CRAY®
THE SUPERCOMPUTER COMPANY

The Fusion of Supercomputing and Big Data

Peter Ungaro
President & CEO

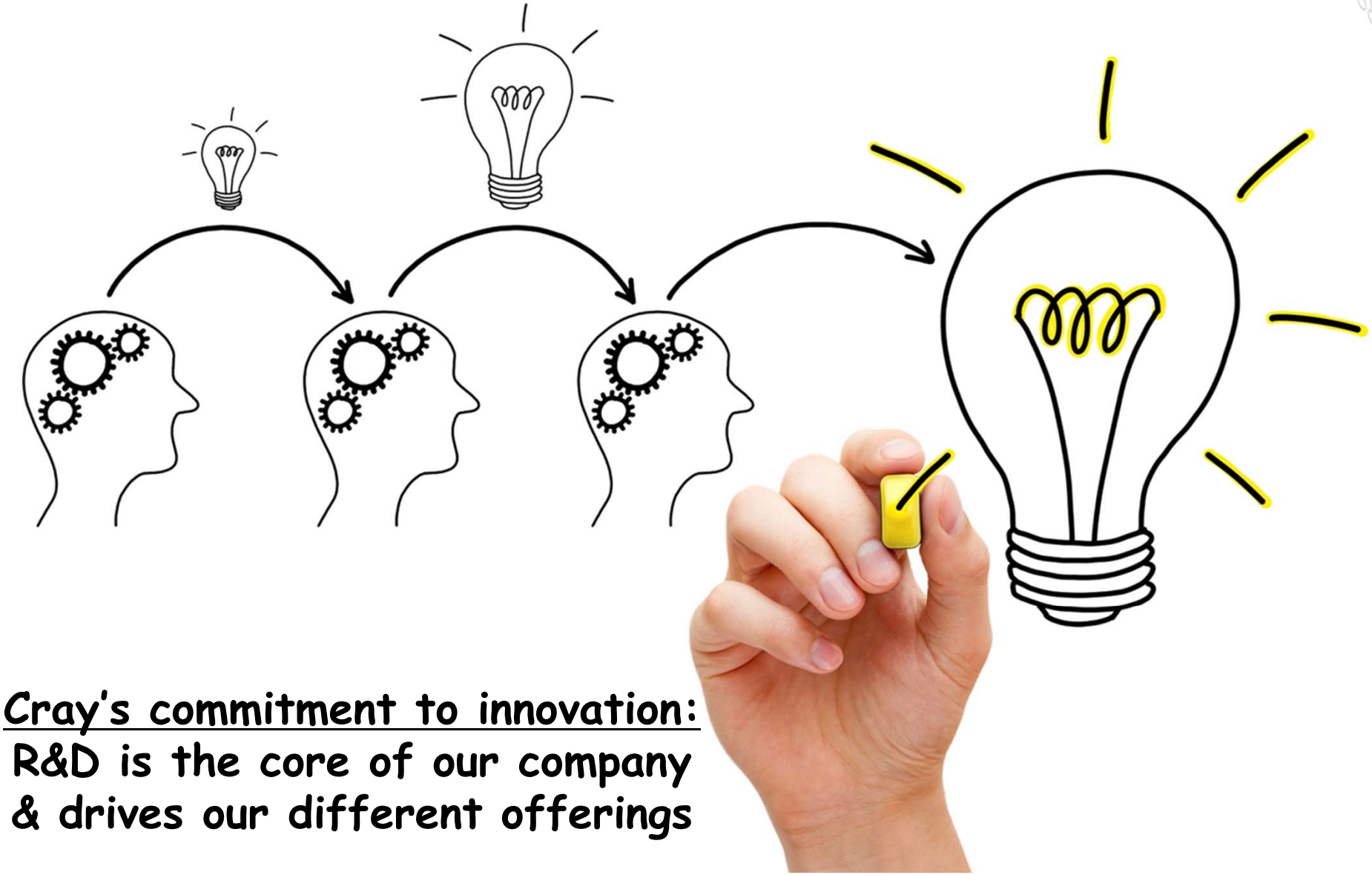
“The Supercomputing Company”

Supercomputing

Big Data

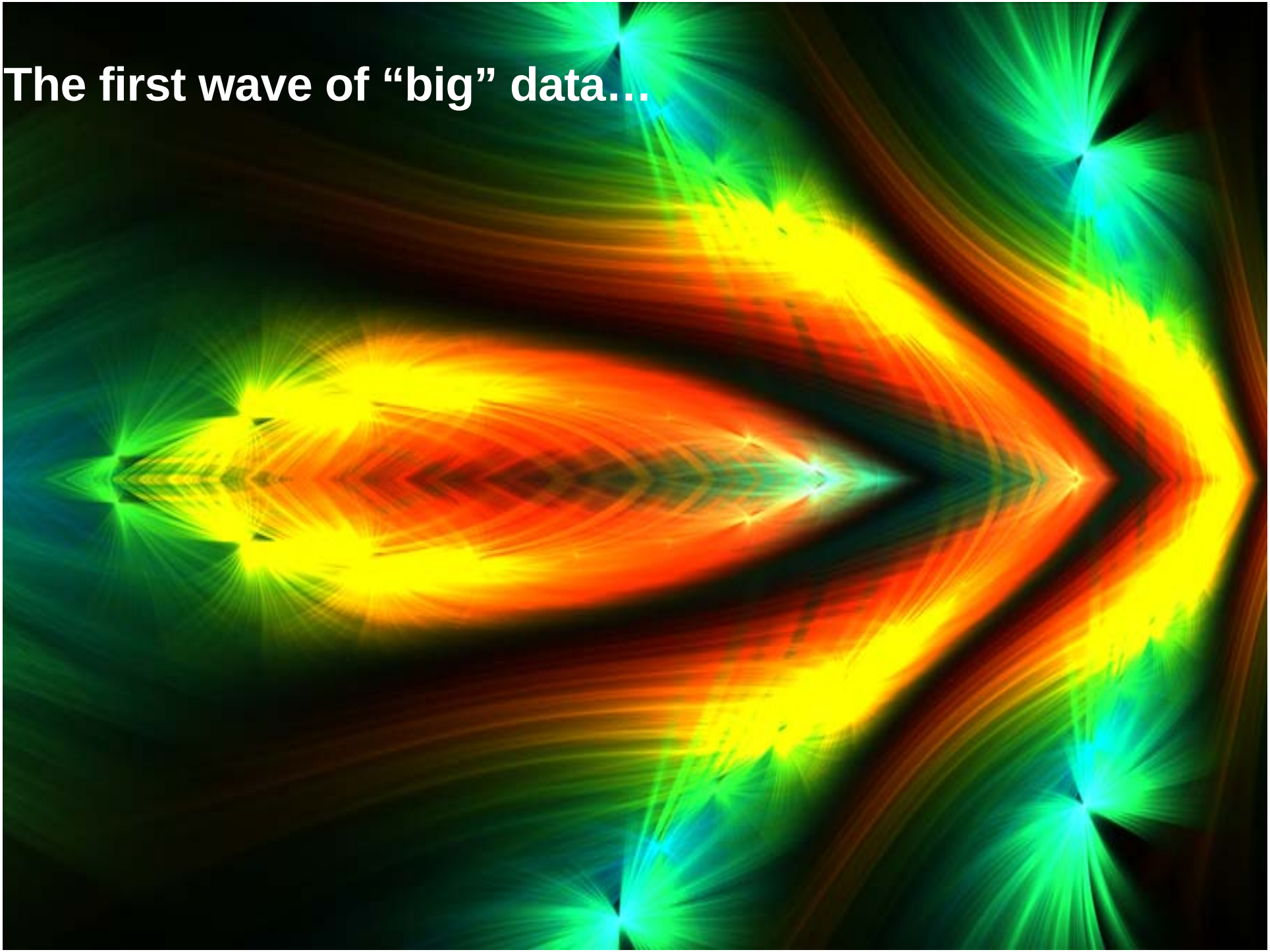
Because some great things never change…

One other thing that hasn't changed….



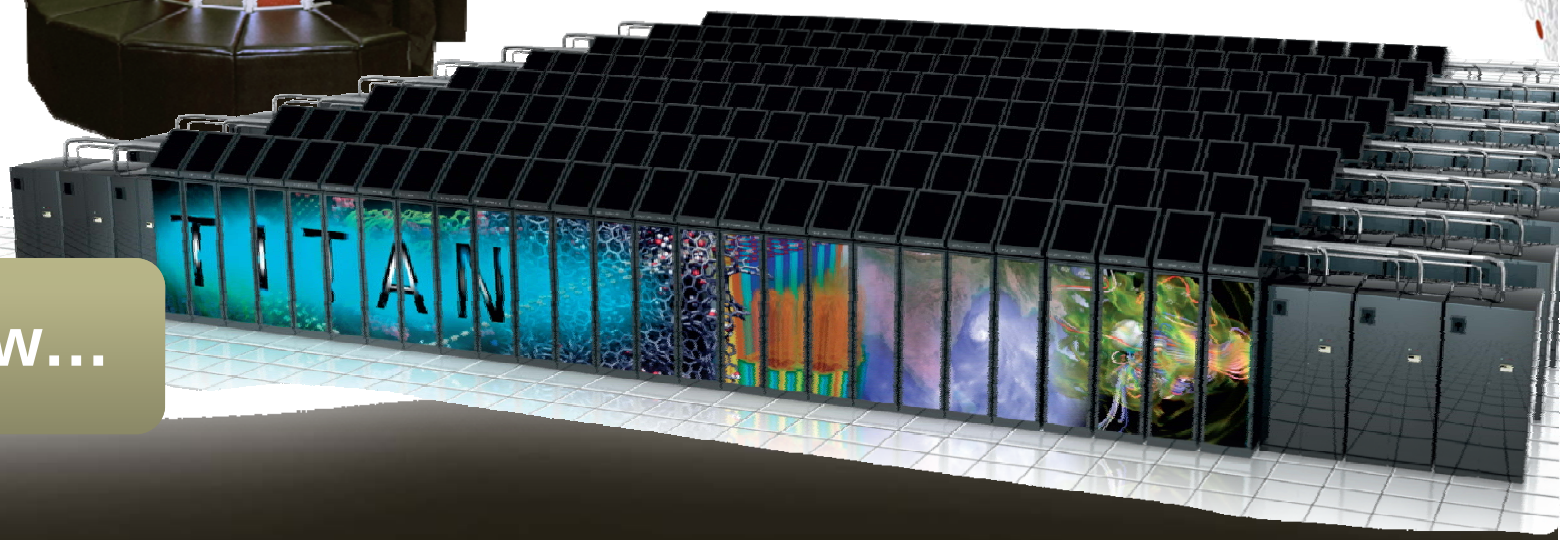
Cray's commitment to innovation:
R&D is the core of our company & drives our different offerings

The first wave of “big” data…



The current wave of Big Data….

Cray then…



Cray now…

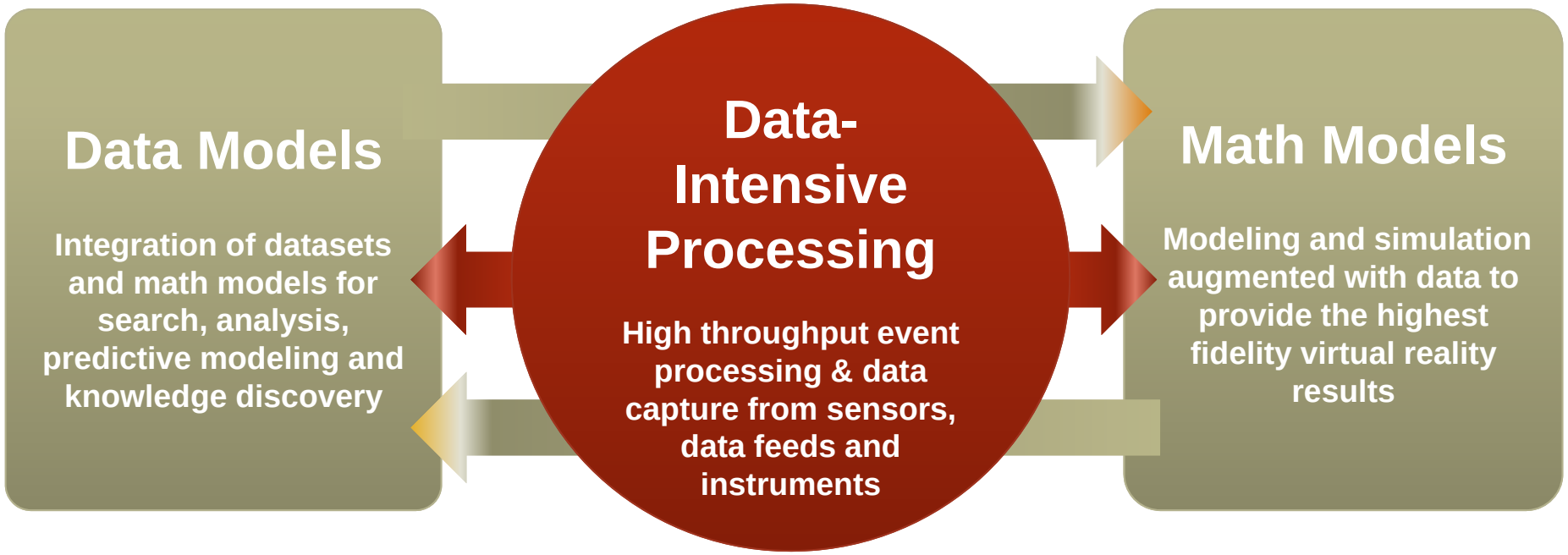
and there is more to come….

Cray's Vision:

The Fusion of Supercomputing and Big & Fast Data

Modeling The World

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

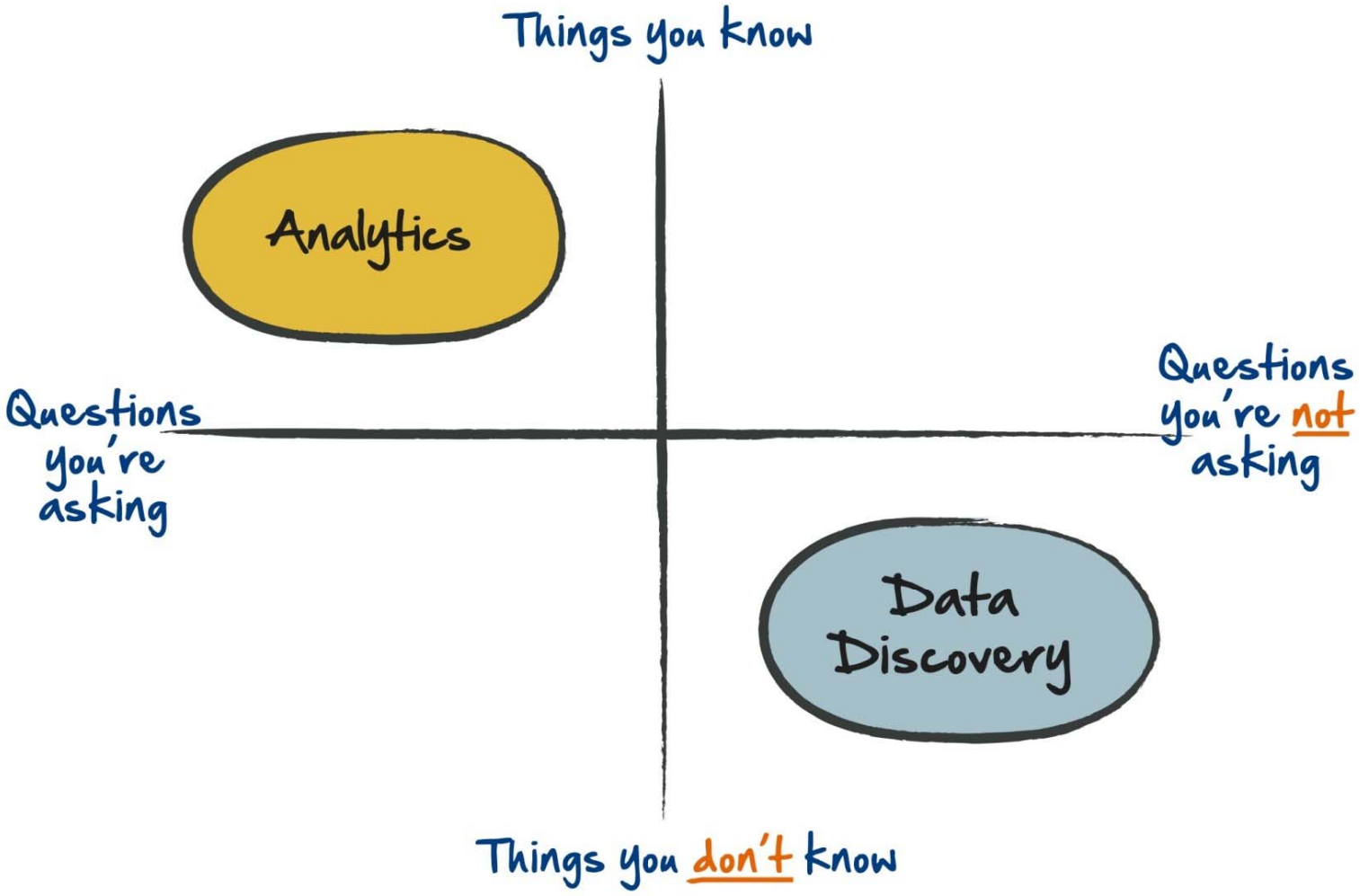


Compute | Store | Analyze

Data-Intensive Processing is driving the need for advanced architectures

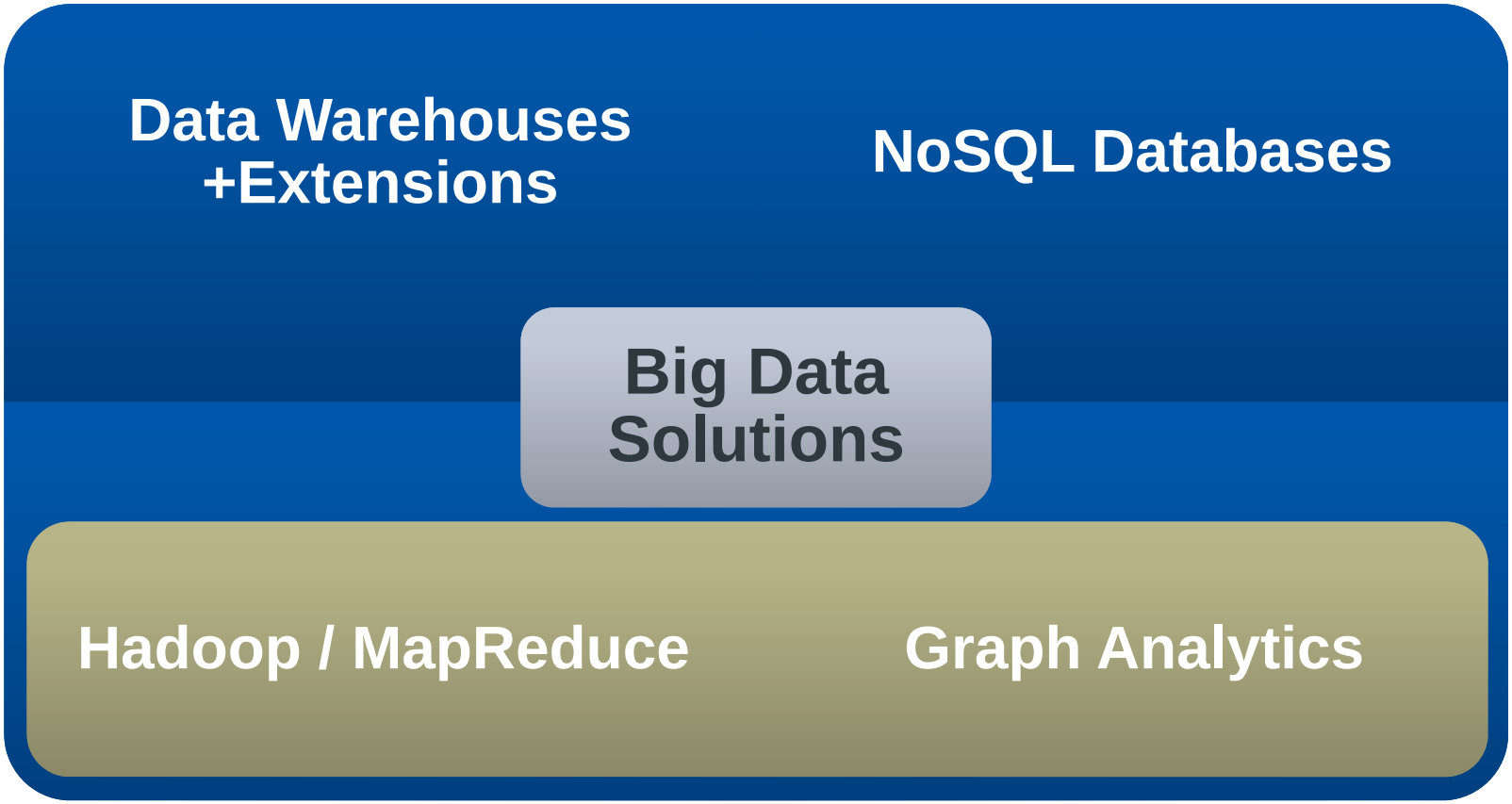
Source: Eric Green, Director, National Institute of Health: NextGen 101 Workshop

Big Data's Methods for Analysis



"Take all these different data sources and put them together and then help me find something about the data that I don't already know…"

Solutions for Advanced Analytics



These solutions can overlap, but also can be very complementary as each has strengths & weaknesses

System Architecture Differences…

Supercomputing

- Scalable computing w/high BW, low-latency, Global Mem Architectures
- Highly integrated processor-memory-interconnect & network storage
- Minimize data movement – load the “mesh” into memory
- Move data for loading, check-pointing or archiving
- “Basketball court sized” systems

Large-scale Data Analytics

- Distributed computing at largest scale
- Divide-and-conquer approaches on Service Orientated Architectures
- Maximize data movement-- Scan/Sort/Stream all the data all the time
- Lowest cost processor-memory-interconnect & local storage
- “Warehouse sized” clouds

Applications Differences…

Computational requirements far outweighed by requirements of memory hierarchy

- Working data set doesn't fit in memory (Capacity limited)
- Floating point or integer unit stalled for data access (Bandwidth limited)
- Highly irregular, cache-unfriendly, data access (Latency limited)

On a given platform, applications are either

- Memory-bound (capacity, bandwidth, latency)
- Interconnect-bound (bandwidth, latency)
- I/O-bound (capacity, bandwidth, latency)

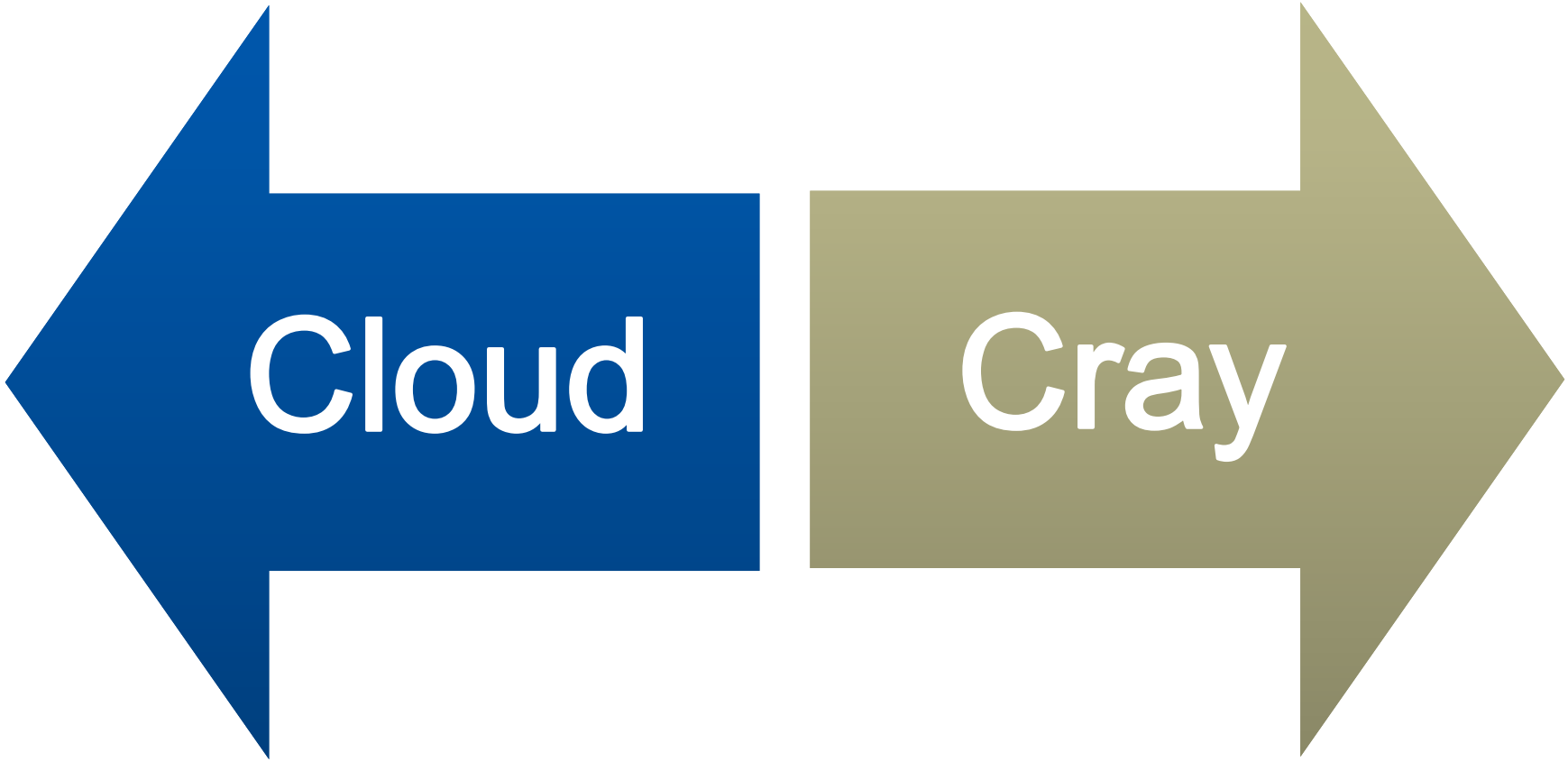
Mirror opposite of compute-bound problems

- FLOPS matter less – no easy path to higher peak computing rates

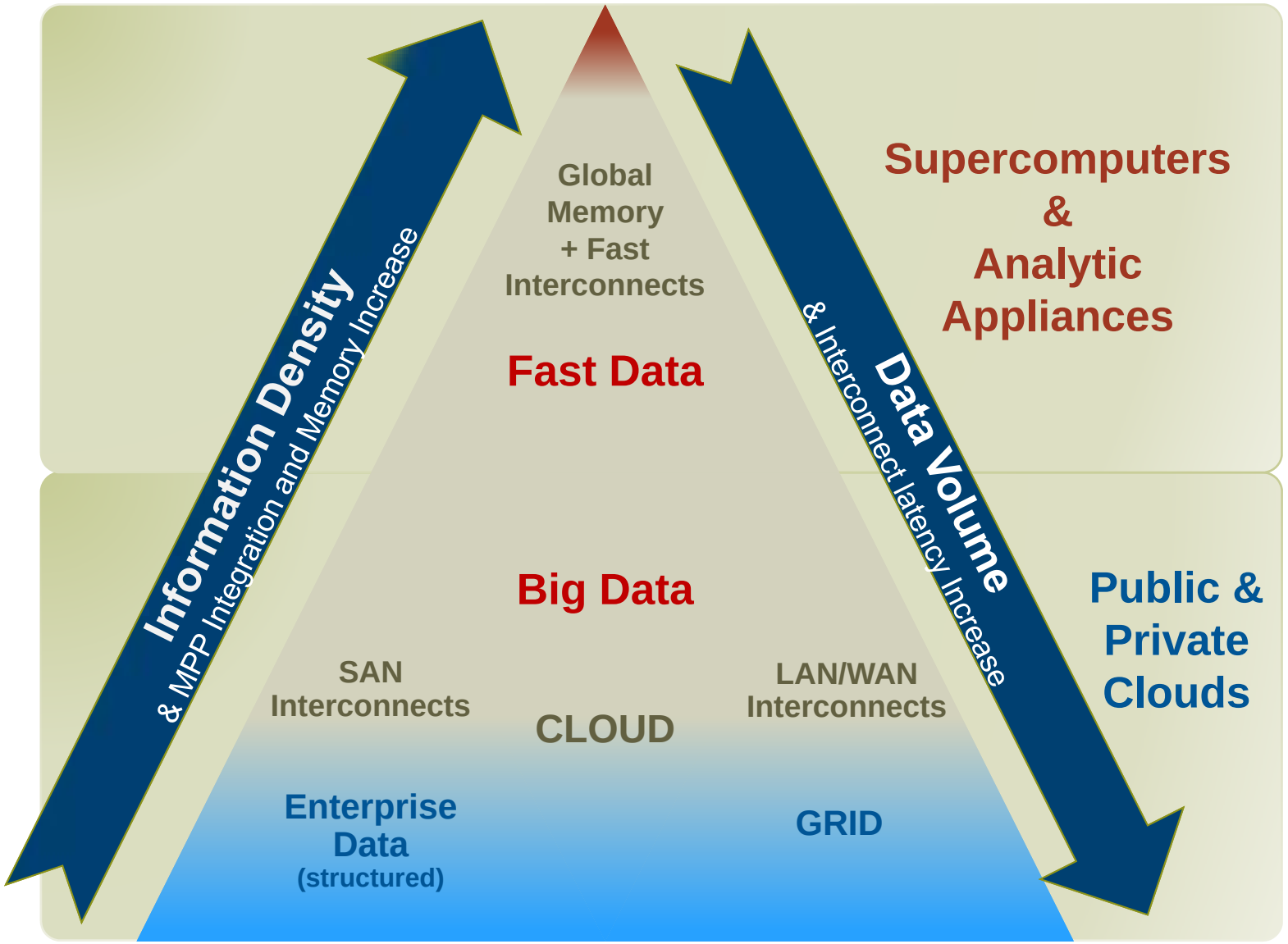
Infrastructure requirements are quite different

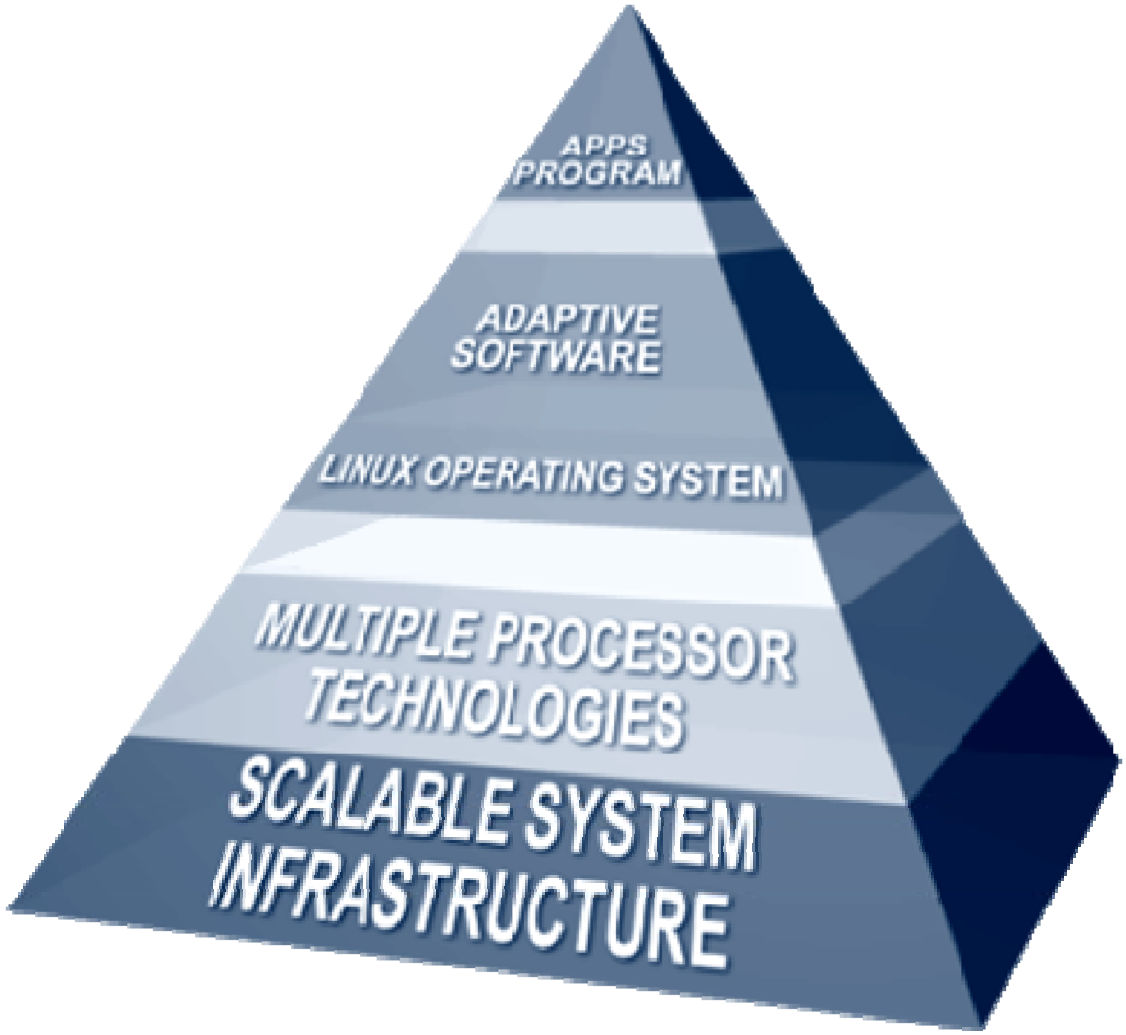
- Must have specialized nodes and adaptive runtimes to handle differences

Can’t “The Cloud” Do This?



Big Data → Fast Data

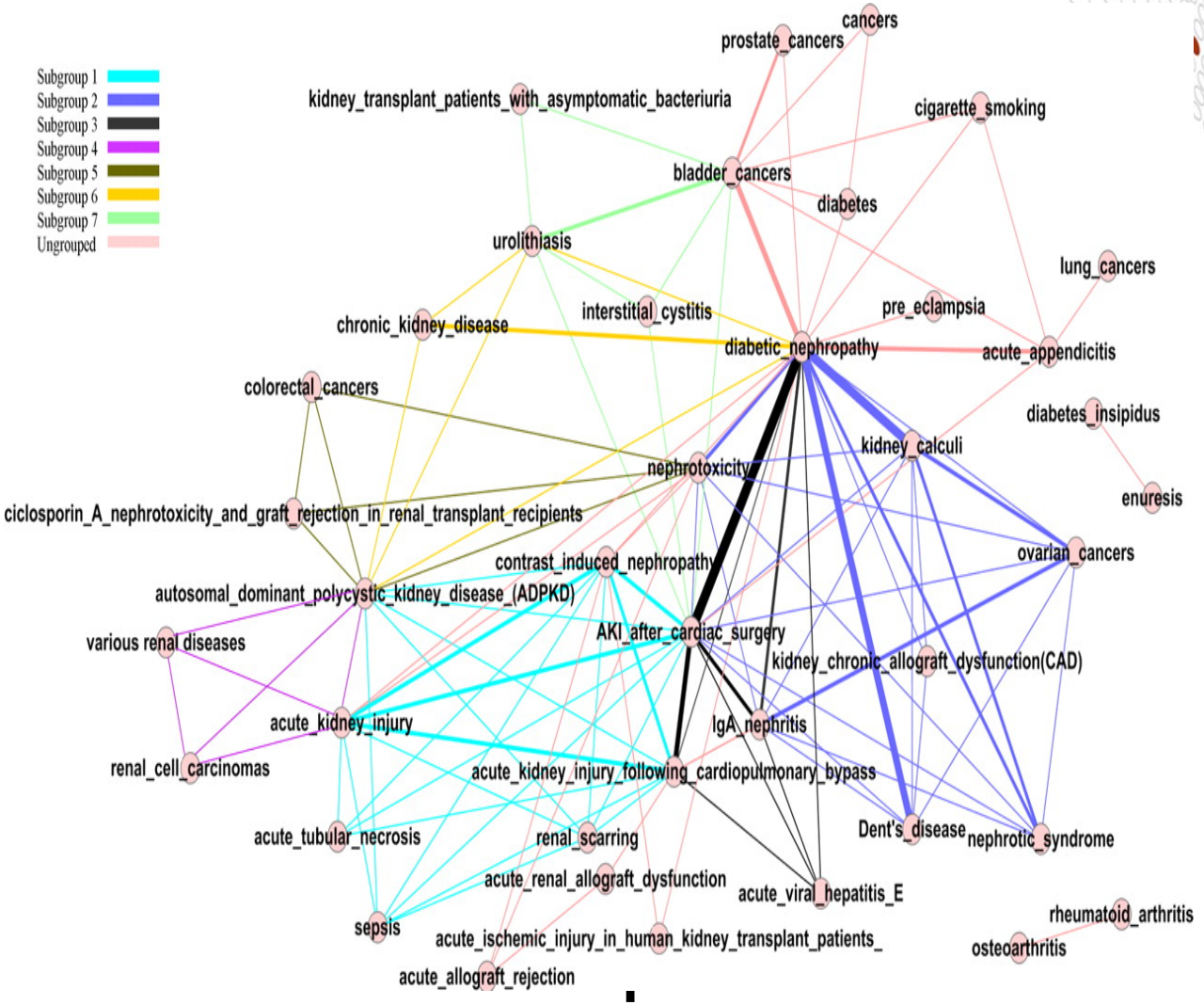




Extending Adaptive Supercomputing to Big Data Workloads

Big Data Appliance for Real-Time Data Discovery

YarcData
A DIVISION OF CRAY INC.

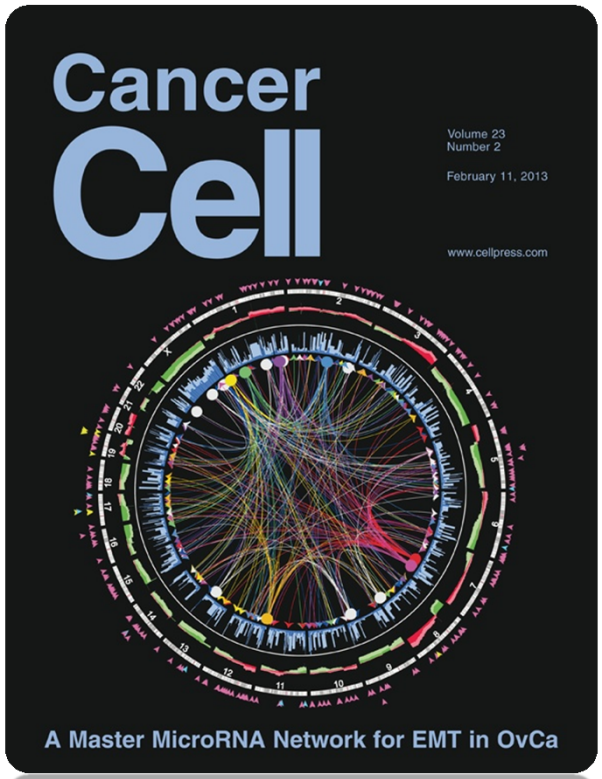


Discover new drug re-purposing opportunities

Drug Discovery:
Finding New Treatments for Cancer

Goal	• Use a Systems Biology approach to discover new Cancer treatments by understanding cancer formation & growth at a molecular level
Data sets	• TCGA (The Cancer Genome Atlas), DrugBank, Pathway Commons, Human Protein Reference Database, Domine, …
Technical Challenges	• Volume and velocity of data; probabilistic and temporal inter-relationships; extremely complex, ad-hoc queries frequently requiring the addition of new data sources
Users	• Research Scientists
Usage model	• In-silico analysis to identify biological pathways leading to formation and growth of cancer at a molecular level; identify compounds known to inhibit those mechanisms as possible drug candidates
Augmenting	• Existing data warehouse, Statistical Analytics

INSTITUTE FOR Systems Biology



Cycle of Discovery in Cancer Research…

Load or Augment Required Data from many diverse data sources

DrugBank Open Data Drug & Drug Target Database · FDA · Pathway Commons · Human Protein Reference Database · DOMINE Database of Protein Domain Interactions · TCGA · Others …

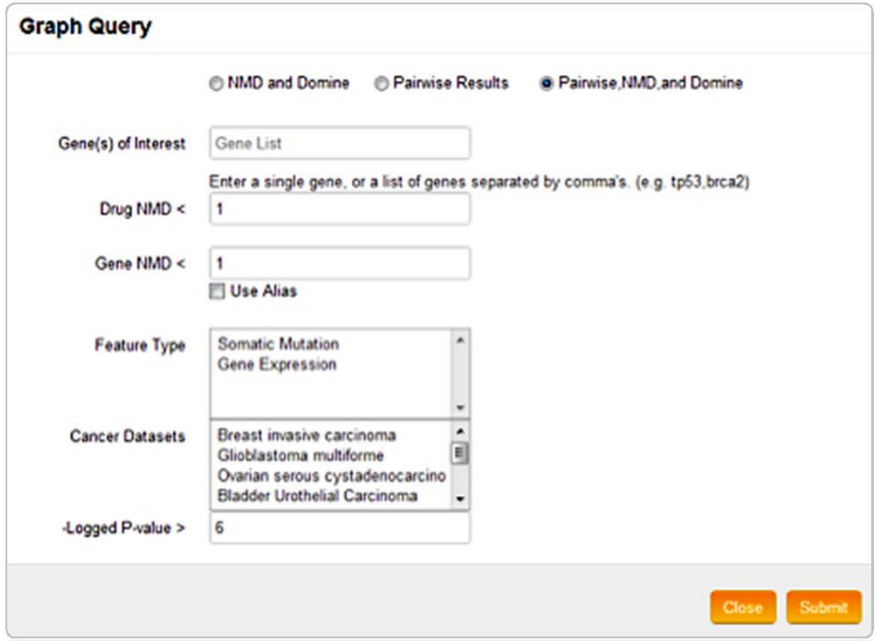
Validate Hypothesis against entire dataset – 5.3 billion relationships

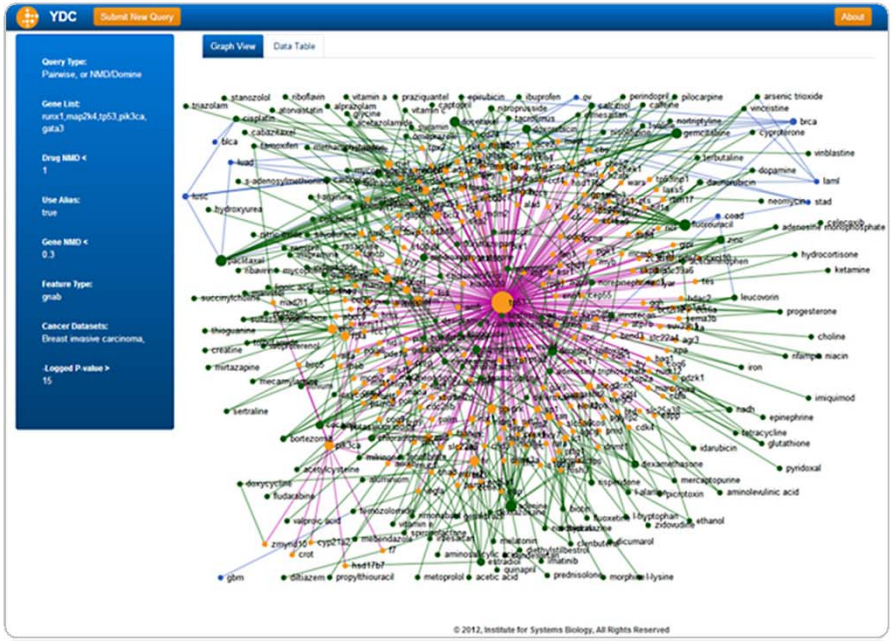
Formulate Partial Hypothesis – Statistical analysis used to assist

INSTITUTE FOR Systems Biology

"In the amount of time it takes to validate one hypothesis, we can now validate 1,000 hypotheses – increasing our success rate significantly."

Flexibility in Workflow…





Hypotheses can be composed in SPARQL – or using a Web based query composer GUI

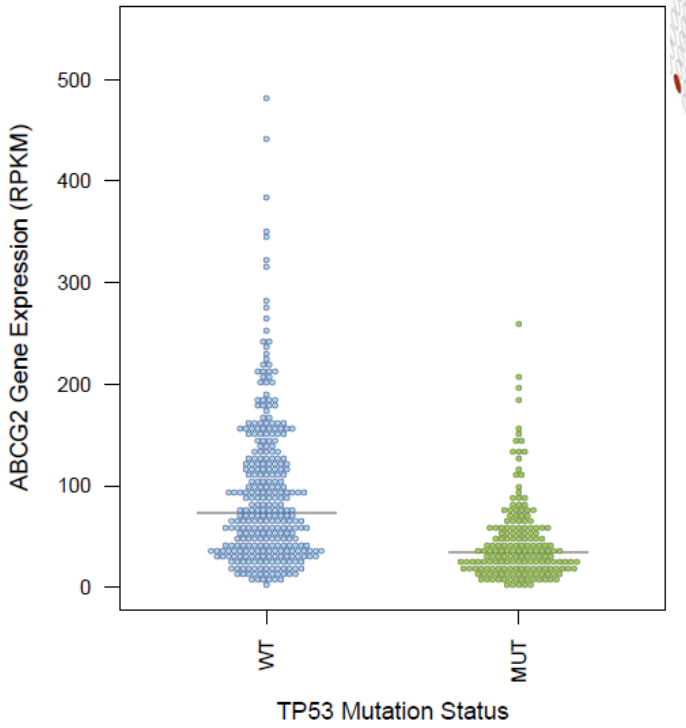
Results can be viewed in a number of ways…

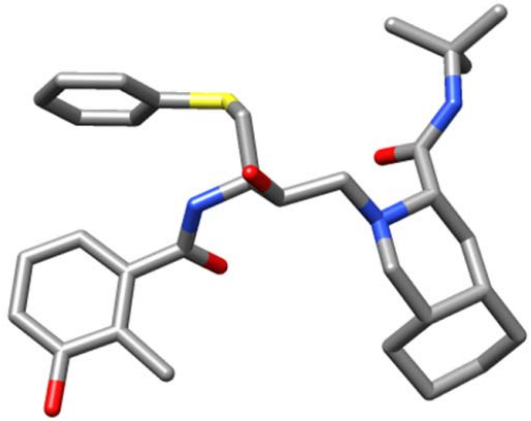
An “Urika” Moment…

Repurpose HIV drug for Cancer

- TP53 is frequently mutated in most tumor types
- ABCG2, also known as Breast Cancer Resistance Protein (BCRP), is associated with TP53 mutation in TCGA breast cancer data
- Nelfinavir, an HIV protease inhibitor, also binds ABCG2 and many other proteins
- High-throughput cell line screening of breast cancer cells recently identified Nelfinavir as a selective inhibitor. “It can be brought to HER2-breast cancer treatment trials with the same dosage regimen as that used among HIV patients. “ [Shim *et al.* JNCI 2012]

They discovered this in just 6 weeks!





Purpose-Built, Turnkey Hadoop Solutions

Best Hadoop Distribution

- **Security**– Comprehensive, and fast, encryption
- **Performance** – Faster Hive, Cache acceleration, etc.
- **Management** – Intel Manager for Hadoop Software

Performance of a Cray

- **Proven HPC** – Cray technology & expertise
- **Vast Scale** – Grow to meet any mission requirements
- **Holistic Design** – Balanced compute, networking & storage

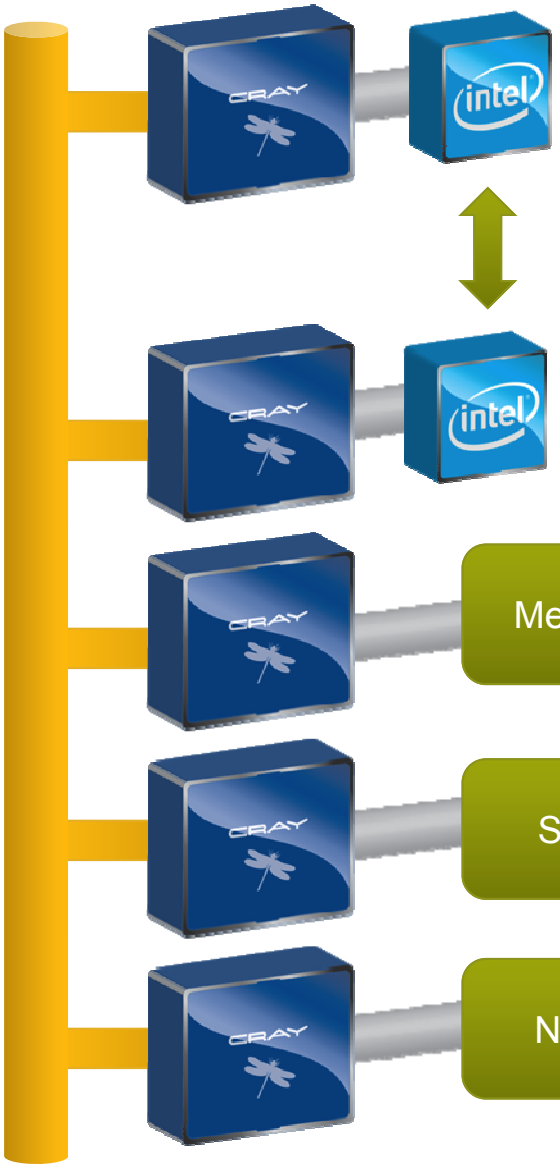
Turnkey Solution

- **Reliable** – Rapid ROI… runs as-advertised
- **Support** – One throat to choke, for the whole stack
- **Maintenance** – Update & evolve, without concerns

High Value Hadoop

- **Performance** – Power to accommodate current & future goals
- **Reliability** – Will meet any challenge, without surprises
- **Maintenance** – Easy to maintain & accommodate change

Adapting to Data-Intensive Computing: Adding Value at the Edge of the Network



Data Processors
- x86
- Many-core
- GPU
- Multi-threaded

Adaptive Supercomputing
- Compilers, Auto-tuning

Operating System
High throughput scheduling, Adaptive Runtime, Network

The most capable interconnects will be key to analytic workloads

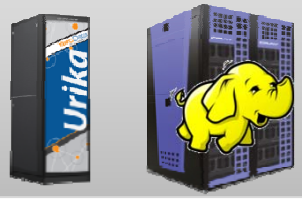
High Performance Memory
- DRAM
- Active NVRAM
- HMC

Storage
- SAS to PCIe
- Tightly coupled Software Stack for RAID
- Very Low Latency

Intelligent I/O
- Protocol Load Balancing
- Specialized Data Ingest
- Policy Based Data Movement Software Defined Network

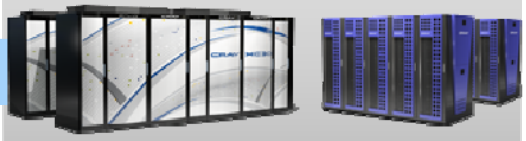
Integrated HPC Environments are the capability that will turn data in to insight and discovery

Advanced Analytics Appliances



Storage & Data Management

Supercomputers



Cray's Roadmap "Fusion"

Data Ingest (Cloud) ***Simulation*** ***Analytics***

Multiple Roadmap Steps:

Combine workflows & data into a single system

Aggressive local and global memory capabilities

Tightly integrated software stack (& runtimes) across all 3 capabilities


Our Vision…
Build a world-class integrated supercomputing environment that enables transformational computing across a broad set of science, engineering and advanced analytics (big data) applications


FUTURE AHEAD
No matter what
the future holds…


CRAY
THE SUPERCOMPUTER COMPANY