Skyport Intro

Folker Meyer, folker@anl.gov

Wolfgang Gerlach, wgerlach@mcs.anl.gov
Andreas Wilke, wilke@mcs.anl.gov
Tobias Paczian, paczian@mcs.anl.gov
Travis Harrison, teharriso@mcs.anl.gov
William Trimble, trimble@anl.gov

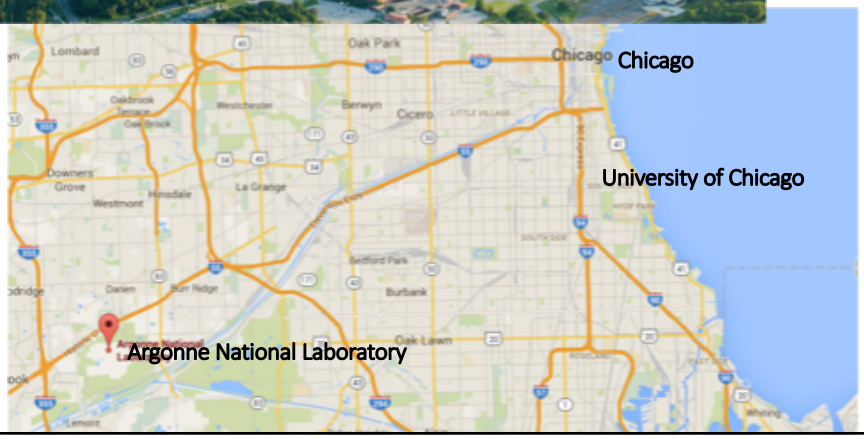
Argonne National Laboratory and University of Chicago

do this now, if you want to follow along later:

- git clone --recursive https://github.com/MG-RAST/Skyport2.git
- cd Skyport2
- source ./init.sh

- [sudo -E] docker-compose up
- (If sudo is required for docker-compose, use with option-**E**)

Argonne National Laboratory



Outline

- Day1:
  - Introduction and Skyport boot strap (interactive)
  - Submitting jobs, starting more AWE worker nodes (interactive)
- Day2:
  - Short introduction CWL
  - Creating a more sophisticated workflow (interactive)
    - Docker and CWL
- Day3:
  - Configuring the system for real world use
  - Data persistence, etc.

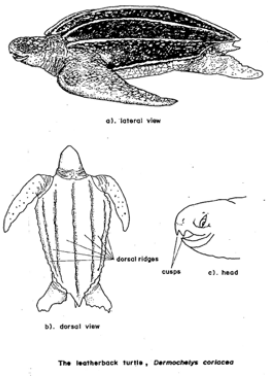
What is Skyport

- Cool name
- Set of services that allow reproducible, data parallel computing with almost zero per client setup cost
- Components (ready made for you to customize – use)
  - SHOCK – RESTful object store
  - AWE – RESTful workload manager
  - Auth
- Example use cases:
  - DOE KBase, MG-RAST, NIH Patric

My background

- CS, then large scale bioinformatics and later “cloud”
- MG-RAST (Meyer et al, BMC Bioinformatics, 2008)
- RAST (Aziz et al, BMC Genomics 2008)
- Distributed execution for thousands of users via interactive web portal 346480346480
- Portal also does data archival and data integration
- Security is always required DOE insists

Biology



http://www.oneocean.org/ambassadors/track_a_turtle/biology

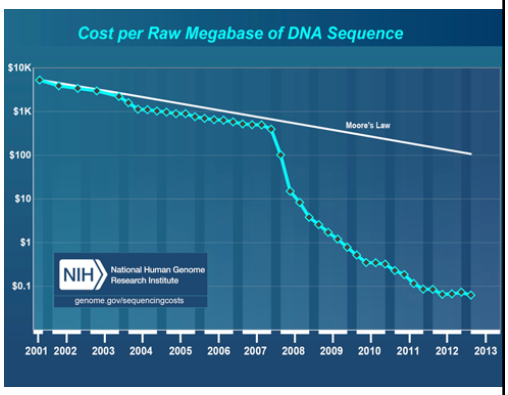
http://www.the-aps.org/education/

http://www.ferrum.edu/majors/biology.jpg

These are: "`biology.png`", "`biology.gif`" and "`biology.jpg`".

Genomics revolution has changed bio-medical research

- Now data processing is ubiquitous
- Prior landscape had genome centers with supercomputers
- Today sequencing is democratized
- New paradigm has solution providers for specific domains/tasks
  - Metagenomics: e.g. MG-RAST or European Bioinformatics Institute
  - Genomes: RAST or European Bioinformatics Institute or NCBI
  - …
- Cost is key factor in new ecosystem

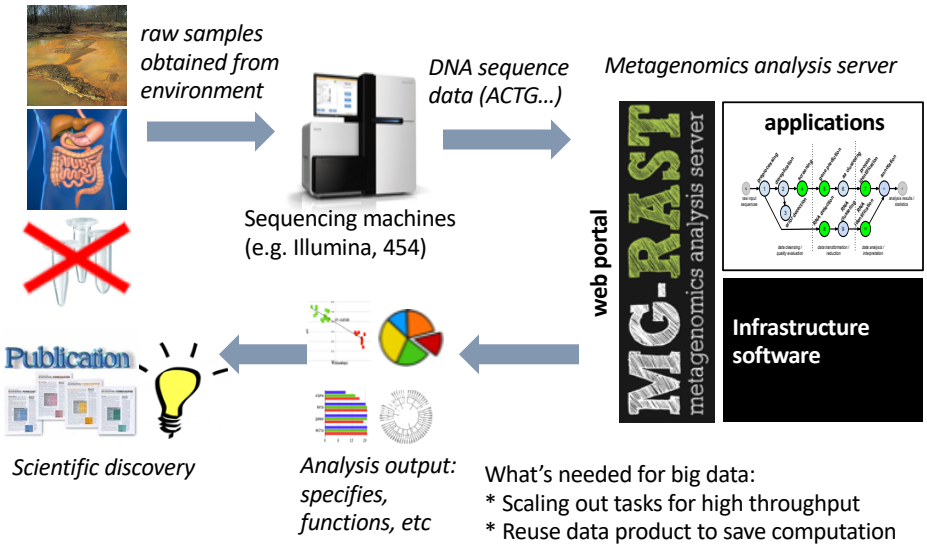


Source: NHGRI

$30k sequencing cost ~= 1,000,000 US$ naïve analysis

8

Typical procedure of metagenomics



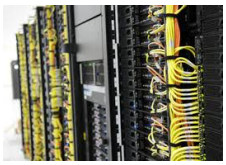
What's needed for big data:
* Scaling out tasks for high throughput
* Reuse data product to save computation

Algorithm choice has highest impact on cost, infrastructure 2nd highest

9

Why Skyport?

- Keep cost low
- Scaling beyond locally available resources
- Computing is (almost) everywhere
- Setup cost for additional resources were excessive
- Multiple "cloud" like systems on the rise
- MG-RAST (~500 nodes today)
  - Originally NFS + SGE (~40 nodes)
  - Possible to add systems but expensive
    - Adjust NFS server settings + Poke holes in firewalls
    - NFS is slo ide and de



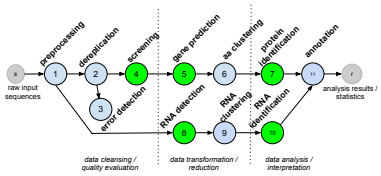
EC2

openstack

What is our niche?

- Simple to set up scale out architecture
- Large scale data re-use to save resources

- Large scale data parallel computing with minimal setup
  - MG-RAST ~20k users web portal for environmental sequence analysis
- Relatively high CPU to IO ratio
- Object store can “learn” the structure and semantics of data files
  - Writing back indices is MUCH faster
  - Plug-in architecture, requires coding in GO-lang
- So far computational biology
  - No hard limitation to that domain
- Real life application



Meyer et al, Briefings in Bioinformatics, 2017

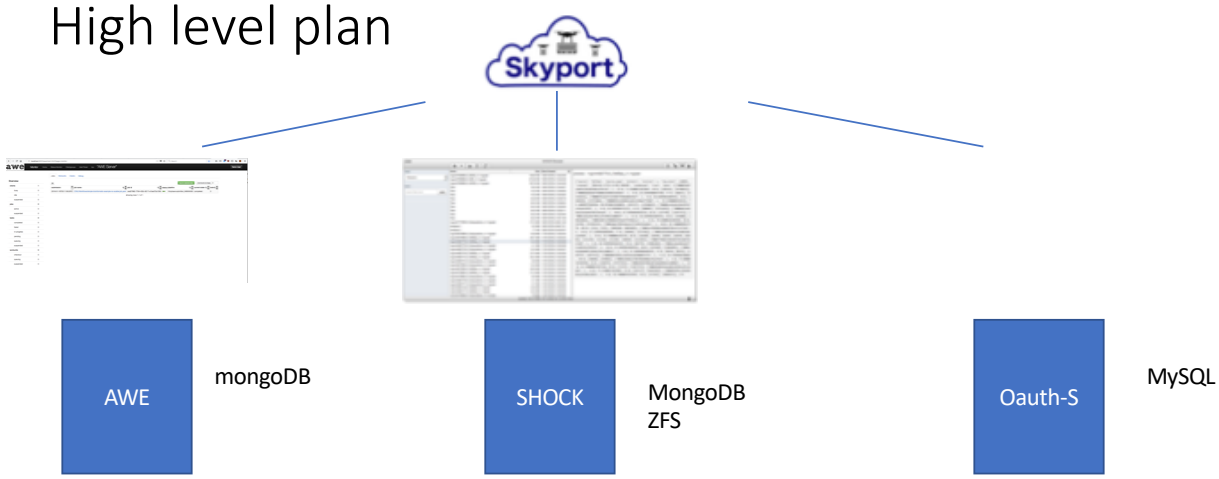
Design Goals

- Support distributed execution with minimal setup and zero conf per client
  - Means no shared file systems
- Support standard workflow language and reproducibility
- Efficient execution
- Allow analysis of large data volumes (avoid local client limitations, e.g. disk size)
- Support most systems
- Limit IO cost
- secure

Design

- RESTful
- Use SHOCK object store
  - Provide subsets of files via object store to limit IO
  - Handle format conversion
- Use AWE resource manager
- Use Oauth2
- Use CWL
- Use containers (Docker/lxc and singularity)
  - Assumes one container per workflow, can be one container per step

High level plan



SHOCK object store

Active storage optimized for streaming data analysis

- Object store
  - MongoDB and POSIX file system
  - REST API
  - Currently with ZFS (moving so S3-API)
- Optimized for streaming data analysis
- Programmable, plug-in architecture
  - data **subset access**
  - on the fly format conversion
- Index driven
- Data semantics aware

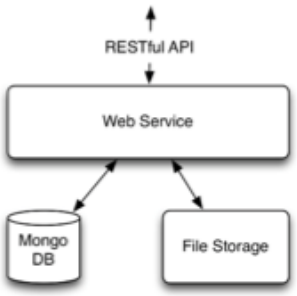


Fig. 1. Design overview of the SHOCK server.

Bischof et al, BDC 2015

16

SHOCK object store // active store cont'd

- Simple data **and** metadata store
  - Metadata ➔ MongoDB, Data ➔ file system
  - Supports arbitrary metadata (also including industry standards…)
- Provides RESTful API
- Provides authentication
- Knows object semantics
  - E.g. create subset, stream file in different format e.g. JPEG->PNG
- Compare to S3: also RESTful, but know object semantics

SHOCK cont'd

Why a new system? Aren't there enough already? ☺

TABLE I
SURVEY OF EXISTING TECHNOLOGIES FOR OUR DESIRED FEATURE SET.

	iRods	S3	NFS	Lustre	Gridftp	HDFS
AWAN	+	+	-		+	-
AAB	-	+	-	–	*	-
FCD	-	+	-	+	++	-
INDEX	-	-	-	+	limited	-
Pluggable	-	-	-	-	-	+
ADD	+	-	-	-	+	-
SEARCH	+	-	-	na	-	-
LTS	+	+	na	na	+	?
ZC	-	+	–	–	-	–

AWAN=across wide area networks, AAB=across administrative boundaries, FCD=fast content delivery, INDEX=index data objects and subsets, PLUGGABLE=extend set of pluggable file parsers ADD=avoid data redundancy, SEARCH=searchable metadata, LTS=long-term storage, ZC=zero config

18

Shock cont'd

SHOCK is something unique

- Hybrid between DB and file store
- Supports multiple indices
- Supports virtual files and subsets

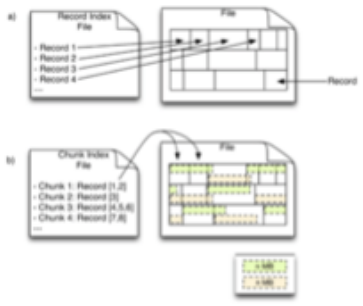


Fig. 2. Graphical depiction of Shock node indexes that reference Shock node data files and provide fast access to regions (i.e., subsets) of the data file.

19

SHOCK cont'd

Good performance with whitebox hardware

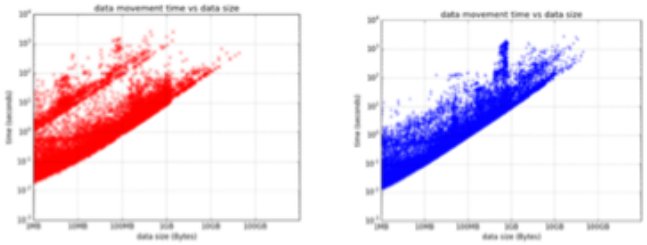


Fig. 3. Upload (red, on the left) and download (blue, on the right) speeds to the MG-RAST Shock server displayed for different data set sizes. This plot represents the load on the production SHOCK server for the MG-RAST system in the period March 1–13, 2015. Some of the upload operations are slower because their creation (subset nodes) requires some computation.

20

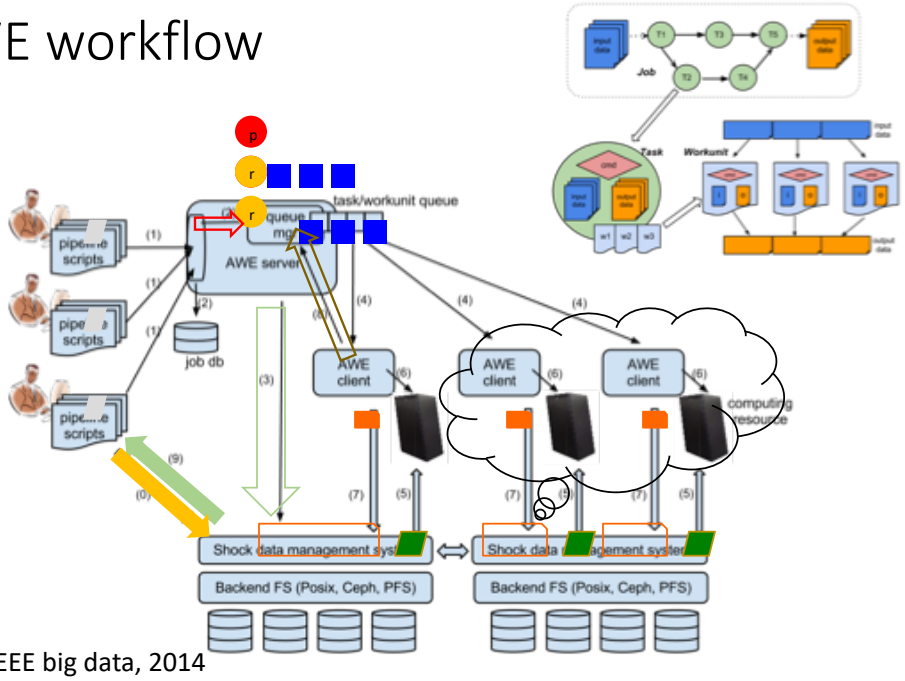
AWE – yet another (RESTful) resource manager

- AWE == Argonne workflow engine (another was taken)
- Compare to Apache Airflow
- Manage complex workflows and allow clients to subscribe to individual tasks
- Allow parallelization of tasks where suitable
- As of v0.65 support subset of all CWL functions
  - Complete implementation is ongoing, compliance test

AWE -- implementation

- RESTful
- GO-lang
- Plug in architecture
- Used in DOE KBase, MG-RAST, NIH PATRIC
- Used with 300-600 worker nodes
- Production since 2012

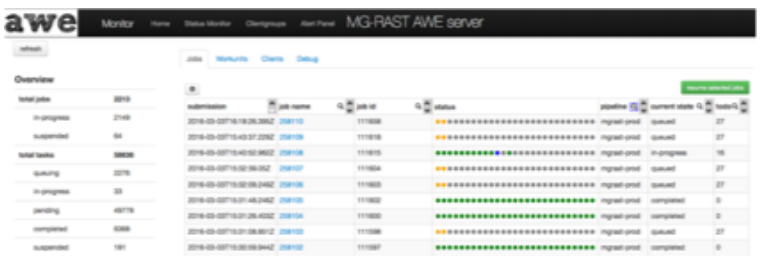
AWE workflow



Tang, et al, IEEE big data, 2014

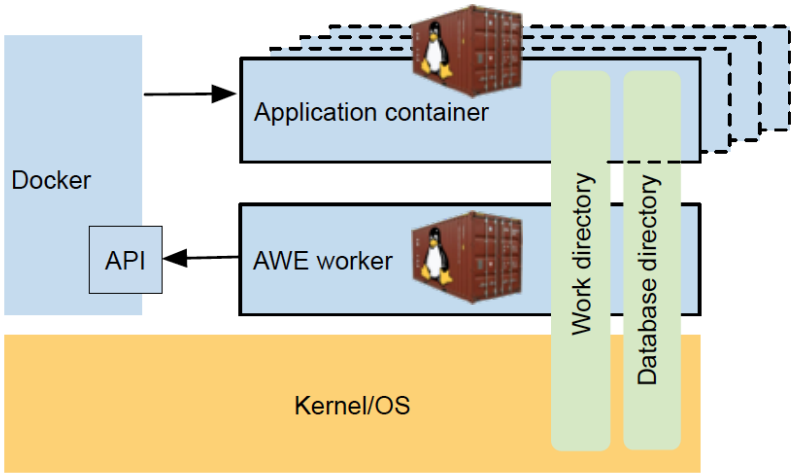
AWE2

- Rudimentary workflow language (v1)
  - V2 supports CWL (currently in compliance testing)
- Clients register with AWE server
- Clients bind DHCP style to task (3 strikes and you are out rule)
- Web based mgmt UI



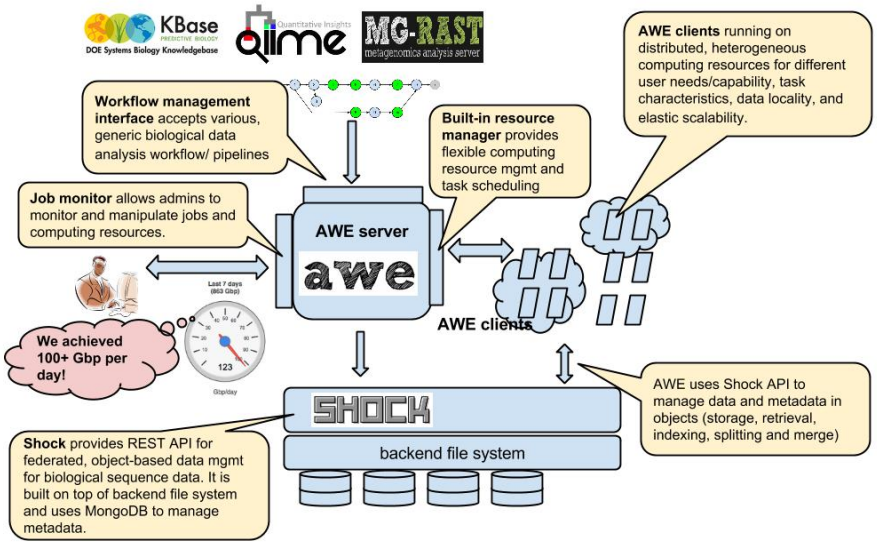
26

Using containers to capture execution environment



Gerlach et al, Cloud Engineering (IC2E), 2015
Gerlach, et al, Proceedings Data-Intensive Computing in the Clouds, 2016

Architecture



First step(s)

- git clone --recursive https://github.com/MG-RAST/Skyport2.git
- cd Skyport2
- source ./init.sh
- [sudo -E] docker-compose up
- (If sudo is required for docker-compose, use with option-**E**)
- Use Firefox to go to http://localhost:8001