UNICLOUD

Docker Use at Unidata

Julien Chastang (UCAR, Boulder, CO USA), Ward Fisher, Sean Arms

January 14, 2016

Outline

What We do at Unidata

Five-year (2013-2018) plan and the Cloud

What is Docker?

Microsoft Azure Grant

Goal of Unidata Server (Motherlode) in the Cloud

"Dockerizing" the LDM, TDS, RAMADDA

Lessons Learned, Other Efforts, Future Work

What we do at Unidata

- ▶ End-to-end geoscience data services
  - ▶ IDD (13 TB/day to 263 IDD sites)
- ▶ Tools for scientific data life-cycle
  - ▶ LDM, IDV, TDS, netCDF
  - ▶ RAMADDA, McIDAS/ADDE
- ▶ Community support help desk for all of the above
- ▶ Primarily serving the US academic community

Unidata 2018 Five-year Plan and the Cloud

- *Transition to [cloud computing for geoscience] is vital to our community's ability to pursue research and education in the 21st century.*
- Cloud shifts IT infrastructure (hardware/OS/software) from in-house departments to data centers managed by cloud providers yielding time, cost savings, and <u>new opportunities</u>.
- Advantageous for big data problems coupled with data proximate analysis
- For profit (Amazon, etc.) and non-profit clouds (NSF sponsored XSEDE for HPC and soon cloud computing)

What is Docker?

- Docker is an open source platform for uniformly building, deploying and running software in the cloud.
- Borrows from the notion of standard size shipping containers, but for software
- Linux only, so most applicable (but not limited) to server-side
- Barrier to entry is low and adoption is already wide-spread
- In last nine month, Unidata has made a significant foray into Docker.

Azure for Research Grant

- Unidata colleague Ward Fisher obtained "Azure for Research" Grant
- $20,000 of cloud-computing resources
- Grant allows Unidata to experiment with cloud resources
- Ubuntu Linux VMs with out-of-box Docker support

Motherlode Server

- Unidata server running LDM/TDS/ADDE/RAMADDA serving TBs of data (GFS, Satellite, Radar, etc.)
- Originally "Demonstration" server
- Has become an ad hoc operational server (99.96% uptime) widely used by atmospheric community via IDV, etc.
- Goal: Let's put Motherlode in the cloud!

What does it Mean to "Dockerize" Software

- "Dockerization" is the process of defining <u>minimal</u> OS environments that will allow your application to run
- These environment are codified in <u>Dockerfiles</u>
- Build and share these environments AKA images (e.g., via dockerhub)
- An instance of an image is called a "container"

Dockerizing the LDM

- Important b/c <u>upstream</u> of other Unidata tools
- Difficult part b/c of privileged user requirements
- Other tricky problems like `rsyslog` requirement
- Coordination of file system inside vs. outside container is tricky (e.g., logging)
- Product queue must live <u>within</u> the container

Dockerizing the TDS/RAMADDA

- ▶ TDS and RAMADDA are Tomcat Java web applications
- ▶ For TDS and RAMADDA, use the canonical `tomcat` base image which buys us a lot
- ▶ For TDS, leveraged work of TDS Team
- ▶ For TDS, also must Dockerize TDM for creating index files and TDS notification
- ▶ For RAMADDA, simply tuned Tomcat, and defined data directory

Software and Data Configuration

- LDM: Configure what data you want via `pqact` configuration file
- Data: 13km RR, 1° GFS, GOES Satellite, NEXRAD Composites
- TDS: Configure catalogs, and where to find data on file system (reuse TdsConfig project).
- RAMADDA: Standard web configuration and server-side views

Motherlode in the Cloud

- Running LDM, TDS, RAMADDA on Azure Cloud
- Provision VM
- Download docker images
- Start docker containers
- Whole process takes < 30 minutes, mostly waiting for VM provisioning and Docker image download
- Fully functional template

Lessons Learned

- Can deploy a motherlode-class machine in minutes
- Codification of "dark knowledge", finally(!)
- Must still configure to suit your needs
- Can be expensive, needs realistic business model but resources like XSEDE may be path forward
- Images Available on Dockerhub:
  `https://hub.docker.com/u/unidata/`
- Instructions in extended abstract and online:
  `http://unidata.github.io/Unidata-Dockerfiles/`

Other Efforts and Future Work

- IDV/cloud control, Python, ADDE, GEMPAK, AWIPSII
- Gradually scale up to more voluminous data feeds
- Make more robust and oriented towards operations
- Continue researching a viable business model (perhaps similar to HPC?)

Acknowledgments

unidata

UCAR

NSF