The Abdus Salam
International Centre for Theoretical Physics

United Nations
Educational, Scientific and
Cultural Organization

IAEA
International Atomic Energy Agency

2068-22

Advanced School in High Performance and GRID Computing - Concepts and Applications

30 November - 11 December, 2009

How to choose the right architecture for your computation

S.T. Brown
Carnegie Mellon University
Pittsburgh
USA

Strada Costiera 11, 34151 Trieste, Italy - Tel.+39 040 2240 111; Fax +39 040 224 163 - sci_info@ictp.it

How To Find the Best Resources for My Application?

"WWAKD".

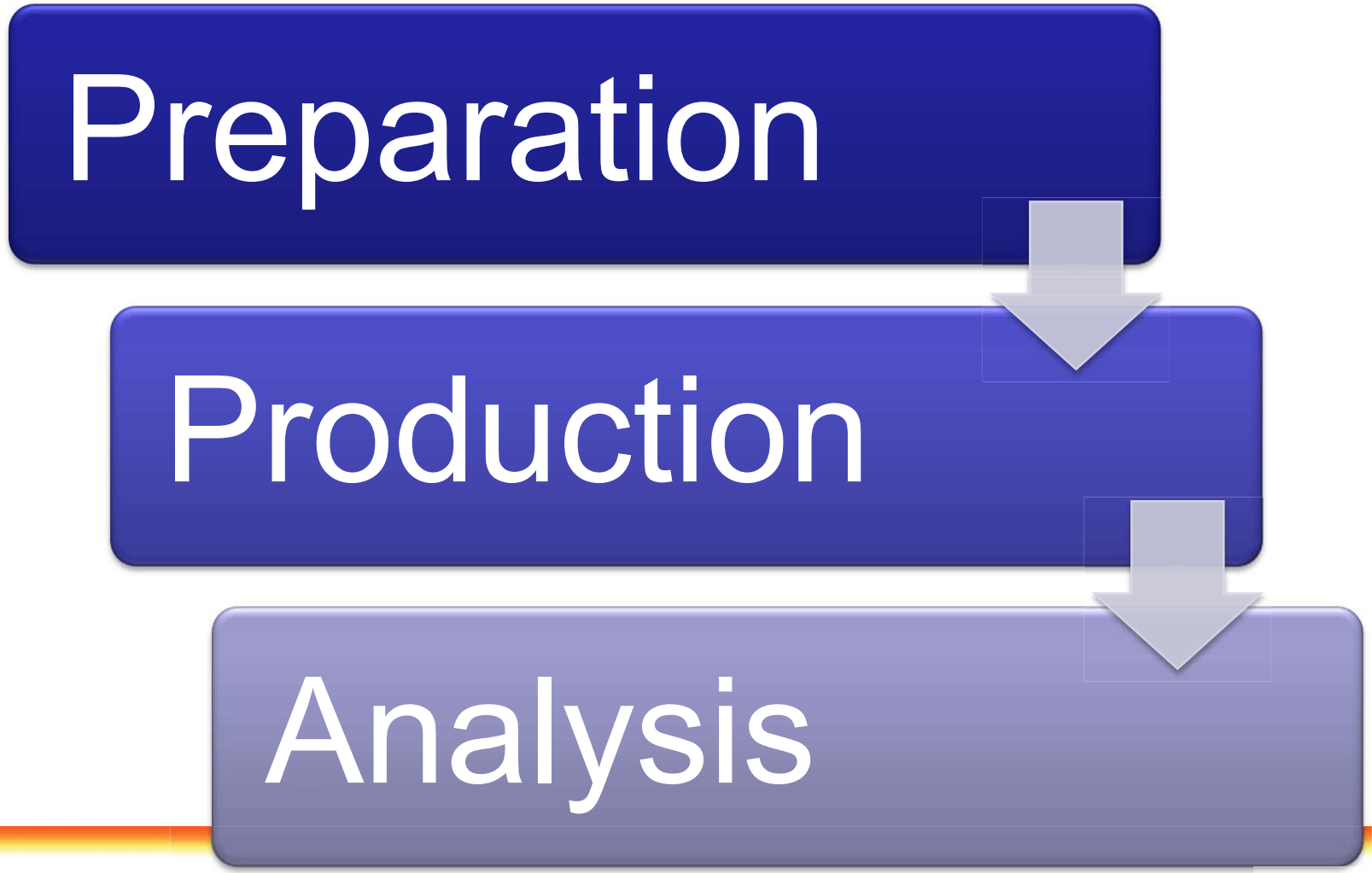
Shawn T. Brown
Senior Scientific Specialist
Pittsburgh Supercomputing Center

Why Should You Care?

Choosing the proper resource makes you more efficient, helping you get your research/work done more quickly and with less hassles.

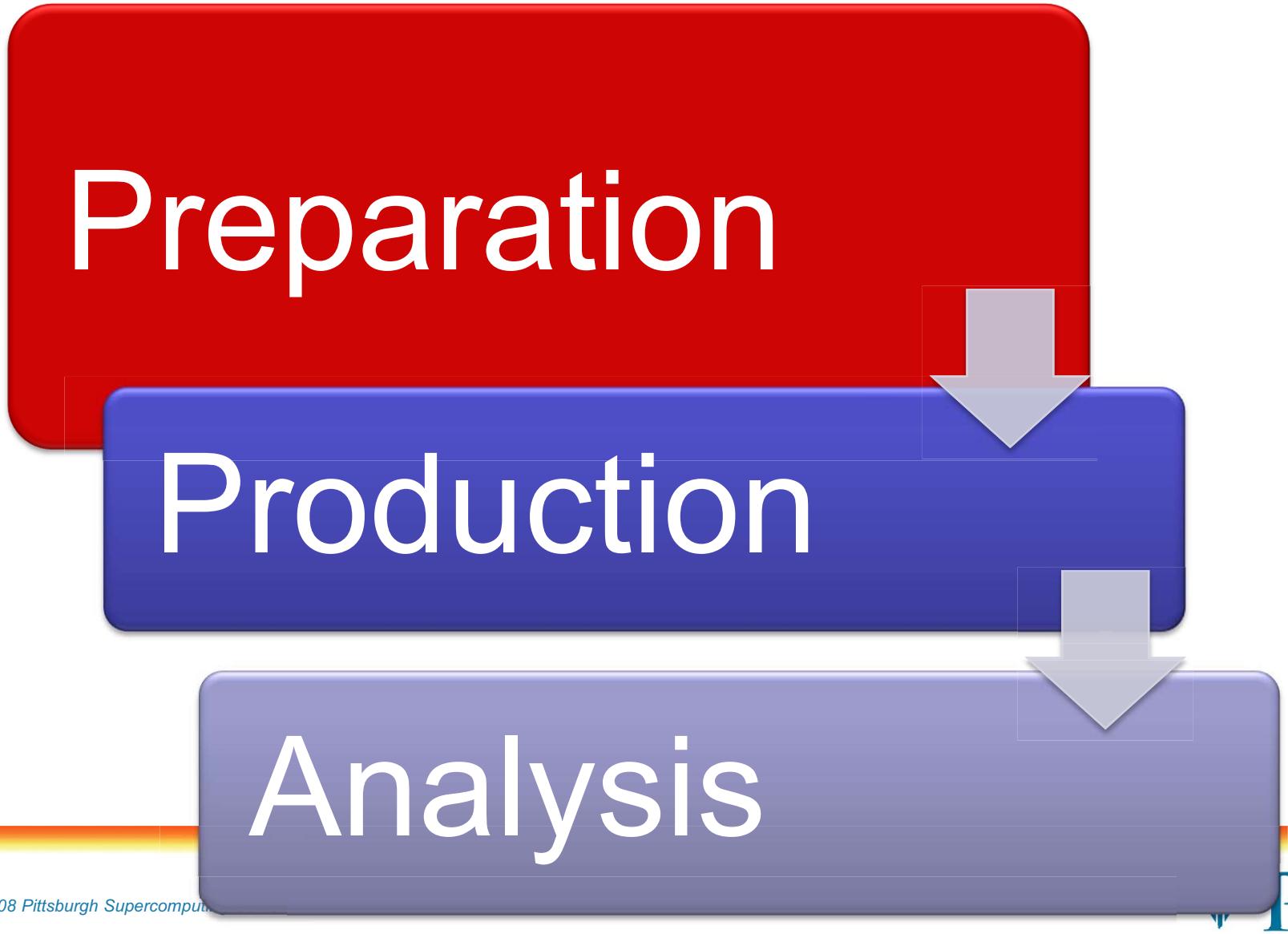
Choosing the wrong resources will hamper your work, hamper the work of others, and waste valuable resources that could be used for the computation for which it was designed.

The Lifecycle of a Computational Research Project



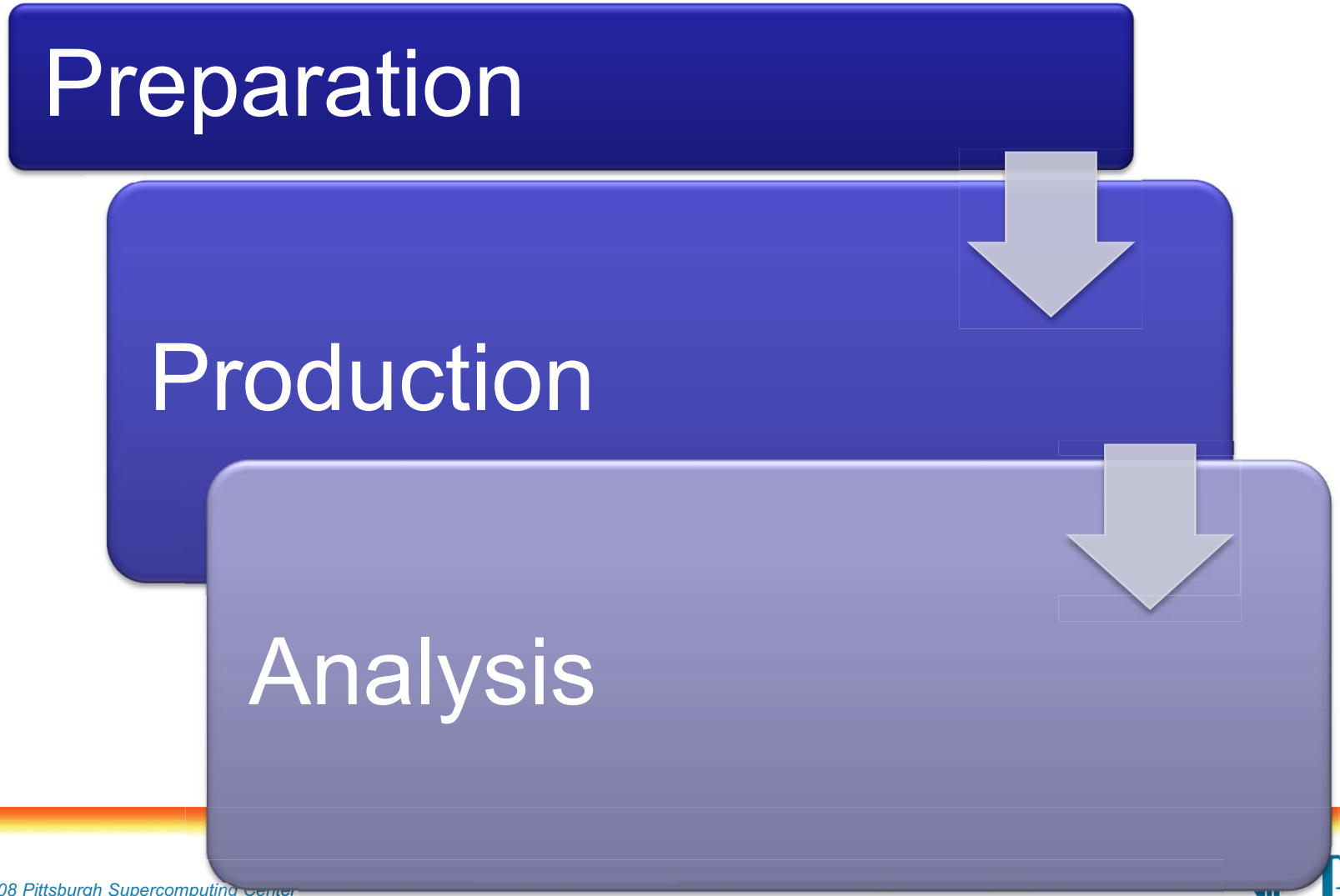
PSC
PITTSBURGH SUPERCOMPUTING CENTER

The Lifecycle of a Computational Research Project: For Beginners



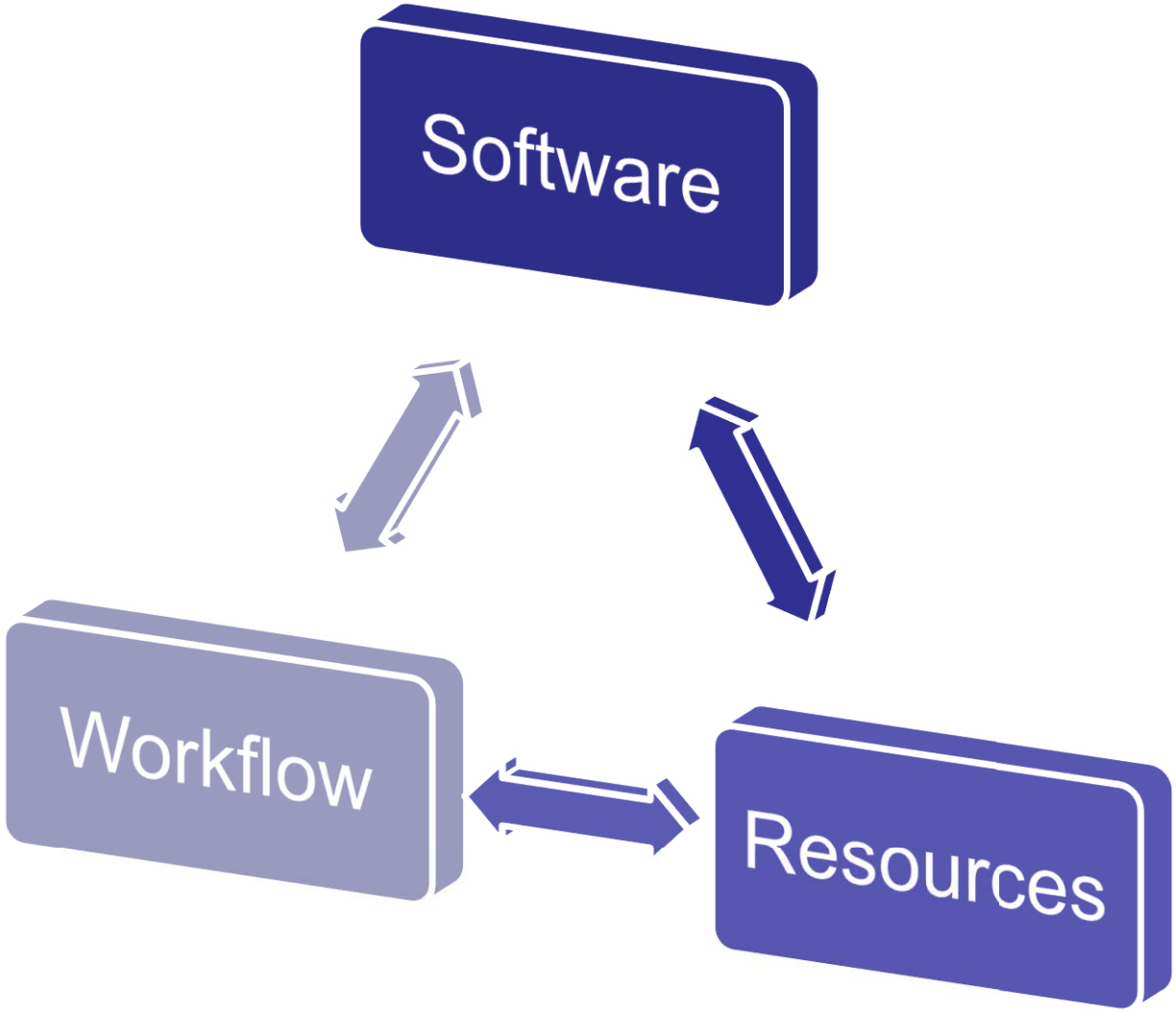
PSC
PITTSBURGH SUPERCOMPUTING CENTER

The Lifecycle of a Computational Research Project: Over Time and with Experience



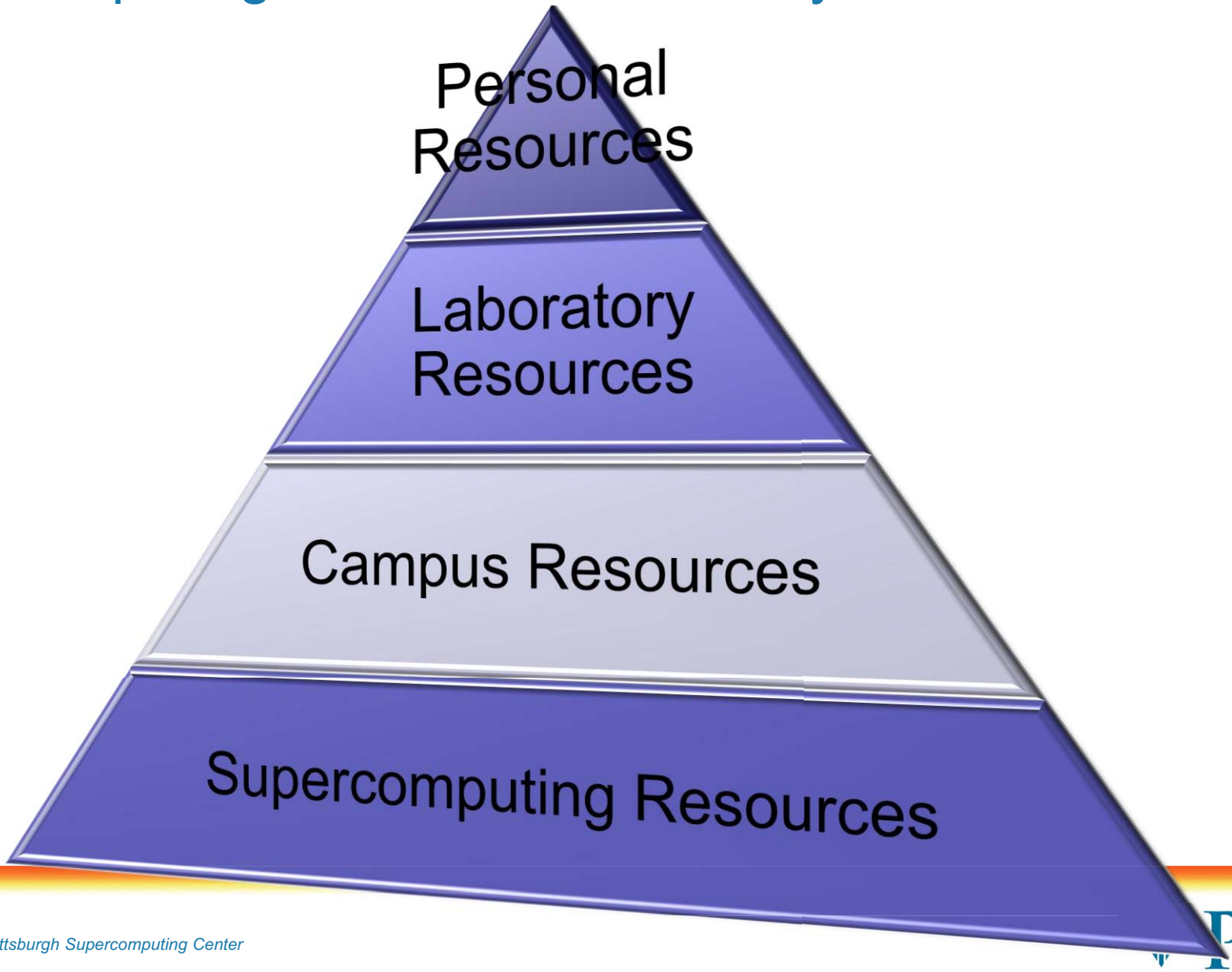
PSC
PITTSBURGH SUPERCOMPUTING CENTER

Preparation



PSC
PITTSBURGH SUPERCOMPUTING CENTER

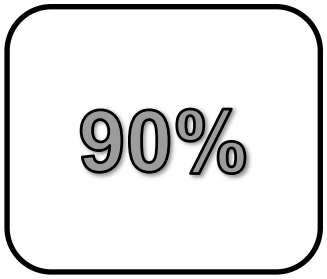
Computing Resources Hierarchy



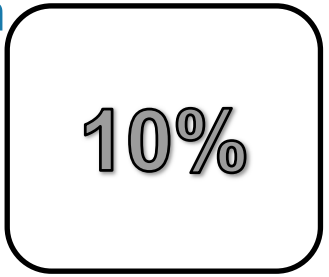
PSC
PITTSBURGH SUPERCOMPUTING CENTER

Capacity vs. Capability

- Capacity computing
  - Creating large supercomputers to facilitate large throughput of small parallel jobs
    - Cheaper, slower interconnects
    - Clusters running Linux, OS X, or even Windows
    - This is what you are learning about this week.



- Capability computing
  - Creating large supercomputers to enable computation on large scale
    - Running the entire machine to perform one task
    - Good fast interconnect and balanced performance important
    - Usually specialized hardware and operating systems



PSC
PITTSBURGH SUPERCOMPUTING CENTER

BigBen: PSC's Cray XT3

4,136 AMD Opteron cores
- 2.6 GHz clock, each 5.2 GFlop peak
- 22 TFlop theoretical peak aggregate

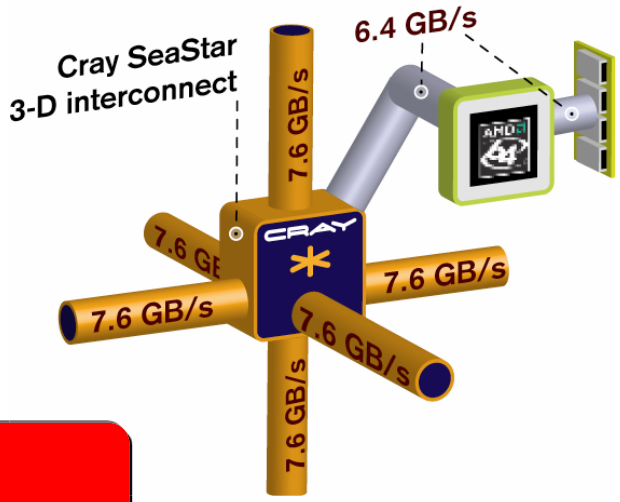
Cray SeaStar interconnect
- extremely high bandwidth:
- 6.5 GB/s sustained
- configured as a 3-D torus

4 TB aggregate memory (1GB/proc)

200 TB disk storage (DDN)

Scalable operating system

11 Million American Dollars
- 1Million / Tflop



EXPENSIVE!

PSC
PITTSBURGH SUPERCOMPUTING CENTER

Choosing the right level of resource

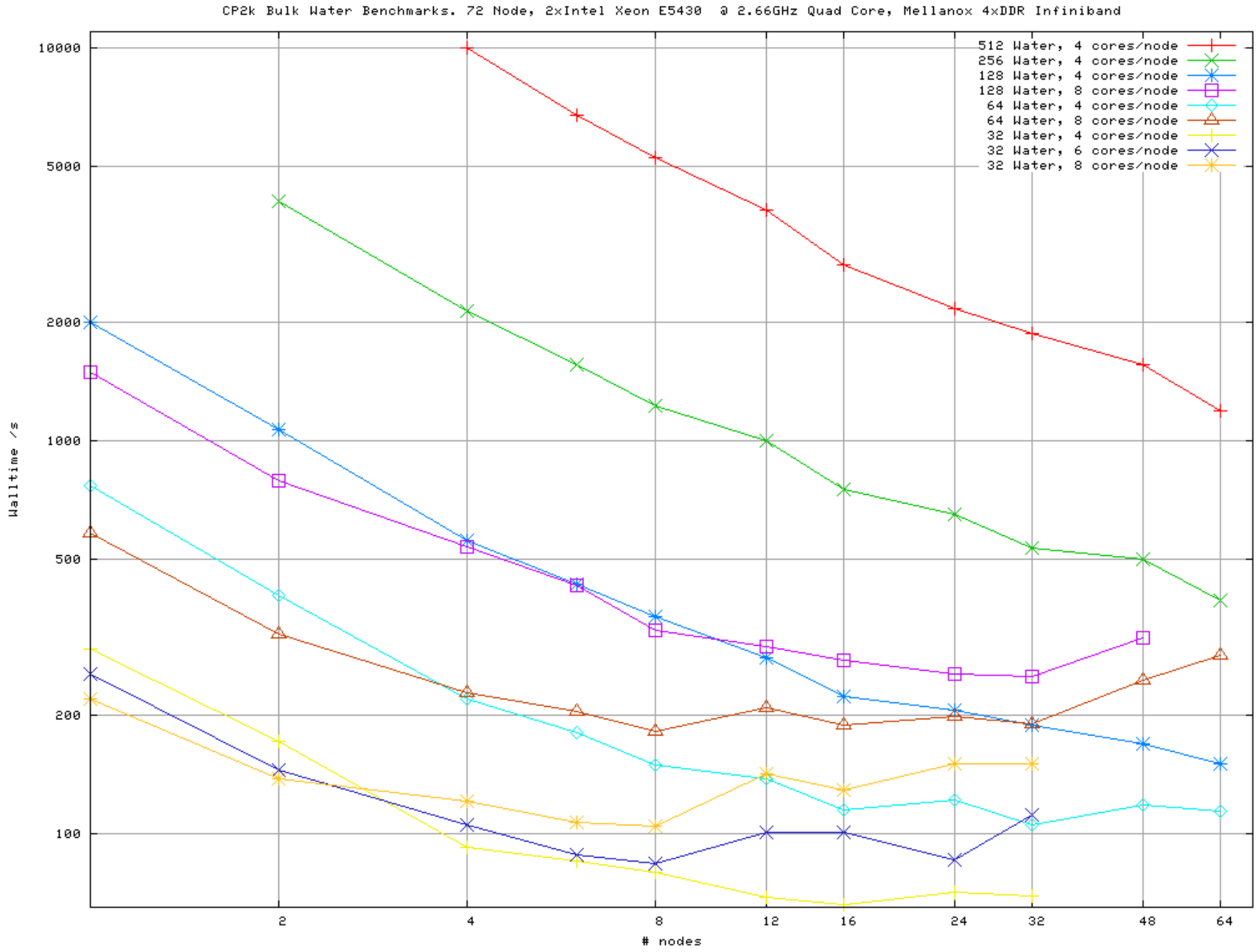
- Many factors go into deciding on a resources
  - Performance
  - Throughput
  - Availability
- There is no shame in not moving all the way to supercomputing resources.
  - Not every problem requires the highest level of the hierarchy.

Software: Things to consider before production.

- Goal: Picking the right software for the task.
  - This part gets easier with more experience
- Problem Size
  - Just because a software package can do a thing does not mean it will do what you want it to.
    - Understand the limitations of the methods and the software you are using.
  - Programs perform vastly different based on different input parameters and problem size
  - Try to reproduce exemplar problems.
    - Make sure you know the code works and you know how to use it.

Preparation

- Performance
  - Benchmarking the software
    - Scalability and raw performance
    - Collect available benchmarks
    - Try various hardware
- Requirements
  - Memory
  - Disk (Type and Space)
  - Parallel requirements
    - Number of Cores
    - Amount of Communication
    - Shared vs. Distributed



PSC
PITTSBURGH SUPERCOMPUTING CENTER

Preparation

- Software Selection
  - If you are using homegrown software
    - Make sure it works!
    - Debug as much as you can before getting into production
  - If you are using 3rd party software
    - Pick the right software for the job.
    - Get to know all of the codes you are considering
    - The best performance is not always the best code
    - Money $$$ (Don't' steal the code, we work hard on this stuff).
    - Make sure that the software is available

What should you be able to answer if Axel asks you!

- What Software are you going to use? Why?
- How does the software scale? What are the primary performance characteristics?
- How much memory? What kind of disk does it need?
- Can the software actually handle the problem size you want to study?

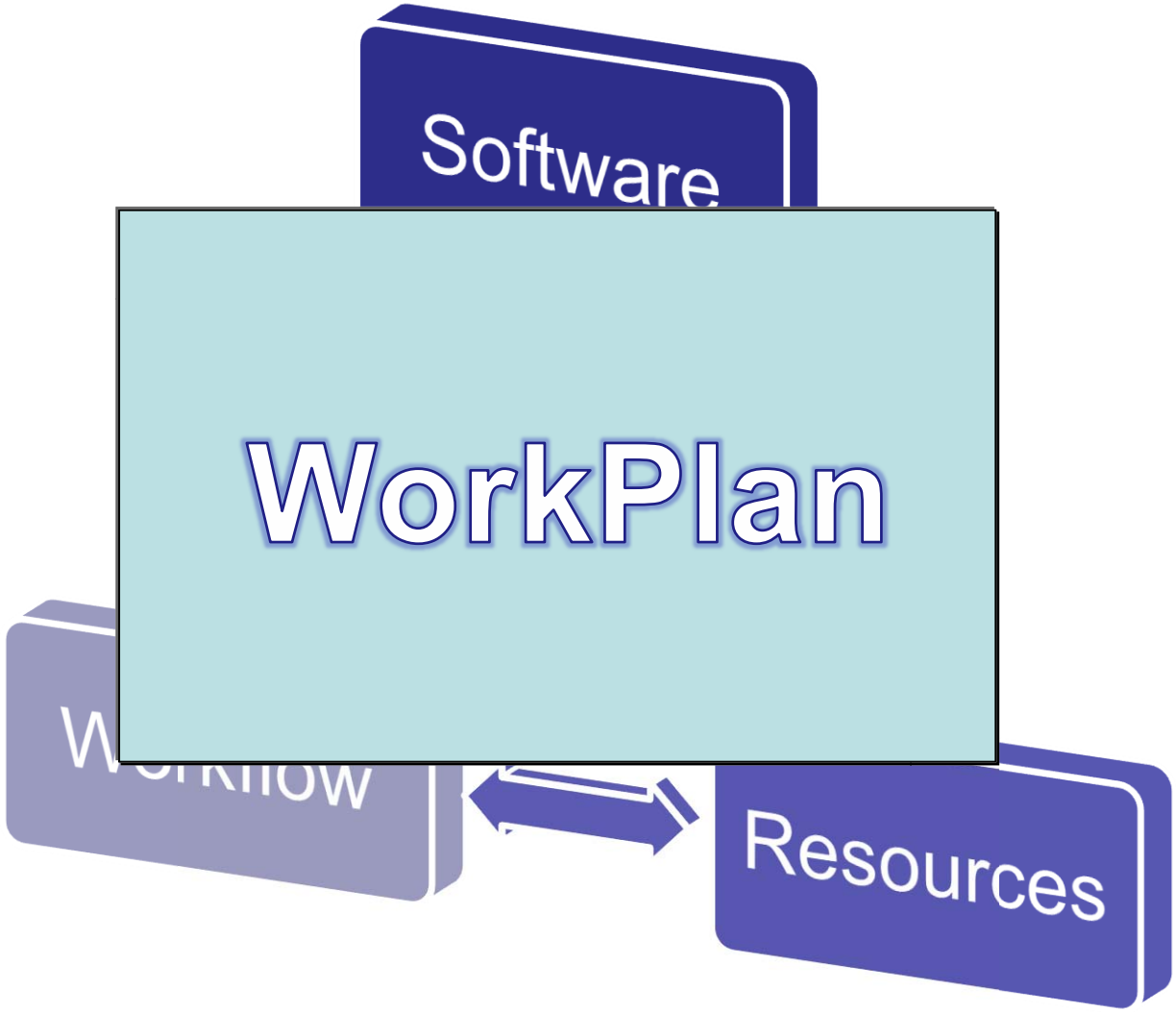
Workflow – Create a Plan

- Armed with benchmarks, and an understanding of the software you are going to use, it is time to come up with a plan
- What should be a part of this plan?
  - How many runs do you think you will need to perform?
  - What resource do you plan on using?
  - At what processor count will you be running?
  - It is alright to estimate, sometimes you just don't know exactly what you need.

Concerns

- Performance vs. Throughput
  - Better performance = less throughput
    - Better performing machines generally are more packed.
    - A balance must be stricken
    - Certain machines are also setup up to cater to certain clientele.
    - Look into the how the queuing structure is set up when making your decision.
    - Leave the capability computing to those who really need it.

Preparation



Other words of wisdom…

- Don’t create a problem that you will never finish.
  - It is actually rather easy to create research that will take way too long.
    - Why do you think a protein isn’t routinely folded in a computer?
- An ounce of preparation is worth a ton of cure
  - Understanding, benchmarking, and planning will save a lot of grief, time and pain.
  - The story of Nilesh.

You need to ask yourself one simple question

- What would Axel Kohlmeyer do?