Hartree Centre
Science & Technology Facilities Council

EMiT Conference – University of Manchester
June 30th 2015

Energy Aware Scheduling on Blue Wonder

Neil Morgan (Hartree), Vadim Elisseev (IBM), Francois Thomas (IBM), Manish Modani (IBM), Terry Hewitt (STFC) , Luigi Brochard (Lenovo)

lenovo. FOR THOSE WHO DO.

IBM

Challenges on the Path to Exascale

- **Total energy consumption & electricity costs of HPC architectures**
- New architectures and their programmability
- Software environments for new architectures
- Optimized numerical libraries for new architectures
- MTBF of hyper scale systems
- etc….

Energy Efficient Computing Areas of Focus

Energy Efficient Hardware

- Latest semiconductor technology
- Energy saving processor & memory technologies
- Use special hardware or accelerators designed for specific scientific problems or numerical algorithms

Energy Aware Management Software

- Monitor the energy consumption of the **compute system** and the **building infrastructure**
- Use energy aware system software to exploit the energy saving features of the platform

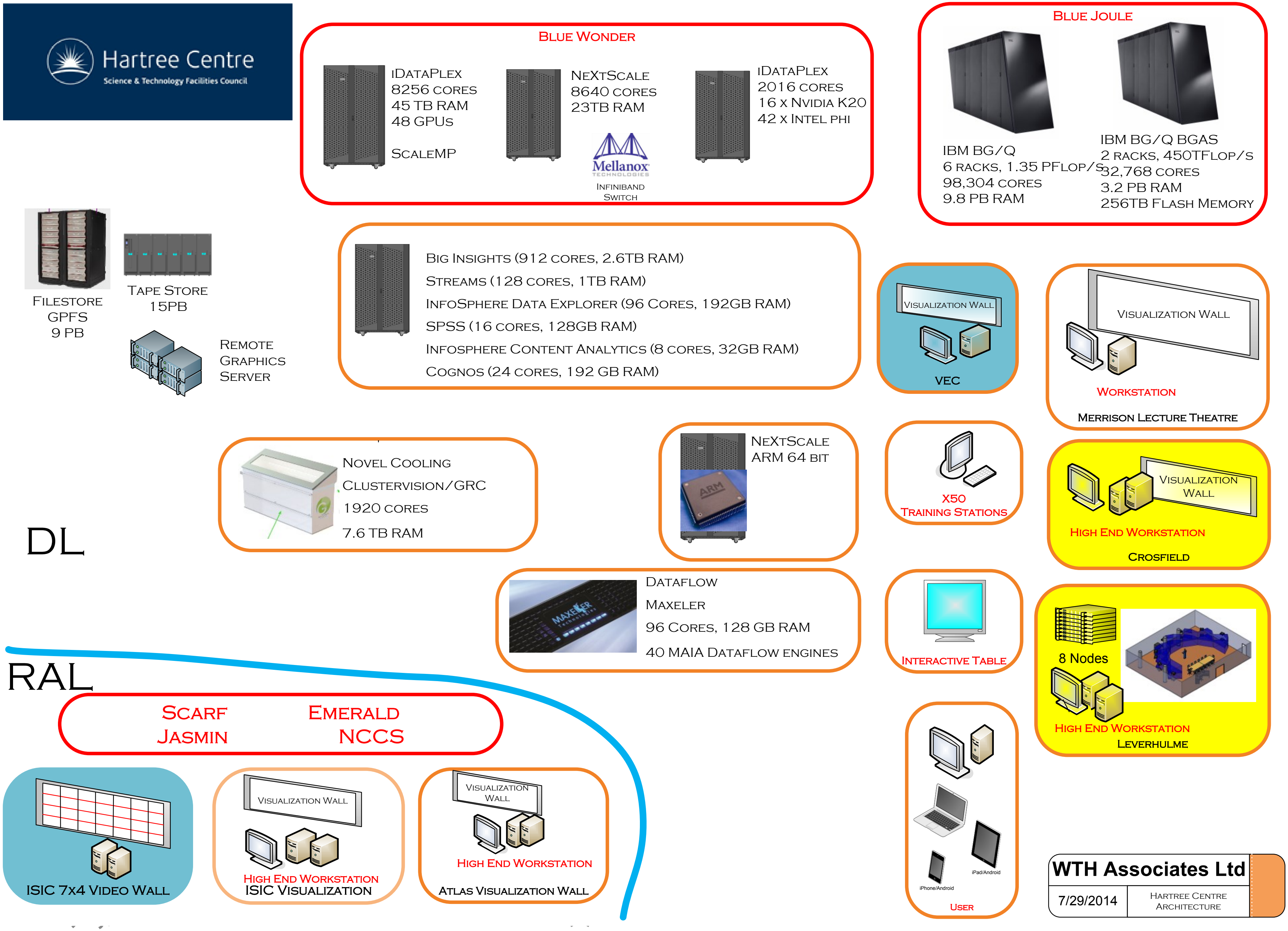
Energy Efficient Infrastructure

- Reduce power loss in the power supply chain
- Improve cooling technology
- Reuse waste heat from systems

Energy Efficient Applications

- Use the most efficient algorithms
- Use best libraries
- Use most efficient programming paradigm

Hartree Centre
Science & Technology Facilities Council



Blue Wonder

iDataPlex
8256 cores
45 TB RAM
48 GPUs

ScaleMP

NeXtScale
8640 cores
23TB RAM

Mellanox Technologies
Infiniband Switch

iDataPlex
2016 cores
16 x Nvidia K20
42 x Intel Phi

Blue Joule

IBM BG/Q
6 racks, 1.35 PFlop/s
98,304 cores
9.8 PB RAM

IBM BG/Q BGAS
2 racks, 450TFlop/s
32,768 cores
3.2 PB RAM
256TB Flash Memory

Filestore
GPFS
9 PB

Tape Store
15PB

Remote
Graphics
Server

Big Insights (912 cores, 2.6TB RAM)
Streams (128 cores, 1TB RAM)
InfoSphere Data Explorer (96 Cores, 192GB RAM)
SPSS (16 cores, 128GB RAM)
Infosphere Content Analytics (8 cores, 32GB RAM)
Cognos (24 cores, 192 GB RAM)

Visualization Wall
VEC

Visualization Wall
Workstation
Merrison Lecture Theatre

Novel Cooling
Clustervision/GRC
1920 cores
7.6 TB RAM

NeXtScale
ARM 64 bit

X50
Training Stations

Visualization Wall
High End Workstation
Crosfield

DL

Dataflow
Maxeler
96 Cores, 128 GB RAM
40 MAIA Dataflow engines

Interactive Table

8 Nodes
High End Workstation
Leverhulme

RAL

Scarf
Jasmin

Emerald
NCCS

ISIC 7x4 Video Wall

Visualization Wall
High End Workstation
ISIC Visualization

Visualization Wall
High End Workstation
Atlas Visualization Wall

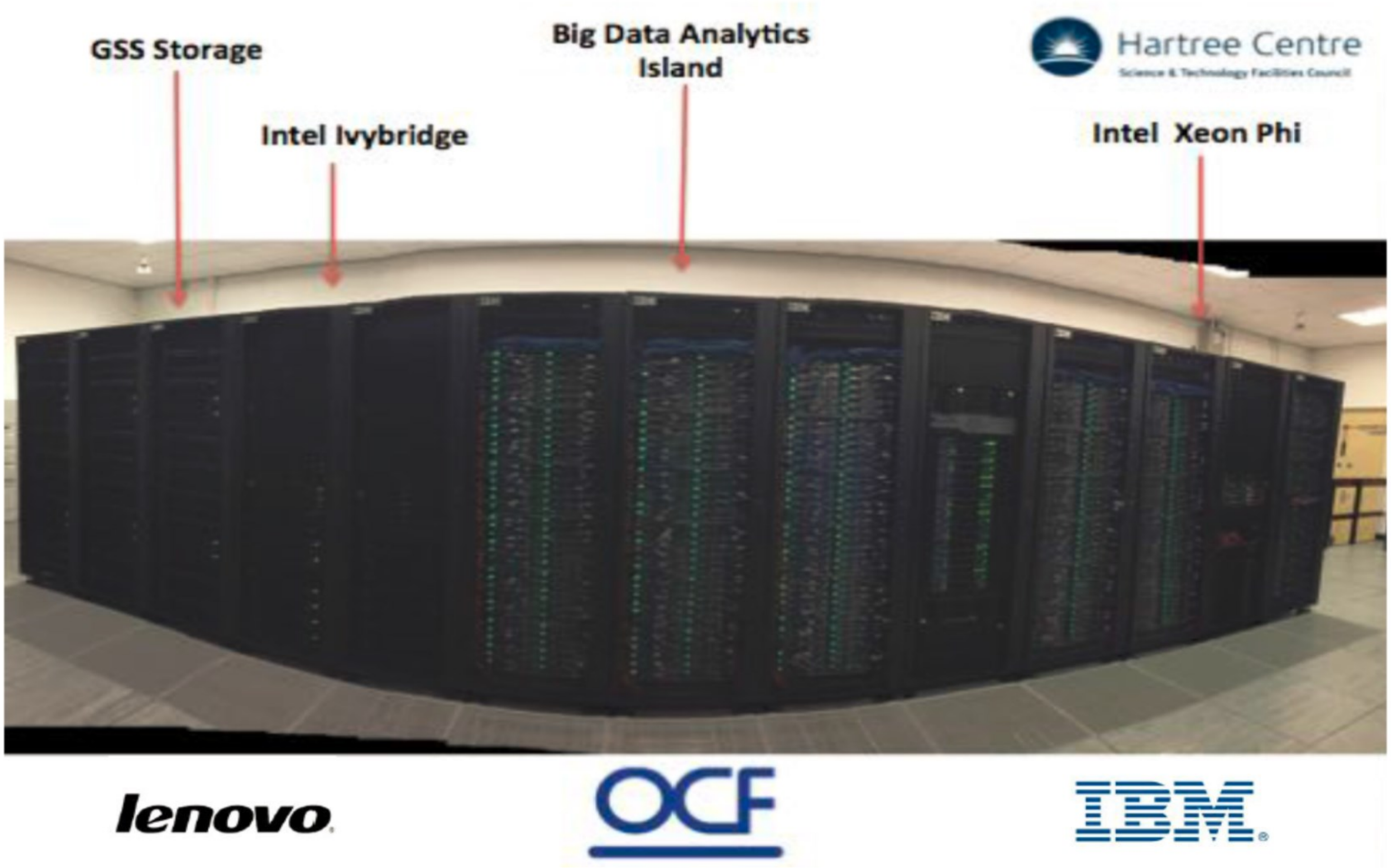
User

WTH Associates Ltd	
7/29/2014	Hartree Centre Architecture

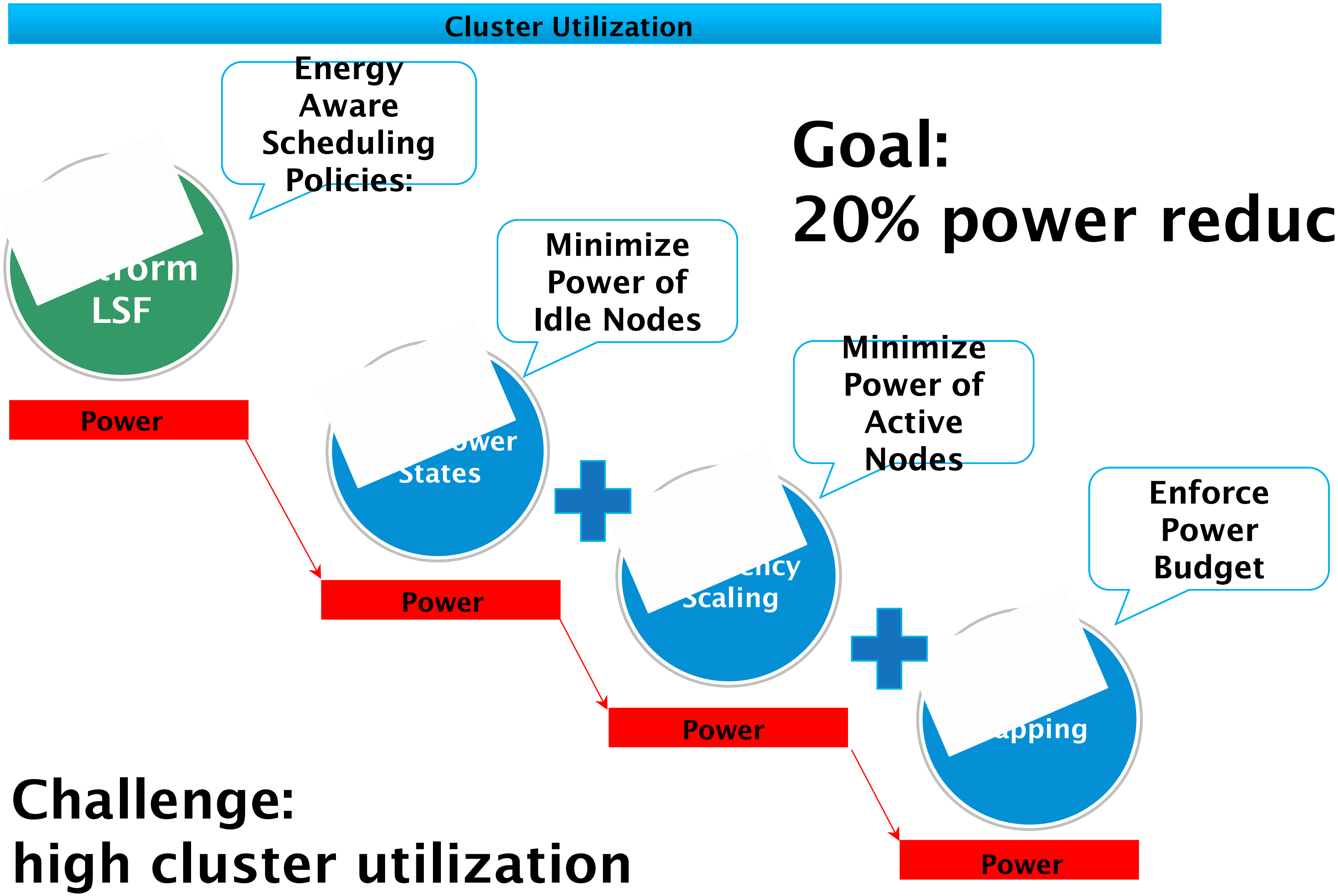
Research Goals

- Validate EAS on the Platform LSF workload management software
- Reduce the Hartree Centre’s £700K annual electricity bill by 20%
- Change users behaviour towards being more energy conscious

Hartree Centre – IBM Blue Wonder



Energy Aware Scheduling for the Hartree Centre



Goal:
20% power reduction

Challenge:
high cluster utilization

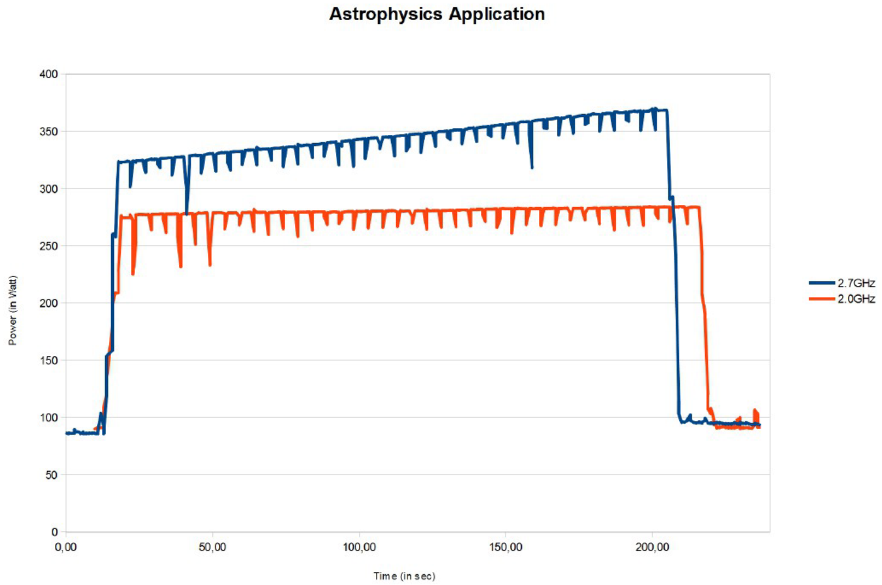
Automatic CPU Frequency Scaling

✧ Reduce Power consumption by decreasing CPU frequency while maintaining acceptable level of applications performance.

✧ We can predict application power consumption and performance at different CPU frequencies.

✧ LSF automatically adjusts CPU frequency based on application profile and EAS policies defined in the cluster:

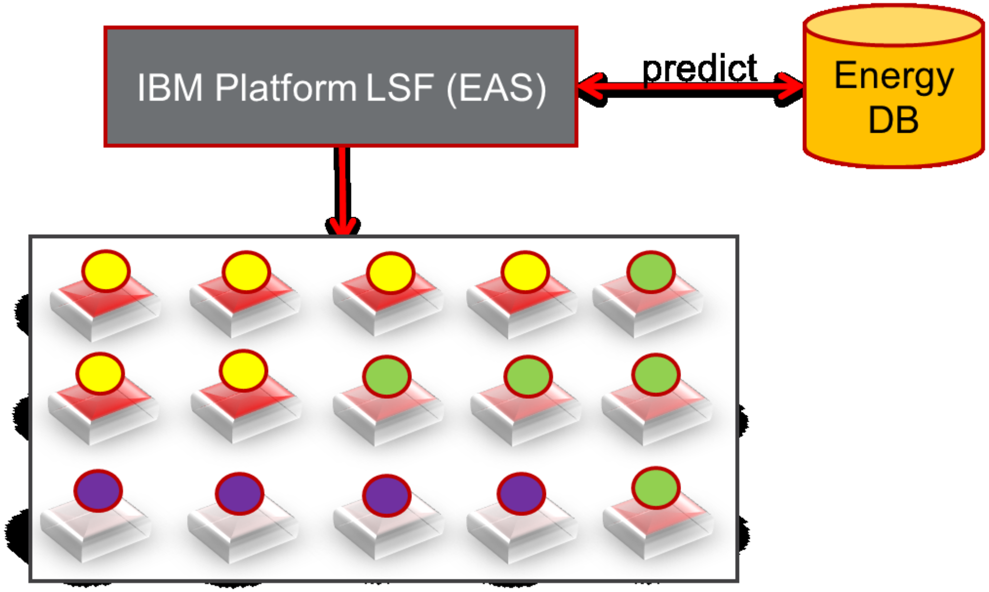
- Minimize Energy to Solution
  - Save energy while allowing maximum performance degradation of X%
- Minimize Time to Solution
  - Allow high performing applications to run at a higher CPU frequency.


Astrophysics Application

Δf=0.7 GHz
ΔPower=-17%
ΔTime=+5%
ΔEnergy=-12%

How LSF Automatically Select Optimal CPU frequency

- **Step I: Learning/Calibration**
  - **LSF evaluates the power profile of all nodes**
  - **calculates coefficients factors**
  - **save them in the energy database**

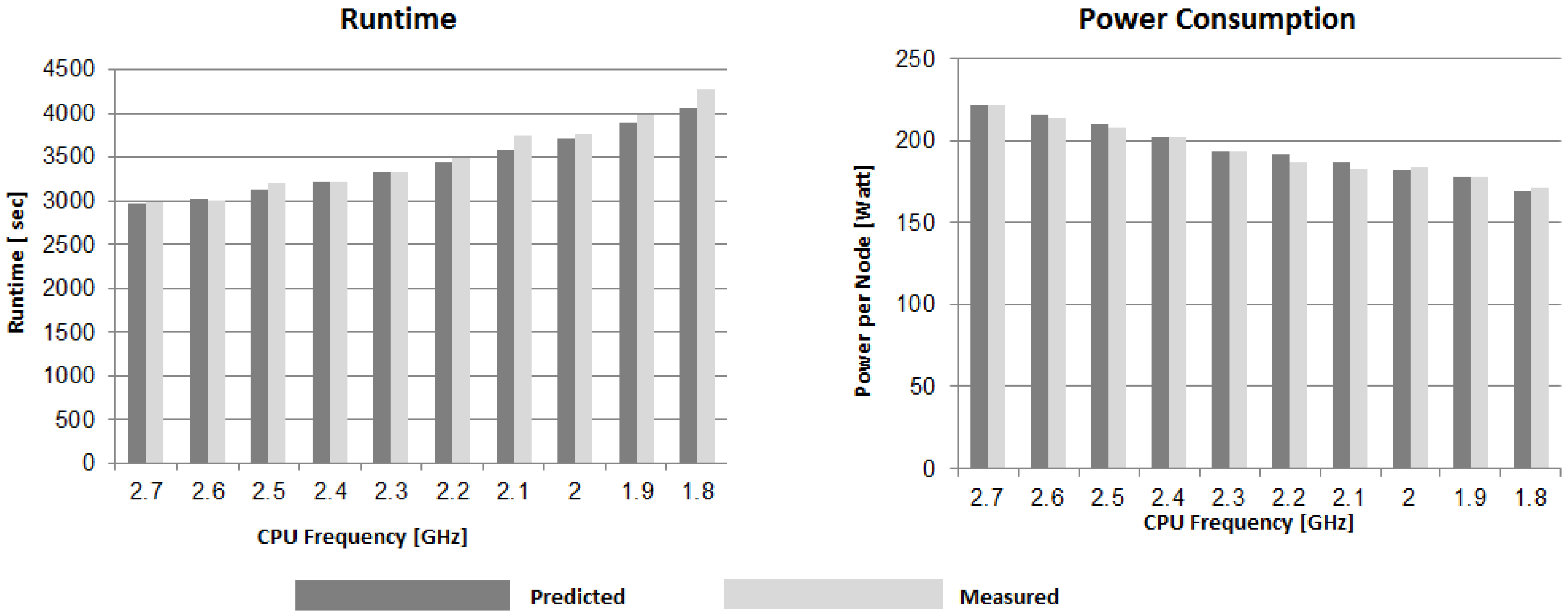


- **Step II: Set Default Frequency**
  - **System administrator defines cluster default cpu frequency (nominal or lower frequency)**

- **Step III: Tag the job first time**
  - **User submits the application with a tag**
  - **runs the job under default frequency**
  - **LSF collects energy consumption, runtime, hardware counters**
  - **Generates predication result and saves in database**

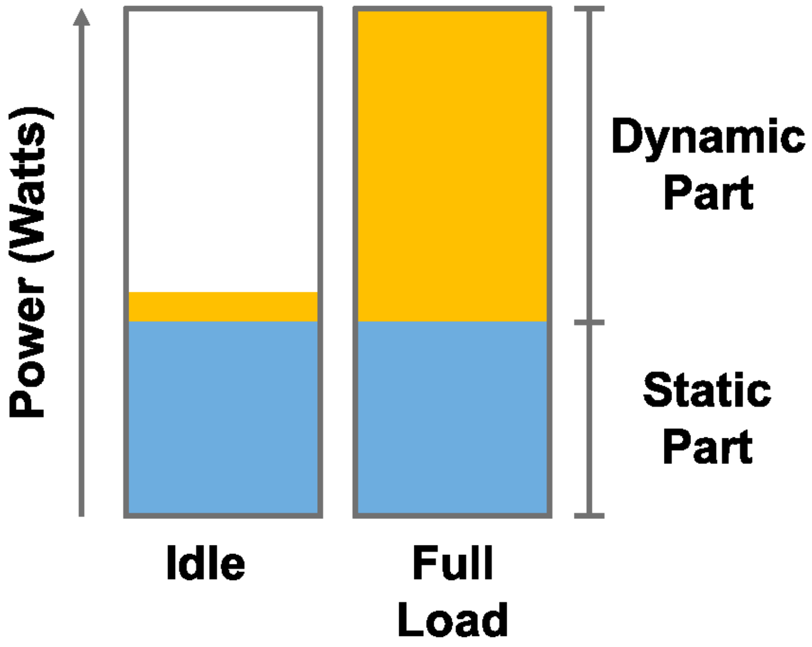
- **Step IV: Use predication**
  - **User re-submits the same application with the same tag and specifies energy policy**
  - **LSF selects the optimal cpu frequency for application based on predication result and policy setting.**
  - **Run the application under selected frequency**

Quantum Espresso Power & Runtime Prediction



Accuracy of the prediction for a 16 nodes configuration is 3.2% or better for the application runtime and 2.7% or better for the application power consumption.

Modeling Compute Node Power Consumption



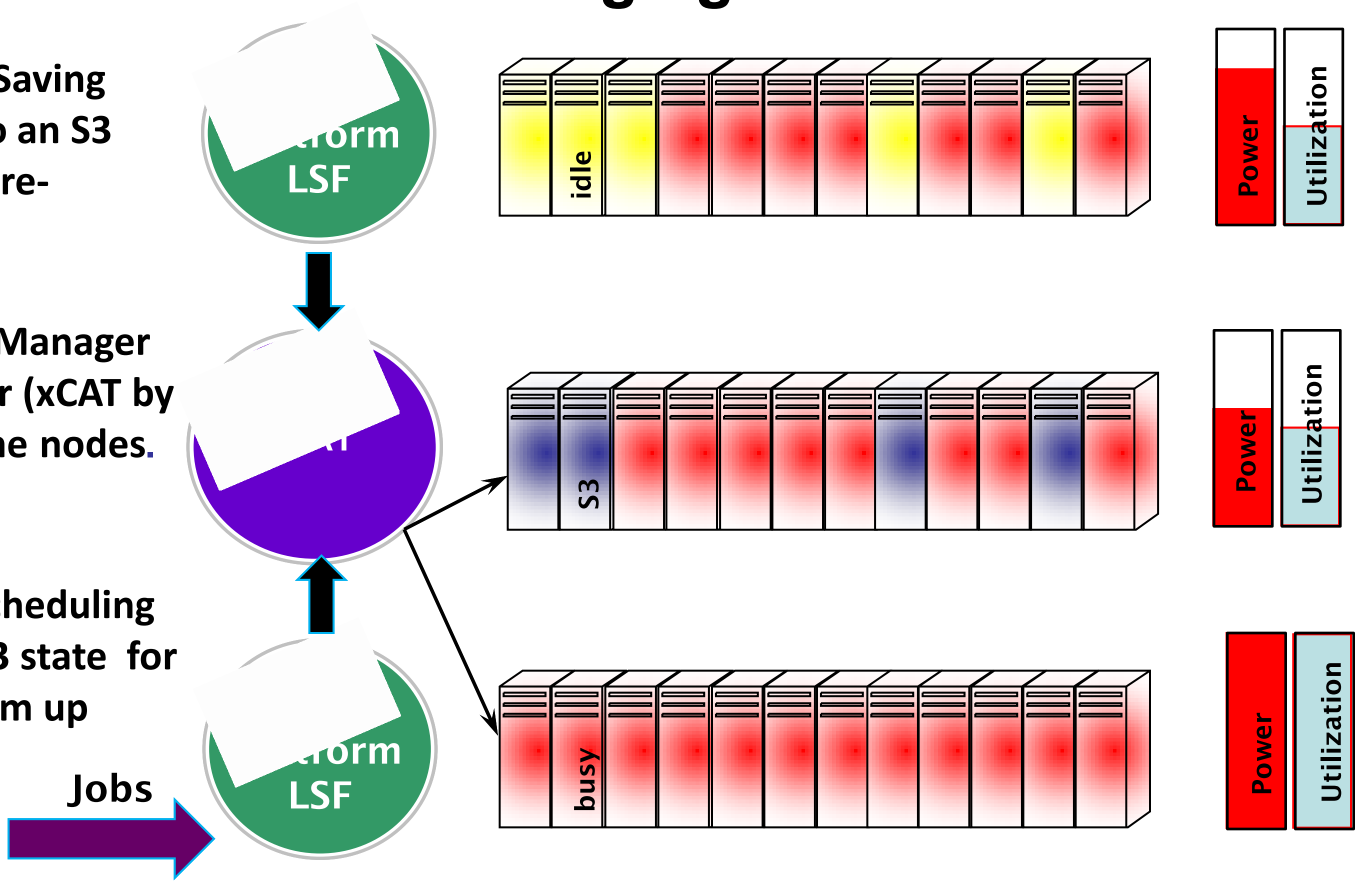
- Simplified Model for Compute Node Power
- Dynamic Part:
  - Processor
  - Memory
- Static Part:
  - South Bridge
  - NIC
  - …

Managing Host Power States

✧ **Policy Triggered Power Saving**
LSF puts idle nodes into an S3 power state following a pre-defined policy.

✧ **Integration with Cluster Manager**
LSF calls cluster manager (xCAT by default) to suspend/resume nodes.

✧ **Power Saving Aware Scheduling**
LSF considers nodes in S3 state for scheduling and wakes them up automatically if needed.

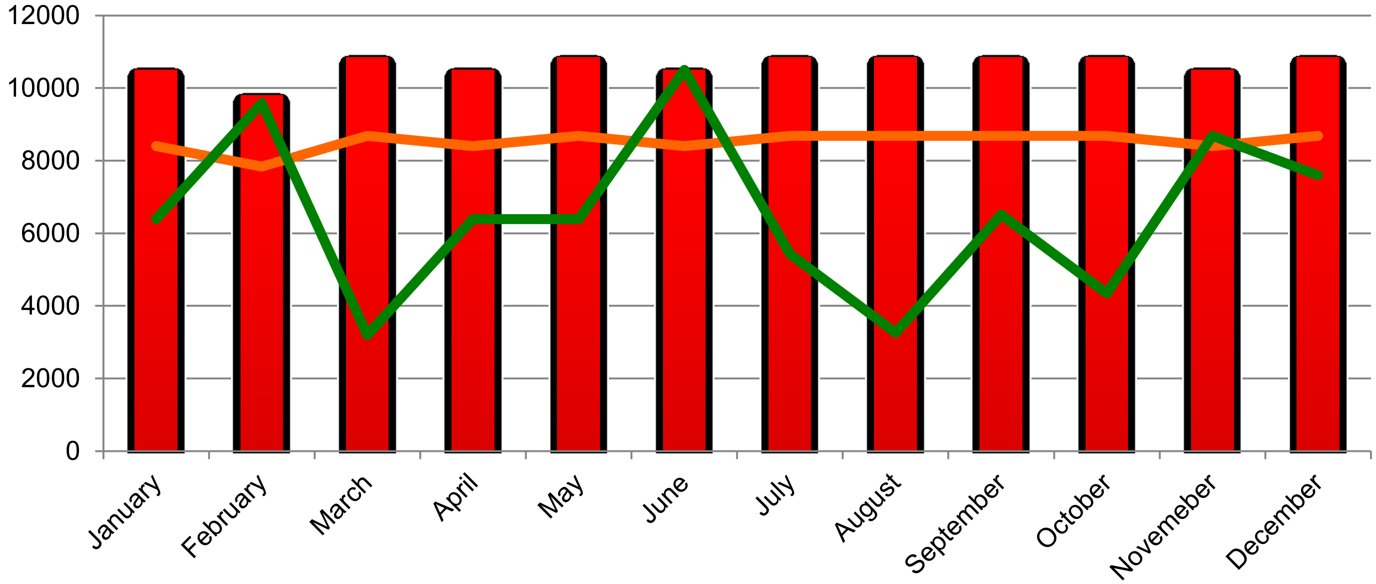


✧ **S3 state saves at least 60W per node**

✧ **We have measured up to 28% power savings from a host power management policy during lower cluster utilization periods**

Next Steps

- Test on a further set of applications
- Use LSF EAS in production
- Implement Power Capping based on predicted and measured power

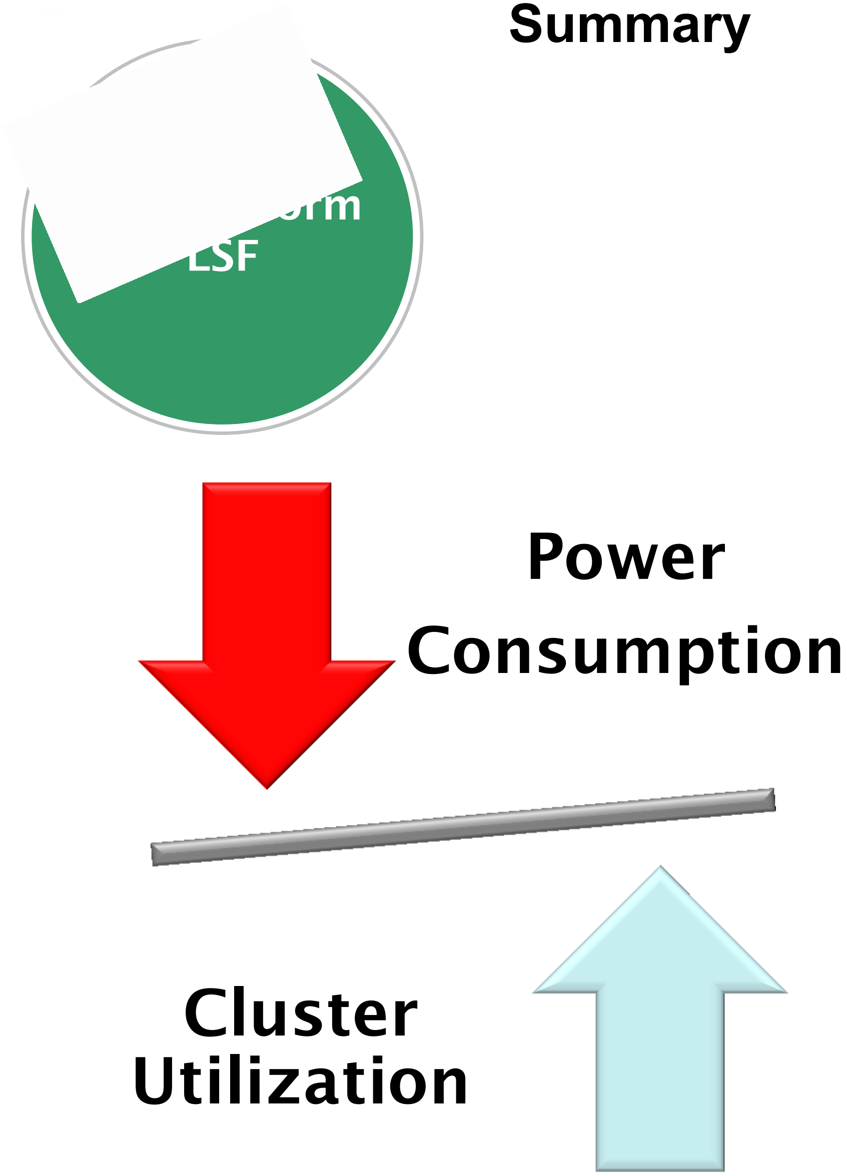


Summary

- ✧ **Energy Aware Scheduling** is a critical element of the Hartree Centre Energy Efficiency strategy

- ✧ Achieve **20%** power consumption reduction from **Energy Aware Scheduling**

- ✧ **Energy Aware Scheduling:**
  - ✧ Managing Host Power States
  - ✧ CPU Frequency Scaling
  - ✧ Power Capping



orm
LSF

Power Consumption

Cluster Utilization

Acknowledgments

- Dr. Luigi Brochard, Dr. Francois Thomas, IBM
- Dr. Axel Auweter, Prof. Herbert Hubert, LRZ

http://www.stfc.ac.uk/hartree

http://community.hartree.stfc.ac.uk

hartree@stfc.ac.uk
01925 603 444

Any Questions?

Hartree Centre
Science & Technology Facilities Council