

Spin: A Docker-Based Platform for Deploying Science Gateways at NERSC

NERSC

Cory Snaveley, Stefan Lasiewski, Annette Greiner, Lawrence Berkeley National Laboratory

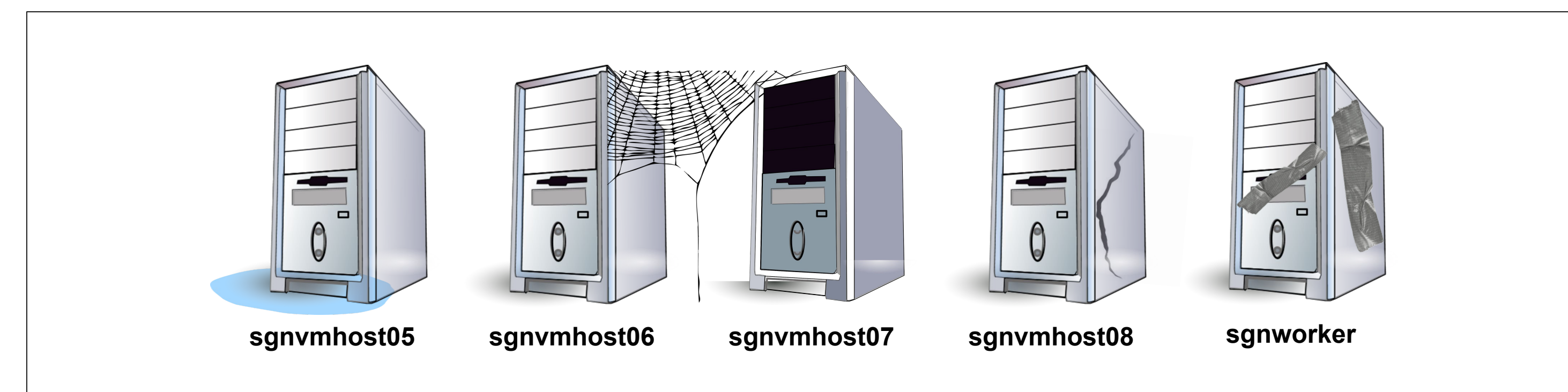
What is Spin?

Spin is a Docker-based, on-premise cloud platform at NERSC that enables researchers to design, build, and manage their own science gateways and other services using container technology.

User Needs

- ✓ better portability/reproducibility
- ✓ quicker initial provisioning
- ✓ less cumbersome configuration
- ✓ improved performance and scaling
- ✓ convenient access to HPC resources

Old Science Gateway Infrastructure

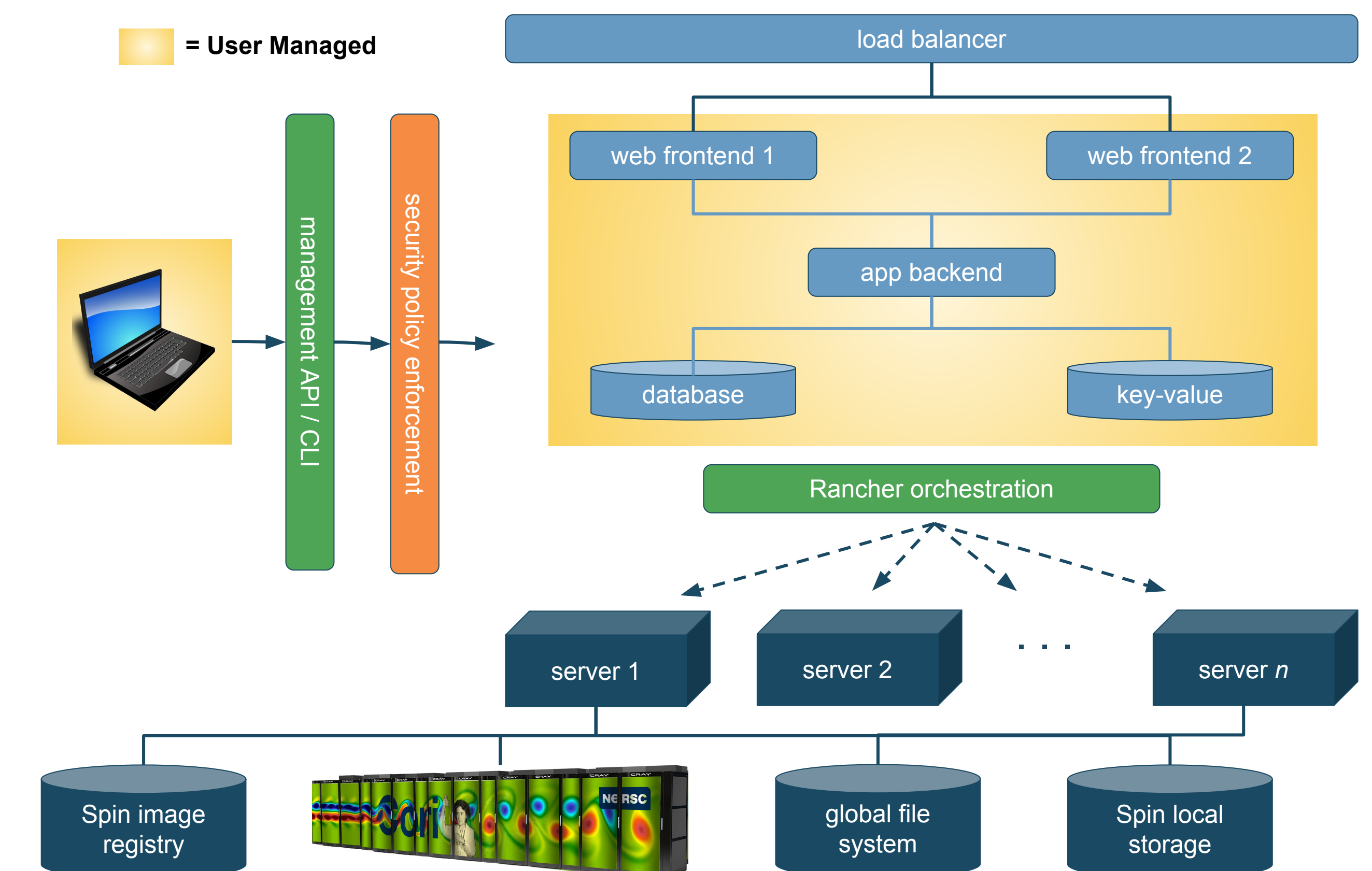


- Shared VMs for less demanding sites
- Individual VM for each more demanding site
- Data stored on selected NERSC global file systems

Challenges

- Services coupled to OS versions; upgrades difficult
- Complicated per-site Apache and Linux configuration
- Many software dependencies, often conflicting
- High-touch user engagement for many sites

New Spin Infrastructure



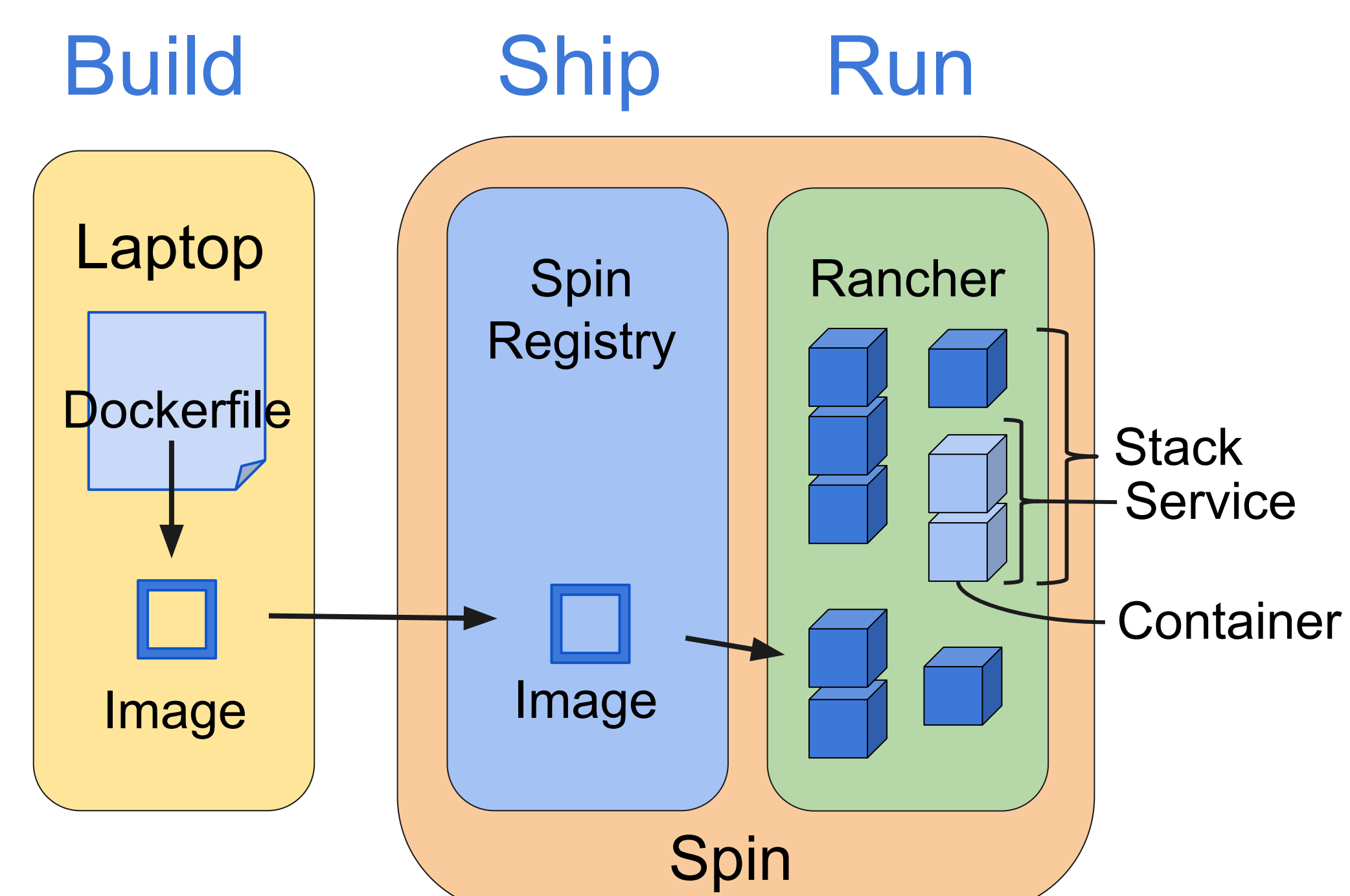
- Users interact with Spin via API and CLI
- Access control and security policy enforced by API interceptor
- Docker stack created for each science gateway
- Rancher orchestration deploys containers to servers
- Data stored on NERSC global file systems or within Spin

Container Concepts

- **Stack** - A group of connected services that implement a distributed application.
- **Service** - An image is deployed to a service and may have one or more **replicas** (containers). Examples: 'web' service, 'db' service.
- **Image** - A lightweight, read-only template that includes everything needed to run one software application service. Example: 'mysql' image.
- **Container** - A running instance of an Image. Example: 'web-1', 'web-2'
- **Volume** - A place to hold persistent data. Containers are normally ephemeral & changes are not retained after a restart.
- **Secret** - A special volume to hold passwords, private certs, etc.

Build, Ship & Run workflow

- A user **builds** an image and **ships** it to the registry via manual or automated workflow.
- The user then configures an application stack to **run** the service (or connected services).
- Test on laptop, test on development, deploy to production



What's Next?

- Rancher 2.0 / Kubernetes
- Pre-built services created on demand
- Simpler integration to batch scheduler (Slurm)
- Graphical UI and improved user experience
- Cleaner security model

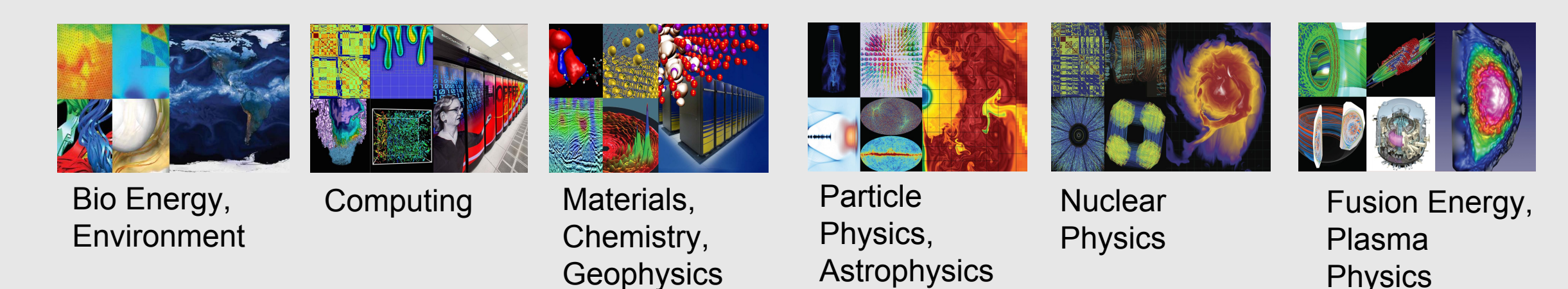
NERSC

The National Energy Research Scientific Computing Center (NERSC) at Lawrence Berkeley National Laboratory is the Primary High Performance Computing and Data Facility for DOE Office of Science Research.

- Over 5000 users from across the world
- 1500 scientific publications annually
- 3 recent Nobel laureates
- Wide range of scientific applications

NERSC is facing a data deluge! Data is becoming increasingly important in the big computing space as experiments and simulation generate massive data sets.

- Light Sources, Telescopes, Sequencers, Particle Detectors, Climate Models



Contact

- <http://www.nersc.gov/users/data-analytics/spin/>
- Spin Working Group: nersc-spin@lbl.gov

