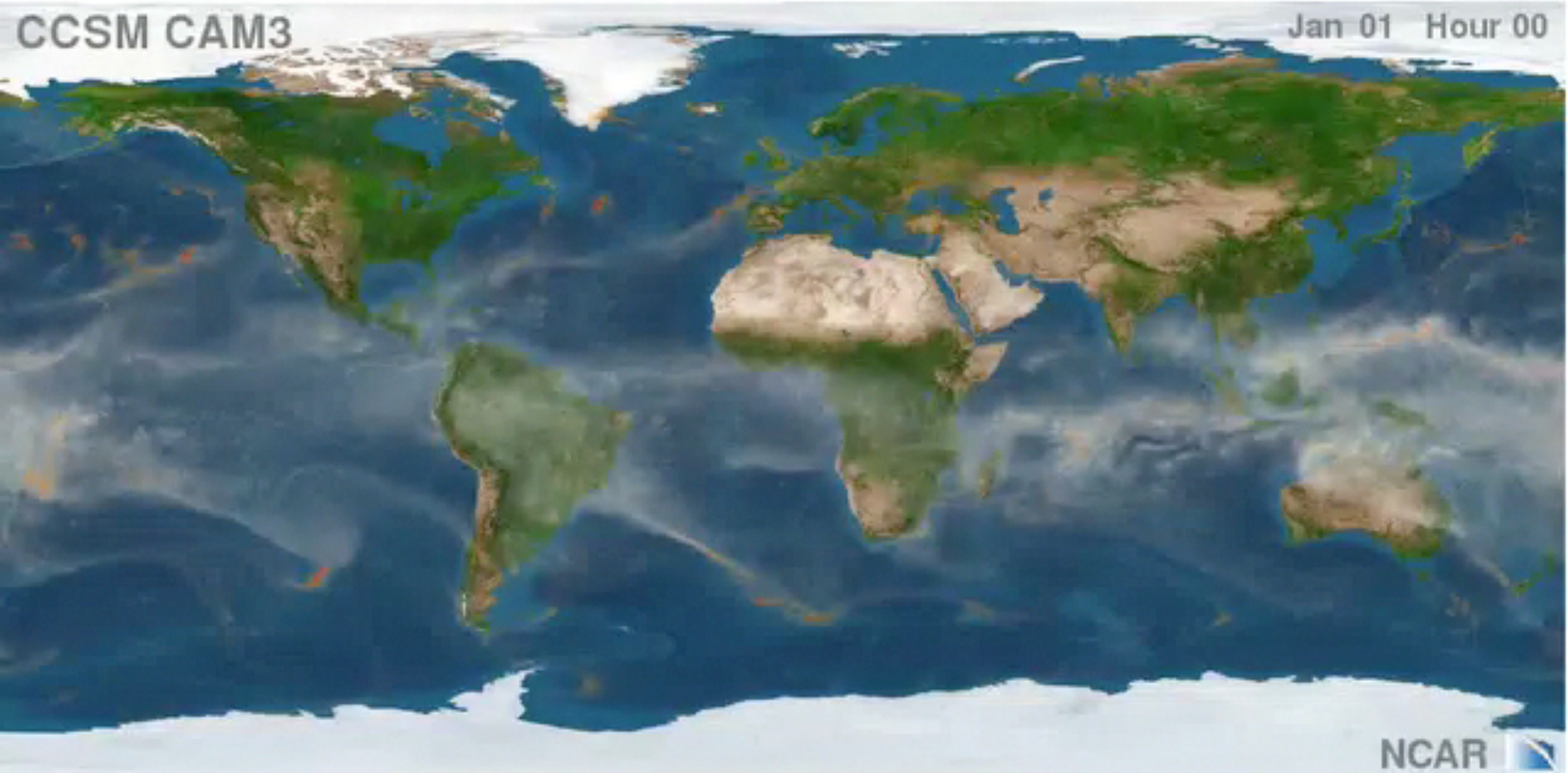
FAST - Big Data for climate science

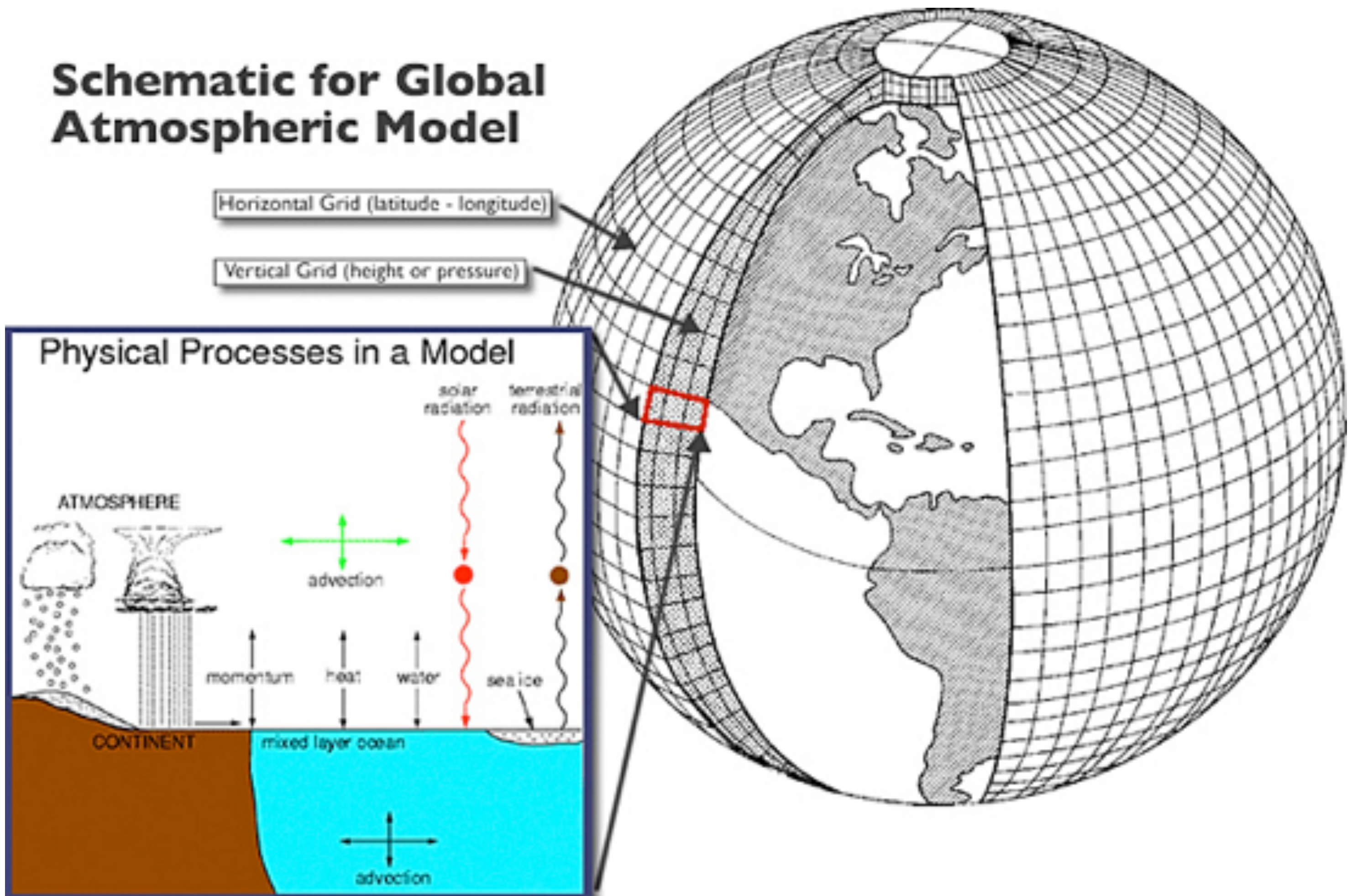
Rodrigo Caballero, Gunilla Svensson (Stockholm University)

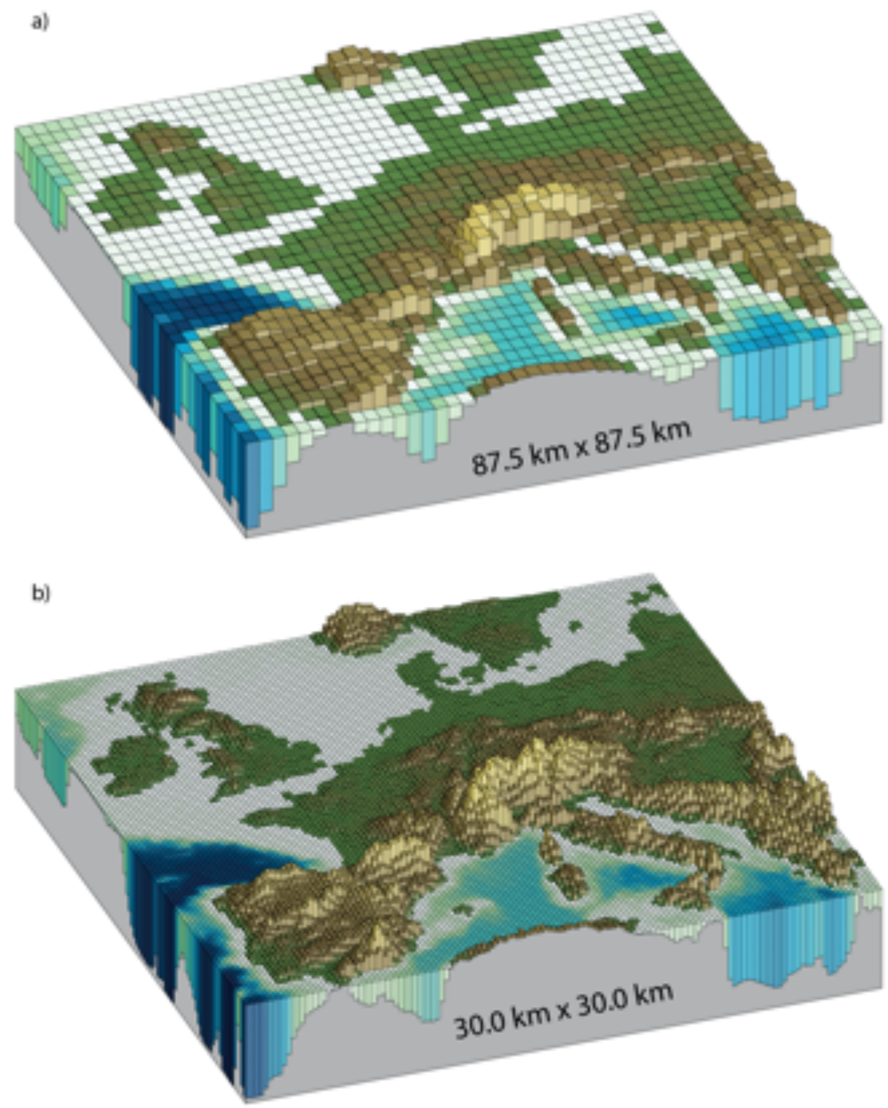
Jim Dowling, **Lars Kroll** (KTH)

Simulating climate



Atmospheric models

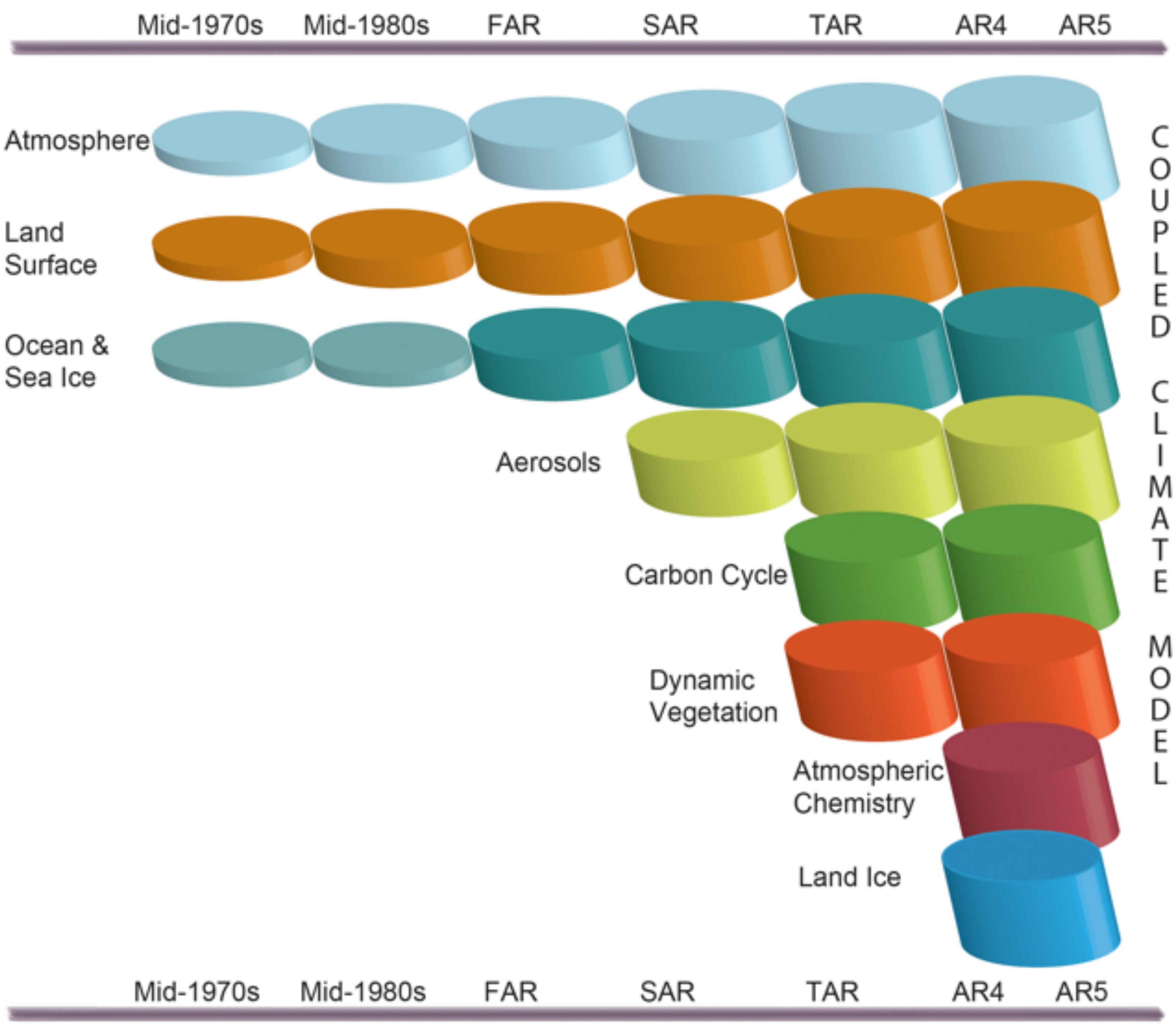




Solve for ***u*, *T*, *q*, *cld*** on 3-D grid.

Where we are now: **50 km** resolution
Where we would like to be: **5 km** resolution => ~1000 greater cost

Climate models



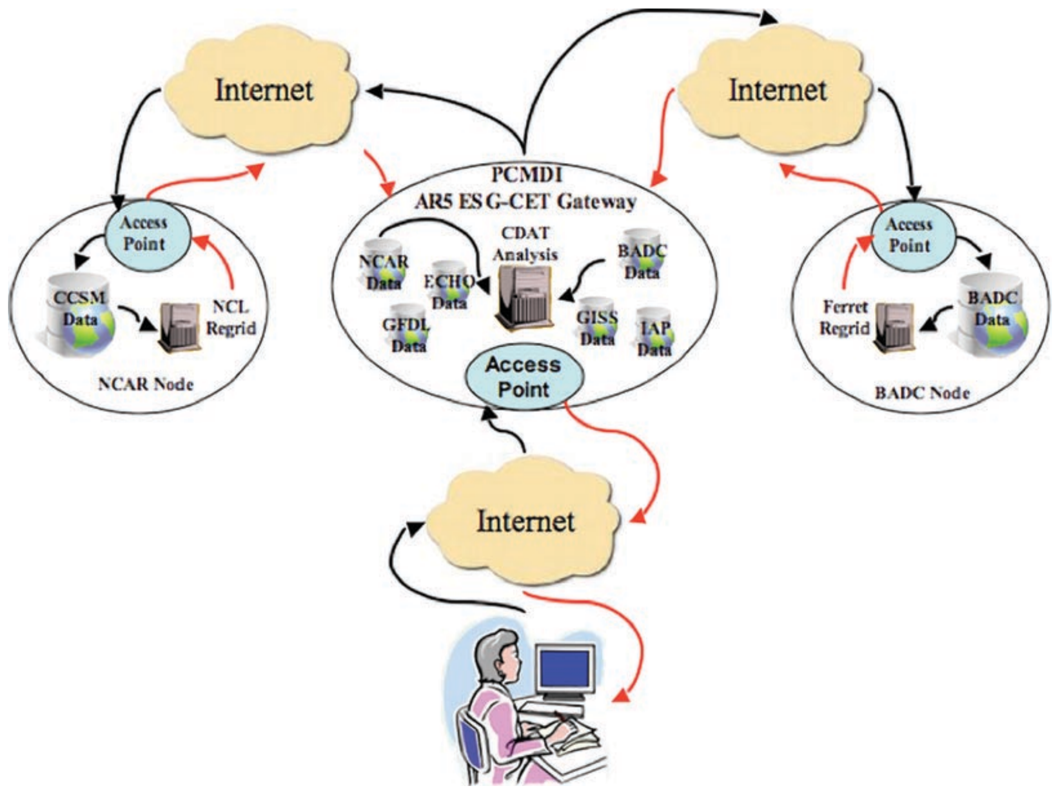
Climate model inter comparison (CMIP)

Community effort to intercompare climate models by conducting simulations using the same standardised protocol across ~30 different models worldwide. Closely associated with IPCC Assessment reports

	CMIP (1996 ONWARDS)	CMIP2 (1997 ONWARDS)	CMIP3 (2005-2006)	CMIP5 (2010-2011)
Number of Experiments	1	2	12	110
Centres Participating	16	18	15	24
# of Distinct Models	19	24	21	45
# of Runs (Models X Expts)	19	48	211	841
Total Dataset Size	1 Gigabyte	500 Gigabyte	36 TeraByte	3.3 PetaByte
Total Downloads from archive	??	??	1.2 PetaByte	
Number of Papers Published	47		595	
Users	??	??	6700	

(Steve Easterbrook)

How CMIP data is stored

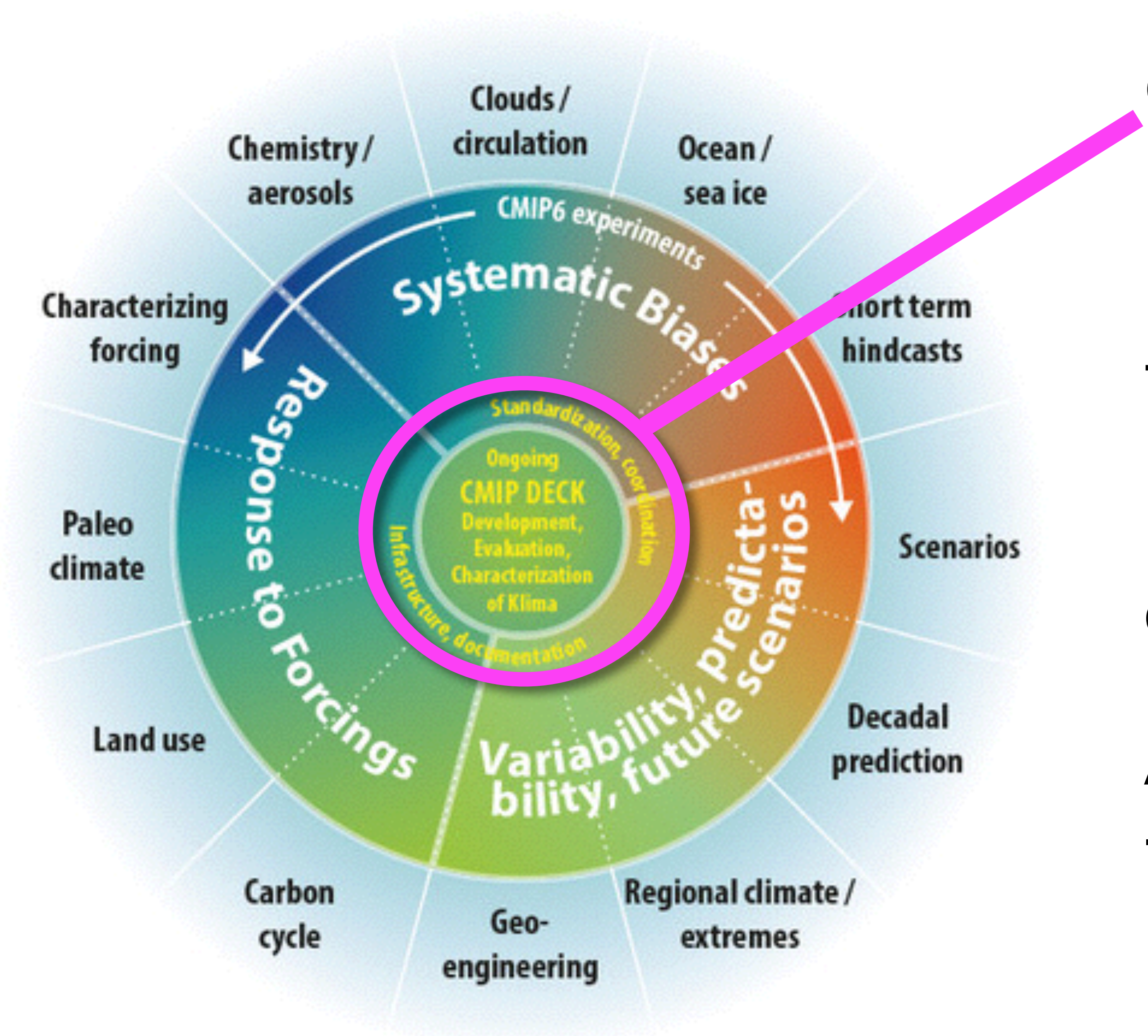


ESGF
Earth System Grid Federation

CMIP output stored on **Earth System Grid**:
distributed system with “smart” search capabilities.

Users download data files and analyse on their own systems

How much data do CMIPs generate?



Current generation: **CMIP6** (2015-2020)

Core set of simulations:
- historical
- next 3 centuries
- range of warming scenarios

Total: 8000 simulation years

If we save 4 snapshots/day: **~ 15 petabytes** only for the atmosphere of 1 model !!

About **30 models**, include ocean, land etc.
-> several **exabytes**

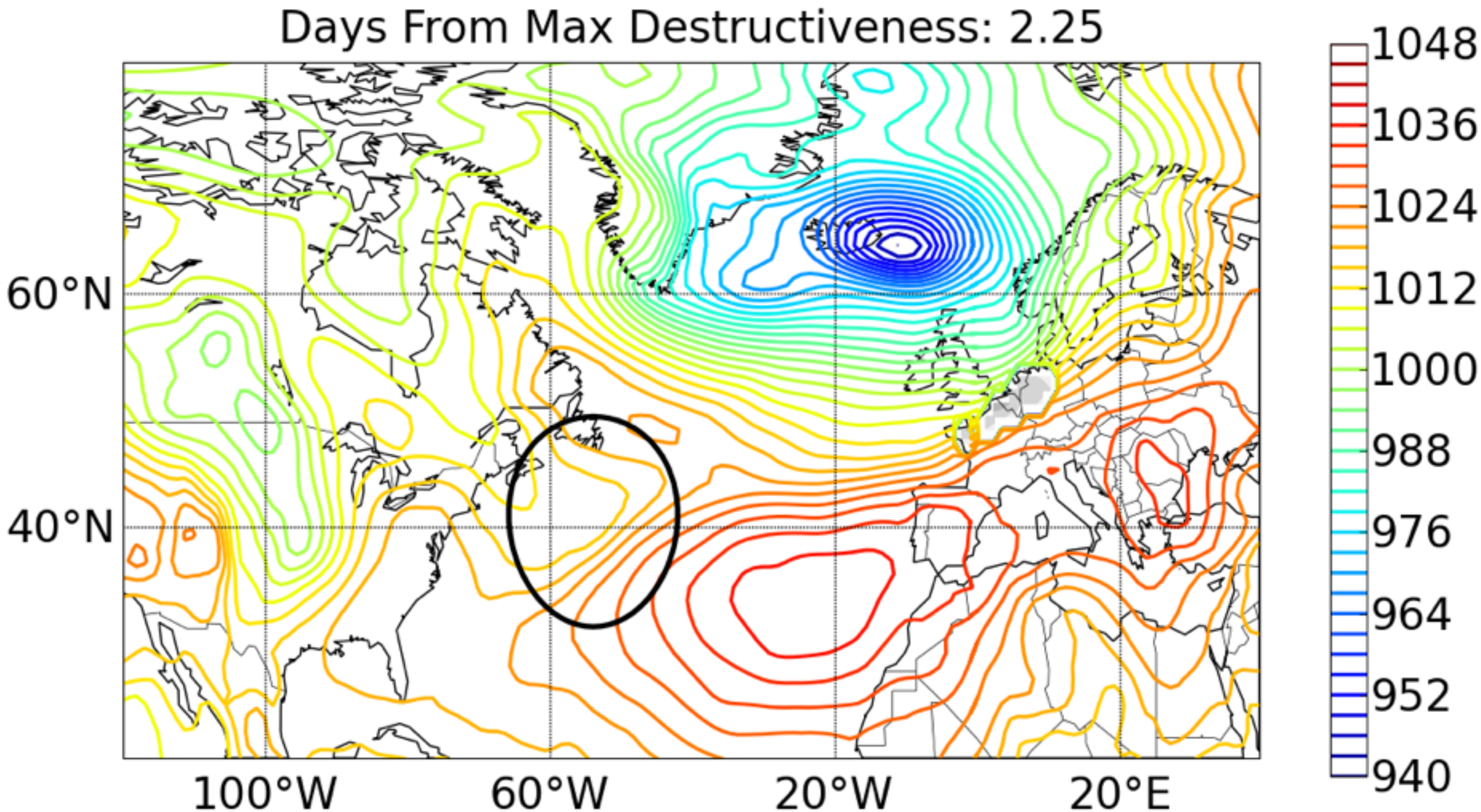
Include non-core MIPS
-> **10-100 exabyte range …**

One way to meet the big data challenge:

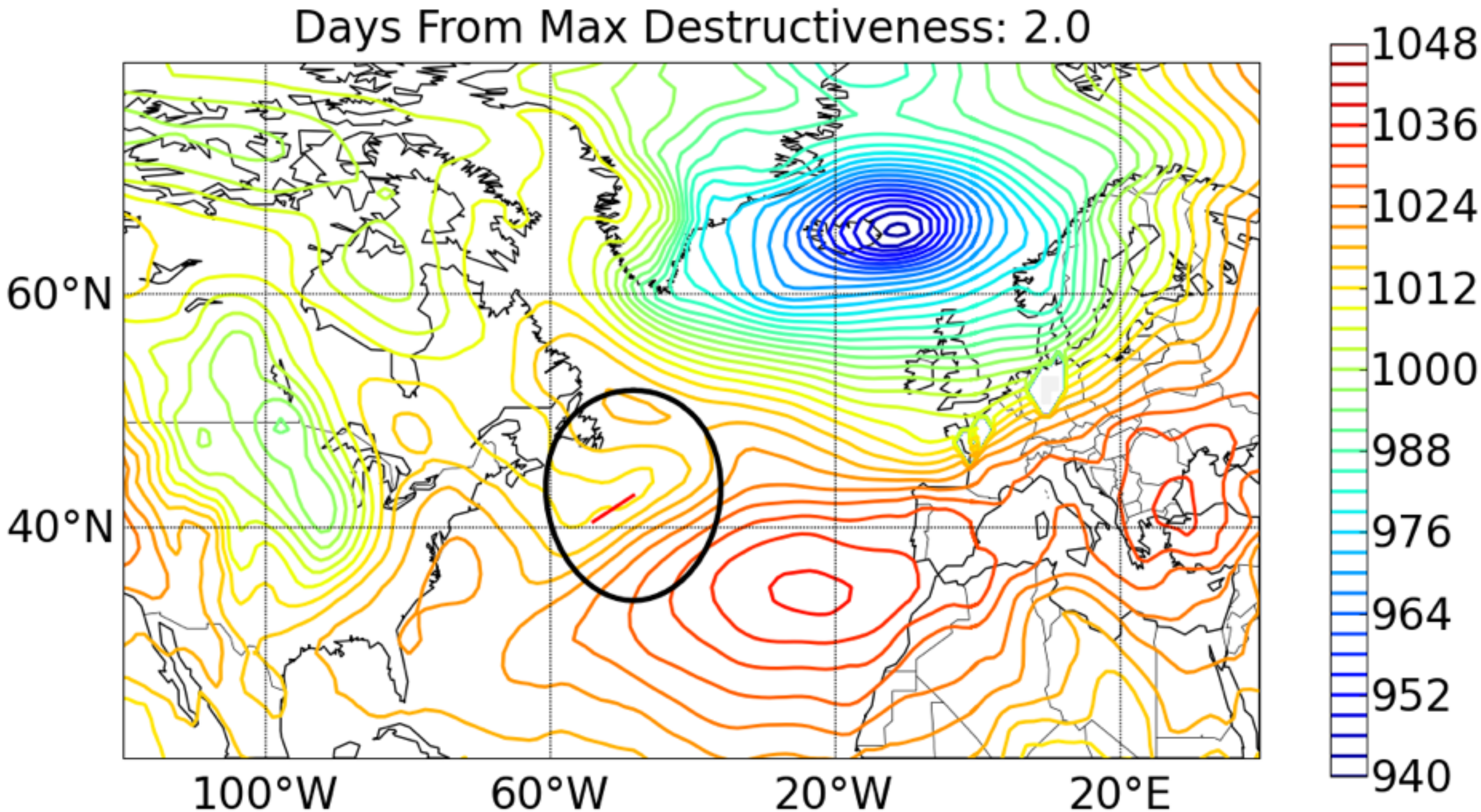
- Generate 10-100 exabytes of data
- Throw 99% of it away.

- Do we really not need that 99%?

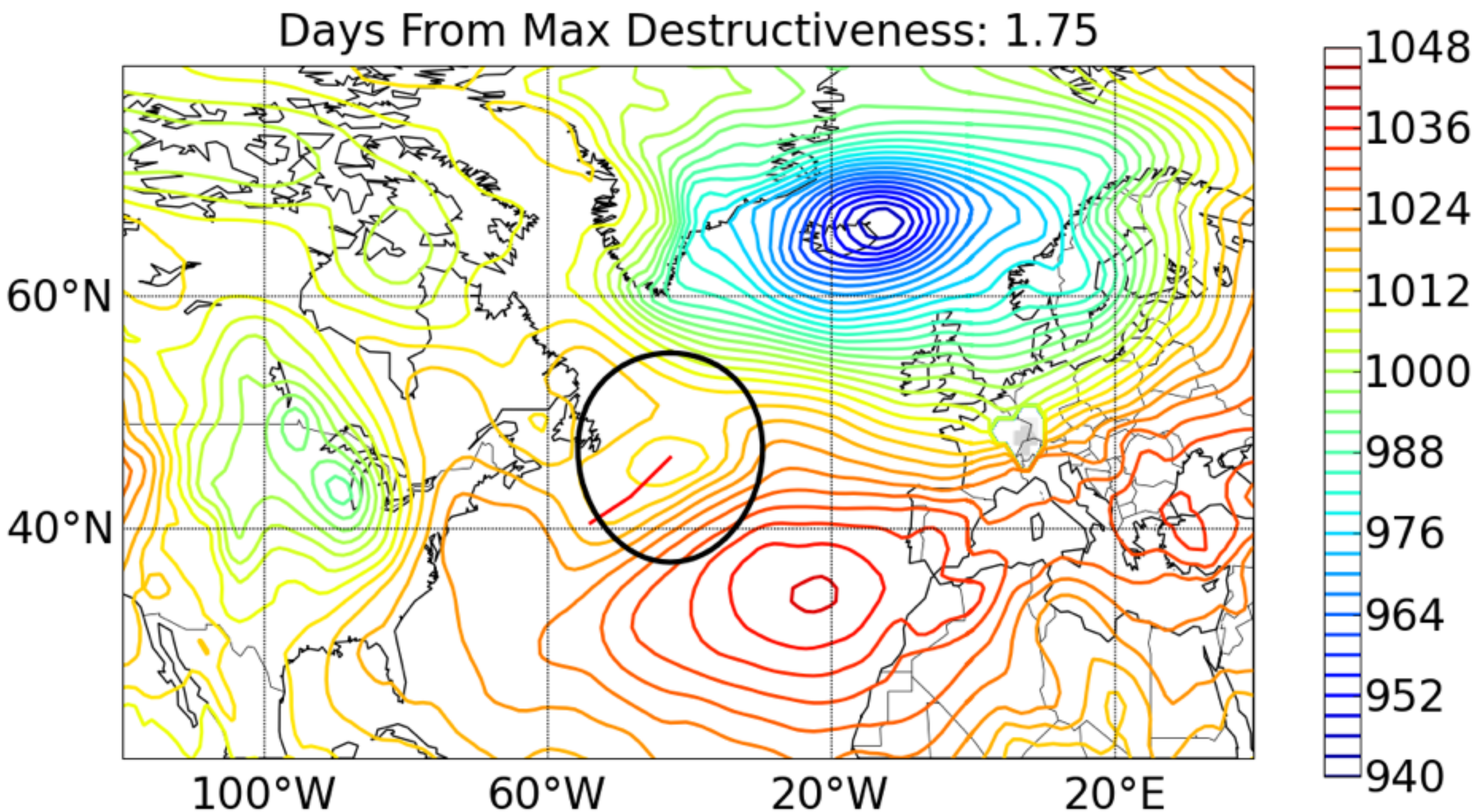
Daria storm, Jan 1990



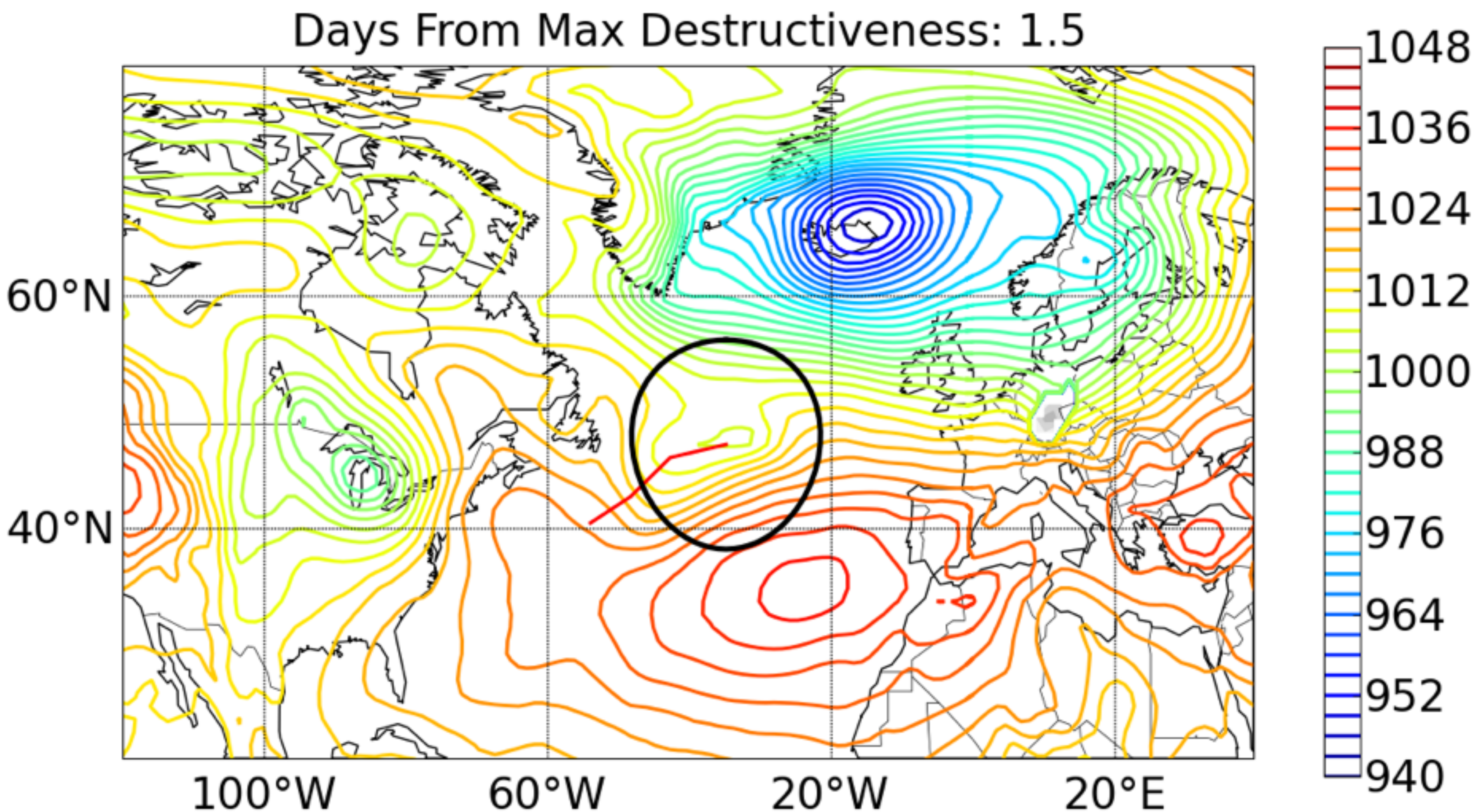
Daria storm, Jan 1990



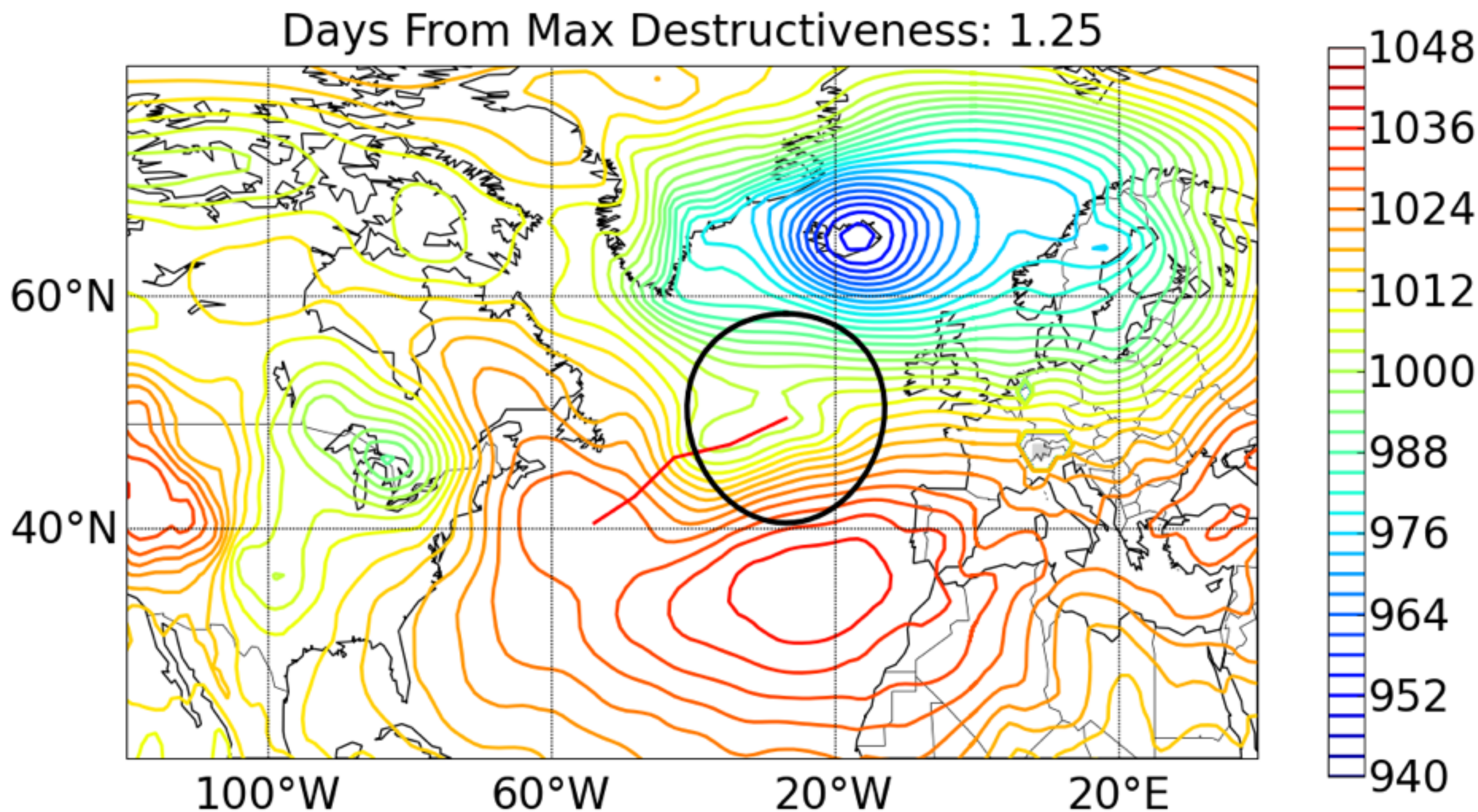
Daria storm, Jan 1990



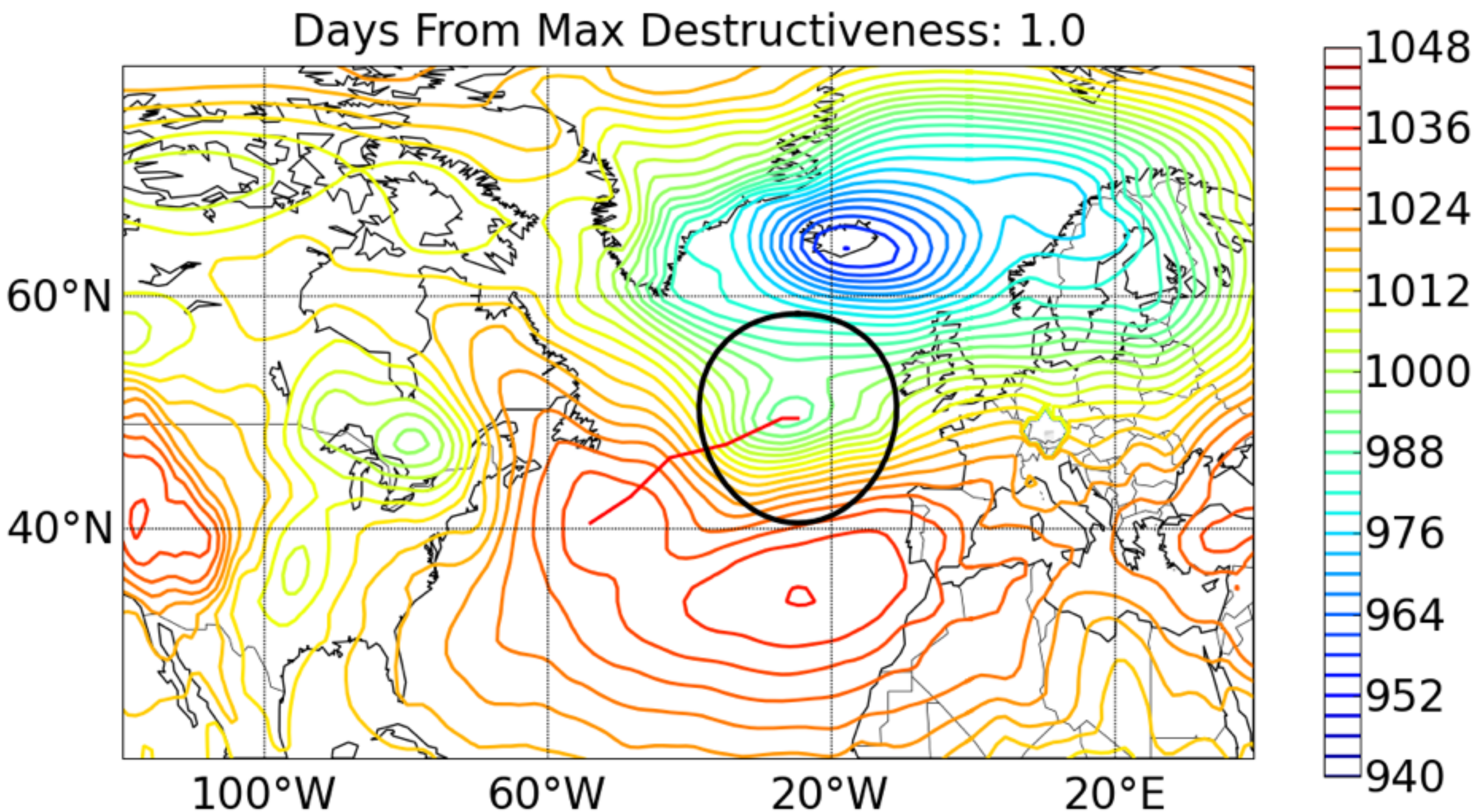
Daria storm, Jan 1990



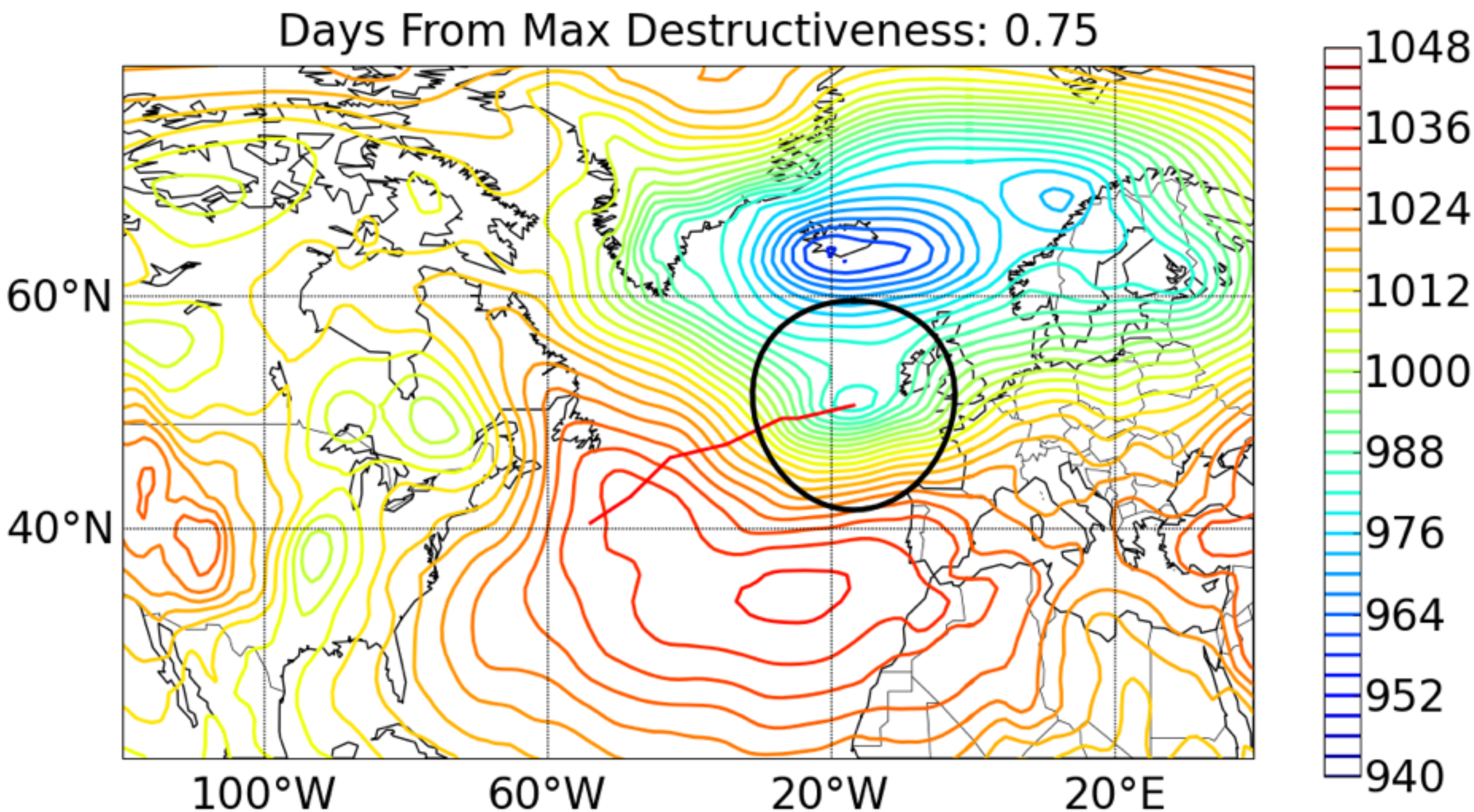
Daria storm, Jan 1990



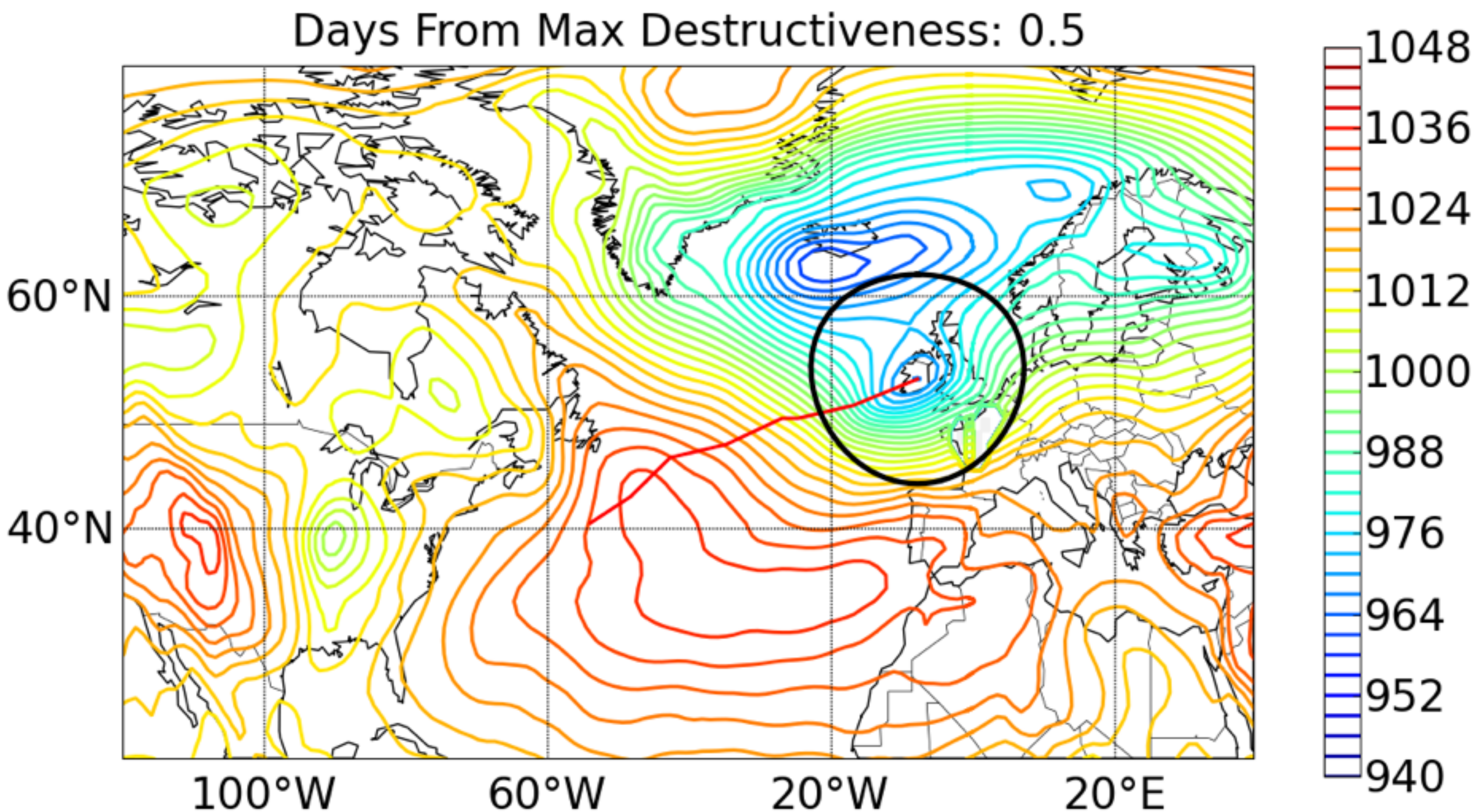
Daria storm, Jan 1990



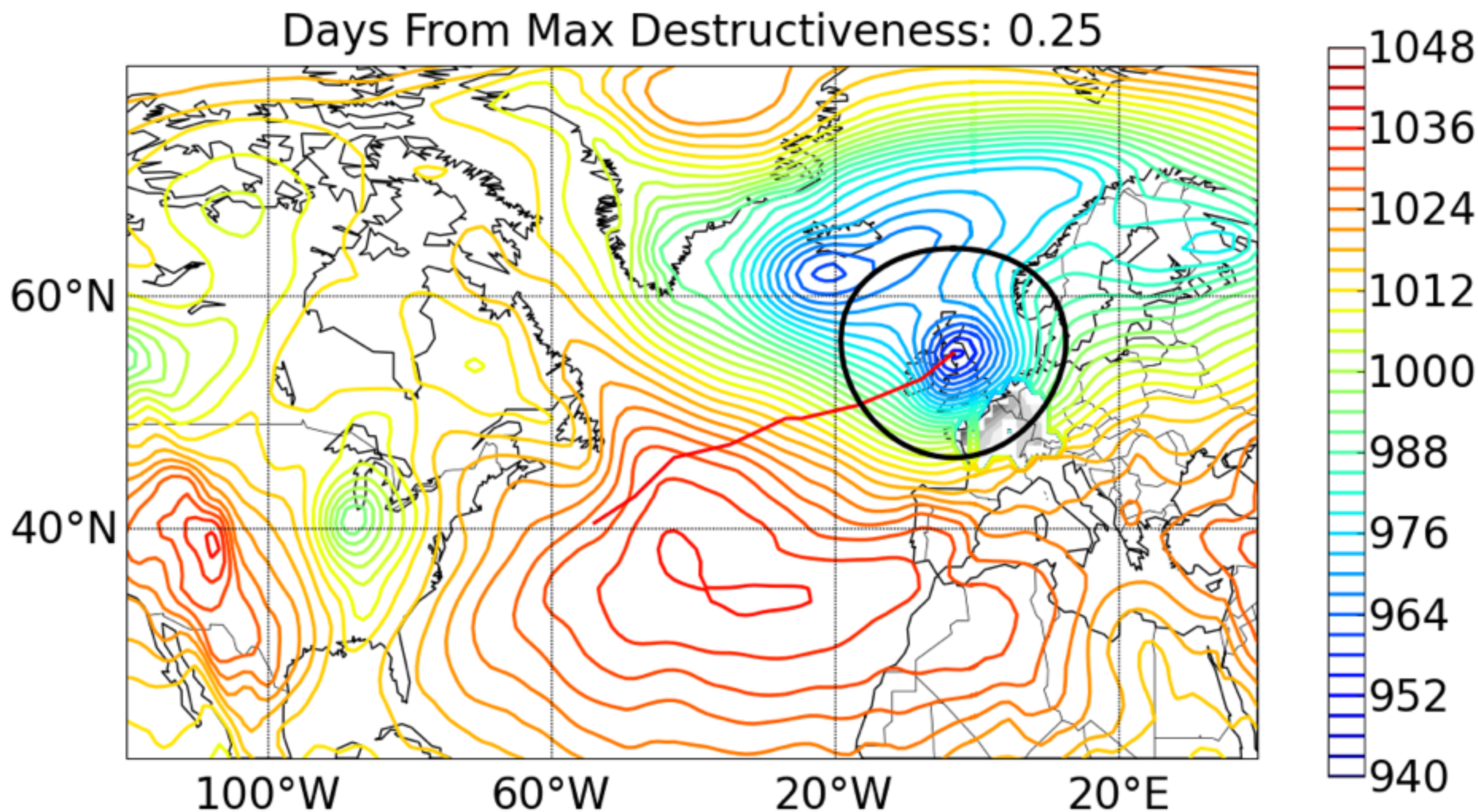
Daria storm, Jan 1990



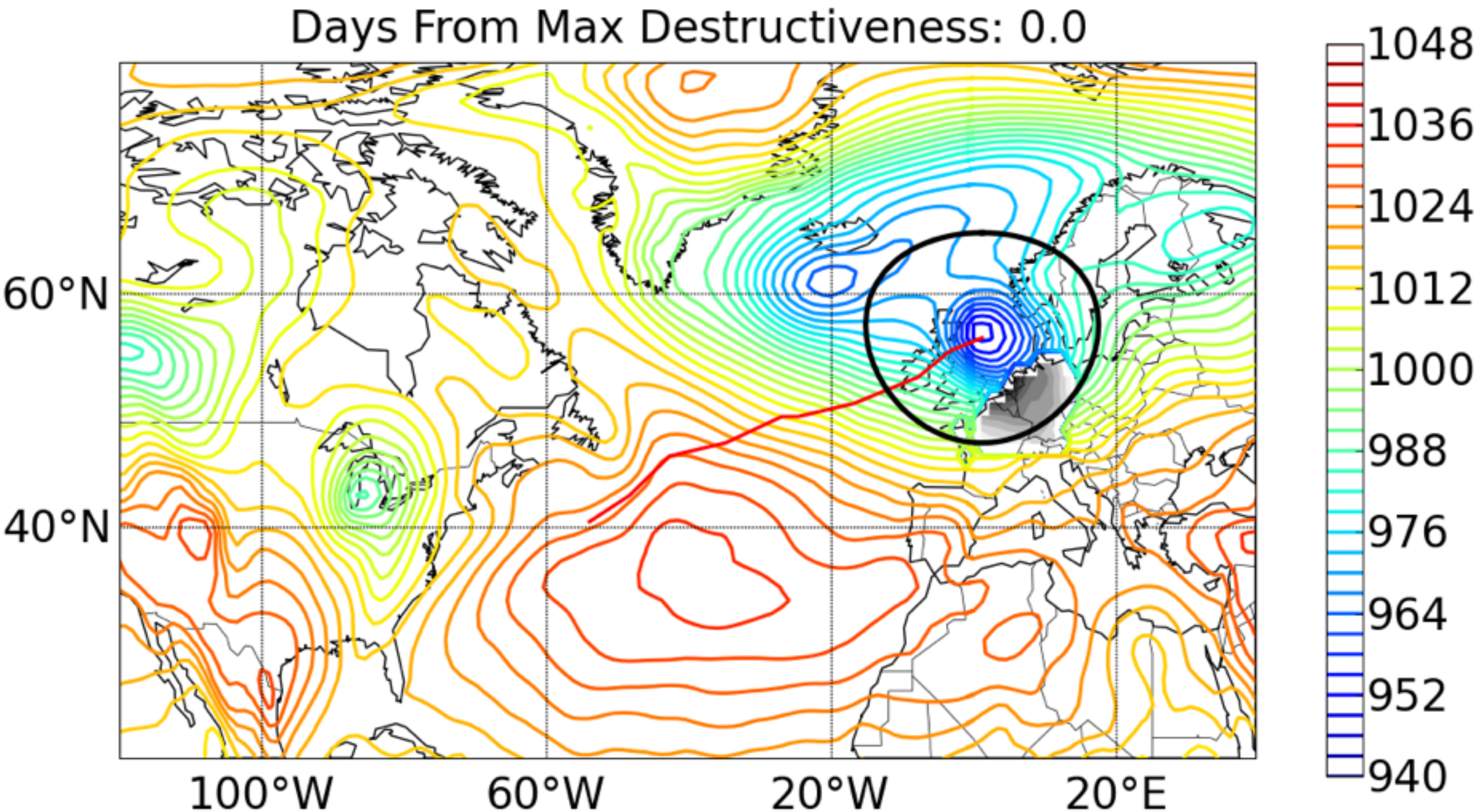
Daria storm, Jan 1990



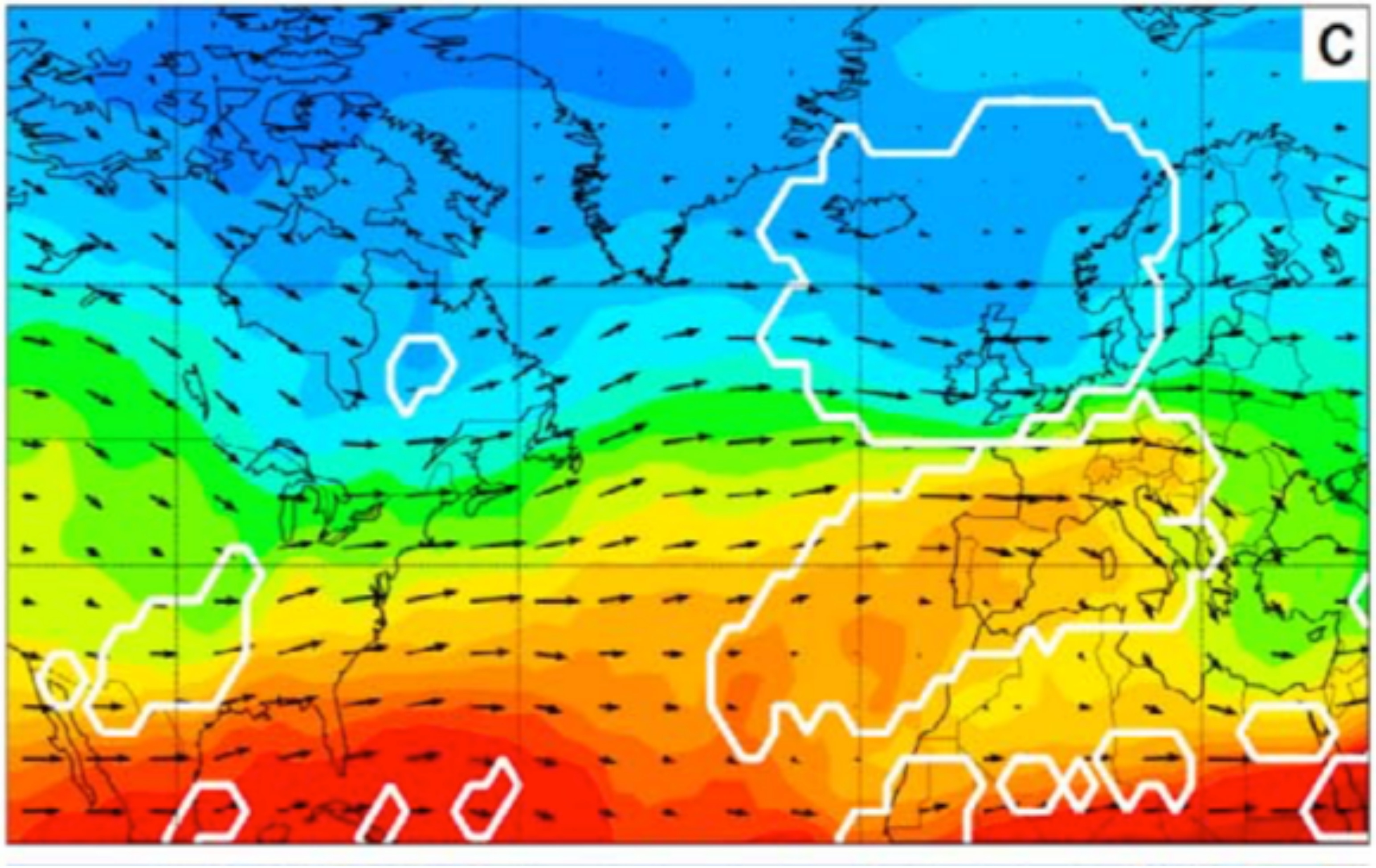
Daria storm, Jan 1990



Daria storm, Jan 1990



Storm-friendly large-scale flow conditions



(Hanley & Caballero 2012)

A 2-pronged strategy:

- Do some analysis **online**
  - things we *know* we're interested in and need high time resolution, e.g. cyclone tracking
  - SSF proposal w/Tino Weinkauf et al.

- Store (lots) of data for analysis **offline**
  - things we *don't know* we're interested in *a priori*
  - needs new tools for storing data on cheap commodity clusters and speeding up analysis
  - FAST MCP

FAST

- **1st stage**: Create a climate data-friendly environment within the hadoop eco-system
  - ie. how do you efficiently store climate model output on hadoop

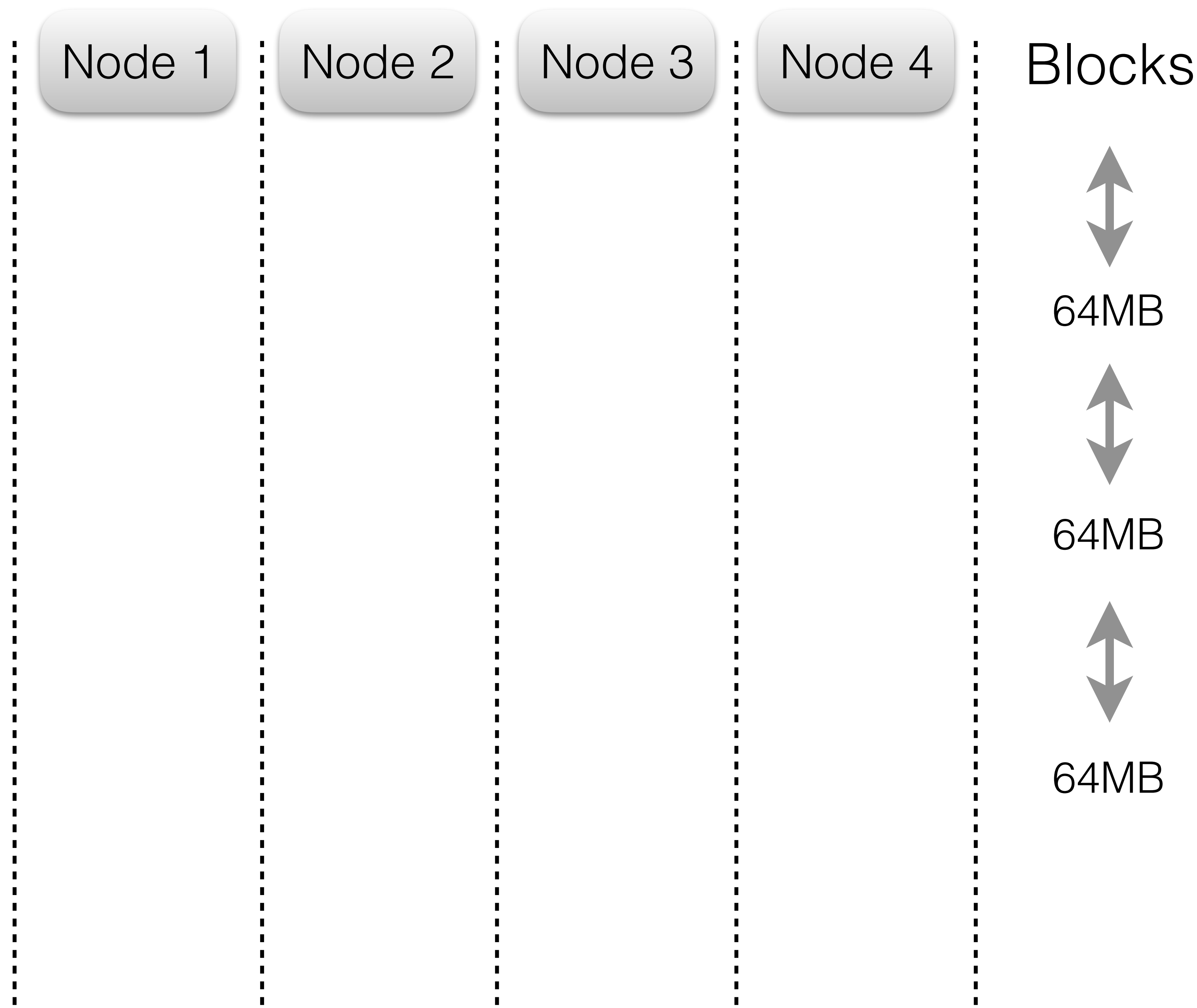
- **2nd stage**: Develop FAST analysis library
  - exploiting the parallelism intrinsic in hadoop

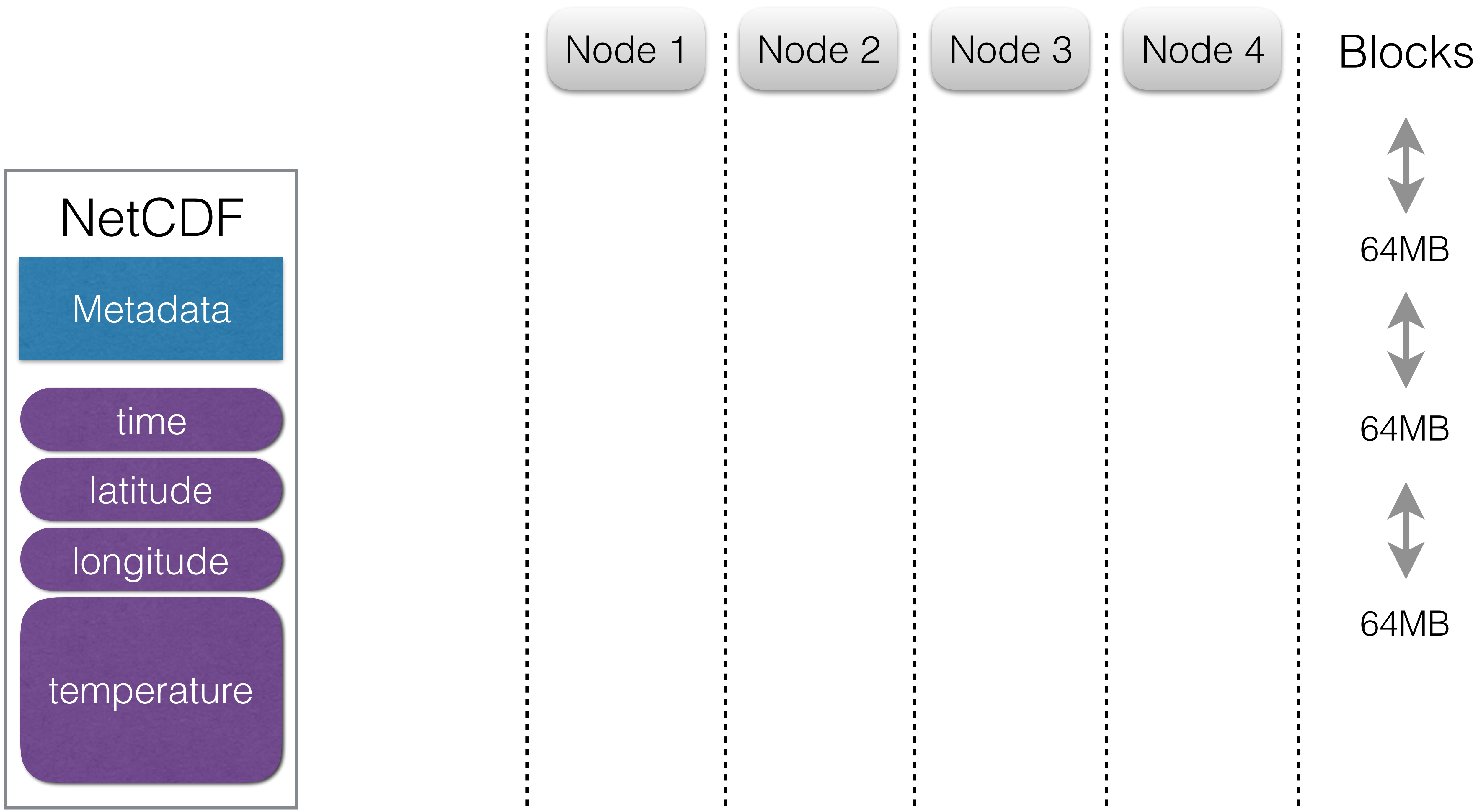
- **3rd state**: Explore deep learning capabilities
  - e.g. families of events

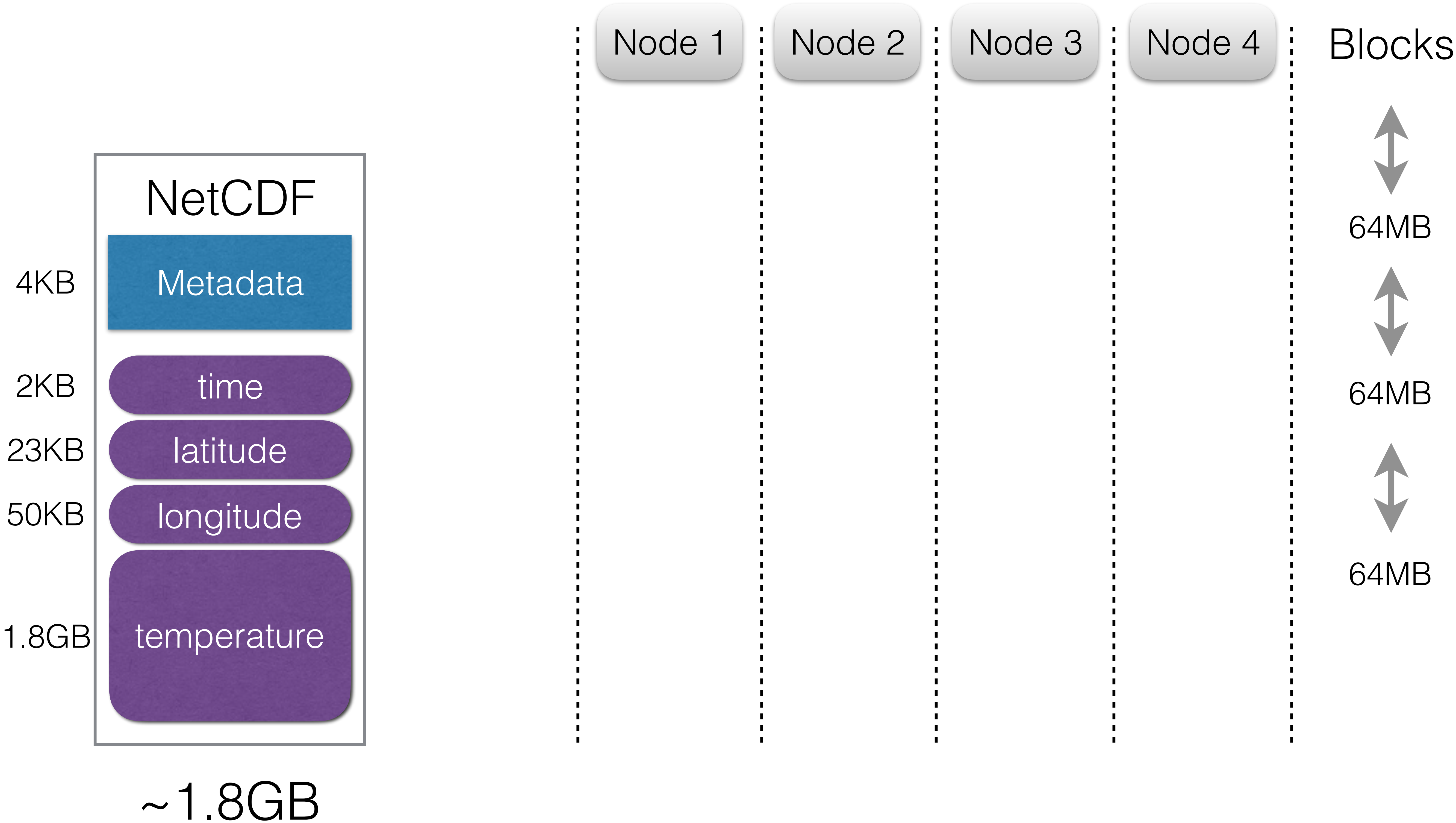
Storage and Processing of Big Data

What is Apache Hadoop?

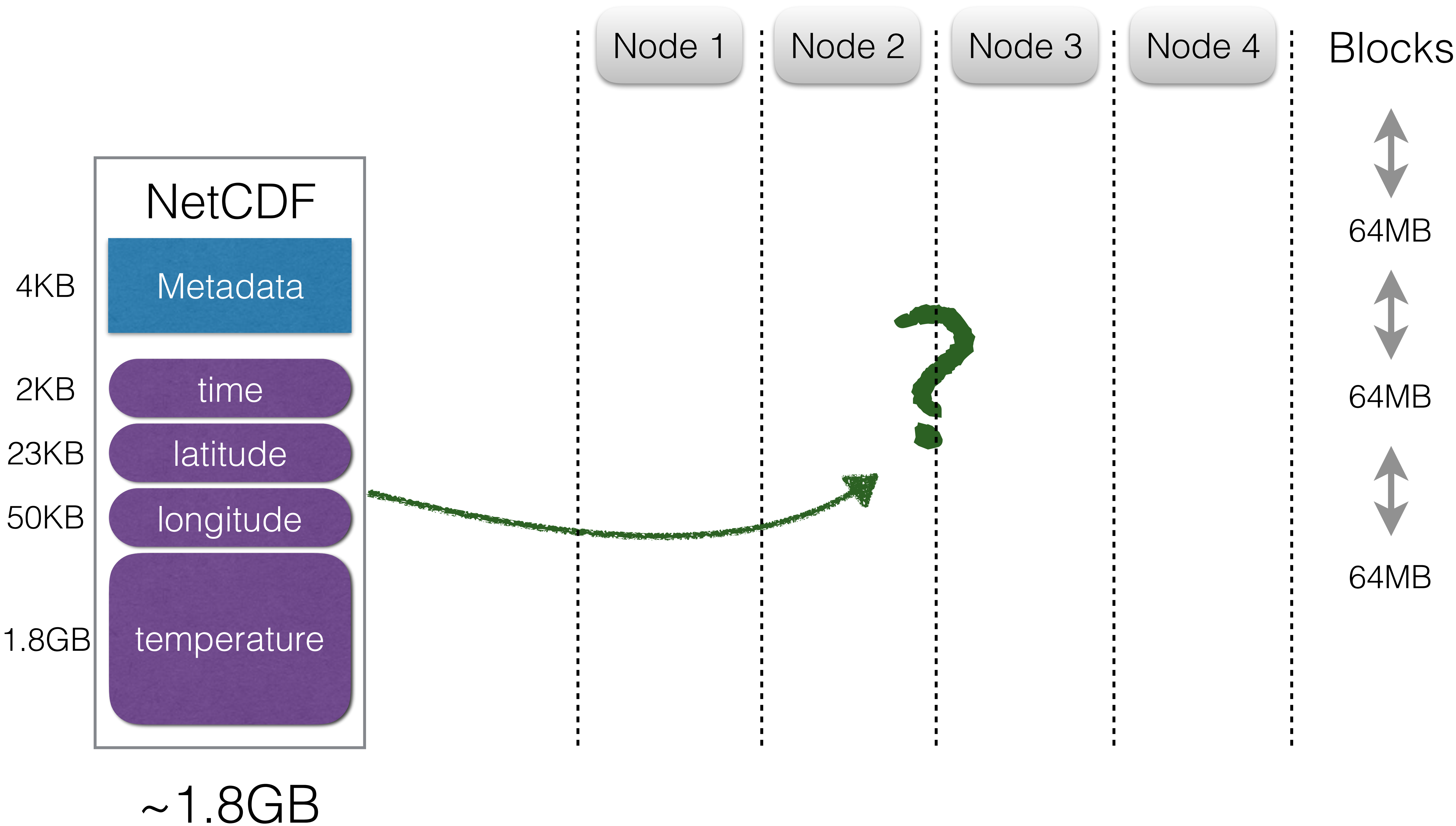
- Huge data sets and Parallel Processing
  - Scales to thousands of nodes on commodity hardware
- Schema-less or with Schema
- Fault tolerant
- Data Locality Aware
- Optimised for analytics: high-throughput file access

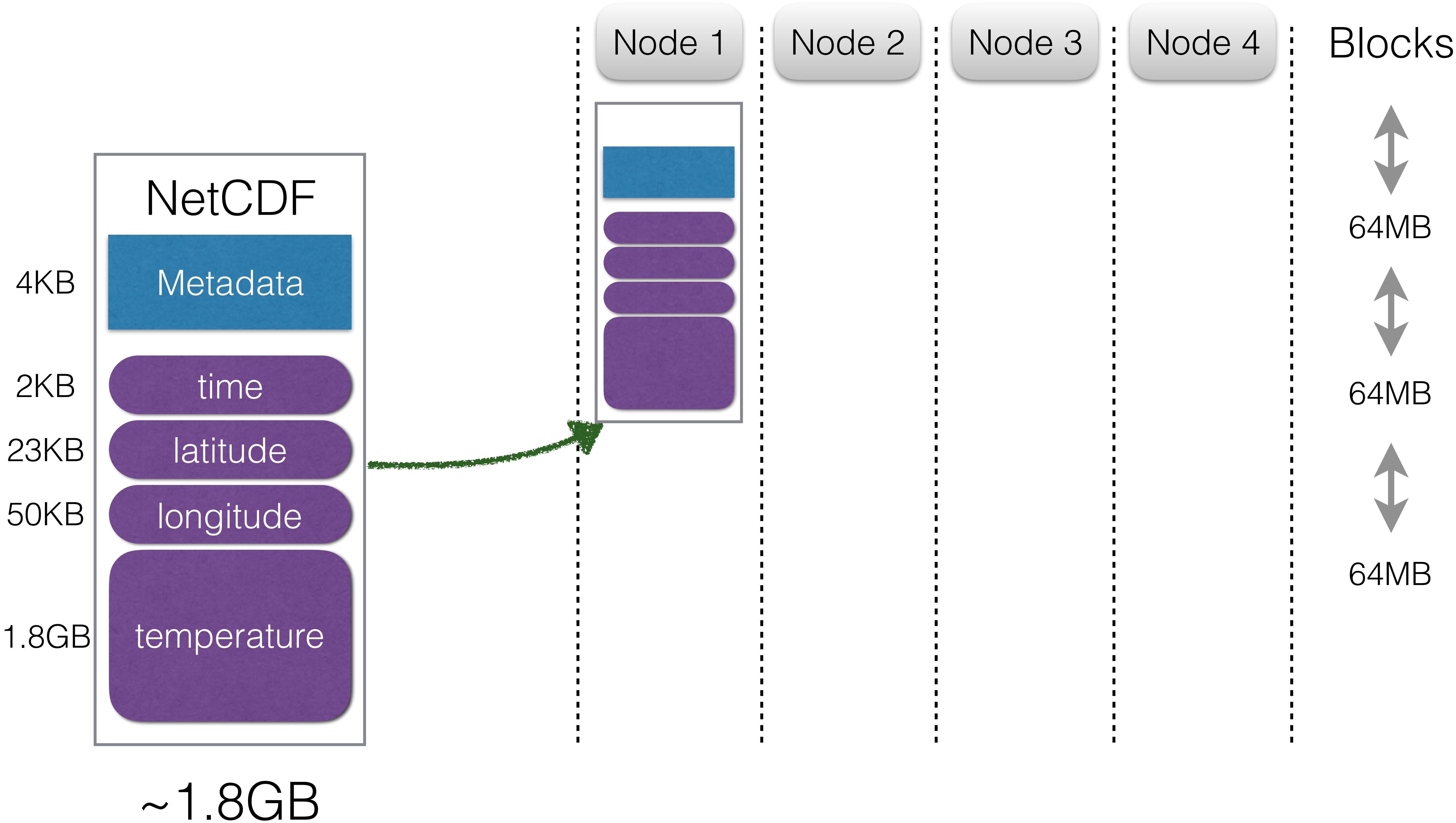

Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB

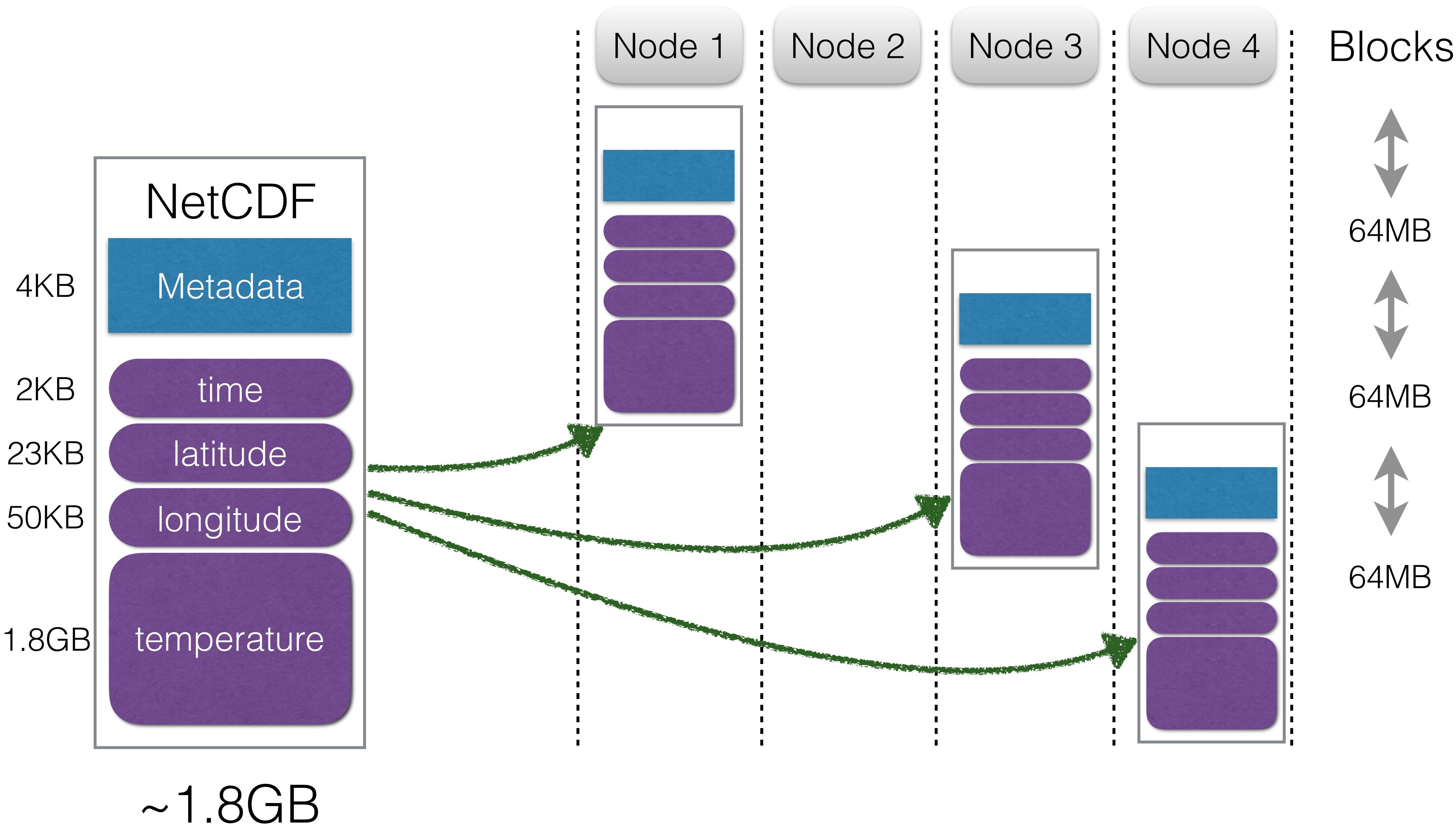

NetCDF
Metadata
time
latitude
longitude
temperature
Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB

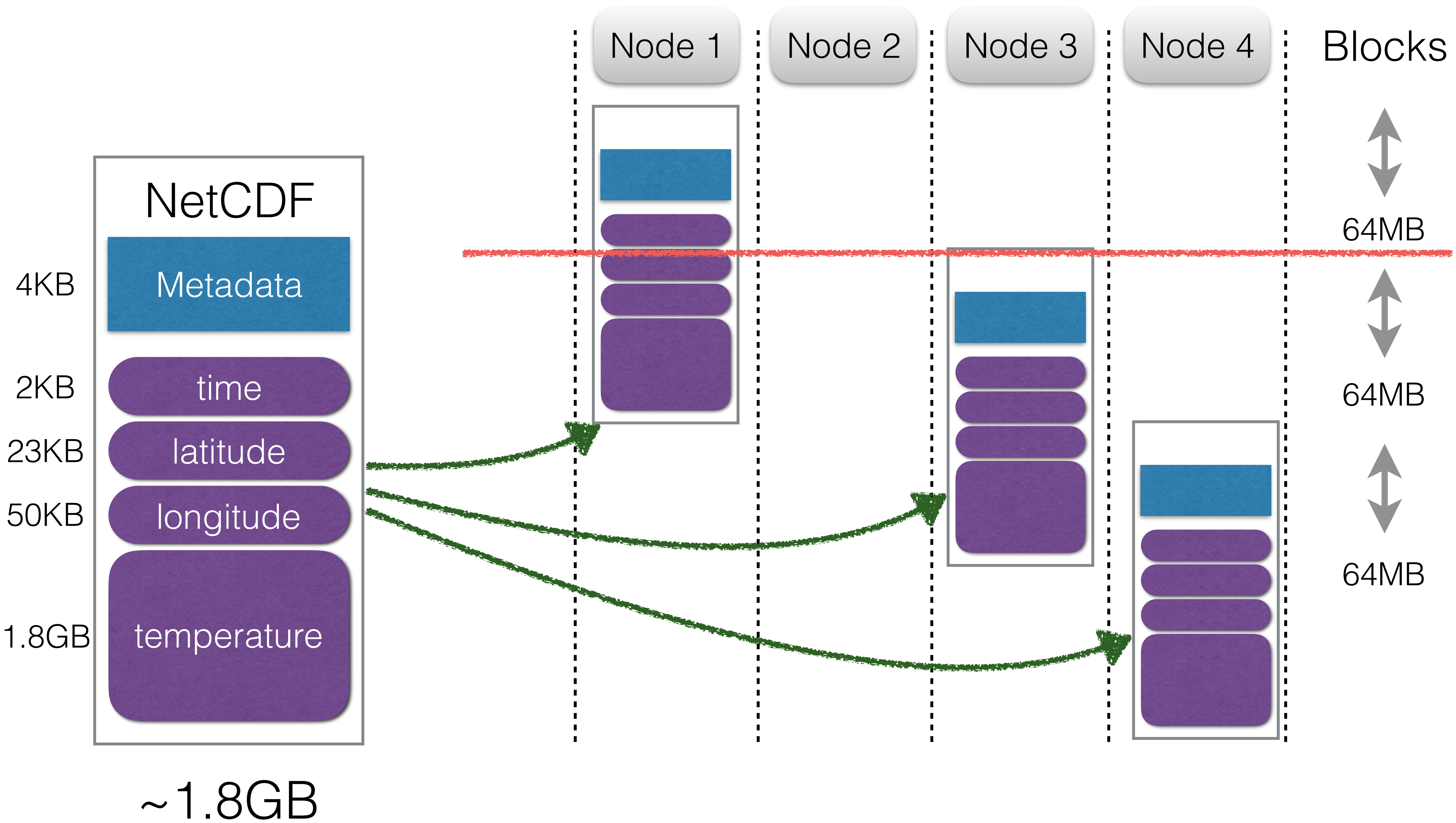


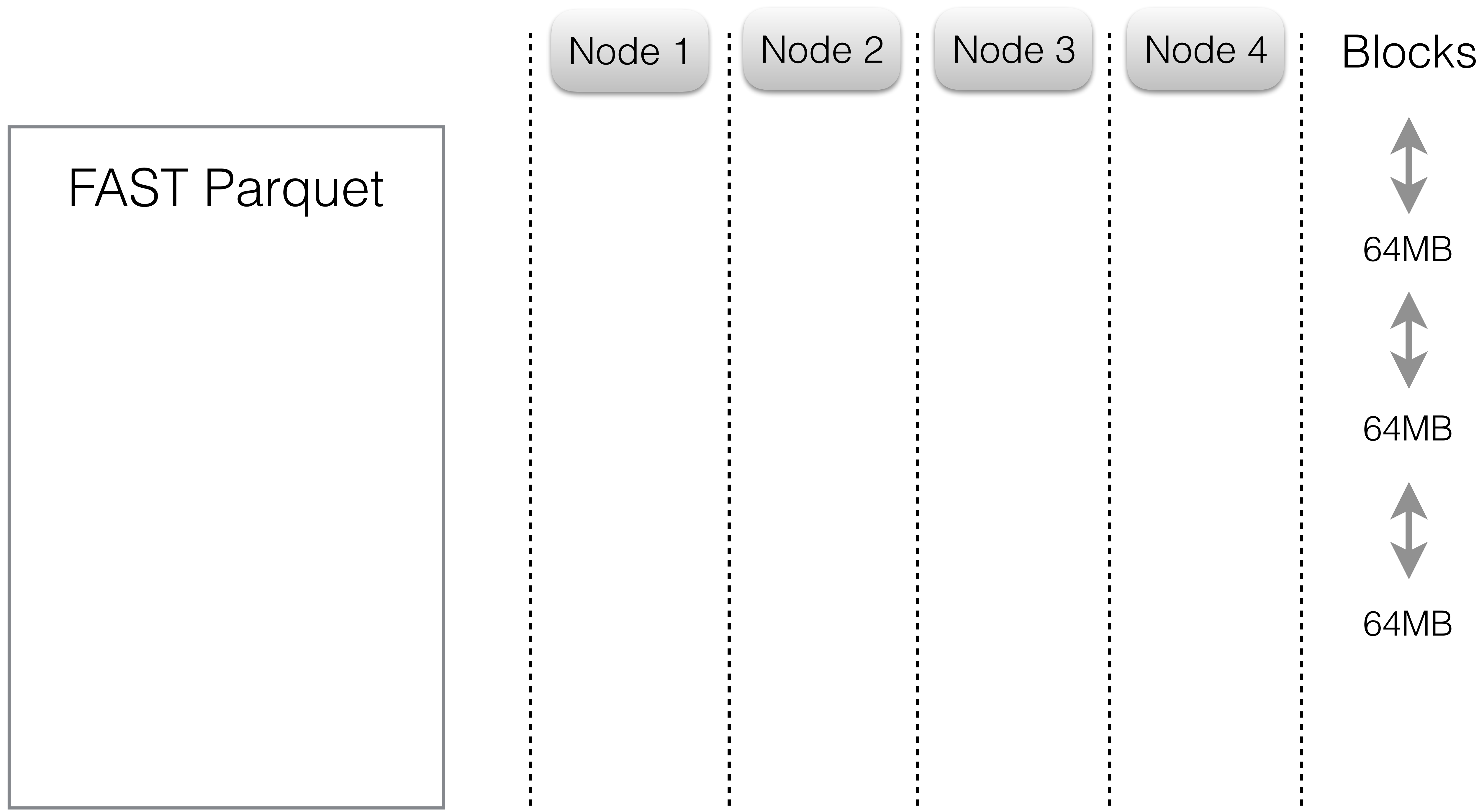
NetCDF
4KB
Metadata
2KB
time
23KB
latitude
50KB
longitude
1.8GB
temperature
~1.8GB
Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB

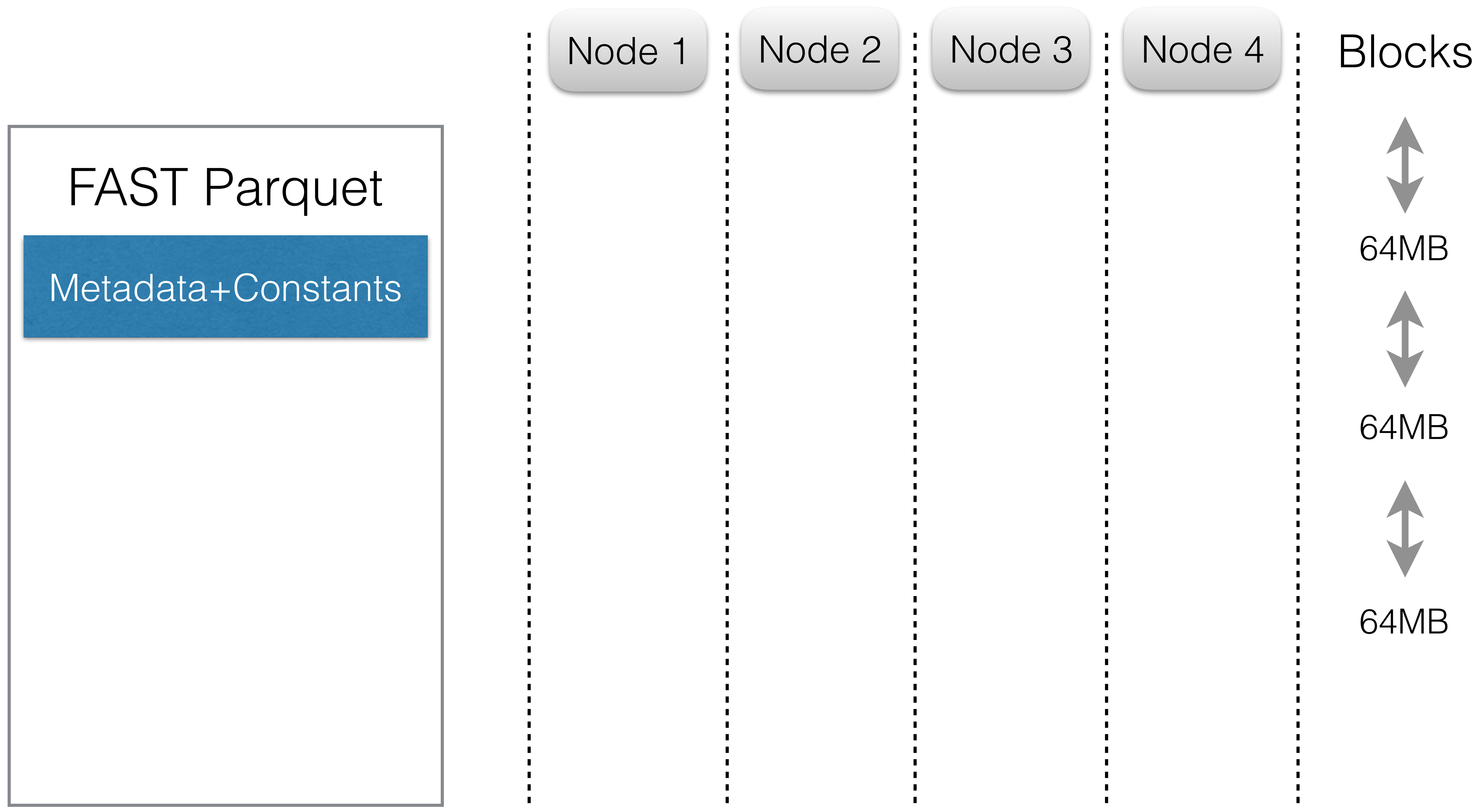

Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB
NetCDF
Metadata
time
latitude
longitude
temperature
4KB
2KB
23KB
50KB
1.8GB
~1.8GB

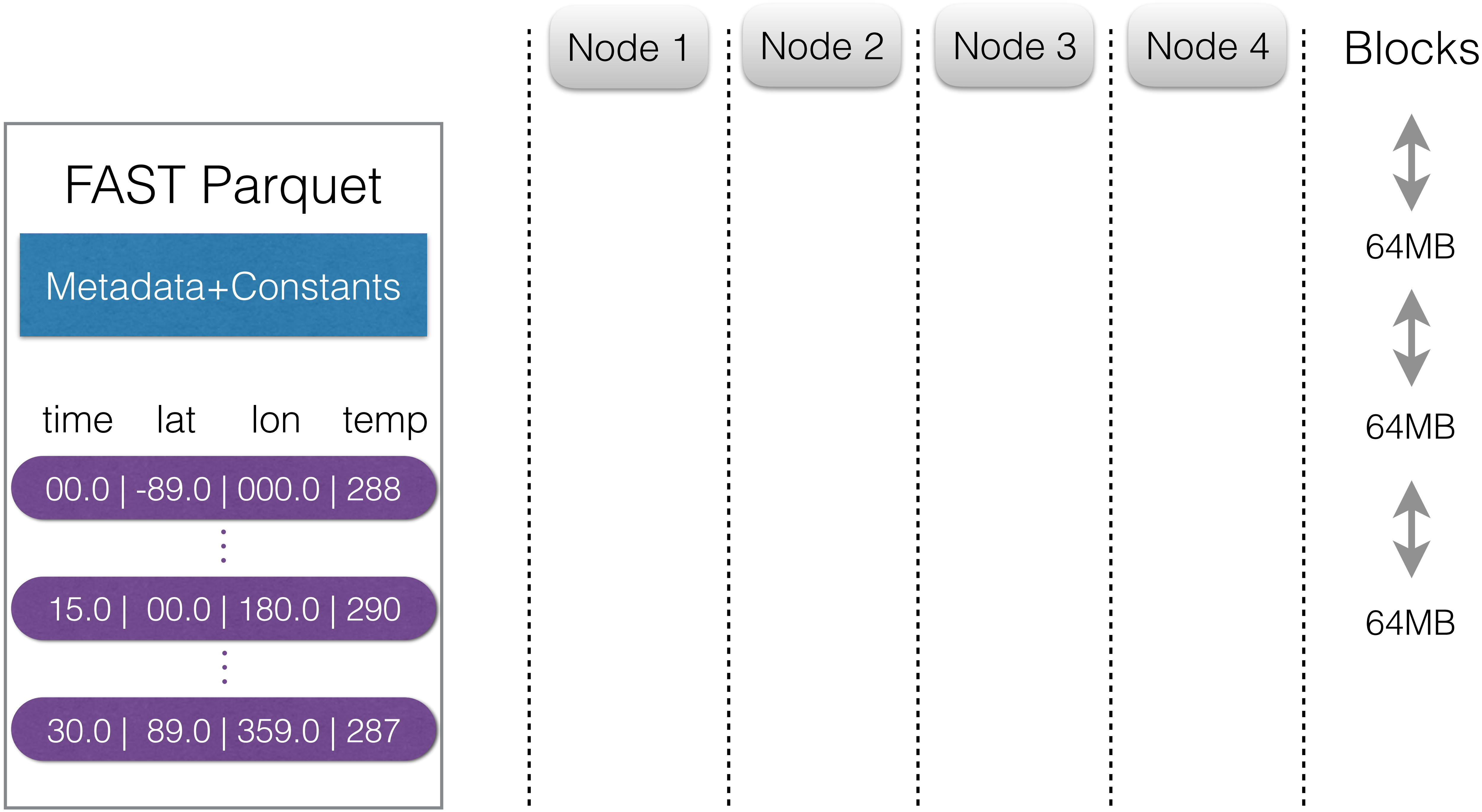

NetCDF
Metadata
time
latitude
longitude
temperature
4KB
2KB
23KB
50KB
1.8GB
~1.8GB
Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB


Node 1
Node 2
Node 3
Node 4
Blocks
NetCDF
4KB
Metadata
2KB
time
23KB
latitude
50KB
longitude
1.8GB
temperature
64MB
64MB
64MB
~1.8GB

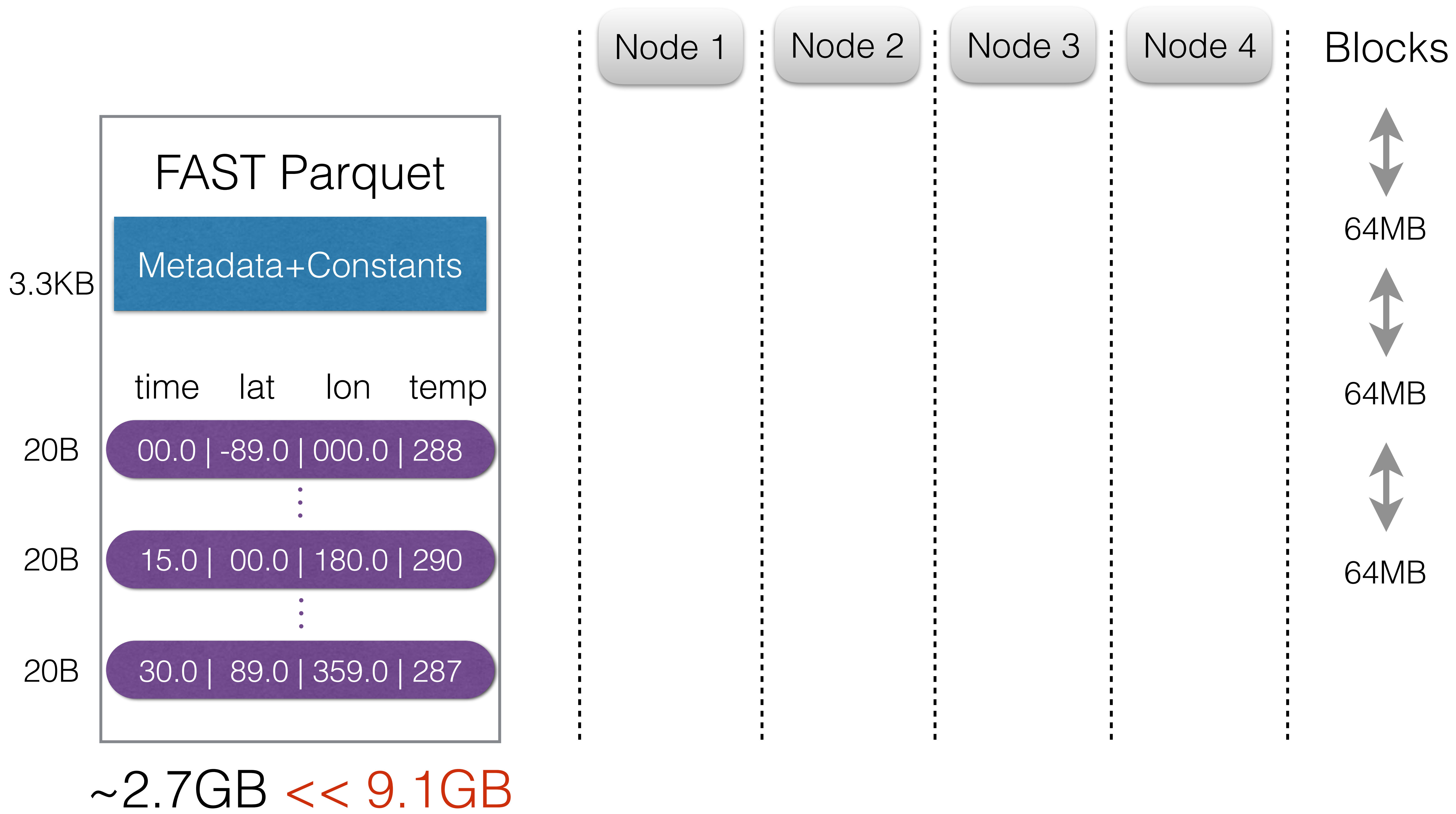

Node 1
Node 2
Node 3
Node 4
Blocks
NetCDF
4KB
Metadata
2KB
time
23KB
latitude
50KB
longitude
1.8GB
temperature
64MB
64MB
64MB
~1.8GB

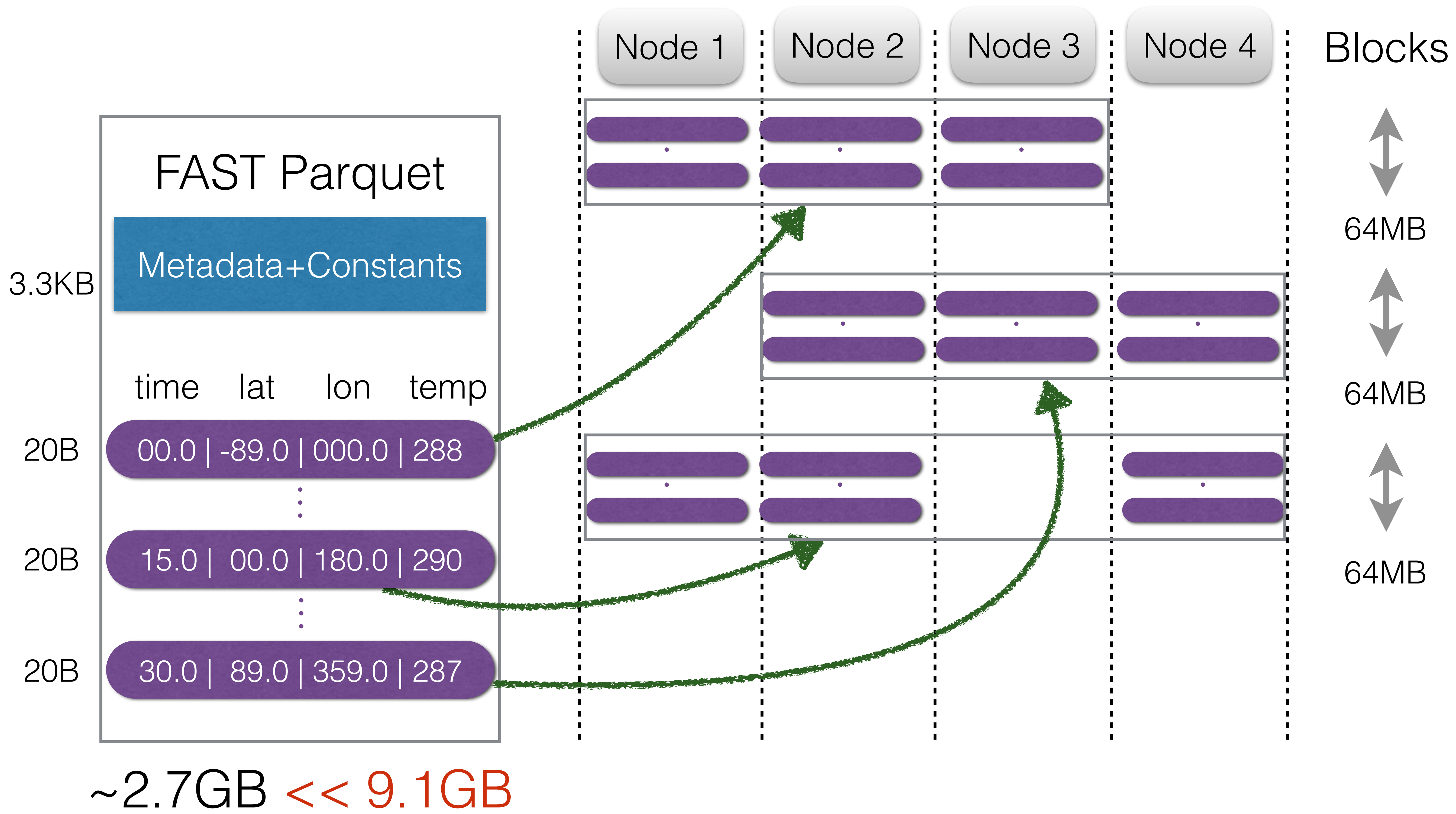

FAST Parquet
Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB

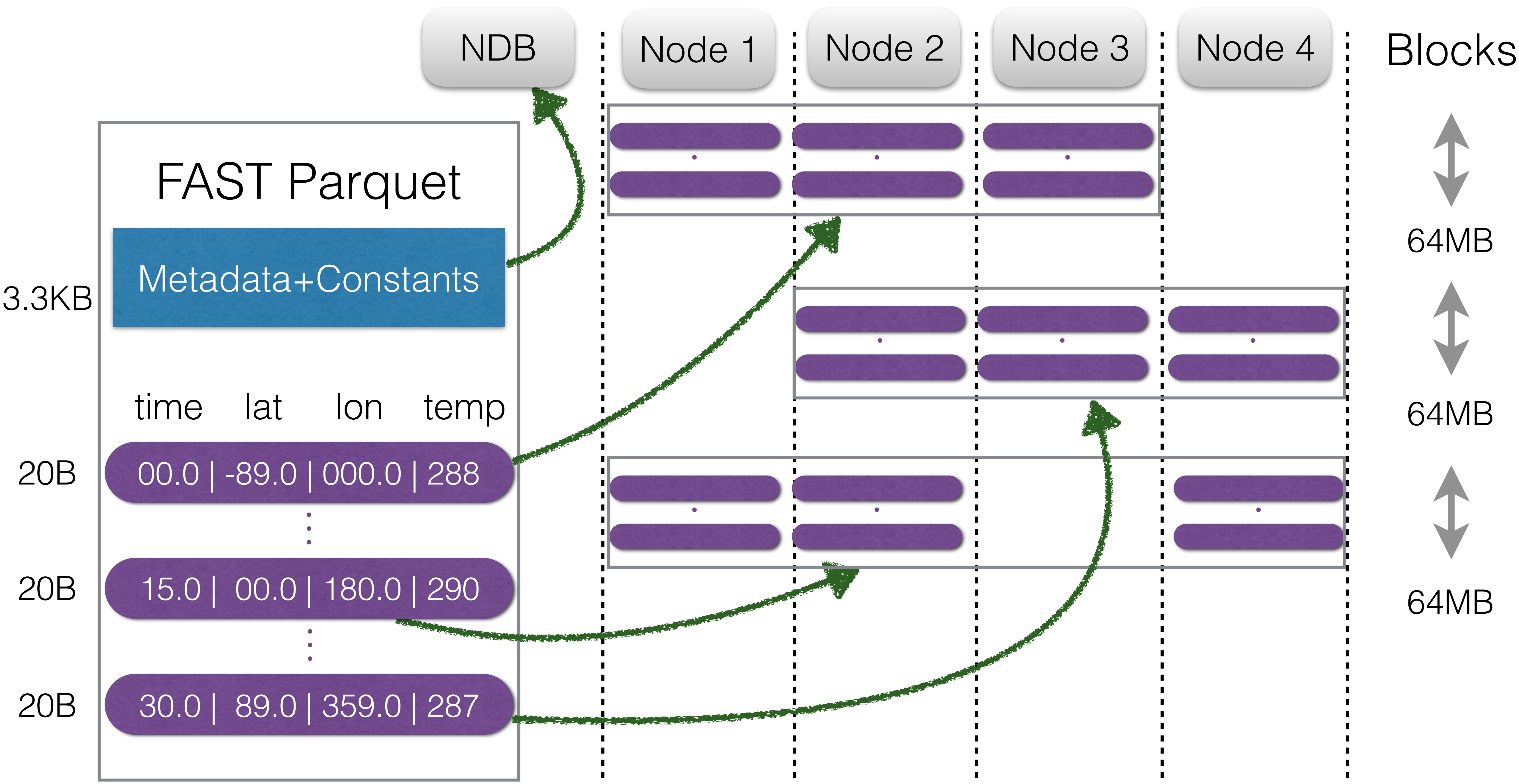

FAST Parquet
Metadata+Constants
Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB



Node 1
Node 2
Node 3
Node 4
Blocks
FAST Parquet
Metadata+Constants
time lat lon temp
00.0 | -89.0 | 000.0 | 288
15.0 | 00.0 | 180.0 | 290
30.0 | 89.0 | 359.0 | 287
64MB
64MB
64MB


Node 1
Node 2
Node 3
Node 4
Blocks
FAST Parquet
Metadata+Constants
3.3KB
time lat lon temp
20B
00.0 | -89.0 | 000.0 | 288
20B
15.0 | 00.0 | 180.0 | 290
20B
30.0 | 89.0 | 359.0 | 287
64MB
64MB
64MB
~2.7GB << 9.1GB


FAST Parquet
Metadata+Constants
3.3KB
time lat lon temp
20B
00.0 | -89.0 | 000.0 | 288
20B
15.0 | 00.0 | 180.0 | 290
20B
30.0 | 89.0 | 359.0 | 287
Node 1
Node 2
Node 3
Node 4
Blocks
64MB
64MB
64MB
~2.7GB << 9.1GB

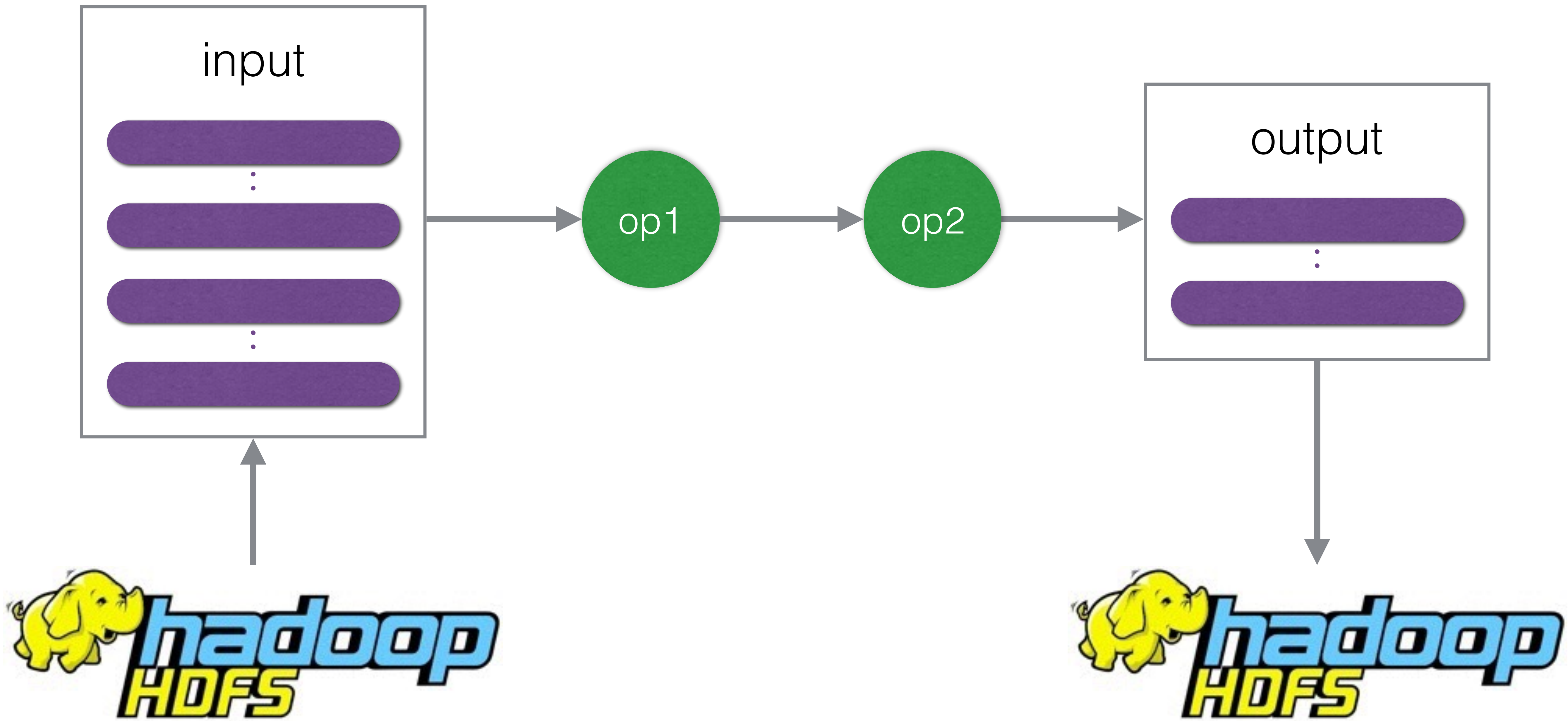

NDB
Node 1
Node 2
Node 3
Node 4
Blocks
FAST Parquet
Metadata+Constants
3.3KB
time lat lon temp
20B
00.0 | -89.0 | 000.0 | 288
20B
15.0 | 00.0 | 180.0 | 290
20B
30.0 | 89.0 | 359.0 | 287
64MB
64MB
64MB

~2.7GB << 9.1GB

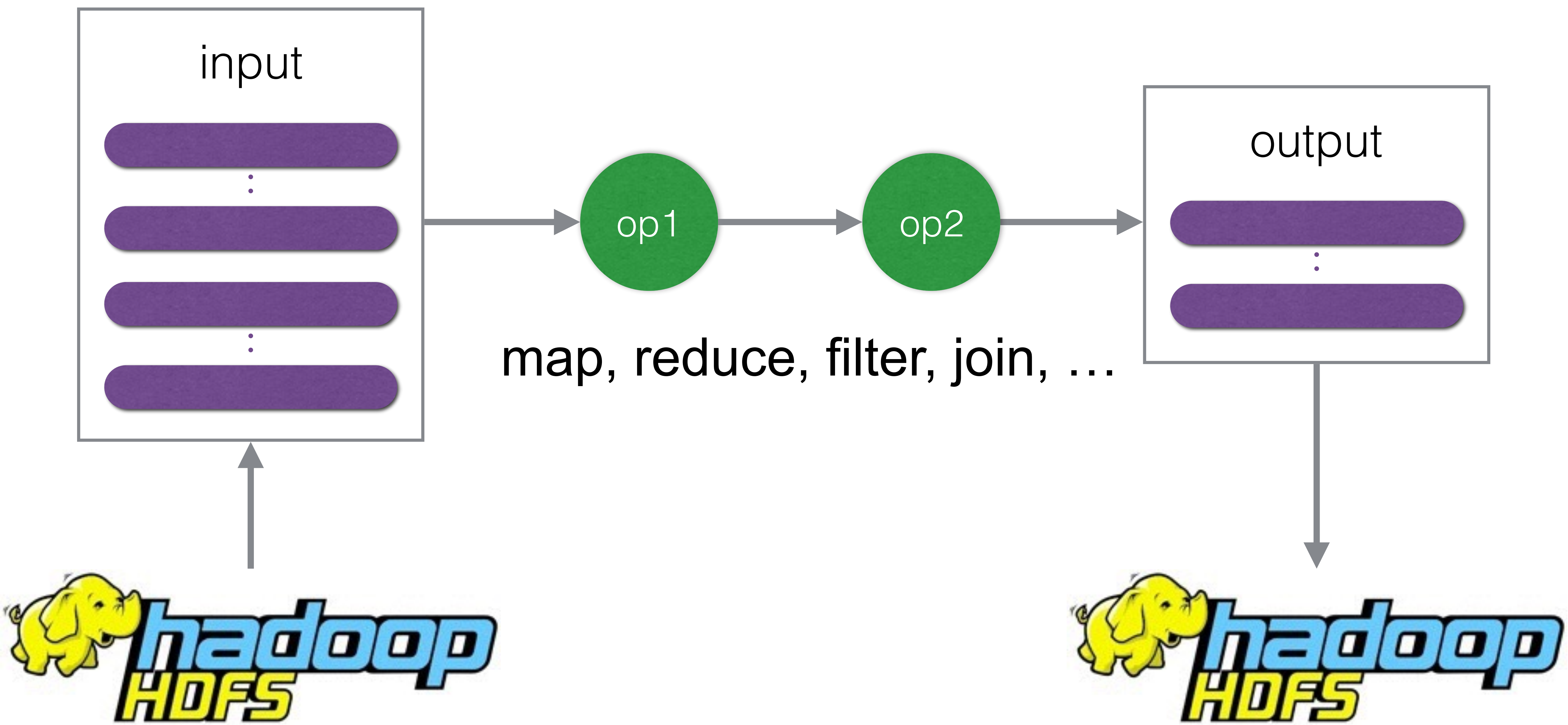
FAST Parquet

- Data re-organised into a row schema with a column for each variable (resolving coordinates)
- Variables with lower dimensionality are duplicated, but Parquet optimises duplicates away by compressing columns
- Every partition is self-describing (schema duplicated)
- Every row is independent (no cross referencing)

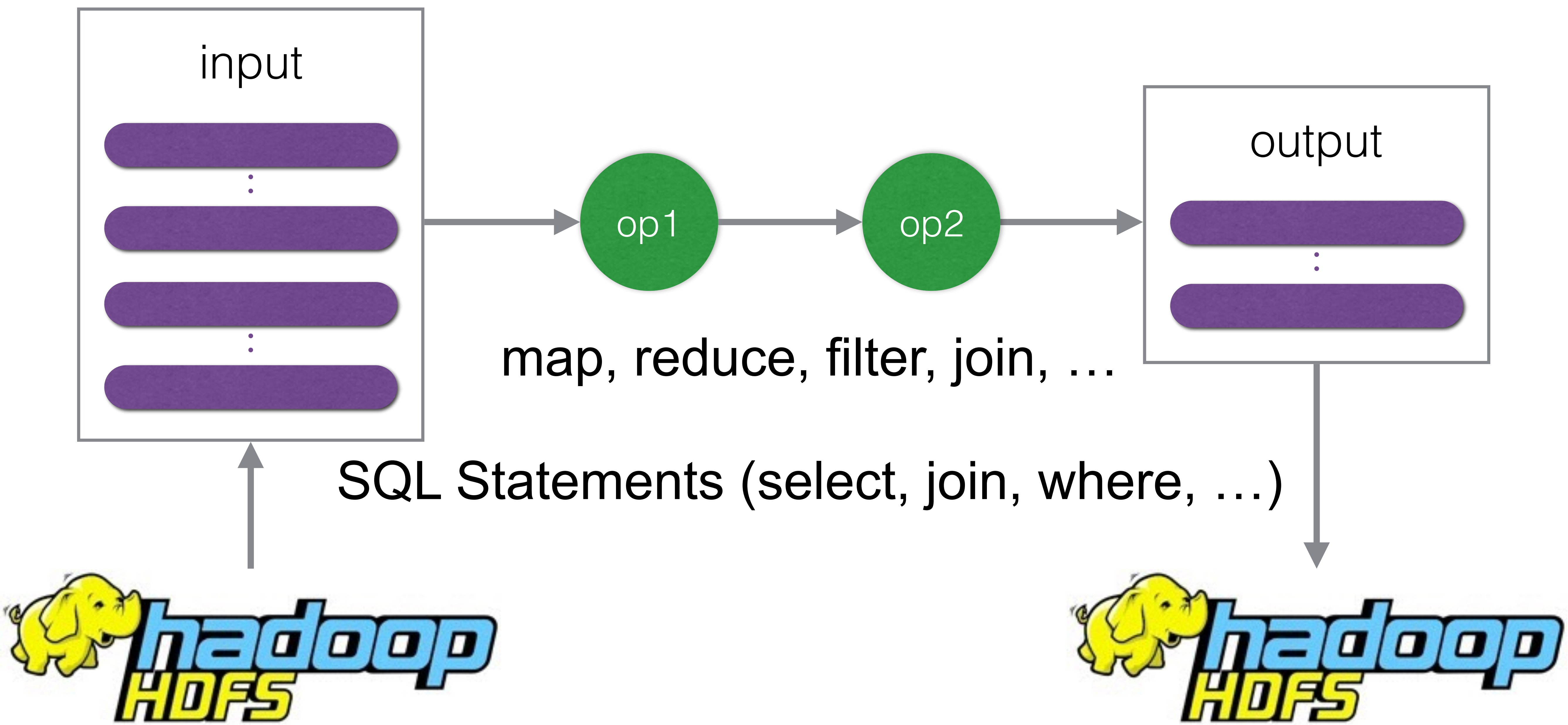
Data Processing in Apache Spark



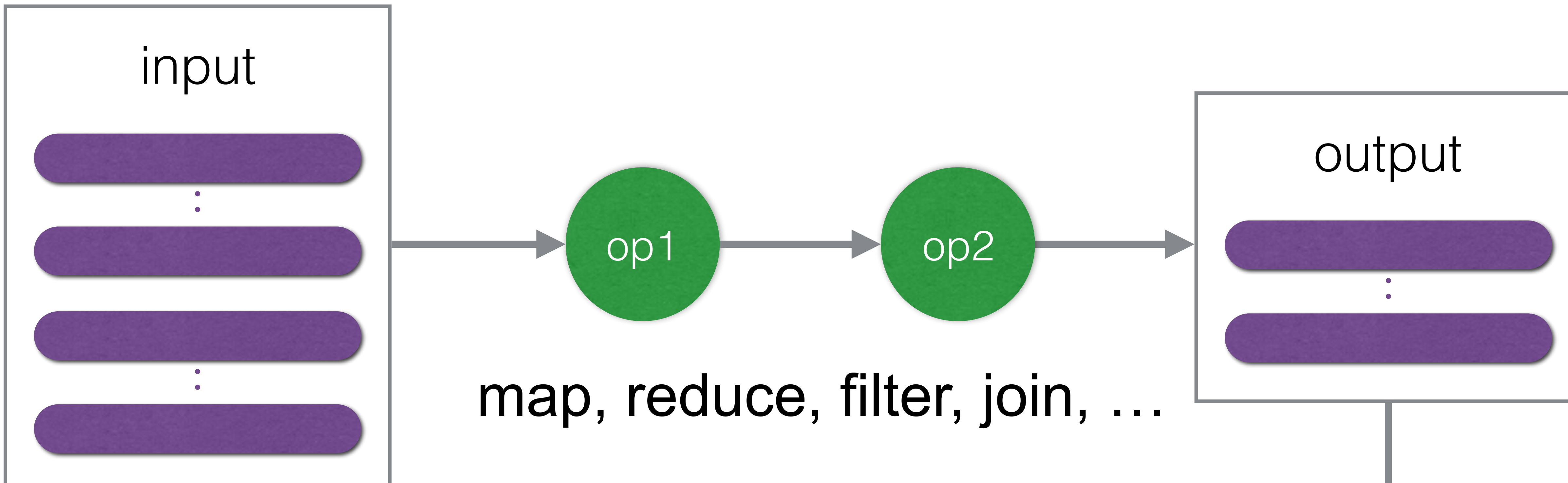
Data Processing in Apache Spark



Data Processing in Apache Spark



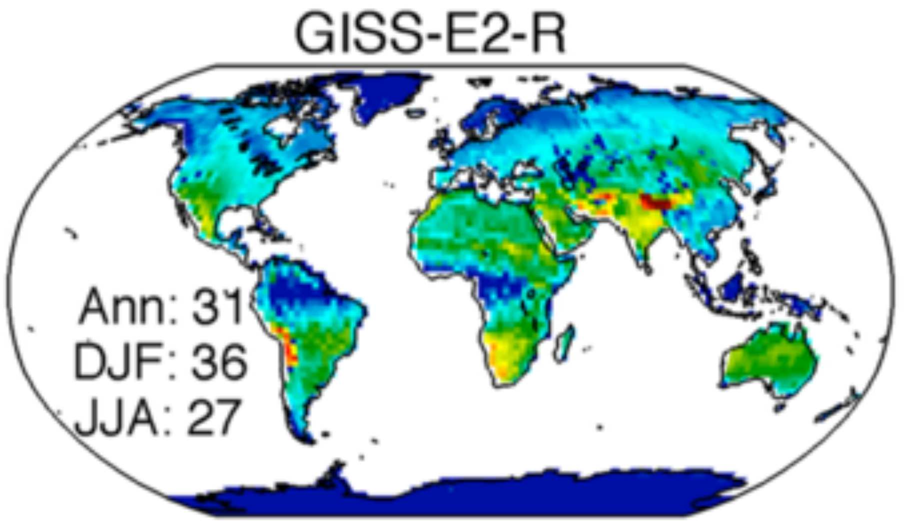
Data Processing in Apache Spark



SQL Statements (select, join, where, …)
Matrix operations (add, multiply, …)

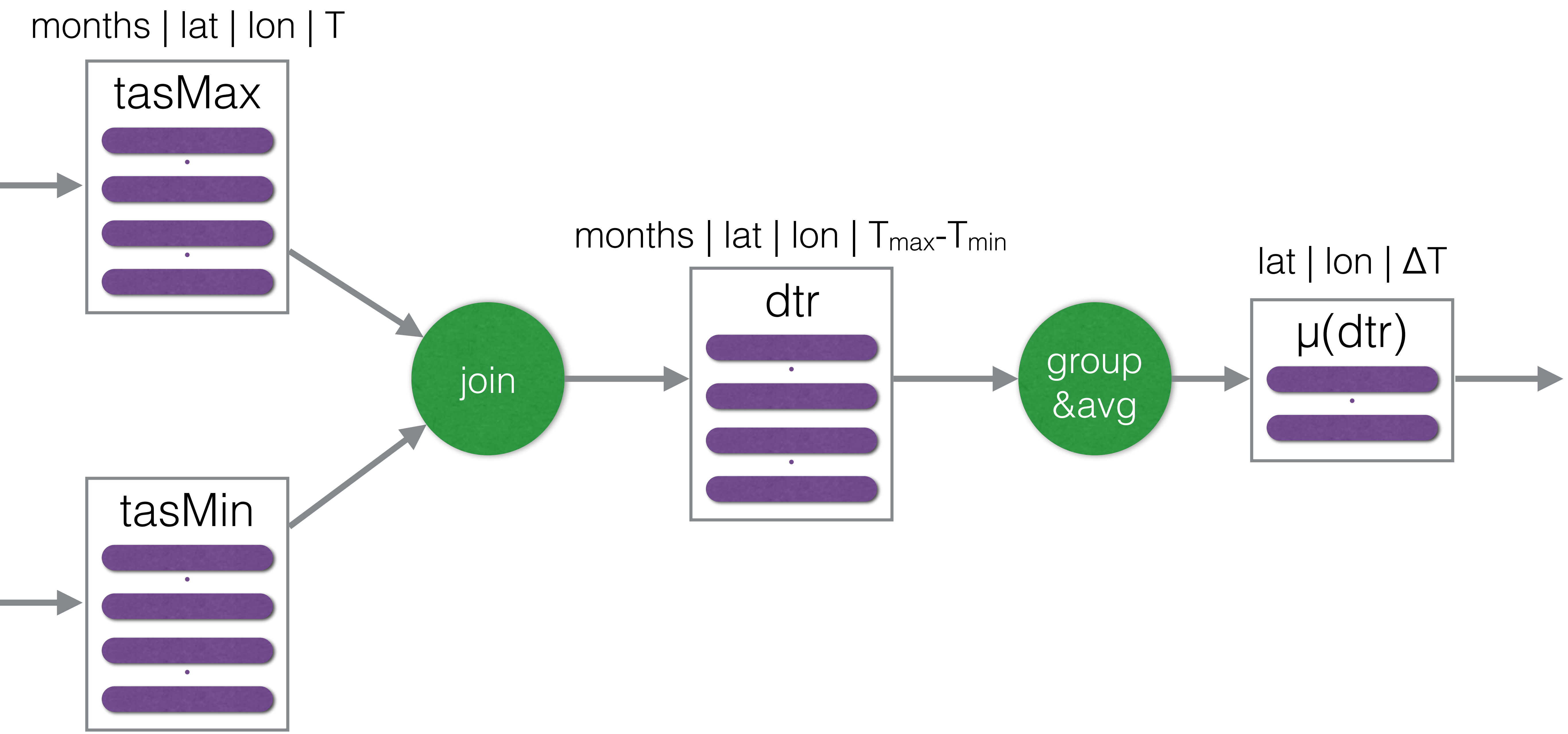
Diurnal Temperature Range

- J. Lindvall and G. Svensson, 2015: The diurnal temperature range in the CMIP5 models. *Climate Dynamics,* 44 (1-2), 405-421.

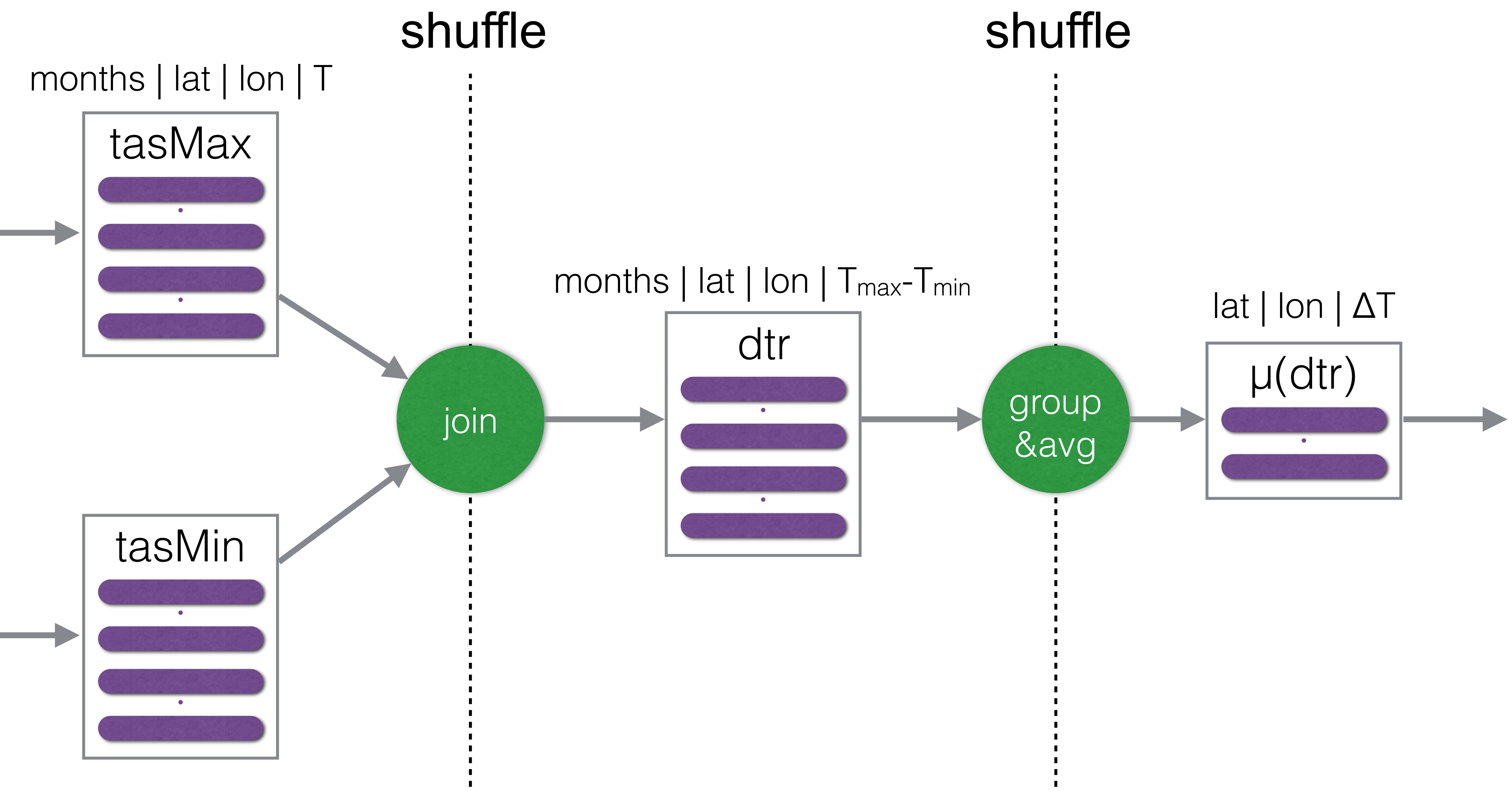


- Reproduce the analysis using Spark and Parquet.

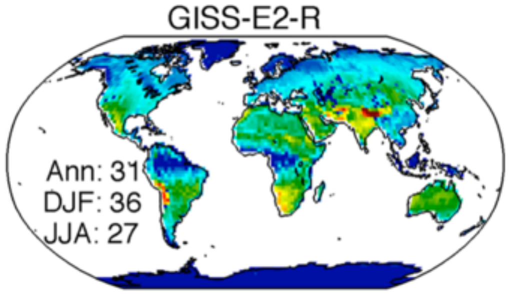
Diurnal Temperature Range



Diurnal Temperature Range



Diurnal Temperature Range



- J. Lindvall and G. Svensson, 2015: The diurnal temperature range in the CMIP5 models. Climate Dynamics, 44 (1-2), 405-421.

- Reproduce the analysis using Spark and Parquet.
- Results:
  - Absolute Error: mean≈5.4×10^{-16}, stdev≈1.4×10^{-15}, max≈5.32×10^{-15} (double precision is ~16 decimal digits)
  - Time: 1min30s for ~64MB (no parallelism, a lot of overhead, ~0.7MB/s)
  - 20GB take around 14min (parallelism of 30, ~23MB/s)
  - Performance improves with larger datasets…working on 650GB at the moment

https://github.com/ClimateFAST

Conclusions

- Analysis of large datasets can benefit from using platforms such as Apache Hadoop and Apache Spark.
- Existing file formats in climate science need to be re-written to exploit data parallel capabilities of Hadoop and Spark.
- The FAST project is developing platform support for scalable climate science analytics on Hadoop and Spark.