



U.S. DEPARTMENT OF
ENERGY



UNIVERSITY OF
CALIFORNIA



BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY



Bootstrapping Institutional Capability

Winter 2012 Joint Techs Conference

Gary Jung

High Performance Computing Services
Lawrence Berkeley National Laboratory

Berkeley Lab

A National Laboratory engaged in open, fundamental research across the sciences.

Our mission is easy: Solve the most pressing and profound scientific problems facing humankind.

- 700M Budget, 4,000 Employees, Thousands of Scientific Collaborators
- 12 Nobel Laureates including 2011 Physics Nobel Saul Perlmutter

A diverse, high-tech facility with advanced computing needs.



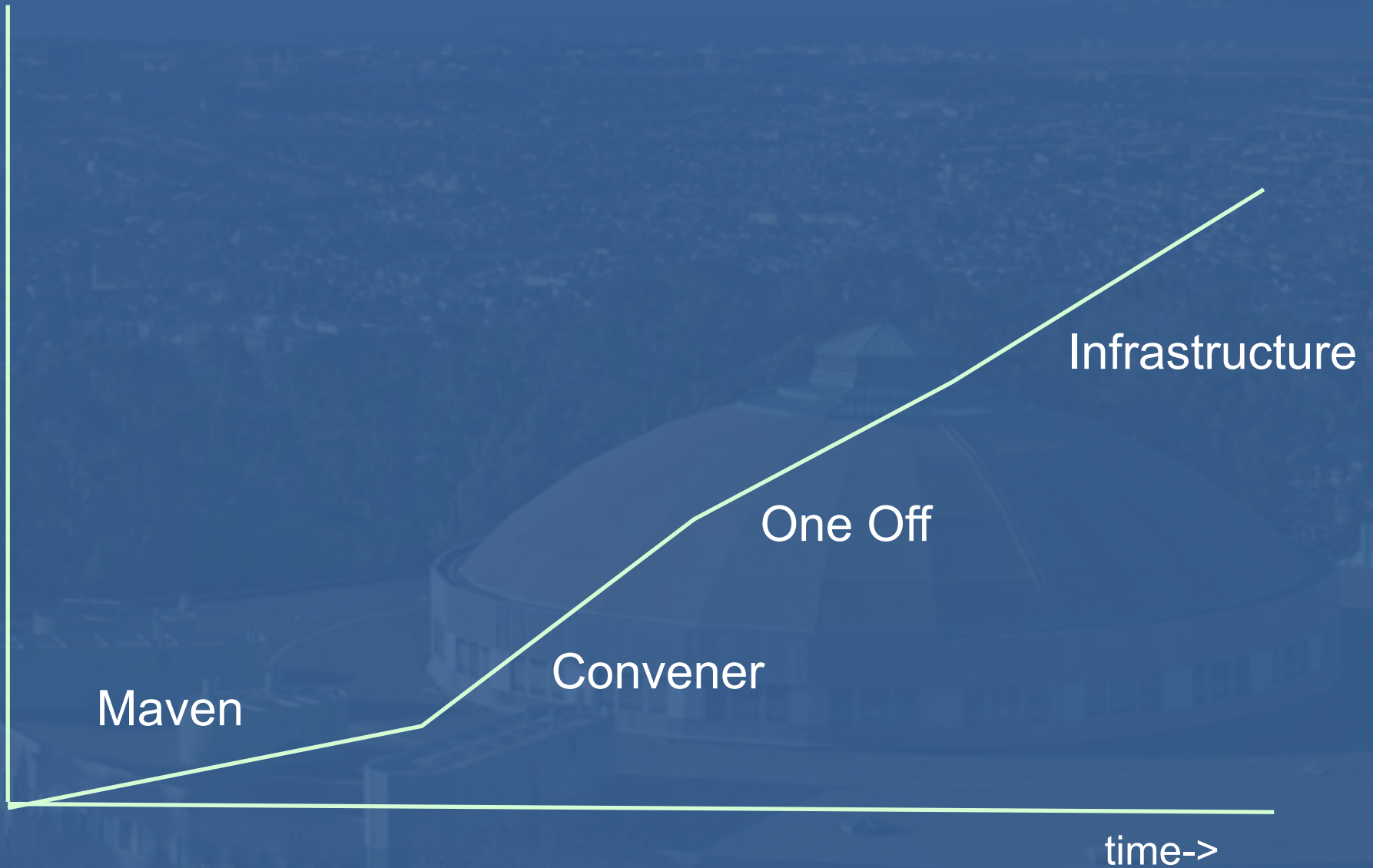
BERKELEY LAB
LAWRENCE BERKELEY NATIONAL LABORATORY



We are IT

Provide IT expertise/infrastructure/
services and leverage economies of scale
to better support science

maturity->



time->

Example: Scientific Computing

- Linux Cluster Support program 2003
 - Support for PI-owned clusters
- Shared HPC Infrastructure 2007
 - Common master/scheduler/home dir storage across all
- Midrange Institutional HPC system 2008
- Cluster Condo program 2011
 - Partner with PIs to grow institutional compute

Based on our experience building scientific computing, we can generalize the issues and project what challenges we might face supporting data intensive science.

Institutional Challenges (we have faced)

These include:

- Funding
 - New or reallocation of funds from traditional IT to scientific computing
 - Issue of what one thing do we want to do this year.
 - Theory of seeding activities and expecting them to become self-supporting
 - Recharge vs indirect overhead funds
- Ongoing management commitment
- Diversity of science/needs
- Buy-in from researchers
- DIY mentality vs central services
- Fully integrated solutions vs self-integrated
- Developing “depth on the bench”
 - Need expertise in scientific disciplines

So how do we support data intensive science?

- Not as tidy. Some requirements sound mutually exclusive. E.g. Big vs small files. Long term storage vs high perf storage.
- Where you compute starts to matter. Turnaround
 - Getting data to the compute and back
 - Crunching through the data
- Building recipes for data analysis requires higher touch
- Cross section of users may differ/overlap
- Long term issues to solve

We don't yet know what infrastructure looks like for data intensive computing.