High-performance GEOS-Chem (GCHP) & Cloud Computing

Jiawei Zhuang
Harvard University

The 1st Regional GEOS-Chem Asia Meeting
May 21, 2018

Behind science – **technical barriers** that GEOS-Chem users are facing

- Compiling and configuring the model require **strong software skills**
- Downloading **large amounts of input data** can take weeks
- Not enough **computational resources** to perform model simulations

The paradox in GCHP development

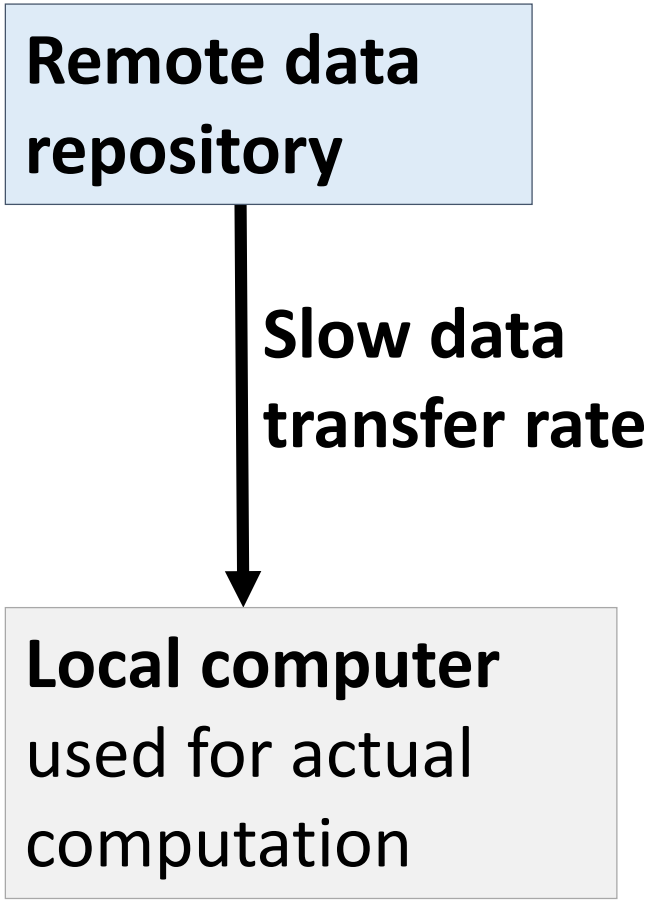
The better the model becomes, the harder to use it for research?

Use advanced software to allow massively parallel simulations → Compiling the model gets much more difficult

Can run at global high resolutions (50 km, 25 km, 12 km…) → Need much more computing power

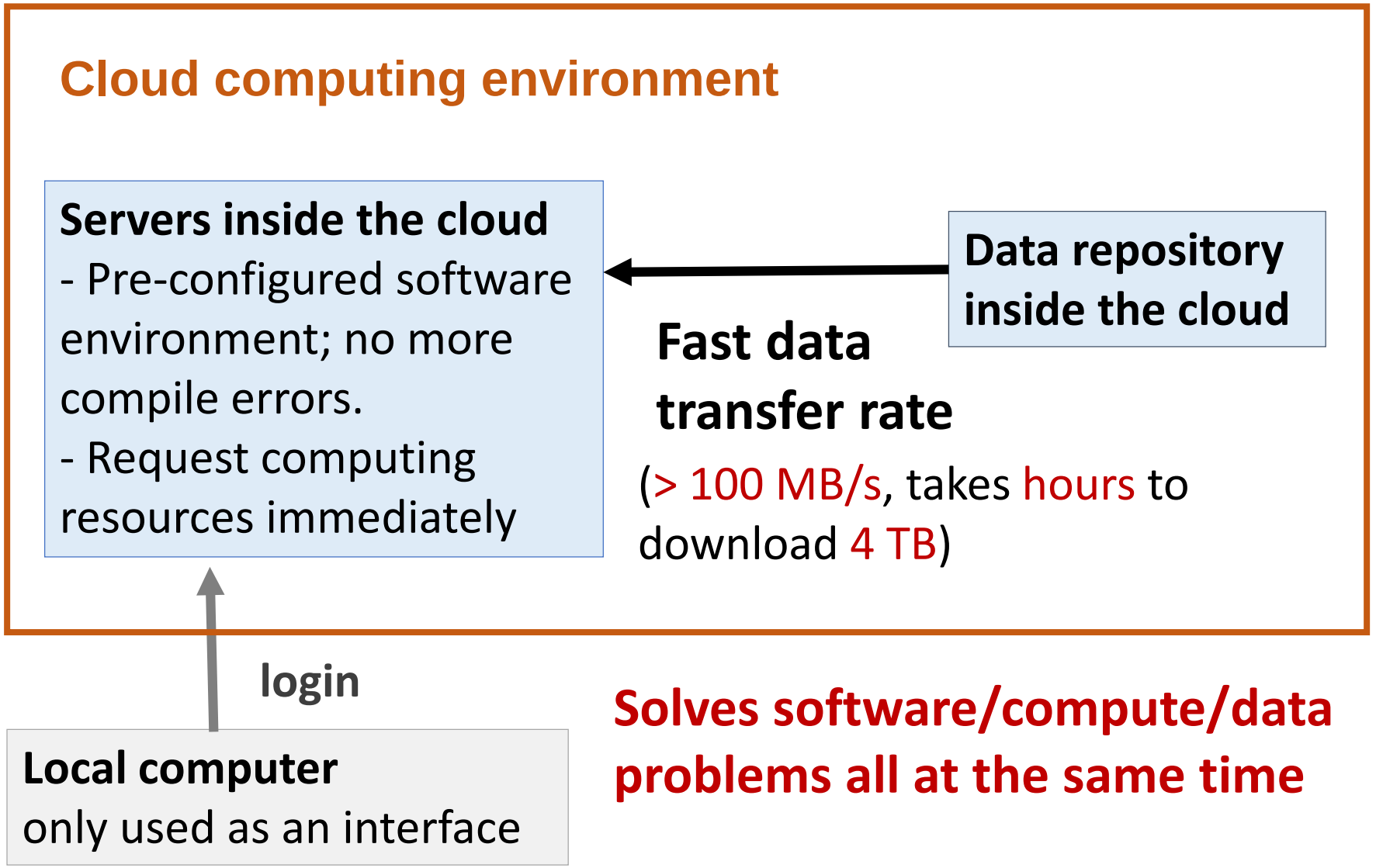
Can read meteorological field at native resolution (~12km) → Downloading data takes much longer

The problem with old paradigm – "move data to compute"



With **1 MB/s** bandwidth, downloading **4 TB** data (1-year 0.25° GEOS-FP metfields) takes **50 days (4×10^6 s)**

Adopt the new paradigm – "move compute to data"



Solves software/compute/data problems all at the same time

All GEOS-Chem data are now hosted freely under the Amazon Web Services (AWS) public data set program

Registry of Open Data on AWS

GEOS-Chem Input Data

climate | weather | meteorological | environmental | air quality

Description

Input data for the GEOS-Chem Chemical Transport Model. Including the NASA/GMAO MERRA-2 and GEOS-FP meteorological products, the HEMCO emission inventories, and other small data such as model initial conditions.

Update Frequency

New meteorological and emission data will be added when available.

License

http://acmg.seas.harvard.edu/geos/geos_licensing.html

Documentation

http://cloud-gc.readthedocs.io

Contact

http://acmg.seas.harvard.edu/geos/

https://registry.opendata.aws/geoschem-input-data/

Why care about cloud computing if you have already GEOS-Chem set up locally

- Get **immediate access** to the latest version of the model and all input data
- **Seamlessly switch** between cloud platforms and local machines, using the “Linux container” technology (think it as a way to set up the model with one click on any computers)
- Make research projects **fully reproducible**

Cloud opens many new research opportunities

- NASA is planning to move ~250 Petabytes of satellite data to AWS cloud
  https://earthdata.nasa.gov/about/eosdis-cloud-evolution
- Lots of Earth science data are already on the cloud
  https://aws.amazon.com/earth/
- Cloud is the go-to choice for training deep neural networks (can the deep learning/AI hype benefit science?)
  https://aws.amazon.com/machine-learning/amis/
- …

GEOS-Chem-on-cloud is ready for use!

- GEOS-Chem-classic is fully functioning
- Data analysis can be performed in the cloud using the free, open-source scientific Python ecosystem
- GCHP also runs on the cloud
  (with some non-critical performance issues to be resolved)

Comprehensive tutorials at cloud-gc.readthedocs.io

Hands-on training of GEOS-Chem, Python and AWS cloud on Wednesday (May 23) 14:30-17:00