

APRU IT Video Conference Session

Cycle Harvesting Grid for Research Computing Challenges and Opportunities

Tan Chee Chiang, Tommy Hor
Computer Centre, National University of Singapore

21 October, 2009

Overview

- Research computing requirements and trends
- How Cycle Harvesting Grid works.
- Strategic role of Cycle Harvesting Grid in Research Computing
- Application examples
- Collaboration opportunities and challenges
- Discussion



Computational Research Requirements and trends

Throughput

Many users execute many serial jobs

Shared-Memory Parallel

Mostly OpenMP jobs with large memory requirement

Distributed-Memory Parallel

Mostly MPI jobs that require more tightly coupled cluster systems.

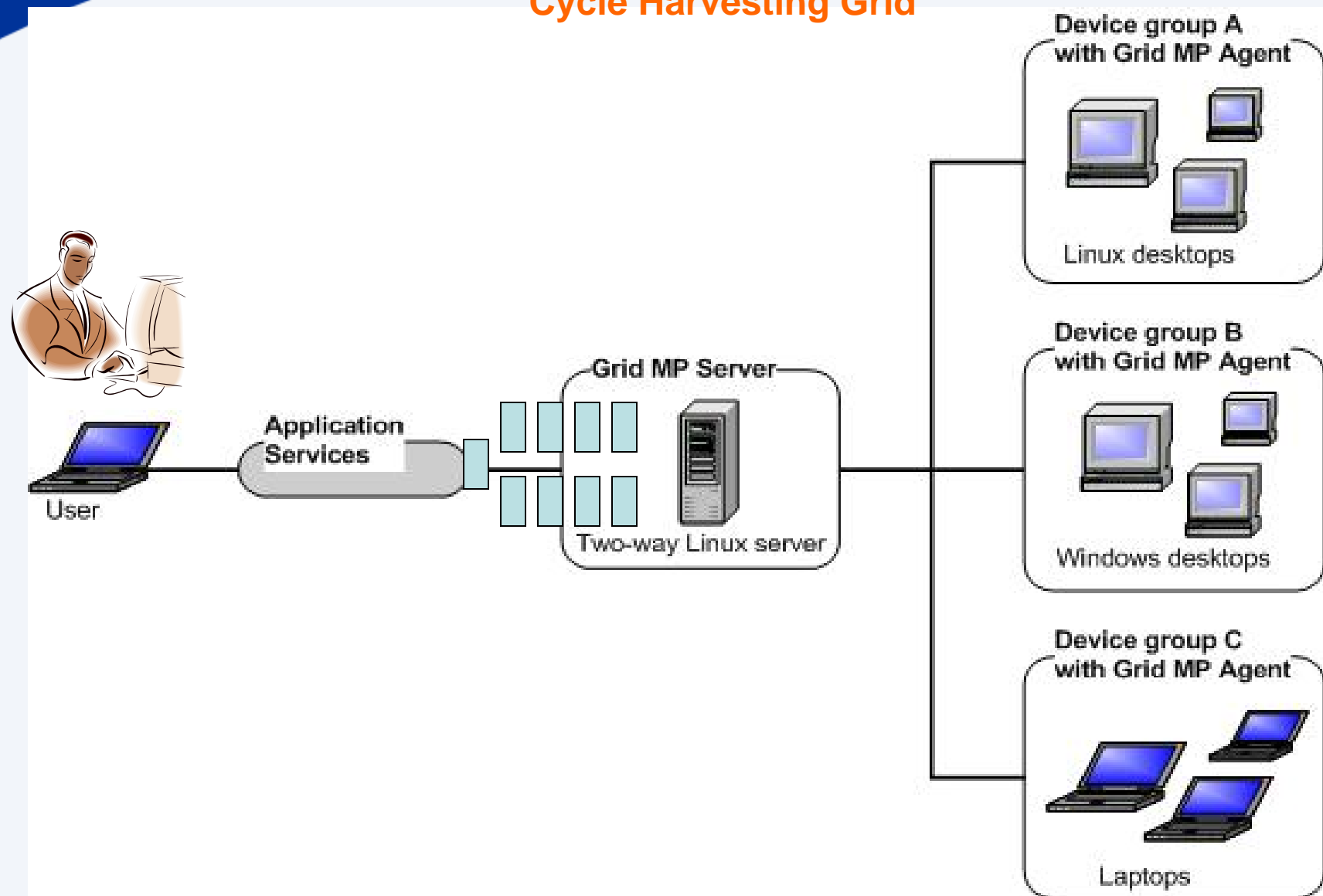
Massively Parallel

Including large-scale data parallel processing and parametric study that can run over loosely coupled distributed systems.

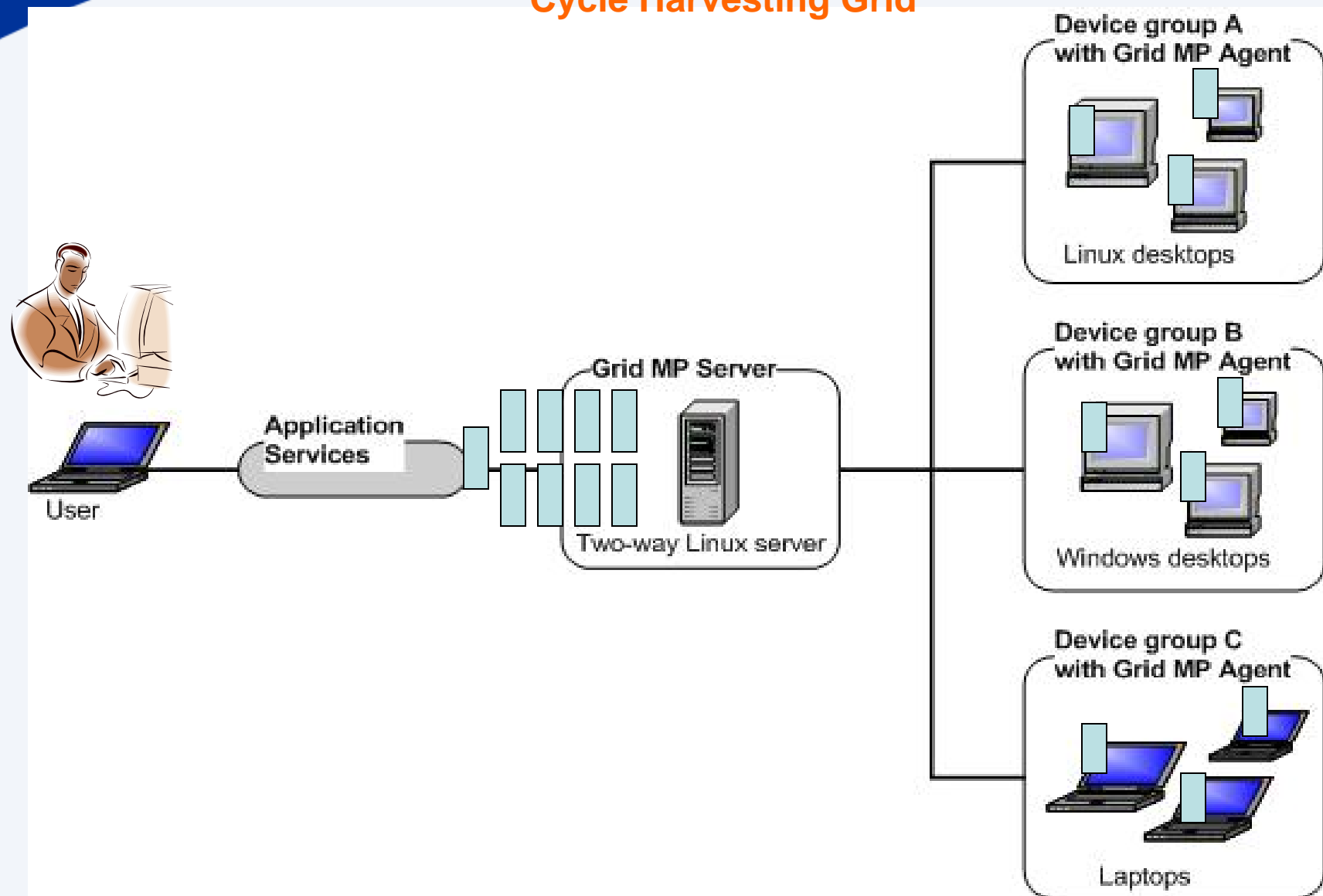
Visualisation/Modeling

Anytime anywhere collaborative support.

Solving Massively Parallel computation with Cycle Harvesting Grid

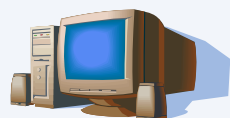


Solving Massively Parallel computation with Cycle Harvesting Grid

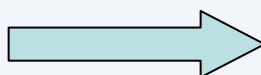


Divide and Conquer Illustration Image Rendering using POV-Ray

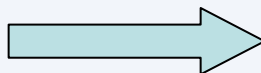
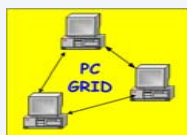
1024x768



Standalone PC



About One day 20 hrs

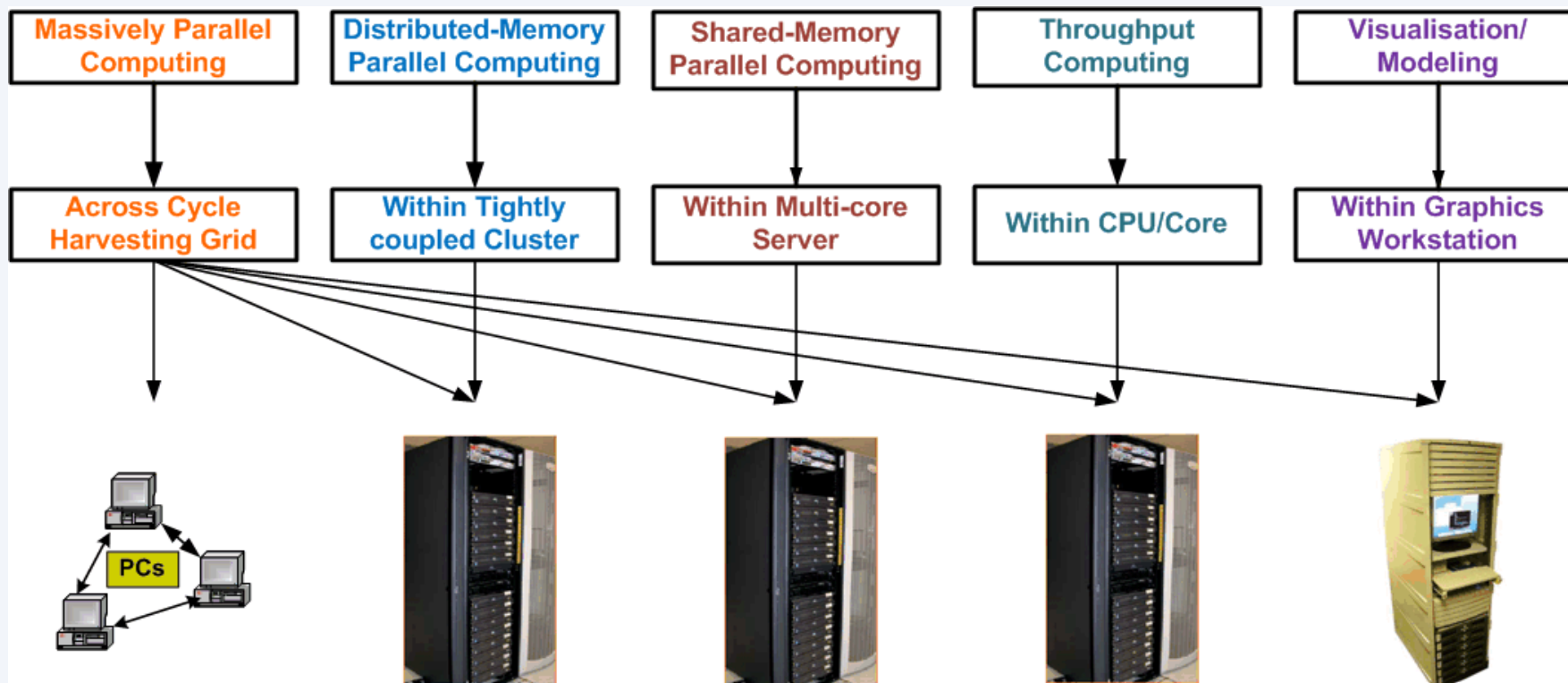


< 2 hours

Strategic Role of Cycle Harvesting Grid in Research Computing

Create new research computing capability through aggregation of existing distributed and diverse compute resources

Maximise compute resource utilisation



1300 PCs and 1250 server CPU cores available today

Future trends and Opportunities

- Increase of data intensive applications in research areas such as Biomedical Sciences, Signal & Image Processing in Science and Engineering, Financial Analytic, Social Sciences etc.
- More servers and desktops with multi-core processors.
- Advances in network performance.



Greater Potential Use of Cycle Harvesting Grid

Current Application - Example 1

NoiseGARCH – Financial Modeling (Matlab)

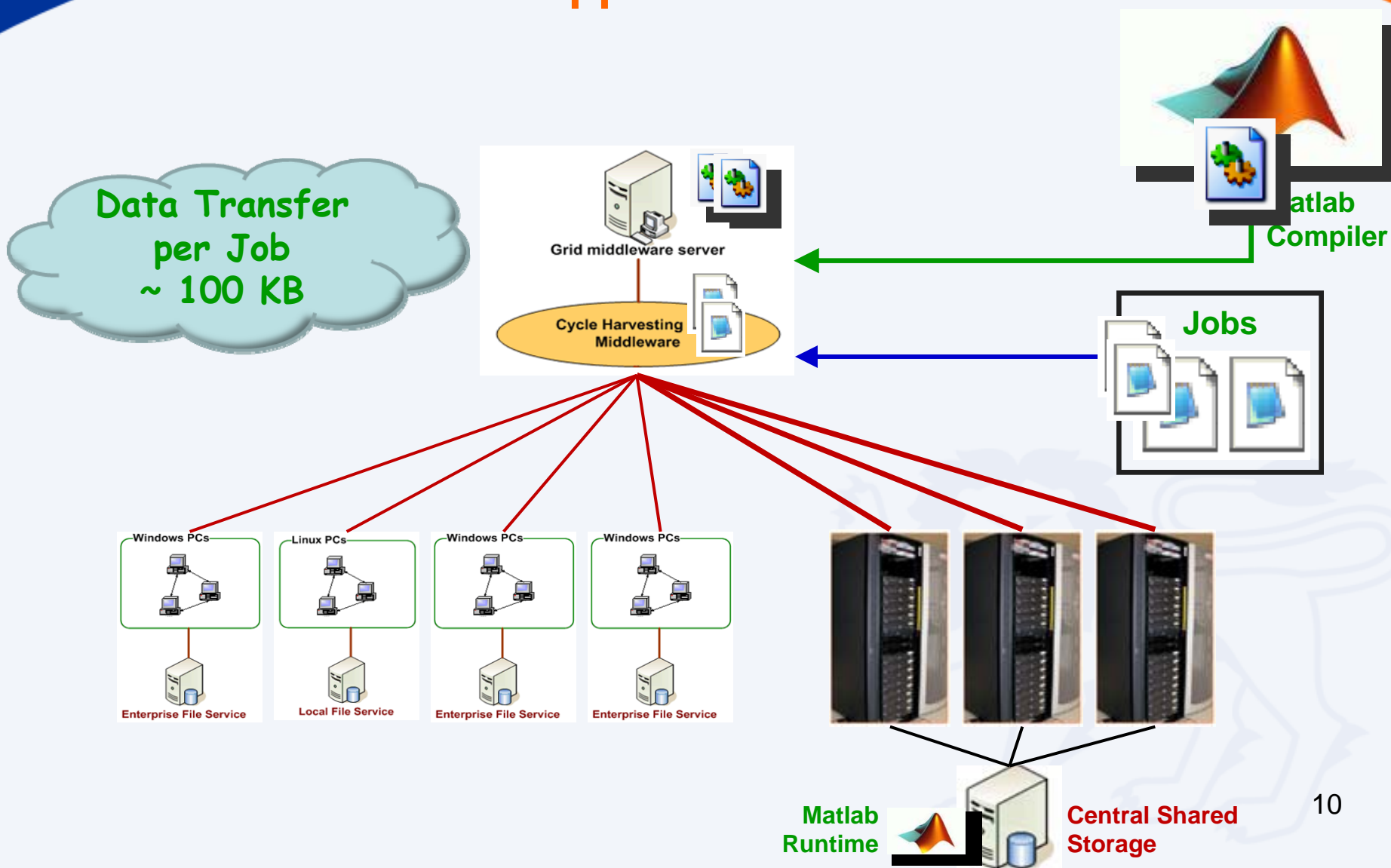
To estimate price model by analyzing the parameters of the model using general autoregressive conditional heteroscedasticity (GRACH) model

No	JOB ID	Work Units	Time to Solution	CPU Time	Speedup Rate
1	10137	100	7h 34m	17d 8h 39m	55.1
2	10149	100	3d 16h	35d 13h 45m	~14.3
3	10150	100	3d 15h 30m	16d 21h 57m	
4	10199	100	18h 52m	30d 22h 1m	39.3
5	10200	100	1d 4h 46m	32d 8h 12m	~76.9
6	10201	100	1d 6h 30m	34d 11h 52m	

Jobs were submitted and run at same time

NOTE: d — day, h — hour, m — minute

Current Application - Example 1 ...Cont Matlab Application on the Grid



Current Application – Example 2

Biochemistry (PHYLIP)

Accelerating Protein Phylogenetic Analysis

No	JOBID	Time to Solution	CPU Time	Speedup Rate
1	10220	1d 21h	105d 13h 35m	56.5
2	10219	1d 6h	68d 8h 9m	55.1
3	10218	16h	38d 17h 2m	58.6
4	10217	8h	17d 2h 35m	51.3
5	10216	1h 51m	4d 7h 58m	56.2
6	10215	31m	1d 1h 51m	50.0
7	10140	7d 14h	386d 12h 18m	50.2
8	10078	1d 15h	84d 11h 46m	52.6
9	10060	23h 38m	53d 20h 6m	54.7
10	10041	2h 30m	5d 10h 4m	52.0

NOTE: d — day, h — hour, m — minute

Current Application – Example 3

Statistical Analysis of Partially Calibrated Biopathway Models – Parametric Study (In-house developed)

Use computational methods to model the bio-pathways, and then estimate and searching the parameters for optimal solutions by collective fitting according to the experimental data.

Job No	Time to Solution	CPU Time	Speedup Rate
1	2 days 6 hours	75days and 12 hours	33.3
2	12 hours	31 days	62

Current Application – Example 4

Patent Image File Processing (In-house developed)

To convert 10-20 Terabytes of image data in raw tiff format from patent database into web-hosting friendly png format.

Job No	Time to Solution	CPU Time	Speedup Rate
1	5 – 8 hours	5 – 8 days	~20

Generated 8.8 CPU years of usage in 3 months

Collaboration Opportunity & Challenges

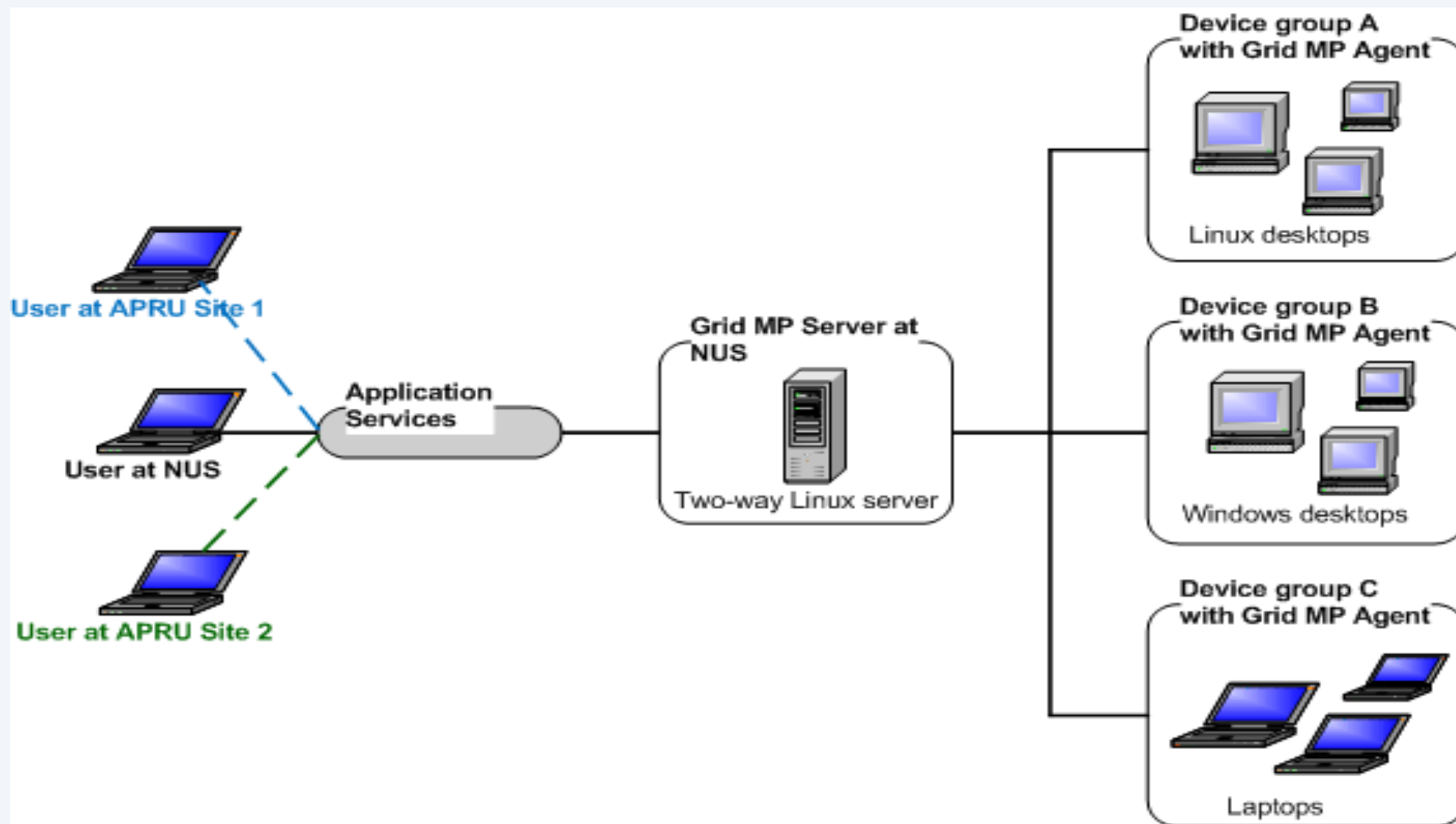
Opportunity

- Create greater capability with more PCs and servers connected.

Challenges

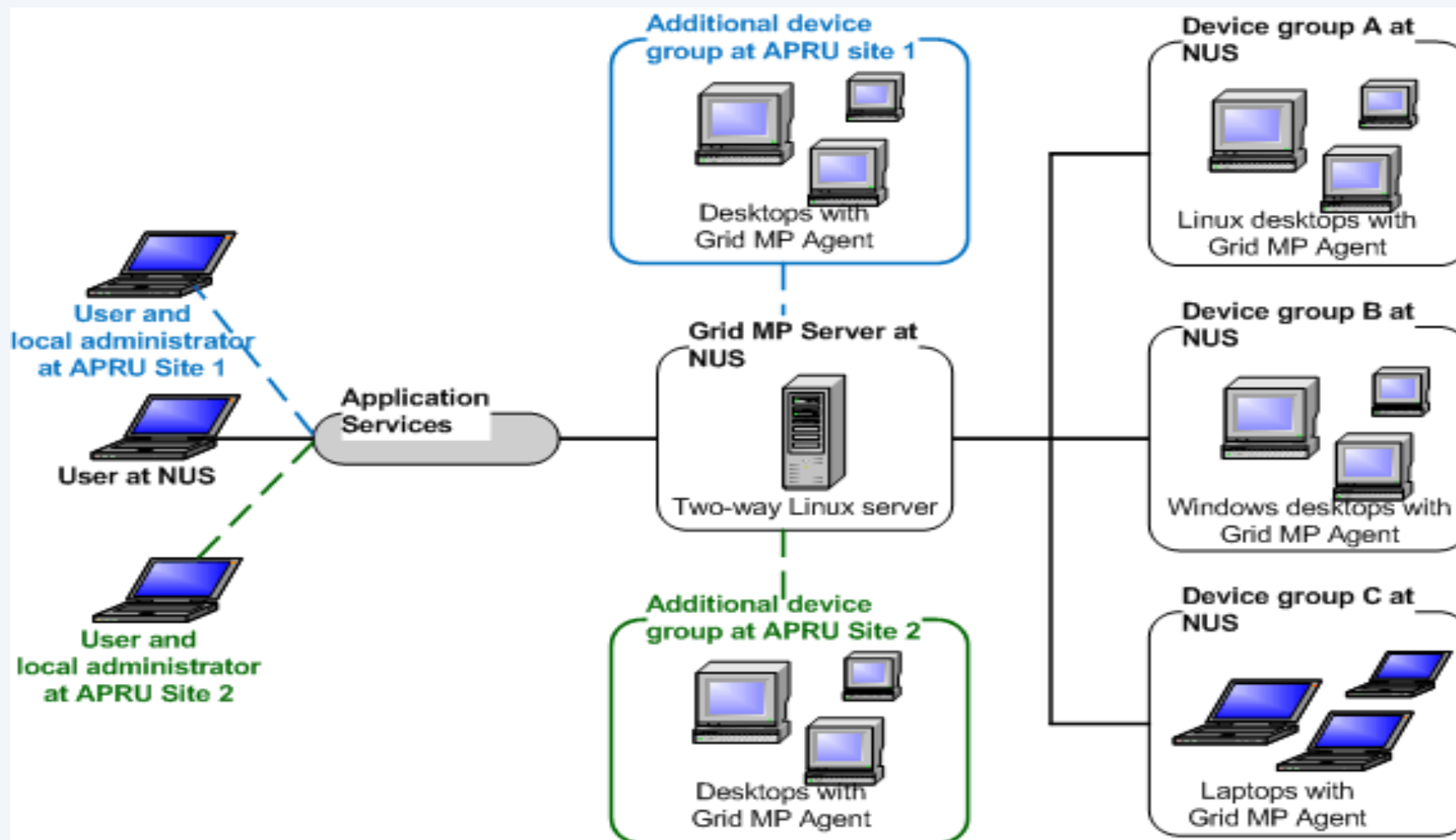
- Grid software licensing
- Network bandwidth and latency
- Convince staff to contribute their desktops

Option 1: Running Jobs on Single Site



--- Inter-site link

Option 2: Running Jobs Across Sites



--- Inter-site link

Discussion

