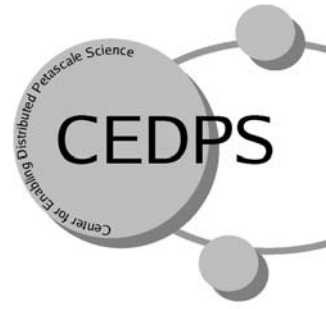# SciDAC Center for Enabling Distributed Petascale Science

Argonne National Laboratory

Fermi National Accelerator Laboratory

Lawrence Berkeley National Laboratory

University of Southern California

University of Wisconsin

www.cedps.net

PI: Ian Foster

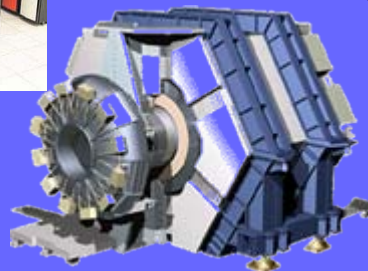Project Director: Jennifer Schopf

The Petascale Data Challenge

Remote distributed users

- DOE facilities generate **many petabytes** of data (2 petabytes = **all** U. S. academic research libraries!)

Massive data

- **Remote users** (at labs universities, industry) need data!

DOE facilities

- **Rapid, reliable access** key to maximizing value of $B facilities

Bridging the Divide (1): **Move Data to Users When & Where Needed**

"Deliver this 100 Terabytes to locations A, B, C by 9am tomorrow"

- **Fast**: >10,000x faster than usual Internet

- **Reliable**: recover from many failures

- **Predictable**: data arrives when scheduled

- **Secure**: protect expensive resources & data

- **Scalable**: deal with many users & much data

Bridging the Divide (2): **Allow Users to Move Computation Near Data**
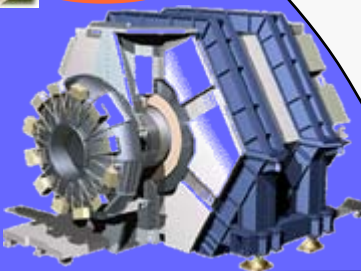
"Perform my computation F on datasets X, Y, Z"

- **Science services**: provide analysis functions near data source
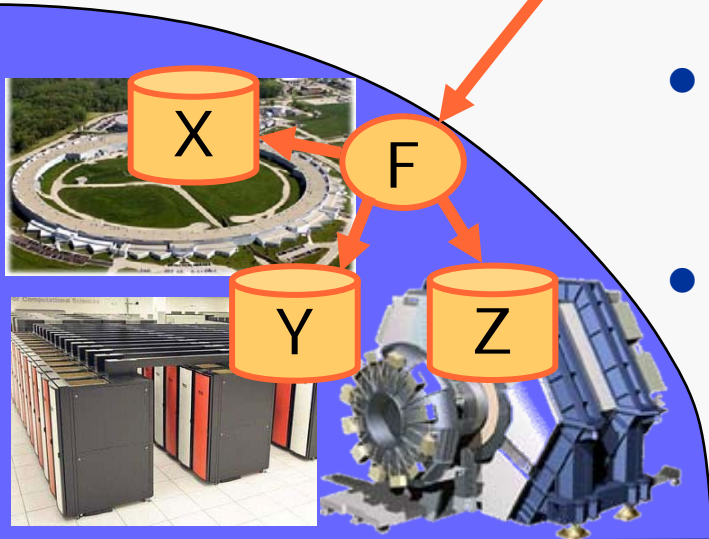
- **Flexible**: easy integration of functions

- **Secure**: protect expensive resources & data

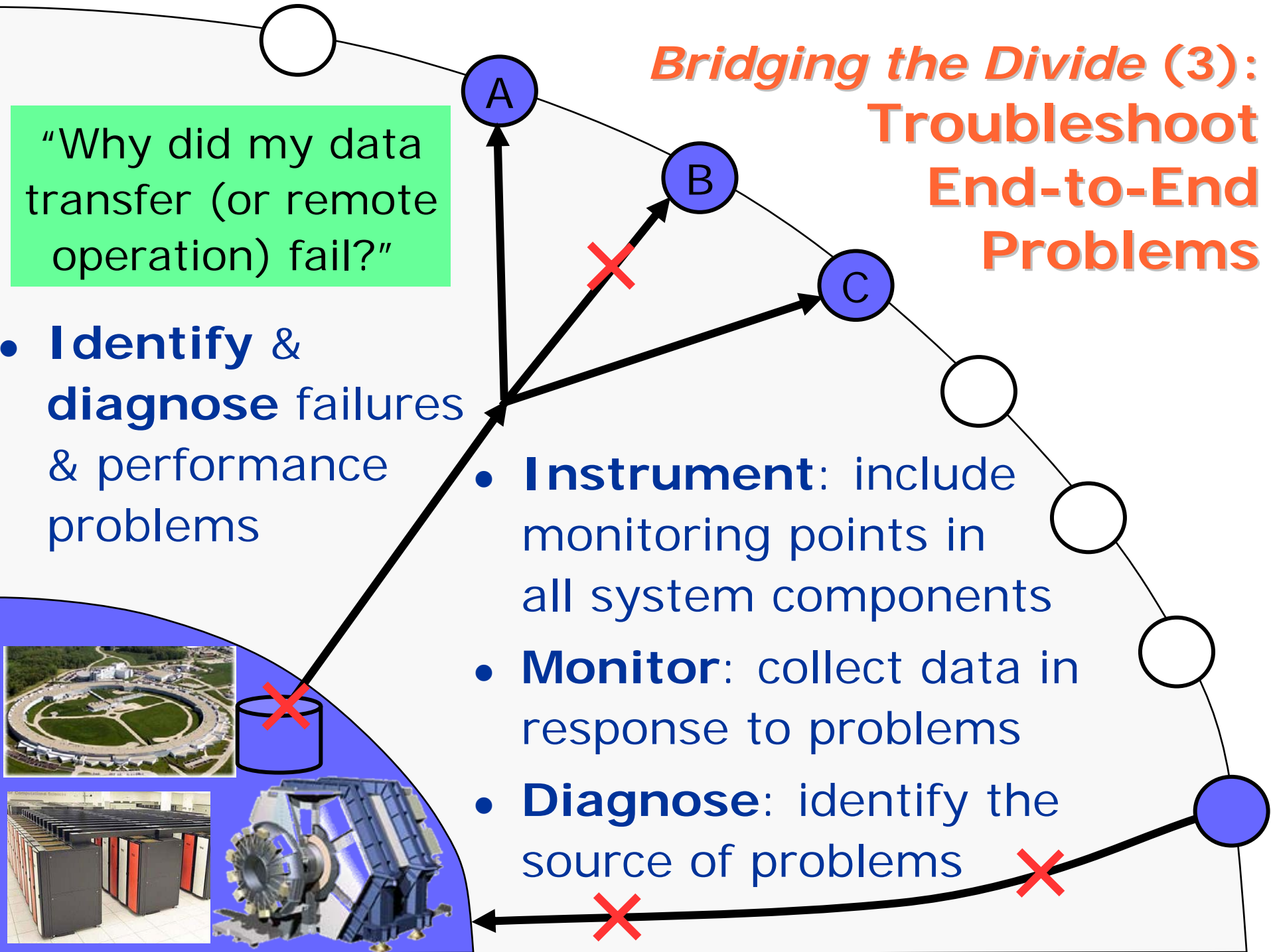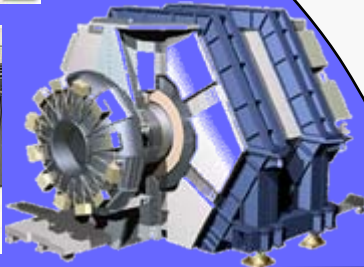- **Scalable**: deal with many users & much data

"Why did my data transfer (or remote operation) fail?"

- **Identify** & **diagnose** failures & performance problems

- **Instrument**: include monitoring points in all system components

- **Monitor**: collect data in response to problems

- **Diagnose**: identify the source of problems

Overview

- **Work with OSG**
  - Data Area: Get dCACHE and GridFTP wto work more closely (Dan/Gene can speak more to this)
  - Troubleshooting Area: OSG has said they are interested in picking up the centralized logging work when it's stable
  - Services: Keahey working with STAR, open for discussion with other applications

- **CEDPS All Hands Meeting March 12-13 at ISI**
  - http://www.cedps.net
  - Project management plan
    http://cedps.net/wiki/images/2/20/CEDPS_management_plan.pdf