

# Polyglot, Event Driven Computational Science Using the Actor Model

Joe Stubbs

Texas Advanced Computing Center

University of Texas, Austin

# What is TACC?

**Mission:** To enable discoveries that advance science and society through the application of advanced computing technologies.

- High performance computing (HPC)
- Cloud & high throughput computing
- Data intensive computing
- Visualization
- Scientific software development & optimization
- **APIs and tools - Agave Platform**
- Web and mobile applications
- Life sciences
- Training & outreach

# What is Agave?

/systems

/files

/apps

/jobs

- Register storage and compute systems
- Ingest, move and transform data files and folders
- Register applications (binaries) on large systems
- Launch jobs to invoke applications

# What is Agave?

/systems

/files

/apps

/jobs

- Register storage and compute systems
- Ingest, move and transform data files and folders
- Register applications (binaries) on large systems
- Launch jobs to invoke applications

**/notifications**

\*All activities are events that can be subscribed to

# Agave Powers web & mobile apps

BioExtract Server  
data access, analysis, storage, and workflow creation

Send us feedback  
Current User: guest  
[Log In] [Register] [Why register?]

Query Extracts Tools Workflows Groups Help

Discovery Environment

FAST TRACK TO PURE ANNIHILATION AND BLOOD-SUCKLING

Now from PLANTS & ANIMALS

Username: Password: [Log In] [Enter As Guest]

Forgot Password

Associate & Genomic Sequencing

Process Genomic Usage: SAMTOOLS

What iReceptor Generates of vaccir antibody

Arabidopsis Information Portal Alpha

WELCOME!

Explore the Arabidopsis genome

ThaleMine

JBrowse

Science Applications

Store Mac iPhone Watch iPad

iTunes Preview

TACC Mobile  
By THE UNIVERSITY OF TEXAS AT AUSTIN  
Open iTunes to buy and download apps.

Description

TACC Mobile is an application interface to the Texas Advanced Computing Center. It is designed to inform society about the value of advanced computing resources and to help registered users better use TACC's power of TACC directly to you wherever you go!

Public Features Include:

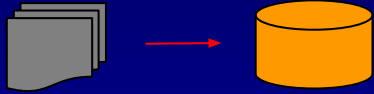
- Feature Stories and News: find out the latest news and announcements
- Events Calendar: see up-to-date information on current events
- Systems & Software: browse a searchable collection of TACC Systems & Software resources
- Connect with TACC: use interactive maps to help guide you to TACC resources

User Features Include:

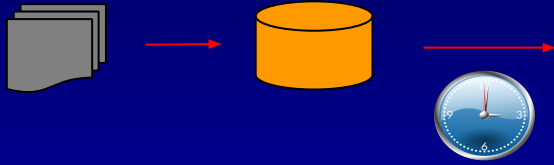
- User News: be the first to hear about important system updates
- Consulting Tickets: help is here when you need it through our ticketing system
- My Jobs: view the status of your HPC jobs and stay notified
- My Projects & Allocations: track of your current allocations
- Systems Monitor: view current system health and status
- Visitation Reservations: browse through upcoming Visitation
- Profile: view and update your profile information

The power of the Texas Advanced Computing Center at your fingertips.

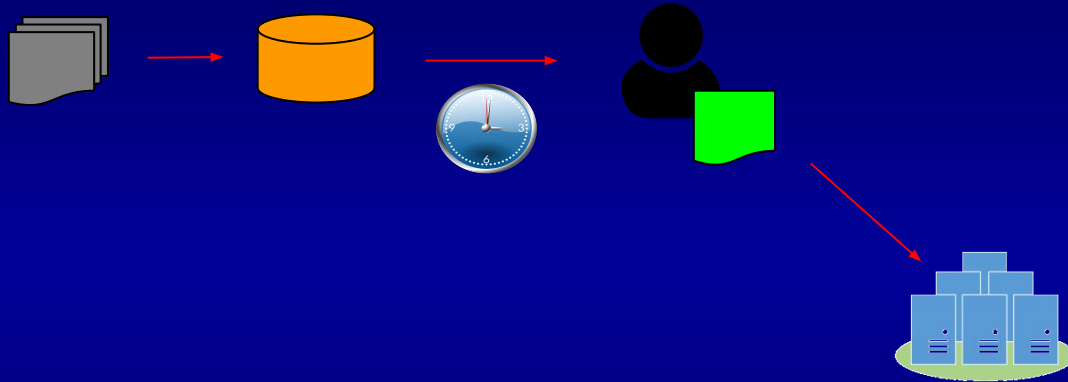
# Computational Science: Traditional Approach



# Computational Science: Traditional Approach

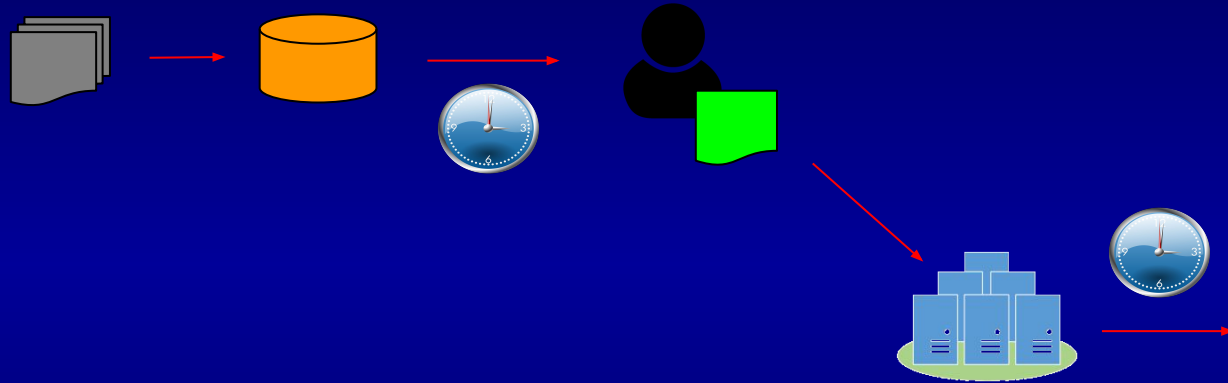


# Computational Science: Traditional Approach

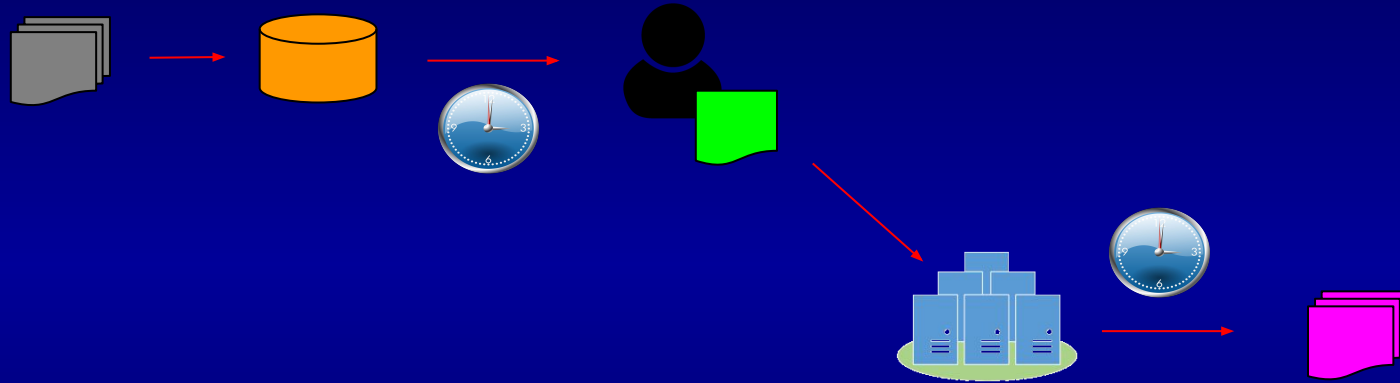




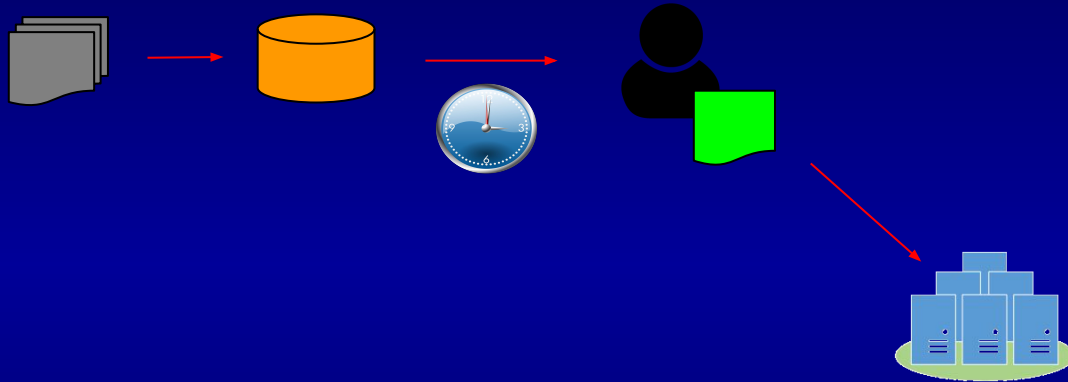
# Computational Science: Traditional Approach



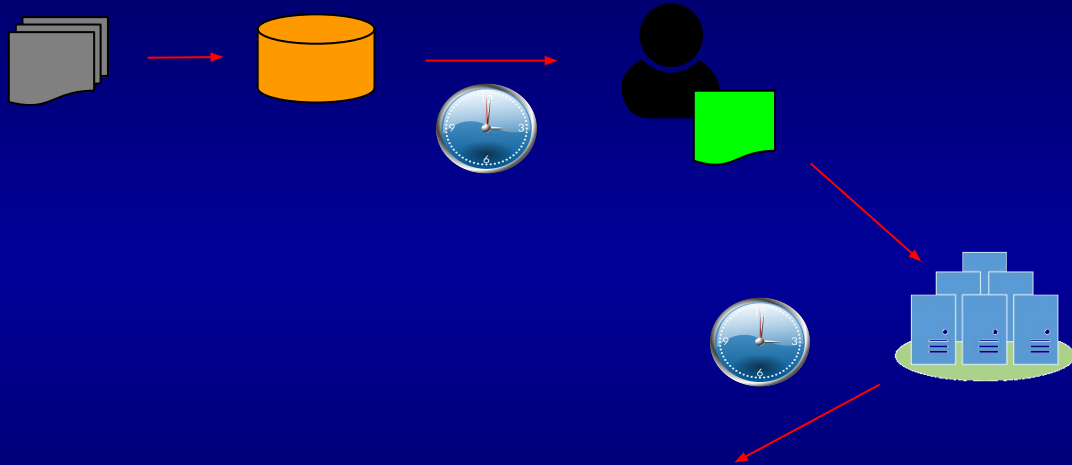
# Computational Science: Traditional Approach



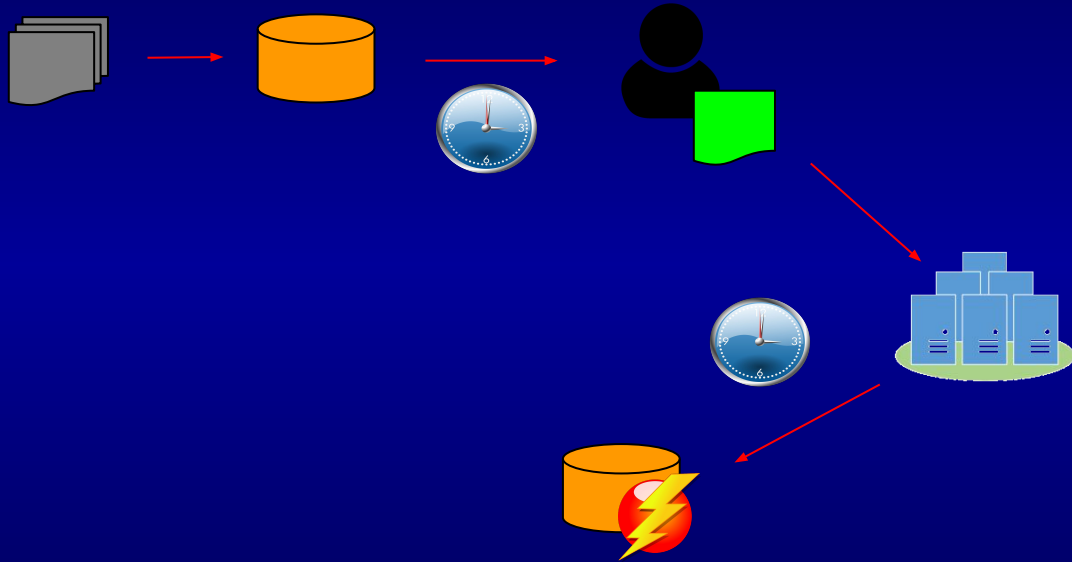
# Computational Science: Traditional Approach



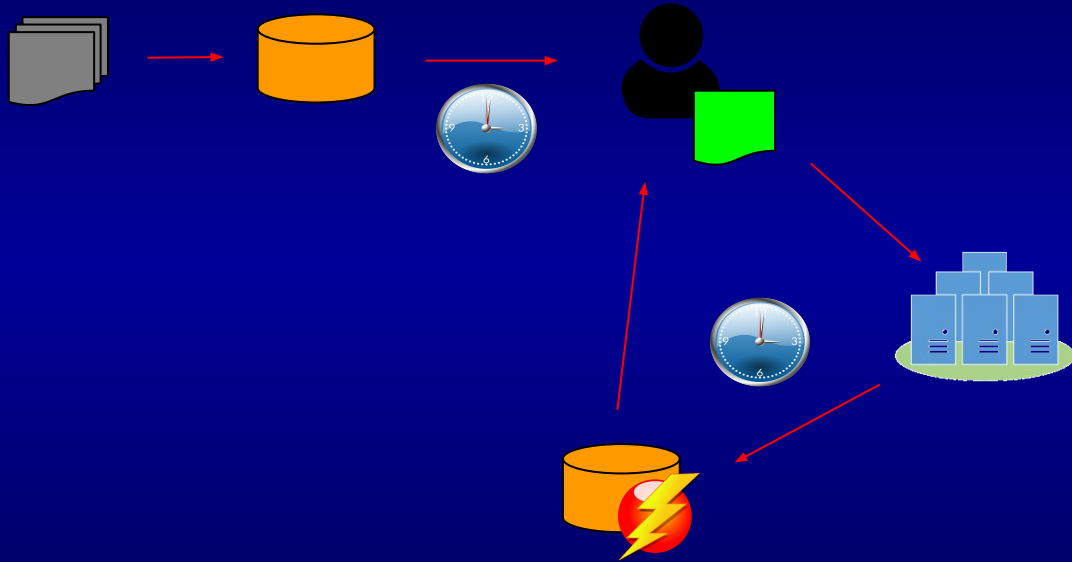
# Computational Science: Traditional Approach



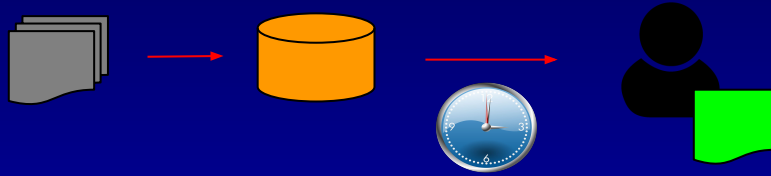
# Computational Science: Traditional Approach



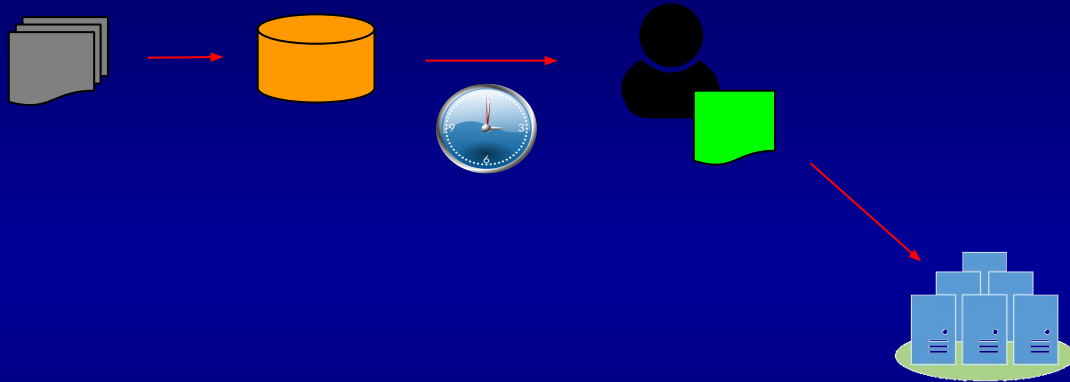
# Computational Science: Traditional Approach



# Computational Science: Traditional Approach

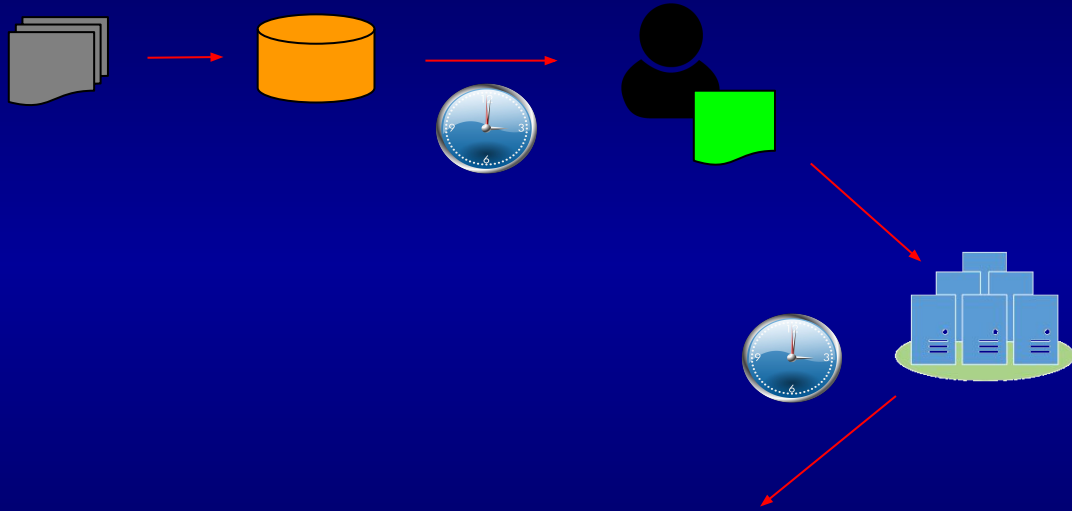


# Computational Science: Traditional Approach

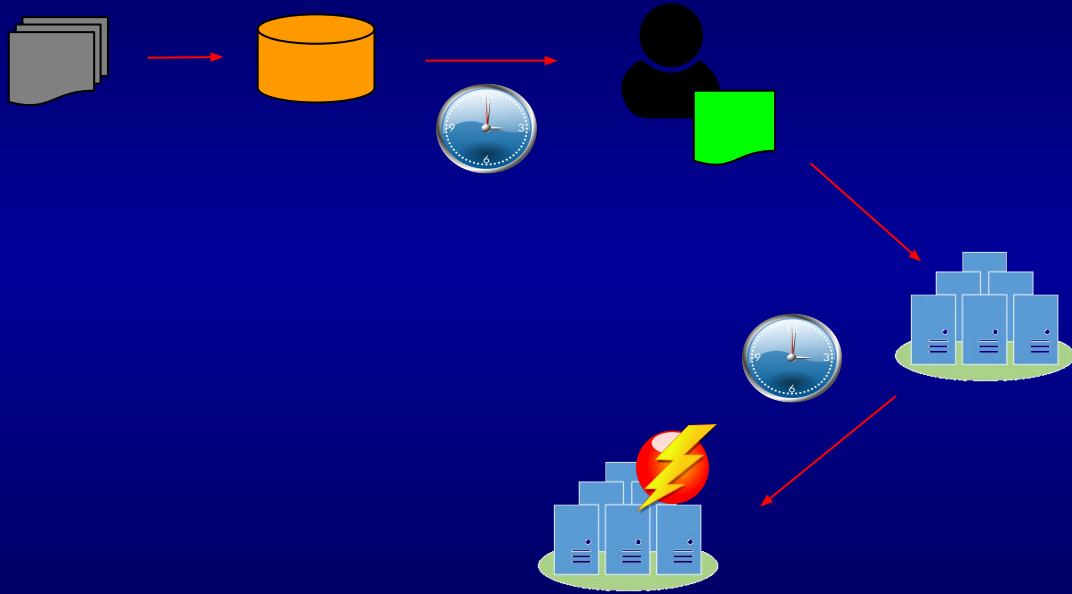




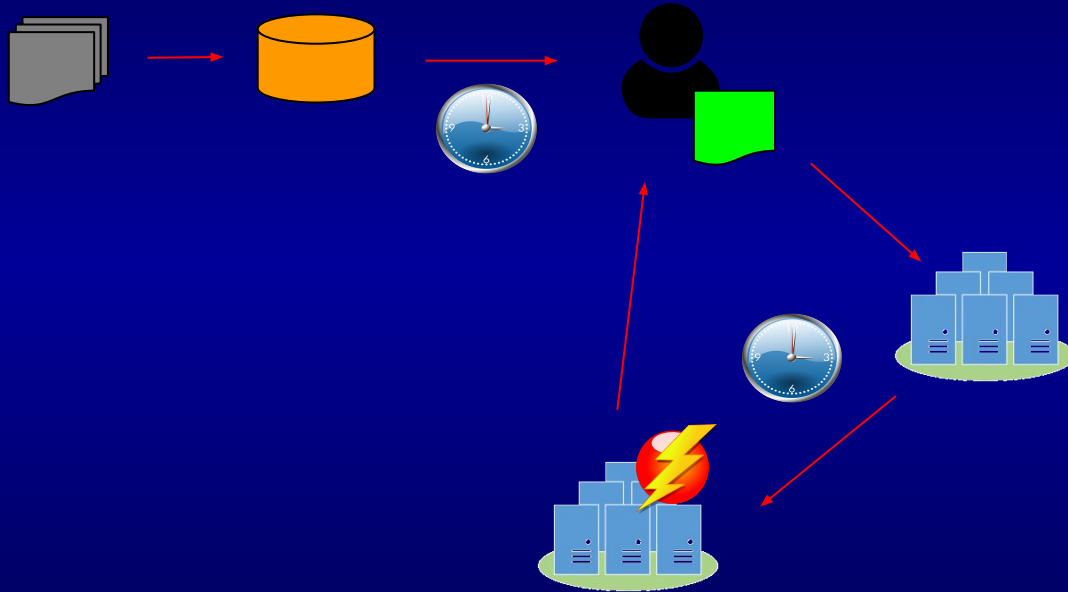
# Computational Science: Traditional Approach



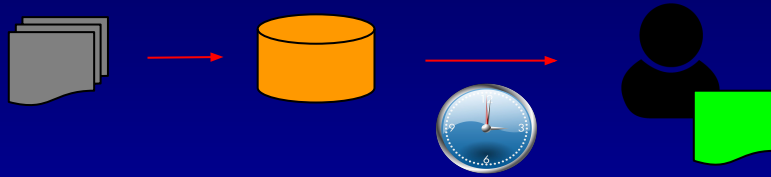
# Computational Science: Traditional Approach



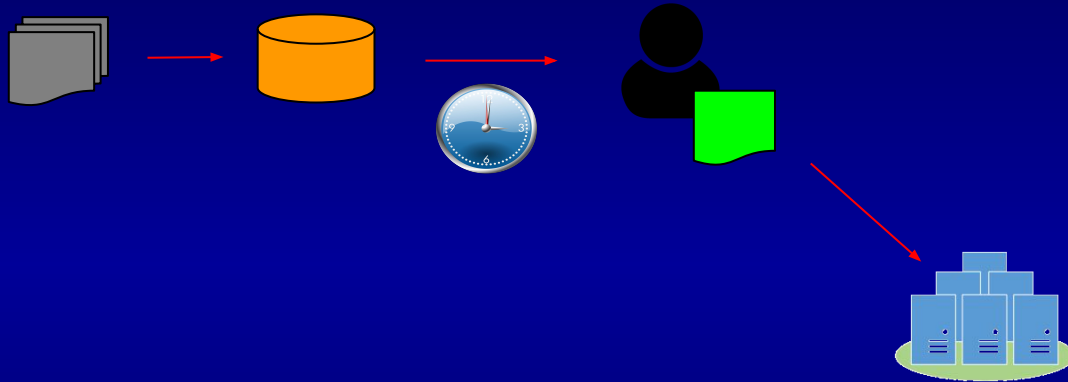
# Computational Science: Traditional Approach



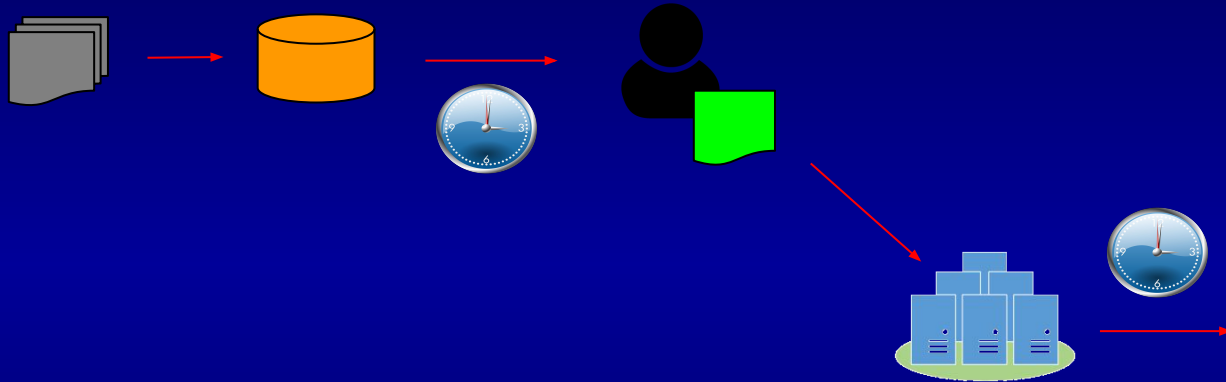
# Computational Science: Traditional Approach



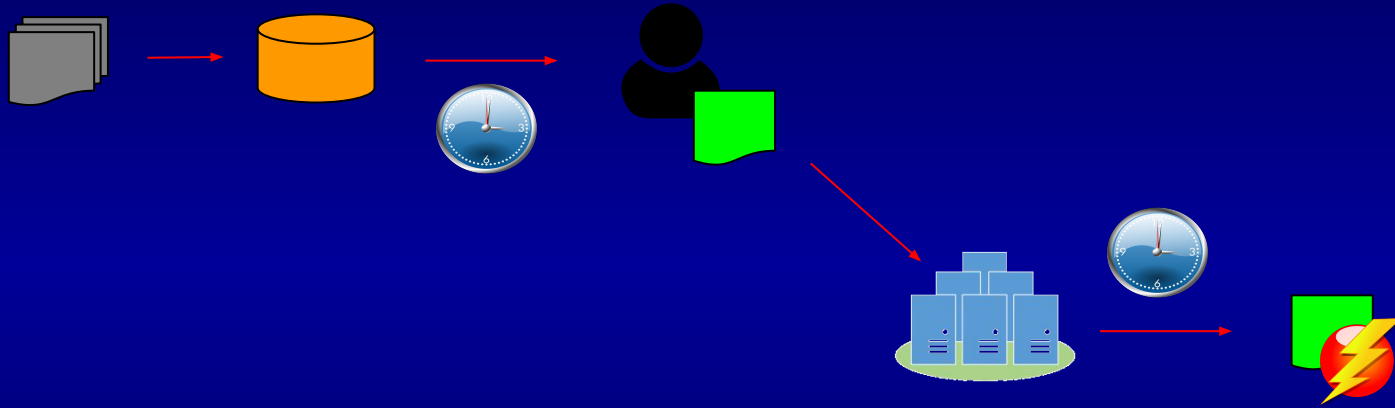
# Computational Science: Traditional Approach



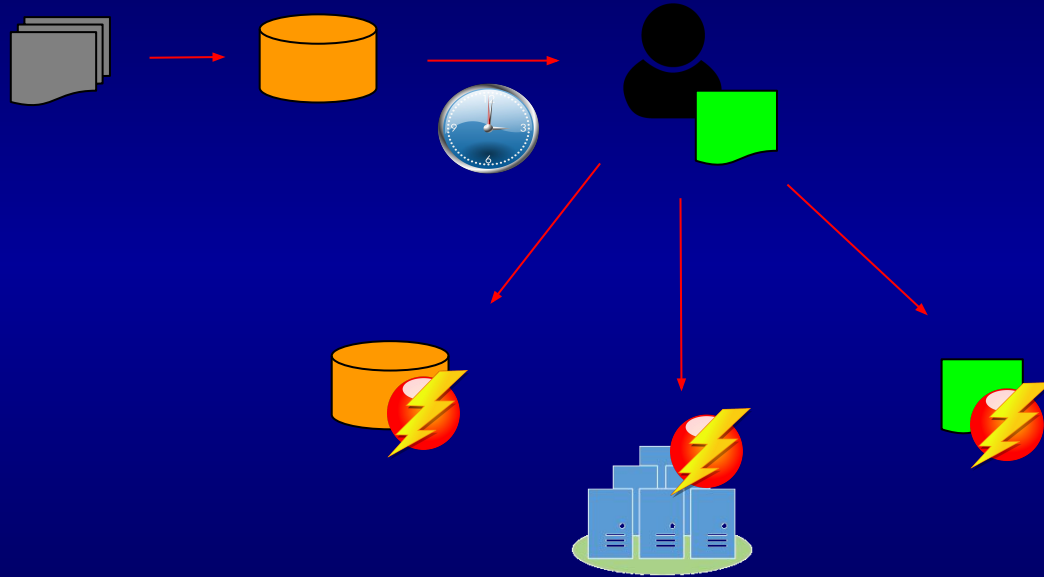
# Computational Science: Traditional Approach



# Computational Science: Traditional Approach

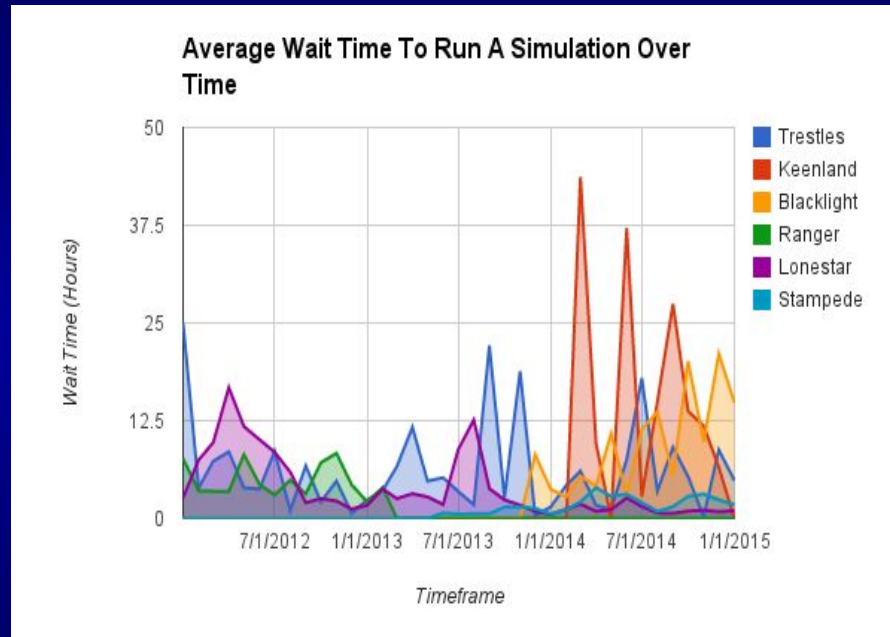


# Computational Science: Traditional Approach



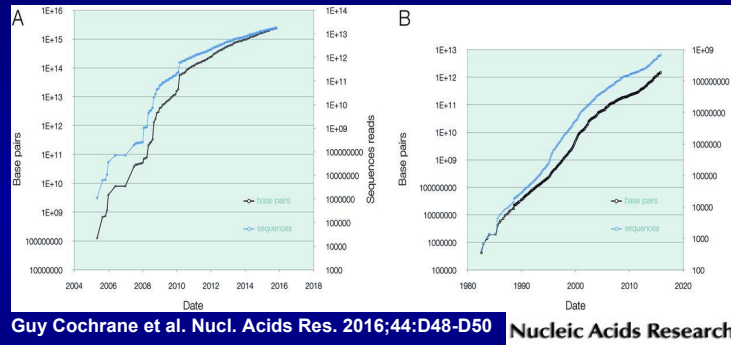


# Computational Science: Traditional Approach



# Meanwhile, Data accumulating...

Cumulative growth in INSDC. (A) Base pairs (black, 2365.5 trillion) and sequence reads (blue, 17.8 trillion) for INSDC raw data.



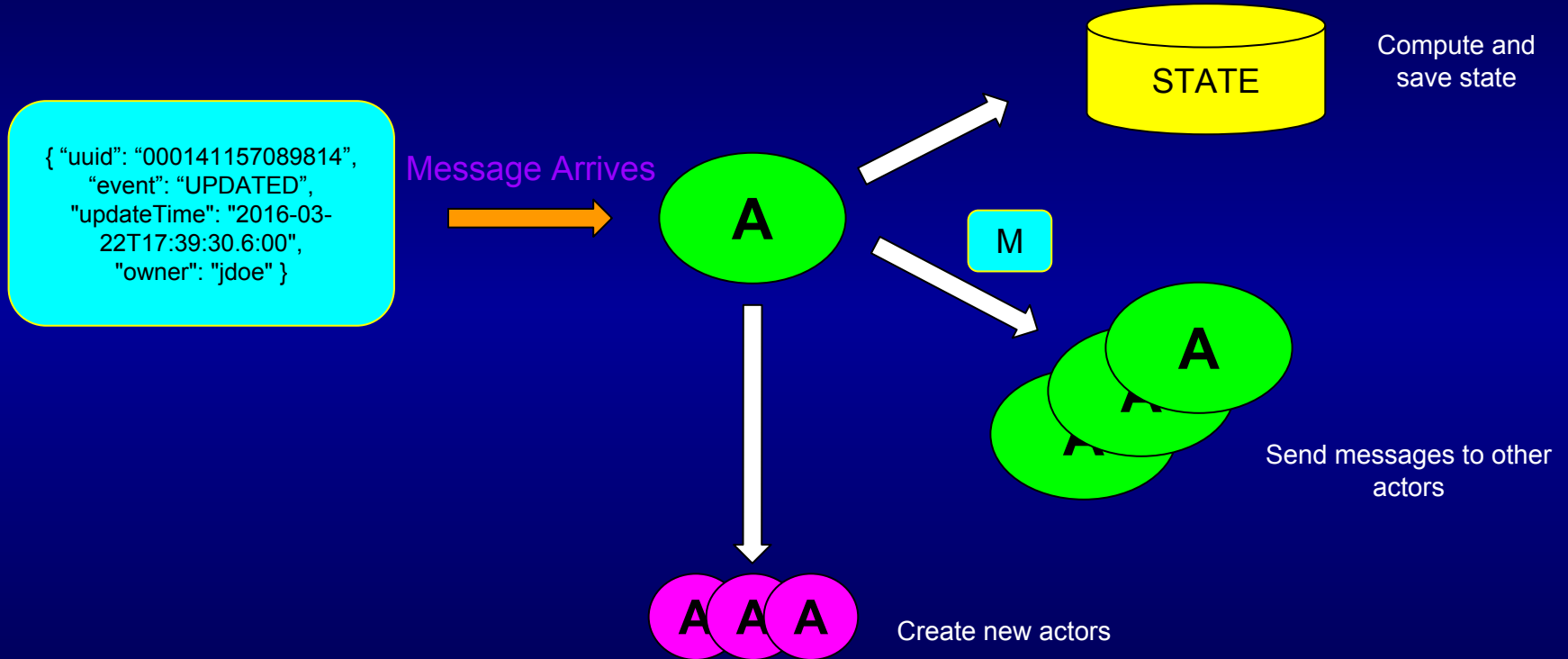
Agave alone moves  
1PB+ data/month

Aggressive purge policy  
of 2 weeks on TACC's  
global /SCRATCH

# Event Based Processing

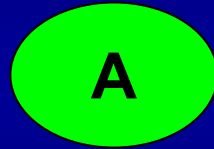
- File or folder appears or is modified on a server...
  - Run a checksum
  - Launch a job to do some analysis
  - Compress the file
  - Move the file to archive storage
- Job completes
  - Job was successful, launch another job
  - Job failed, check inputs and launch again?
- Execution system goes offline for maintenance
  - Submit to a secondary system
- Storage system goes offline for maintenance
  - Submit jobs using data from a different system
- New user signs up for portal/project
  - Bootstrap storage and compute

# Actor Model

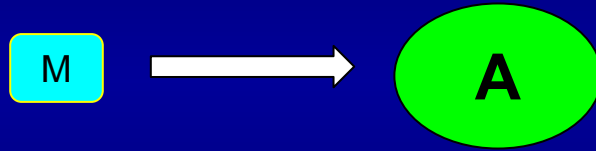


# Actor Model: Inherently Concurrent

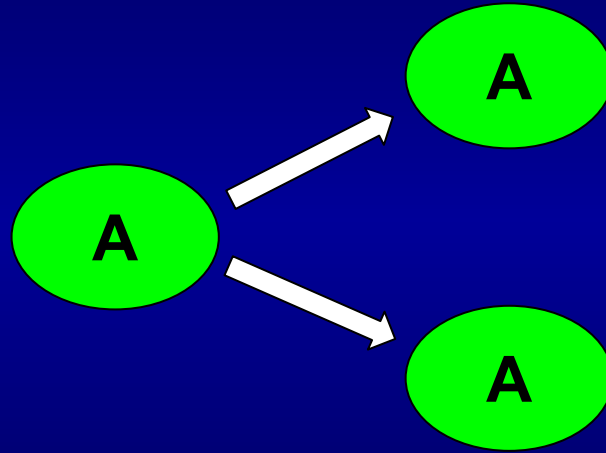
# Actor Model: Inherently Concurrent



# Actor Model: Inherently Concurrent

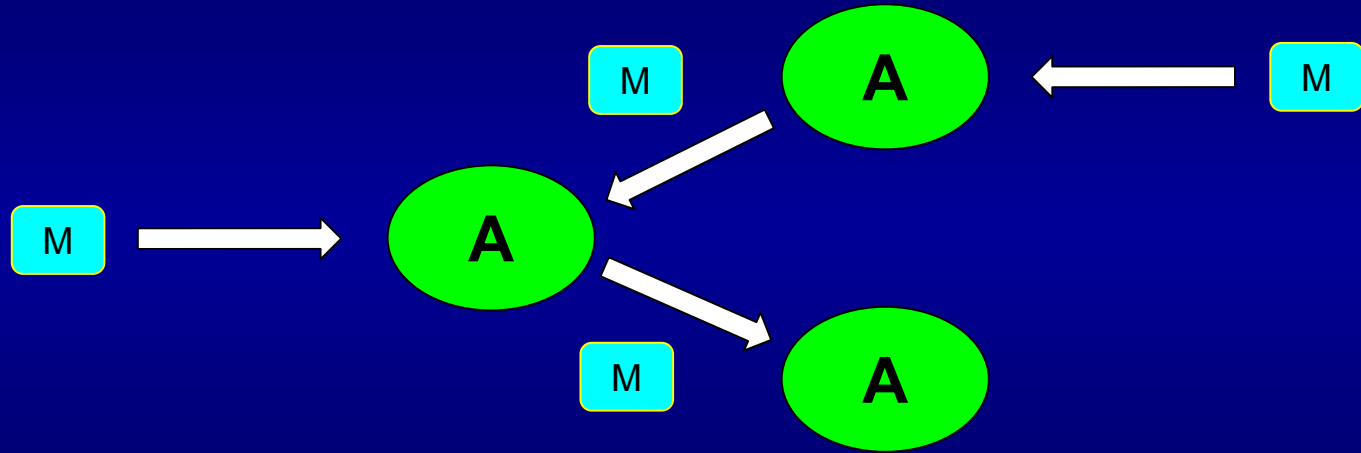


# Actor Model: Inherently Concurrent

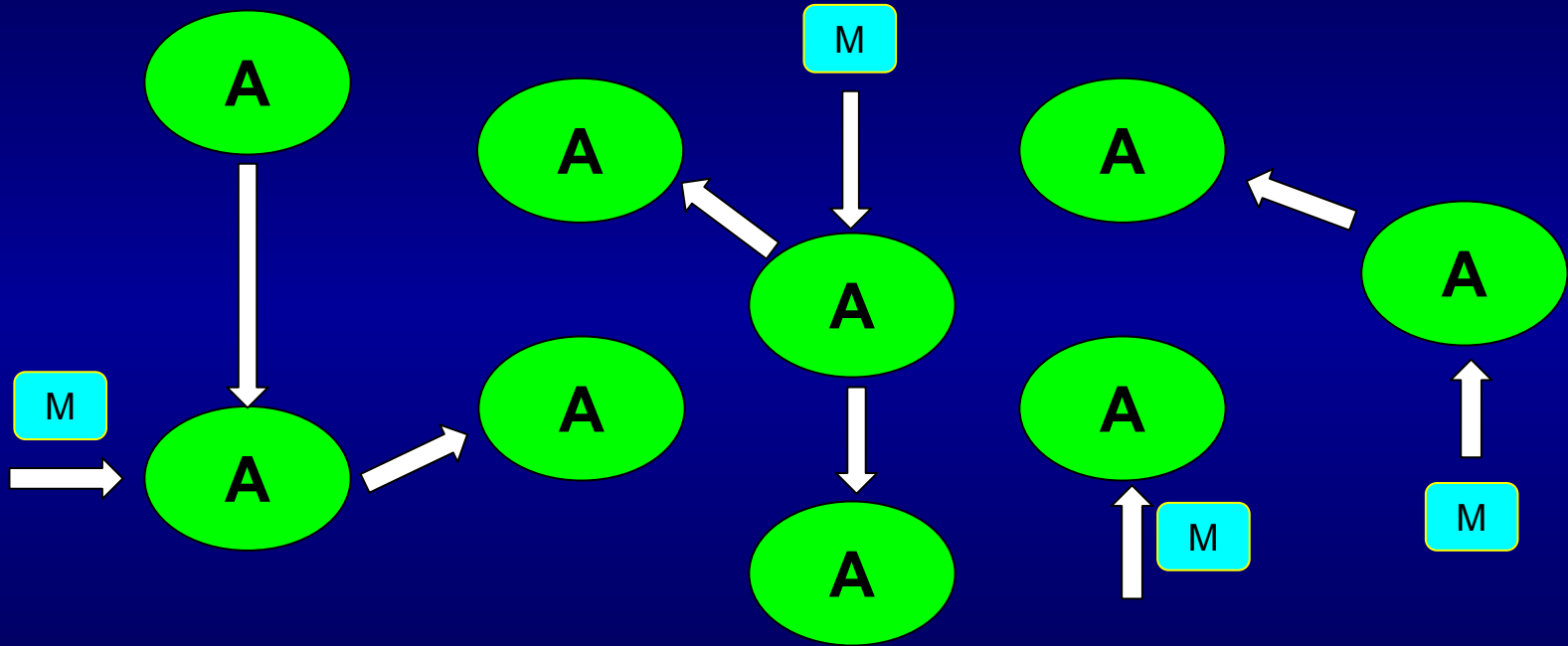




# Actor Model: Inherently Concurrent



# Actor Model: Inherently Concurrent



# User-Defined Actors Via Docker



- Associate an actor with a Docker image.
- Assign the actor's inbox to a unique URI.
- Launch a container from the image in response to a message.

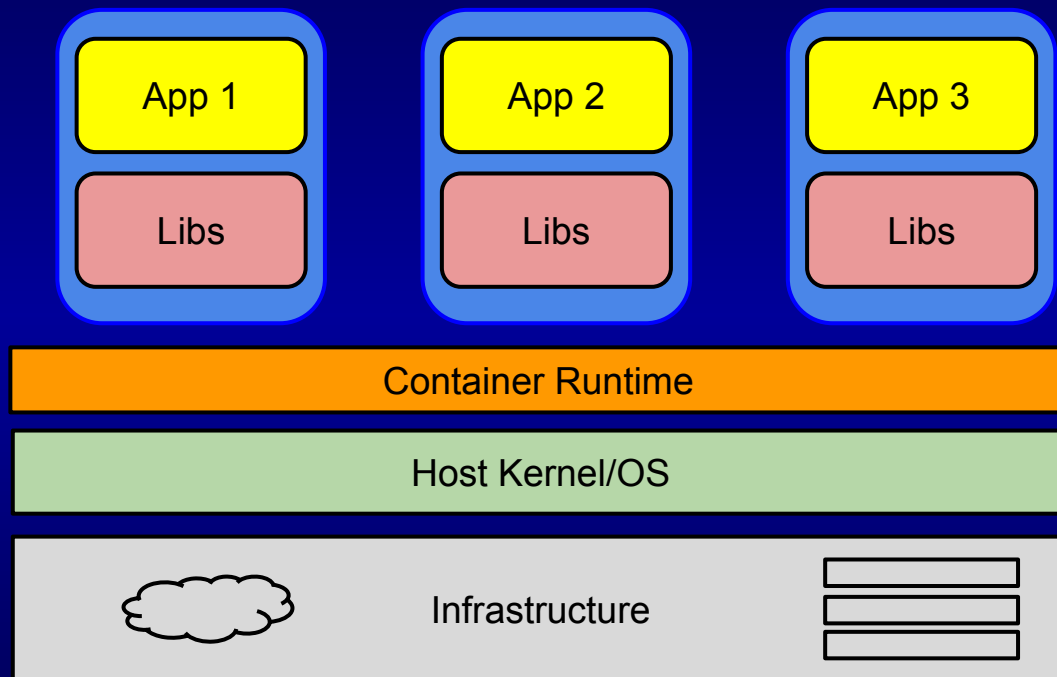
# Containers: Reproducible Environments

Isolated Userland Processes

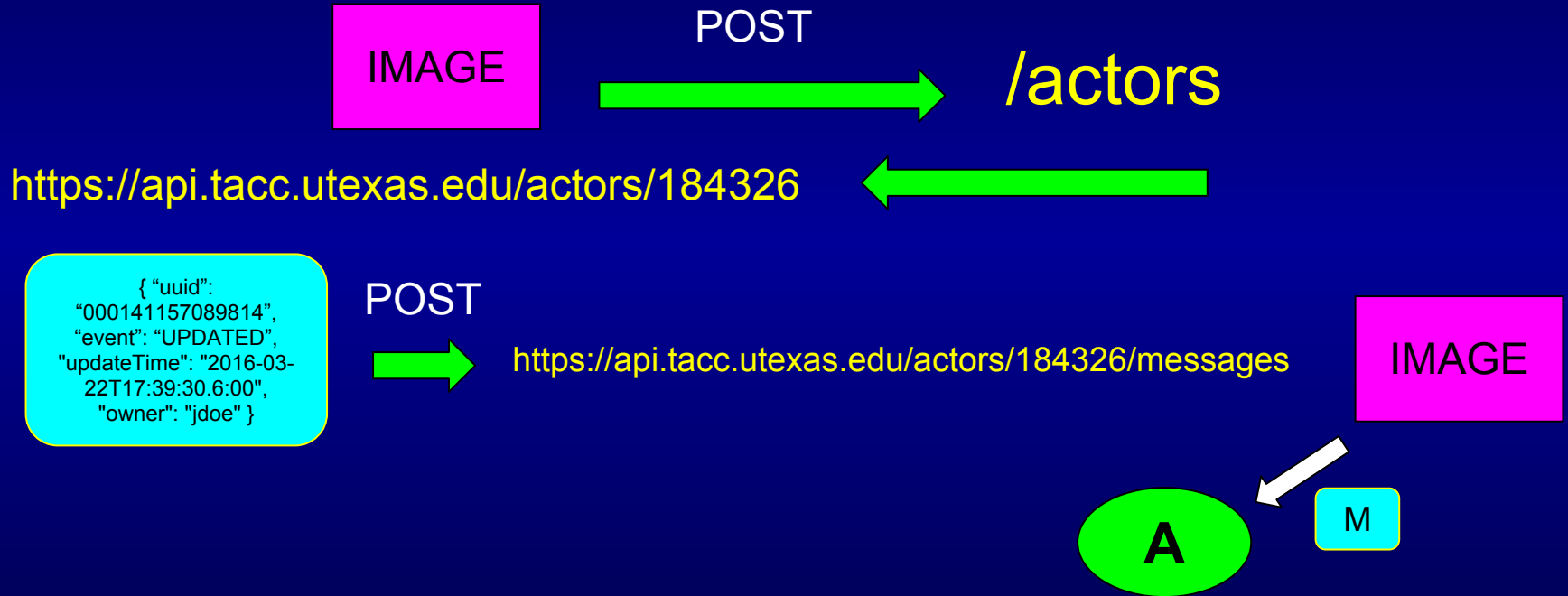
Virtualized:  
Network  
I/O  
CPU and MEM

Containers:

- Include all dependencies
- Ease installation
- Start up in milliseconds



# Abaco: Actor Based Containers



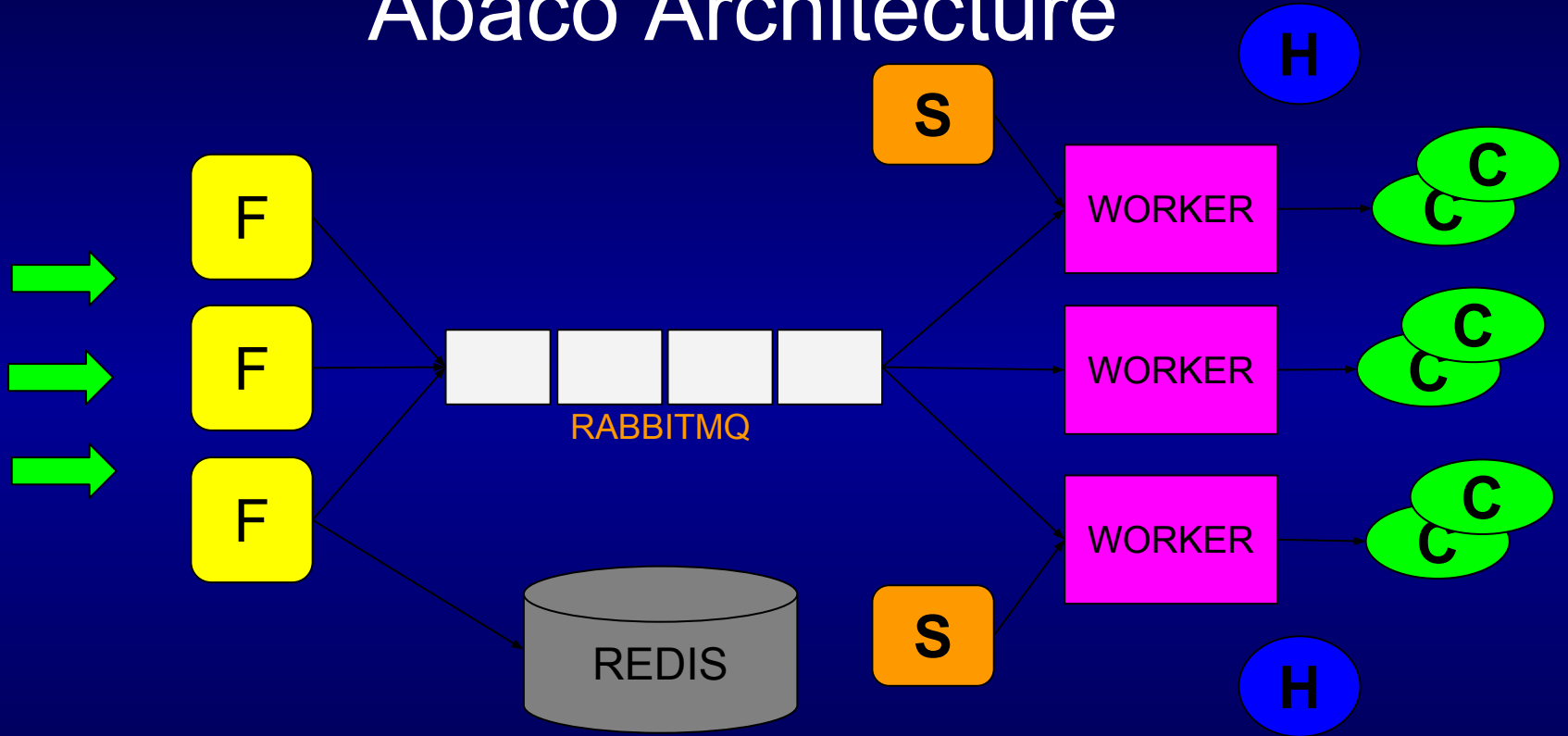
# Abaco: Agave Event Processors

- Notifications API in Agave allows users to subscribe to events
- Event subscriptions can be in of several forms:
  - Email
  - SMS
  - **Web callback** - in this case, Agave sends details of event in message payload.
- Events API coming this summer makes registering subscriptions even easier.

# Abaco: Agave Event Processors

- Create a Docker image to “process” an event.
- Register the image as an actor in abaco.
- Register the actor’s inbox URI as the callback to a notification for the event.

# Abaco Architecture





# Challenges

Abaco in beta, available to select “friendly” users.

Potential issues:

- Accidentally subscribing to “way too many” events.
- Buggy containers “hanging” during execution.

# Early Use Case: GSAF

Genome Sequencing and Analysis Facility: from sequencer to SNPs

- Raw genetic material sequenced by Illumina Sequencer.
- “short reads” files dumped to server - kicks off a chain of events.
  - Initial quality checks
  - Alignment routines.
  - Basic analyses: Single Nucleotide Polymorphism (SNP) calls, etc.
  - Data moved to scientist’s storage system.

# Conclusion

- Massive data collections make real-time processing more and more of a necessity.
- The actor model provides a simple yet robust paradigm for concurrent, event-driven programming.
- Containers can be used to provide portable, reproducible environments.

# Thanks!

## Questions?

Email: [jstubbs@tacc.utexas.edu](mailto:jstubbs@tacc.utexas.edu)

Agave: <http://agaveapi.co/>

abaco: <https://github.com/TACC/abaco>