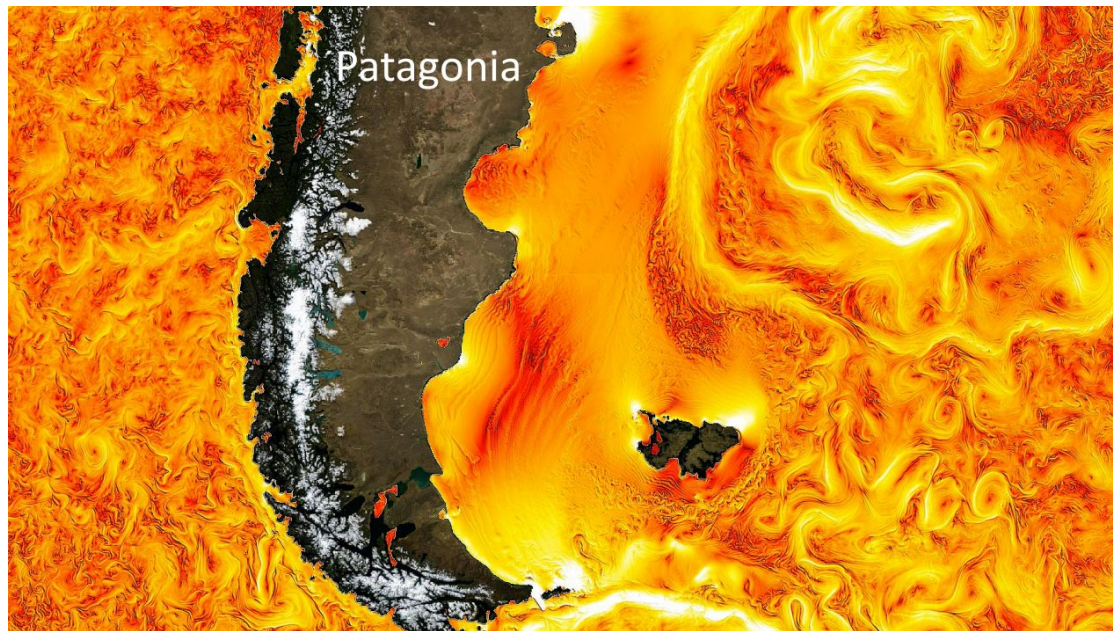
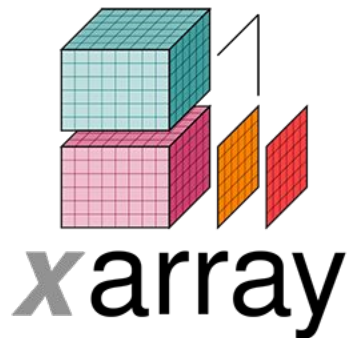


Cubed: Bounded-Memory Serverless Array Processing (in Xarray)

*Tom Nicholas
Tom White



Big science means ***Big*** arrays



▼ Data variables:

U (time, face, j, i_g)

	Array	Chunk
Bytes	7.97 TiB	71.19 MiB

PBs??



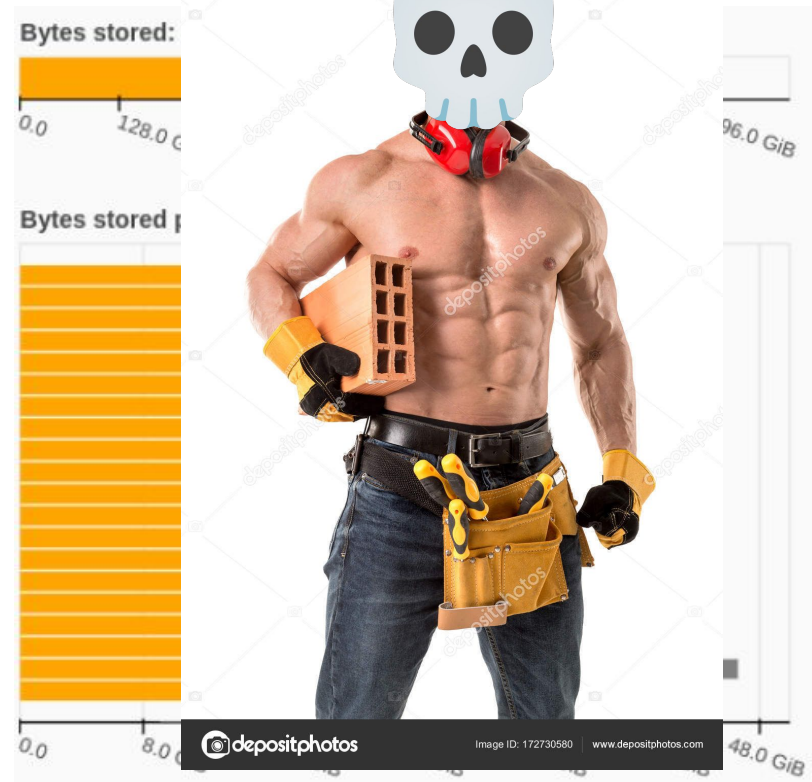
So use `dask.array`!



Dask is great, but it doesn't always succeed...

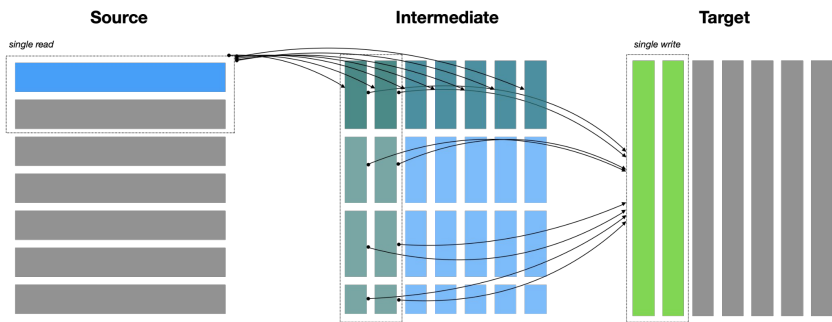
Sometimes unexpectedly exceeds your
RAM budget 😞

**Q: Can we guarantee
distributed array execution
respects RAM constraints?**

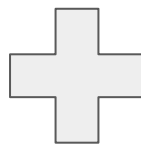


A: Yes! For certain operations...

Rechunker



(Bounded-memory)



Arr

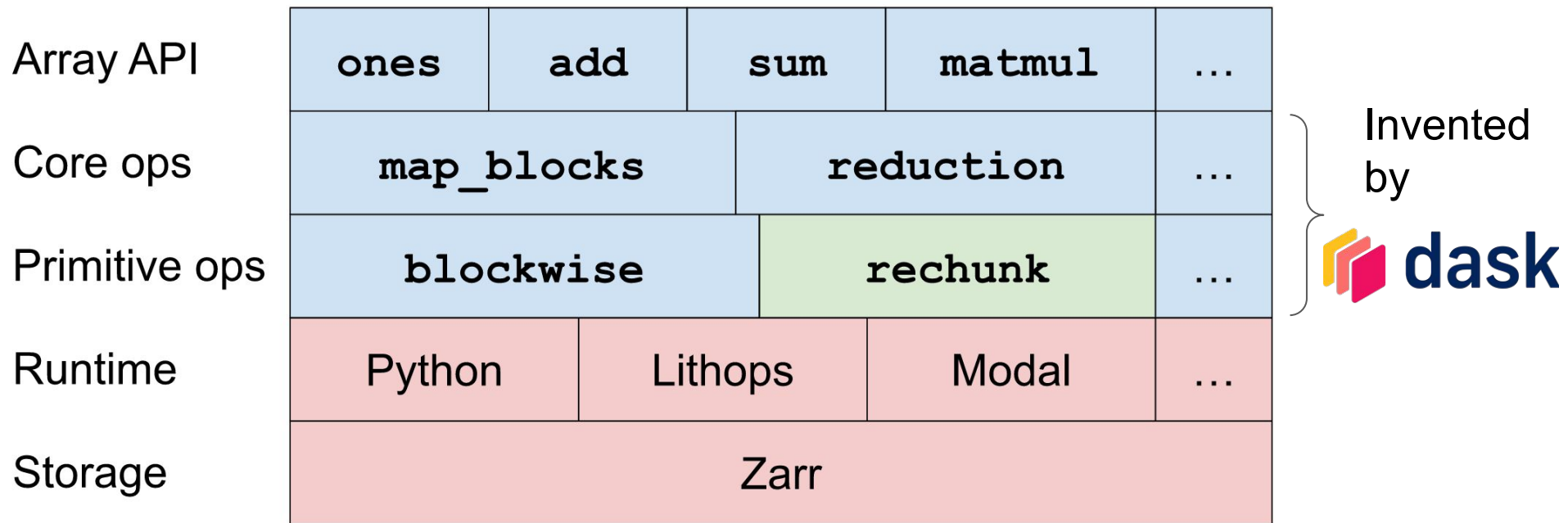


andard



✨ Cubed ✨

Cubed's Design



Serverless execution

Deploy one serverless container per chunk - read from / write to Zarr

							
IBM Cloud	aws	Microsoft Azure	Google Cloud	Alibaba Cloud	kubernetes	OPENSIFT	Coiled A Dask Company
Cloud Functions Code Engine VPC Gen2	AWS Lambda AWS Batch AWS EC2	Functions Container APPs	Cloud Functions Cloud Run	Functions Compute	Batch/Job - OpenWhisk Knative		Coiled Functions
---	---	---	---	---	---		...
Cloud Object Storage	AWS S3	Blob Storage	Cloud Storage	Object Storage Service	OpenStack Swift - Ceph MinIO - Redis - Infinispan		

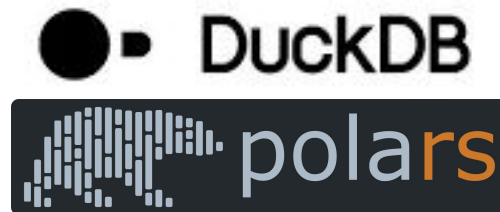


Xarray wraps Cubed OR Dask OR [new things??]

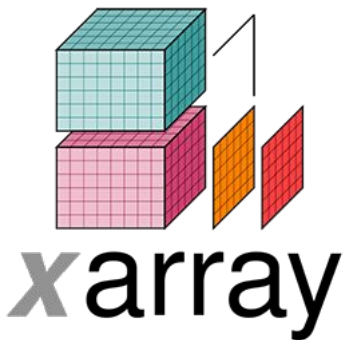
Tabular
data:



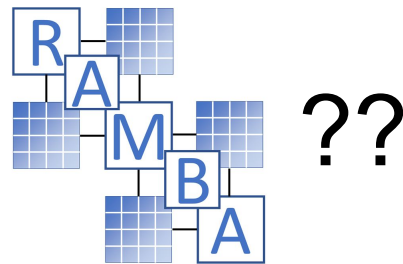
Executes via



Array
data:



Executes via



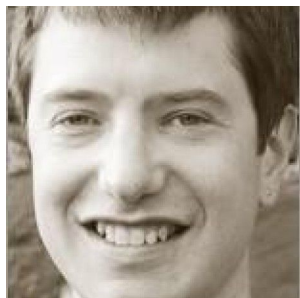
Read the blog post! <https://xarray.dev/blog/cubed-xarray>

Join the discussion!



New Working Group for Distributed Array Computing

■ News & Announcements



Also thanks Tom White
for writing Cubed!

