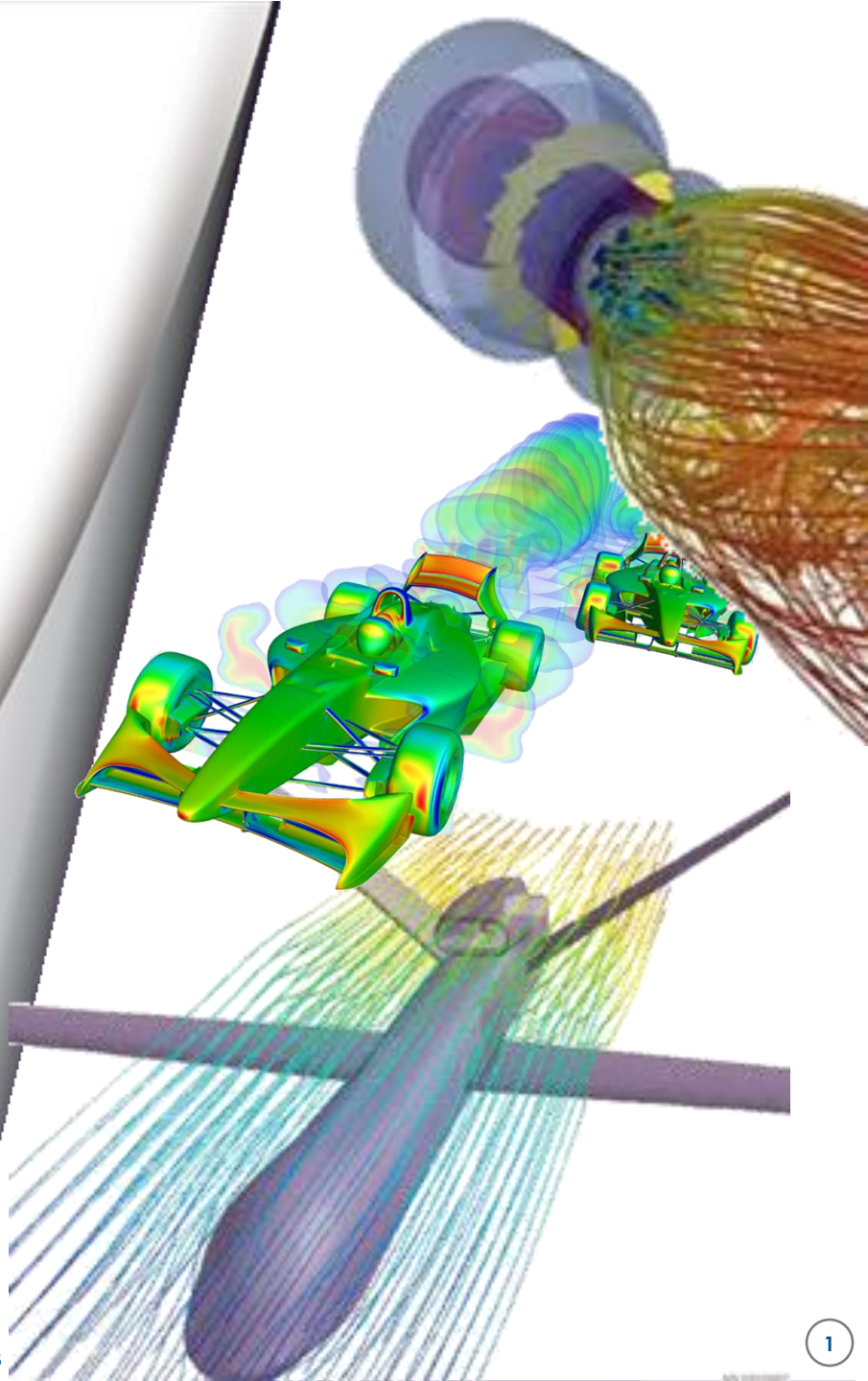
CASC Panel

The Crystal Ball Petascale to Exascale Trends

Barry Bolding, Ph.D.
VP Storage & Data Management
VP Marketing
bbolding@cray.com



CRAY
THE SUPERCOMPUTER COMPANY

Who are We?

Cray Inc.

- Nasdaq: CRAY
- ~1000 employees across 30 countries
- Headquartered in Seattle, WA

Serving

- Fortune 100 Companies
- Governments
- Research Consortia
- Many organizations represented in CASC

Business

- High Performance & Cluster Computing
- Storage and Data Management
- Big Data Analytics Solutions – YarcData Company

Modeling in a Data-Intensive World

Cray Supercomputers solving "grand challenges" in science, engineering and analytics

Data Models

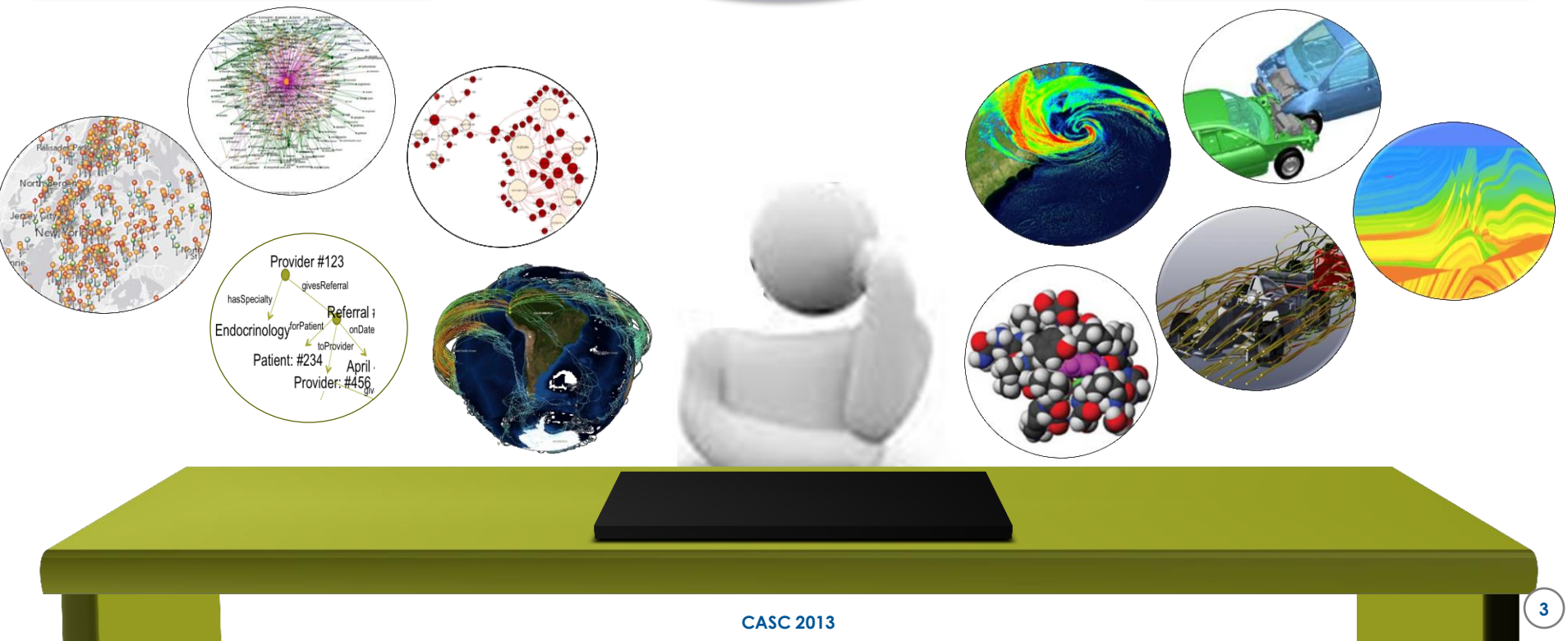
Integration of datasets and math models for search, analysis, predictive modeling and knowledge discovery

Data-Intensive Processing

High throughput event processing & data capture from sensors, data feeds and instruments

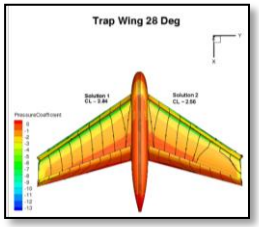
Math Models

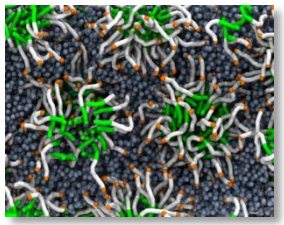
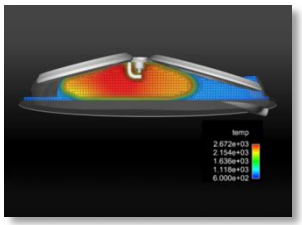
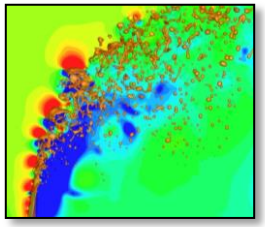
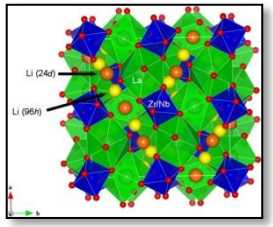
Modeling and simulation augmented with data to provide the highest fidelity virtual reality results

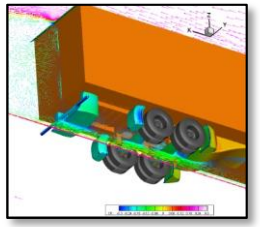


Industrial Innovation Every Day

Aircraft design	Consumer products	Engine cycle-to-cycle variation	Jet engine efficiency	Li-ion batteries	Long-haul truck fuel efficiency
Simulating takeoff and landing scenarios improved a critical code for estimating characteristics of commercial aircraft, including lift, drag, and controllability	Leadership computing and molecular dynamics software advanced understanding of chemical processes that can limit product shelf life	Emerging model of engine cyclic variation will apply thousands of processors to a challenging problem	Accurate predictions of atomization of liquid fuel by aerodynamic forces enhance combustion stability, improve efficiency, and reduce emissions	New classes of solid inorganic Li-ion electrolytes could deliver high ionic and low electronic conductivity and good electrochemical stability	Simulations reduced by 50% the time to develop a unique system of add-on parts that increases fuel efficiency by 7–12%



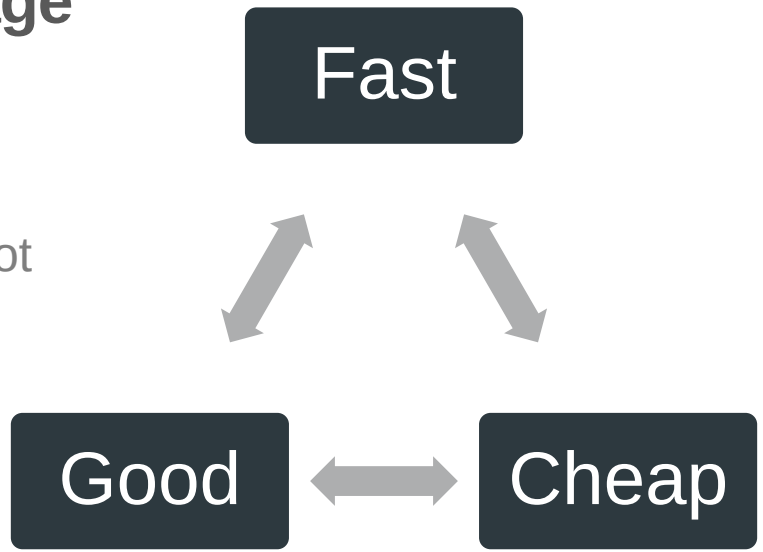


Living in a Data-Intensive World

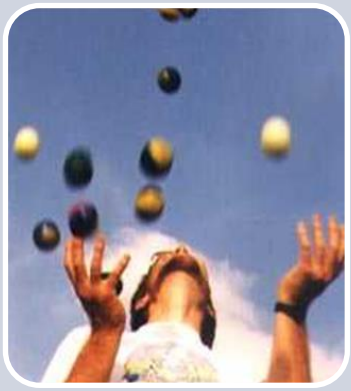
- **Sustainable funding and business models**
  - Flexibility in long customer/vendor/govt engagements
  - Partnership. Means different things to everyone involved
  - Customer/Vendor Engagement.
    - e.g. Cray User Group and Cray Centers of Excellence
- **Can accelerator (GPU/MIC) systems be supported and cost effective at mid-sized HPC centers? Is there a recommended mix of standard systems versus accelerator based systems at such centers?**
    - New technologies will keep coming
      - Open Standards,
        - OpenMP, OpenAcc, OpenSFS, OpenStack, etc.
        - Not all of these will succeed, but some will flourish.
    - There is no magic ratio of acceleration
    - Metrics
      - Sustained performance per KW
      - TCO
      - Time to knowledge discovery
      - SW Development time
      - Support for future data-intensive workflows

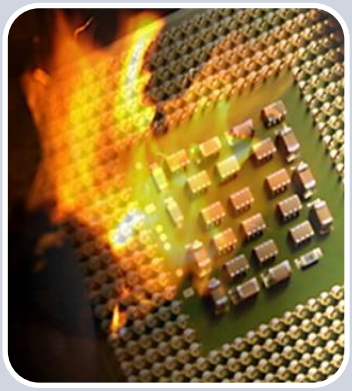
Clouds for a Data Intensive World

- **What is the role, if any, for the public cloud in providing HPC and storage in the future?**
  - Cloud is an access model, it is not a technology definition
    - Makes resources more accessible but not necessarily better for everything
  - Evaluate return on investment
  - Decide on requirements on
    - Data availability
    - Data sharing
    - Data preservation
  - A wide variety of technologies and solutions can exist in a cloud access model.



Key Challenges to Get to Exascale





Power

- Traditional voltage scaling is over
- Power now a major design constraint
- Cost of ownership
- Driving significant changes in architecture

Concurrency

- A billion operations per clock
- Billions of refs in flight at all times
- Will require *huge* problems
- Need to exploit *all* available parallelism

Programming Difficulty

- Concurrency and new micro-architectures will significantly complicate software
- Need to hide this complexity from the users

Resiliency

- Many more components
- Components getting less reliable
- Checkpoint bandwidth not scaling
- Impacts both systems and storage