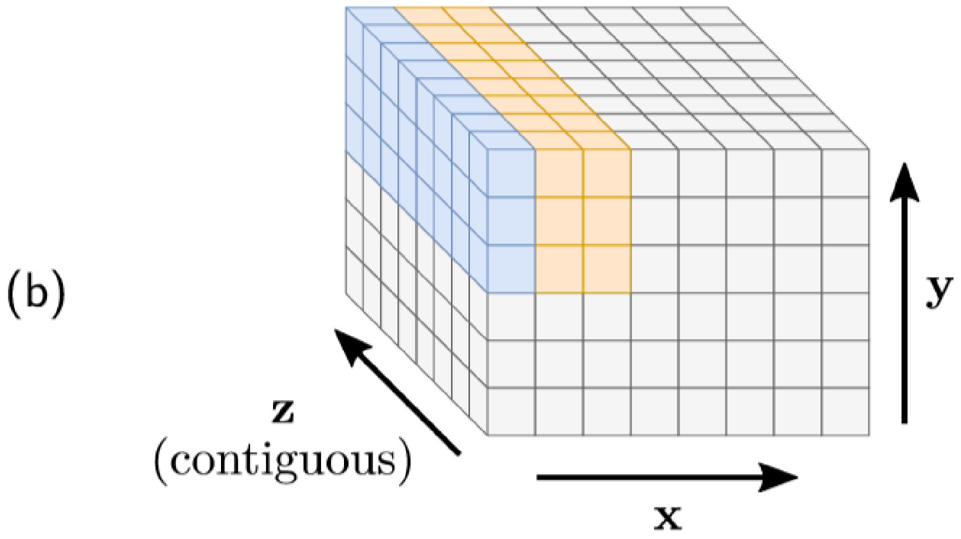
Conclusion

- Schema helps define structure for derived data
- Accelerates common workloads significantly
- Enables efficient and scalable exploration of image cubes

Additional: Region example



(a) 3 x 8 = 24 reads of 3 pixels each



(b) 3 reads of 24 pixels each

Additional: Online resources

- Converter and schema details: https://github.com/idia-astro/hdf5converters
- Benchmark: https://github.com/veggiesaurus/adass_hdf5_benchmark