COMPUTE | STORE | ANALYZE

Supercomputing Trends in Earth System Modelling

ECMWF Workshop on HPC in Meteorology
26 October 2016

Dr. Phil Brown
Earth Sciences Segment Leader

Overview

- **Cray Update**
- **Early KNL results with Earth System Models**
- **Thoughts future trends**
  - System Reliability@Exascale
  - Deep Learning in Weather/Climate
  - Converged Architectures

Cray Solutions for the Earth Sciences

Why Cray ?

- Cray's solutions enable a broader and more detailed range of meteorological services and products
    - Advanced modeling capabilities
    - Shortened research to operations
- Experience delivering and operating world's largest and most complex systems
- Emphasis on total cost of ownership – power, upgradability and efficiency
- Commitment to long-term partnerships delivering significant ongoing value to our customers.

Market Presence

- Broad presence across NWP and climate communities:
    - From Terascale to Petascale
    - Research and operational environments
    - Model development platforms for extreme scale architectures

COMPUTE | STORE | ANALYZE

Cray Growth in Weather, Climate and Oceanography

CRAY

ECMWF

DWD

NAVY DSRC

Met Office

Meteorological Service Singapore
Centre for Climate Research Singapore

Naval Oceanographic Office

KMA Korea Meteorological Administration

South African Weather Service
ISO 9001 Certified Organisation

NOAA

NOAA

Finnish Meteorological Institute

noaaresearch

National Severe Storms Laboratory

INPE

CPTEC

NOAA
NCEP

DMI

AWI

DWD

Jan 2015

July 2015

Aug 2015

Sept 2015

Nov 2015

Dec 2015

Mar 2016

Australian Government
Bureau of Meteorology

CSCS
Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

ECMWF

Cray Growth in Weather, Climate and Oceanography

Since last workshop:

- >89,000 sockets shipped
  - >53 PetaFlops nominal peak
- 89PB of Cray Sonexion
  - >3TB/s nominal IO BW

Nov 2015

Dec 2015

Mar 2016

Two Large XC Systems that will Impact Future Technologies and Applications Throughout the Community

- **Los Alamos / Sandia – “Trinity”**
  - >40 Pflop system, mix of Haswell + KNL
  - 3TB/s / 3PB SSD DataWarp Capability

- **NERSC8 – “Cori”**
  - >40 Pflop system, mix of Haswell + KNL
  - 3TB/s / 3PB SSD DataWarp Capability
  - Transitioning user base to “many-core” processing
  - NERSC Exascale Science Applications Program (NESAP)

HOMME – KNL vs Broadwell

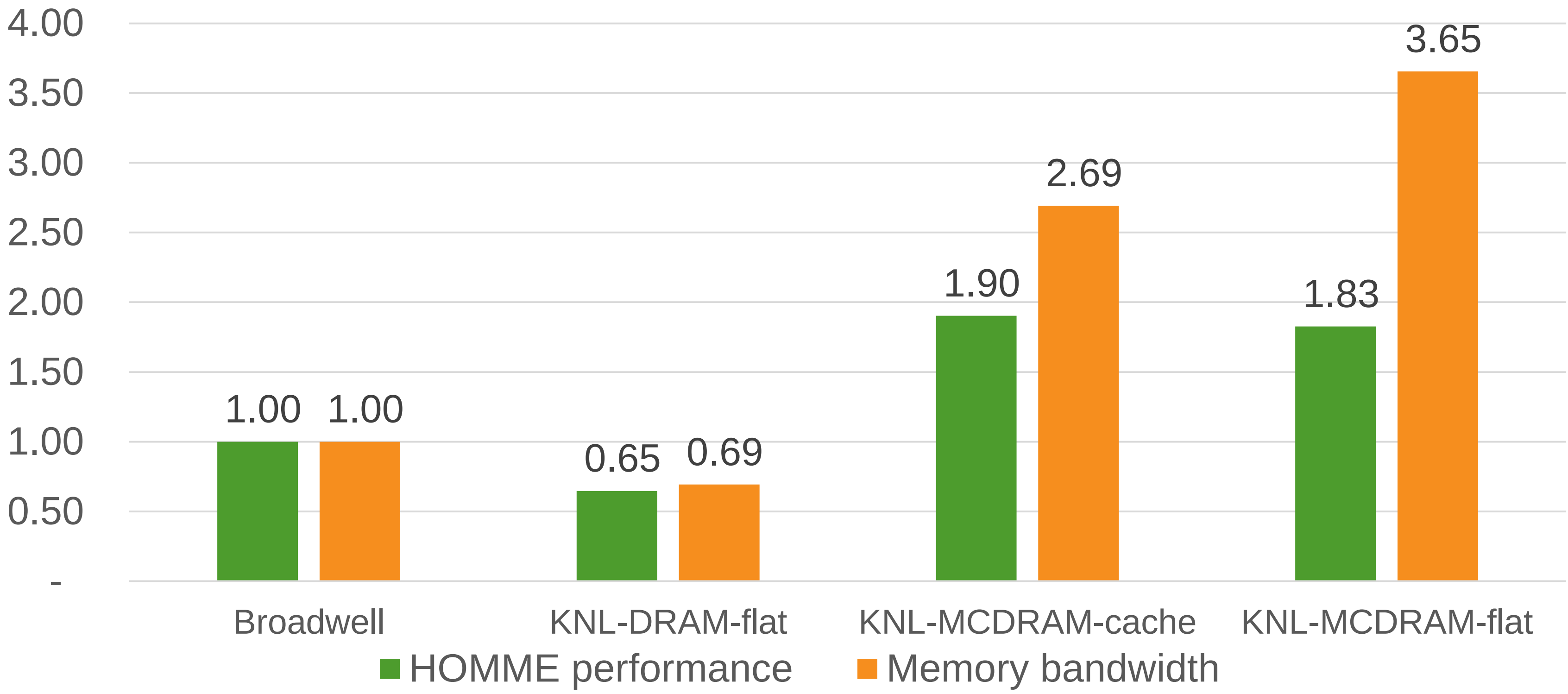
- **Spectral element dynamical core within CAM component of CESM**
- **Performance dominated by advection (2^{nd} order Runge-Kutta), memory bandwidth limited**
- **Key optimizations targeted at KNL implemented by NCAR ASAP group**
  - See John Dennis' talk at the 6^{th} UCAR MultiCore workshop

- **"perfTESTWACCM" benchmark: Baroclinic wave in N. hemisphere**
  - Size NE=8, 70 vertical levels, 135 tracers
  - Runtimes from "prim_main_loop"

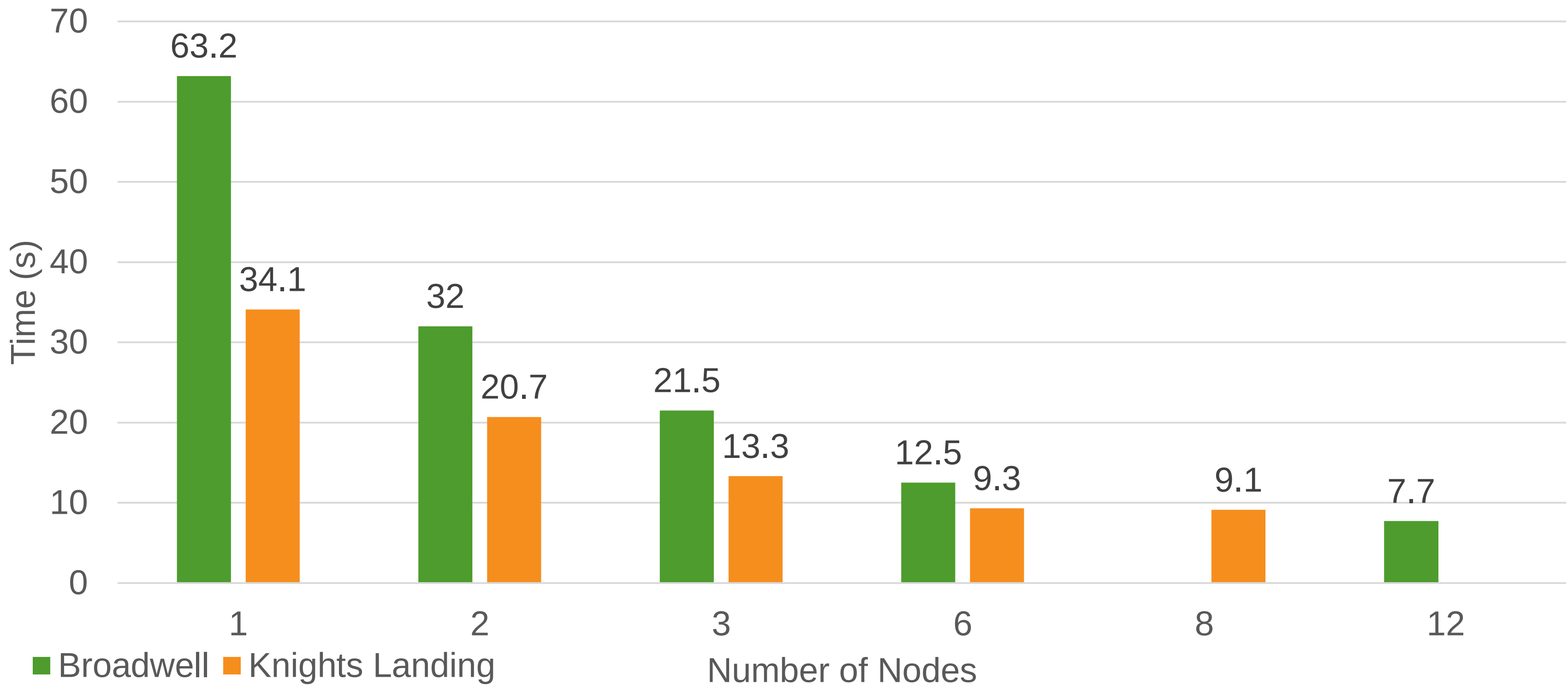
- **Node: Intel Xeon Phi 7250 68-core 1.4GHz, 96GB DDR4-2400, 16GB MCDRAM**
  - MCDRAM bandwidth (quad-flat): 475-490GB/s
  - MCDRAM bandwidth (cache): ~350GB/s
  - DDR4 bandwidth (quad-flat): ~90GB/s
- **Node: 2 x Intel Xeon Broadwell E5-2699 22-core 2.2GHz, 128GB DDR4-2400**
  - DDR4 bandwidth = ~130GB/s
- **Cray XC, with Cray compiler & programming environment**

- **Strong scaling study performed by Marcus Wagner, supported by NERSC CoE**

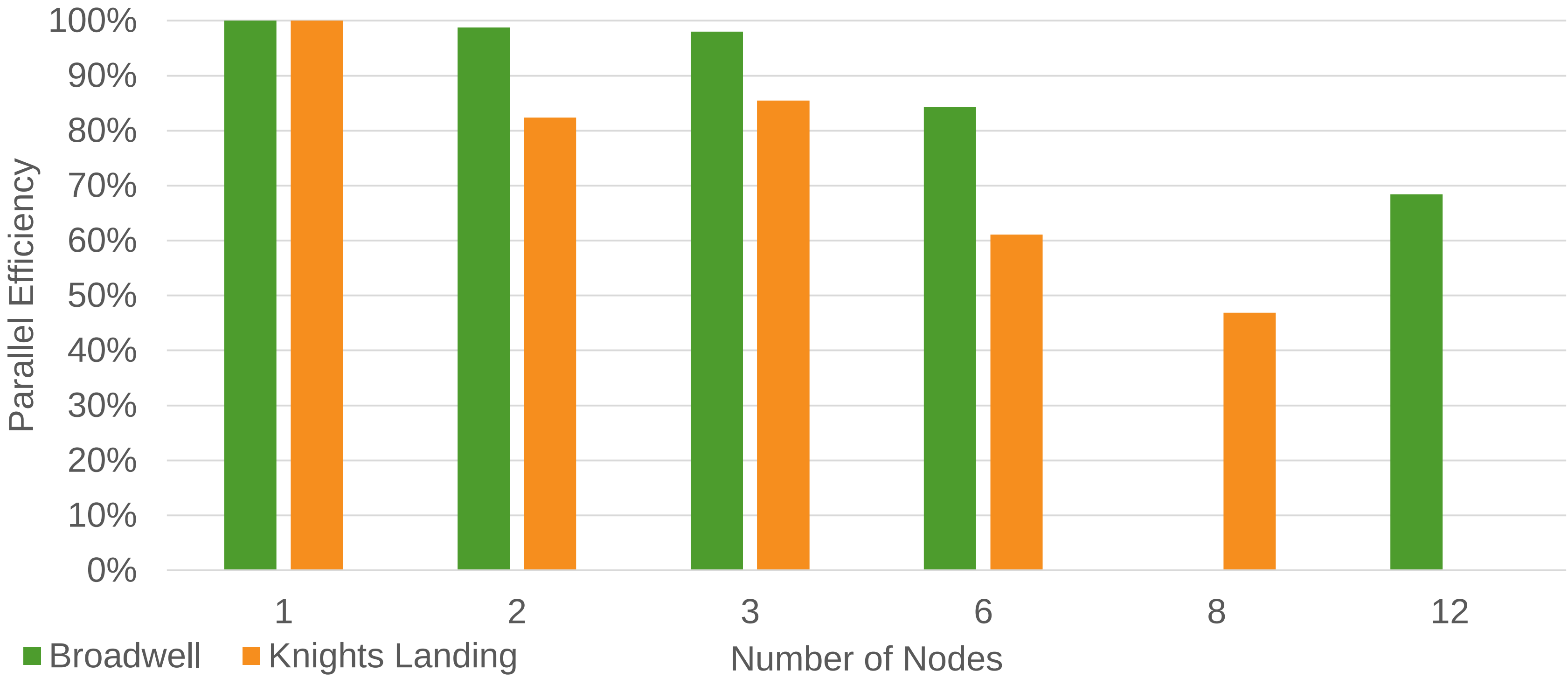
Single Node Results



Multi Node Results - runtime



Multi Node Results – parallel efficiency



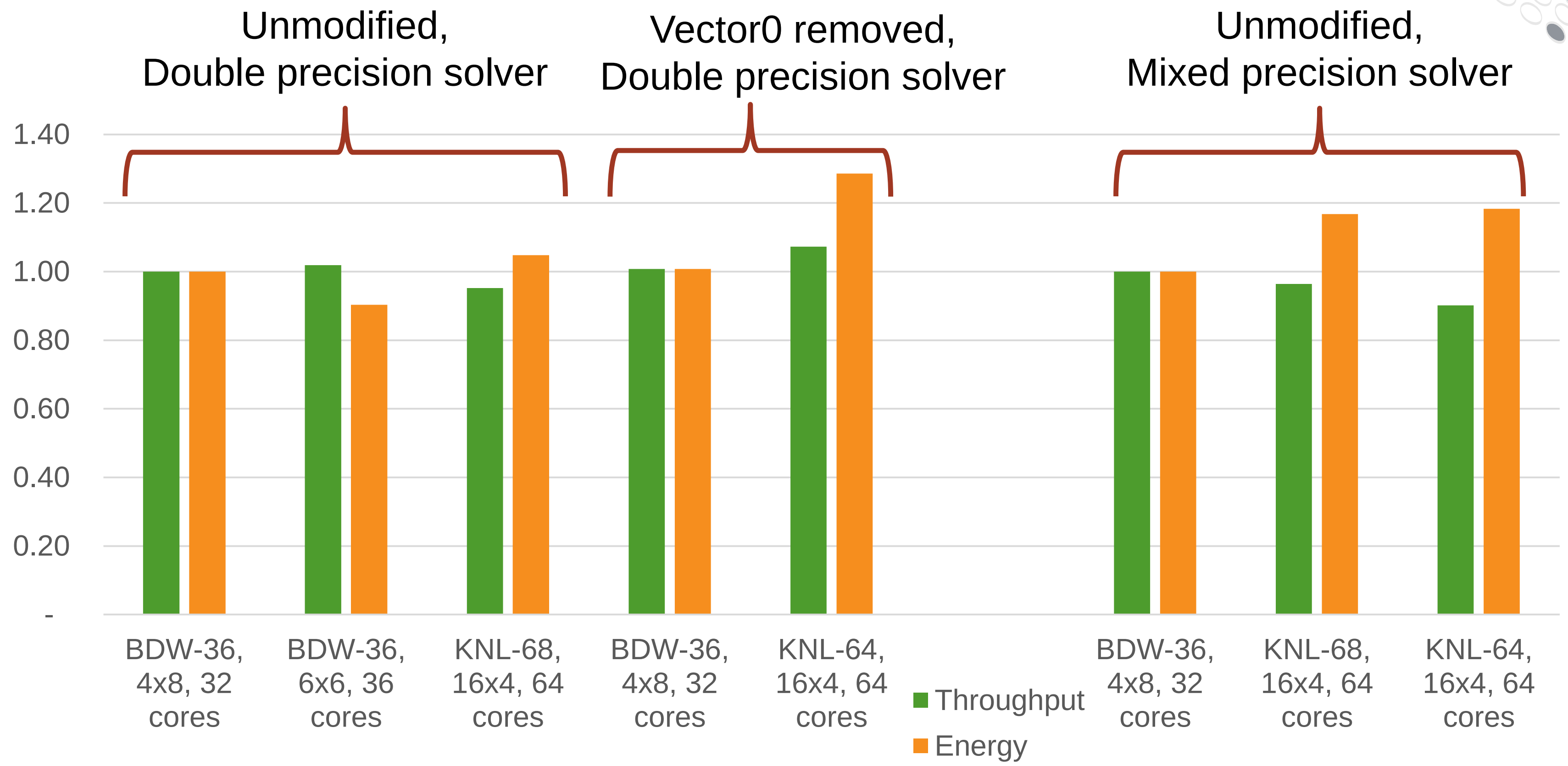
Unified Model – KNL vs Broadwell

- **Unified Model v10.3 / AMIP test case**
  - N96 (135km) Global Atmosphere
  - Memory footprint approximately 26GB
- **Test platforms:**
  - Node: Intel Xeon Phi 7230 64-core 1.3GHz, 96GB DDR4-2400, 16GB MCDRAM
    - MCDRAM bandwidth (cache): ~350GB/s
    - DDR4 bandwidth (cache): ~70GB/s
  - Node: Intel Xeon Phi 7250 68-core 1.4GHz, 96GB DDR4-2400, 16GB MCDRAM
    - MCDRAM bandwidth (cache): ~350GB/s
    - DDR4 bandwidth (cache): ~70GB/s
  - Node: 2 x Intel Xeon Broadwell E5-2695 18-core 2.1GHz, 128GB DDR4-2400
    - DDR4 bandwidth = ~130GB/s
- **Cray XC, with Cray compiler & programming environment**
- **Investigation conducted by Eckhard Tschirschnitz**

Unified Model – KNL vs Broadwell

- **Three cases tested**
  1. Unmodified code, compiled to target KNL/BDW, double precision solver
  2. Unmodified code, compiled to target KNL/BDW removing all “vector0” flags, double precision solver
     - Vector0 disables automatic vectorization
     - Used to avoid some numerical issues observed previously in UM with higher vectorization
     - No issues apparent in this particular test case
  3. Unmodified code, compiled to target KNL/BDW, mixed precision solver
     - Inspired by early KNC work, higher cache efficiency

UM - Throughput & Energy Efficiency



Thoughts on early KNL results

1. **Potential for greater throughput/node**
   - Significant optimization may be required to maximize potential
   - Improving vectorization, threading & cache re-use

2. **Significant potential for higher energy efficiency**

3. **Parallel efficiency reduces faster with scale**
   - Not unexpected due to lower per-core performance
   - May prove problematic where jobs have strict runtime targets which challenge Xeon today

Reliability in the coming years

- **Considering a contemporary 3000 node system**
  - Very high whole system MTTI/availabilities achievable
  - Expect job failures to occur every 5-10 days
    - Caused by uncorrectable soft and hard errors in memory and/or CPUs
    - Scales with number of devices/sockets in system

- **Expect whole-system reliability to remain high, but rate of job failures to increase**

Impact on Weather/Climate models

- **Not ideal for deterministic models**
  - Requires re-run from start or saved state
- **Losing an ensemble member not so impactful**
  - Close coupling of ensembles into single MPI launch not desirable until MPI-resiliency features exist
- **Overall impact modest?**
  - **Providing workflows are engineered to expect & react to failures**

Machine/Deep Learning in Weather/Climate

- **Deep Learning used to describe a family of algorithms related to multi-level neural networks:**
  - Deep Neural Networks
  - Convolutional Neural Networks
  - Recurrent Neural Networks
  - Lots more!

- **Key enabler has been access to compute resources**
  - DL is predominantly FLOP bound
  - Large scale problems rapidly becoming “HPC”-class

- **Delivering “state of the art” results in computer vision, speech recognition, natural language processing etc.**

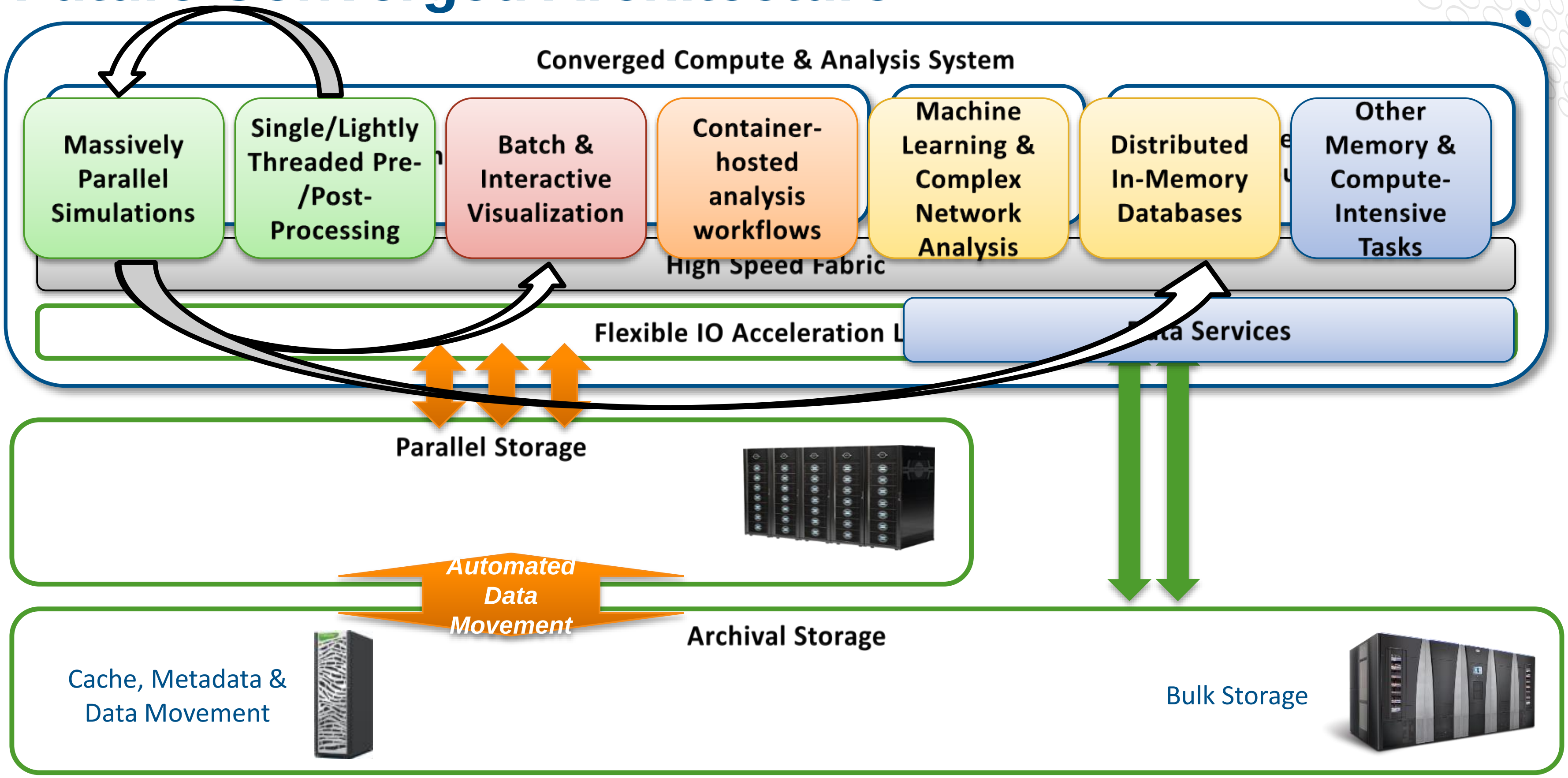
Machine/Deep Learning in Weather/Climate

- **Almost the opposite of a physics/dynamics based model**
  - Arduous to train, but comparatively quick to run

- **Use-cases will be complementary?**

- **Some ideas:**
  - Rapid classifiers for radar/observations
  - Pattern recognition in model outputs
  - Infilling/smoothing model outputs

Future Converged Architecture



Thank you for your attention

CRAY®

THE SUPERCOMPUTER COMPANY