Faculty of Science, Engineering and Technology

Swinburne University of Technology, Australia

24th March 2014

Automated Analysis of Performance and Energy Consumption for Cloud Applications

ICPE 2014

Presented by
Feifei Chen

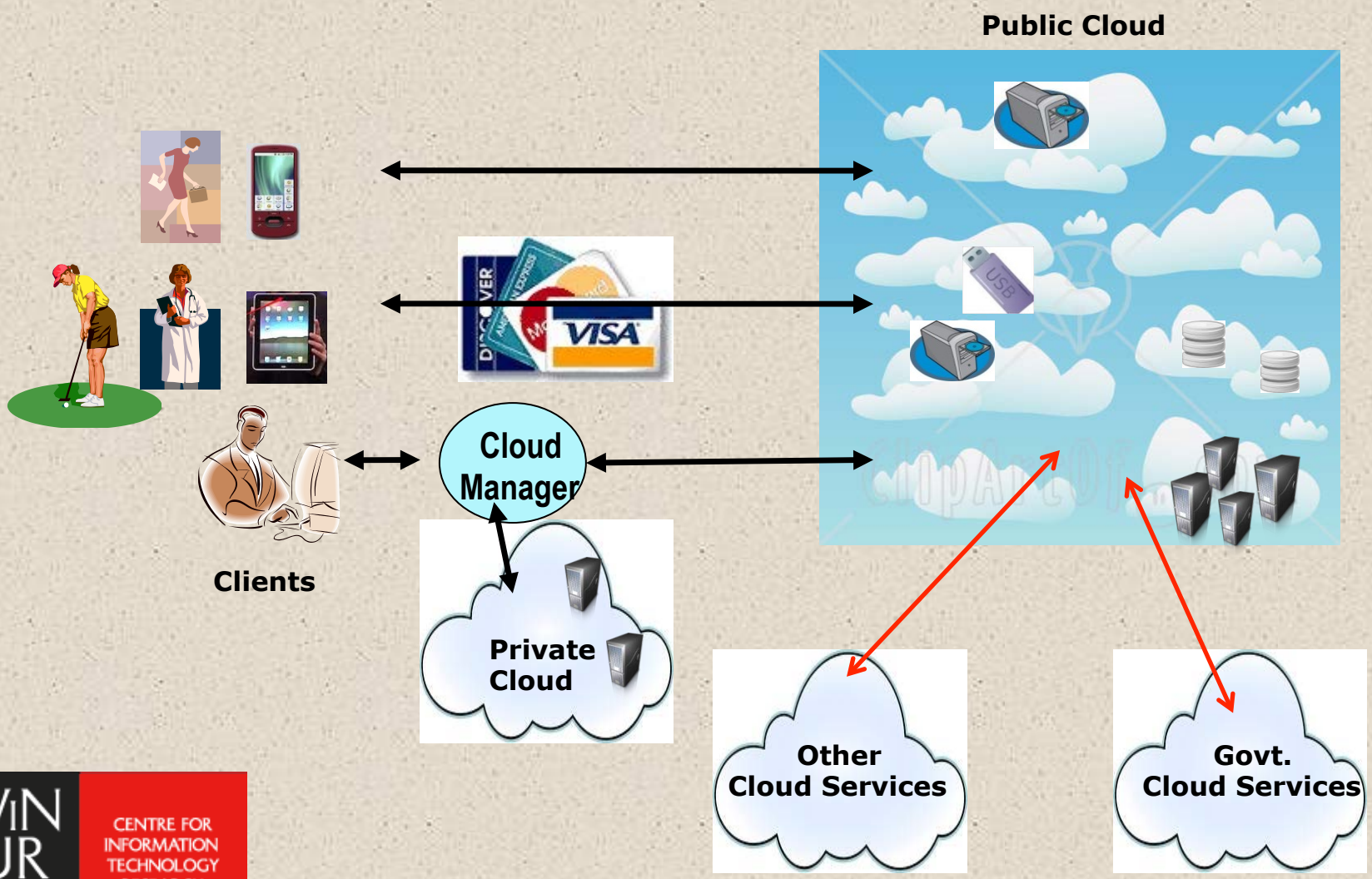
Agenda

- **Introduction**
- **Approach**
- **Experiment Setup**
- **Experimental Results**
- **Conclusion**

SWINBURNE

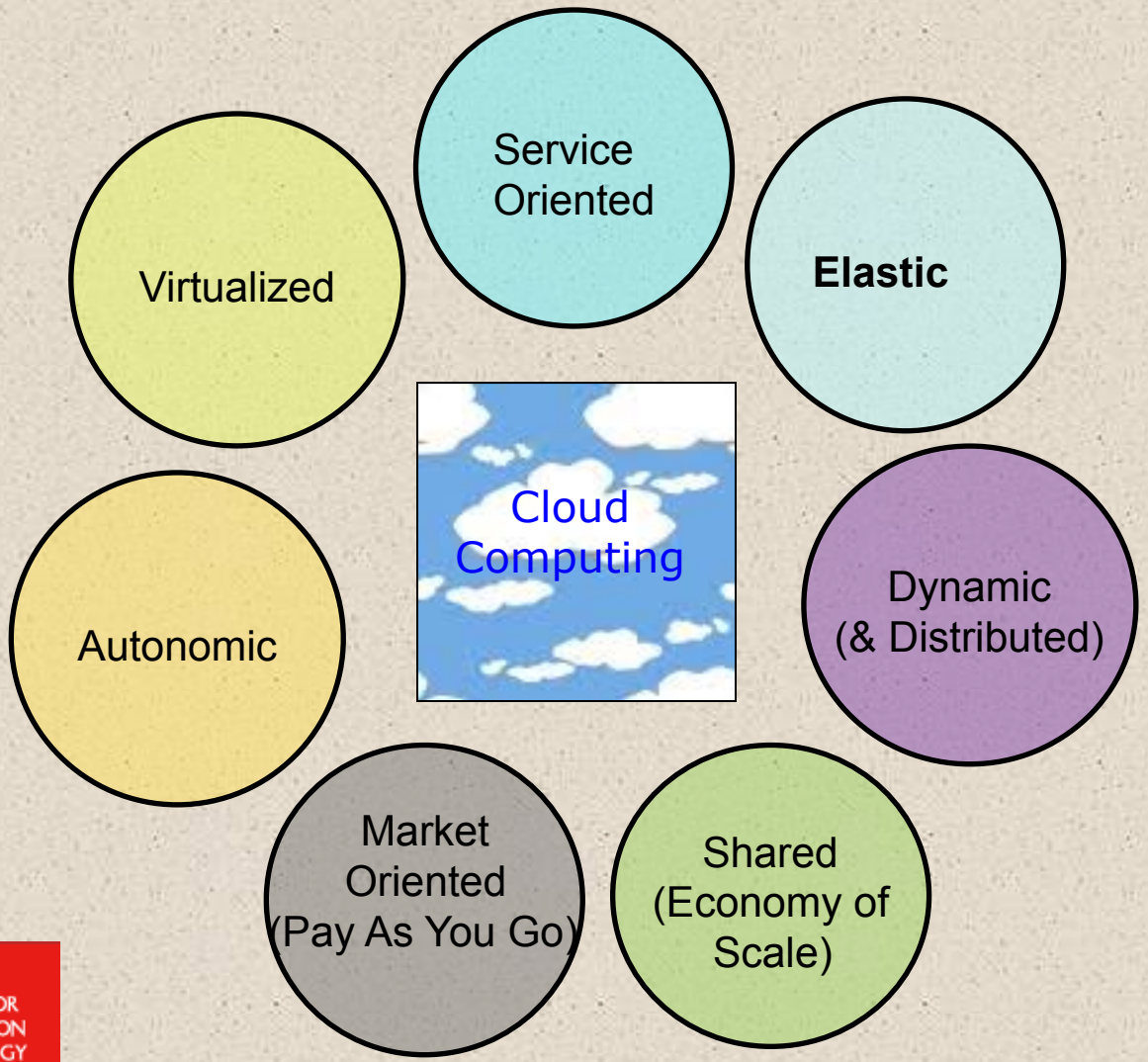
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Clouds offer Subscription-Oriented IT Services: {computation, application, data,..} as a Service (..aaS)

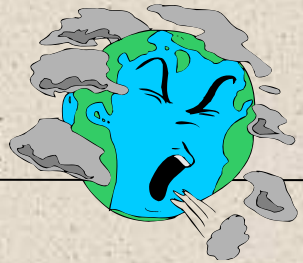


SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

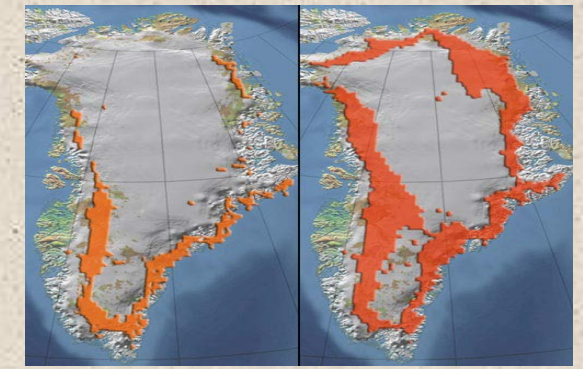
Several Benefits……



Dark side…..



- **High CO2 emissions contribution**



  - New York Times 2012: Data centres use about 30 billion watts of electricity per hour worldwide, equivalent to the output of about 30 nuclear power plants

- **High Operational Cost**

SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

SLA Requirement

- **A key objective of cloud service providers:**

  develop resource provisioning and management solutions at minimum energy consumption while still guaranteeing Service Level Agreements (SLAs).

Understand both system performance and energy consumption pattern

SWINBURNE CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Agenda

- Introduction
- Approach
- Experiment Setup
- Experimental Results
- Conclusion

SWINBURNE

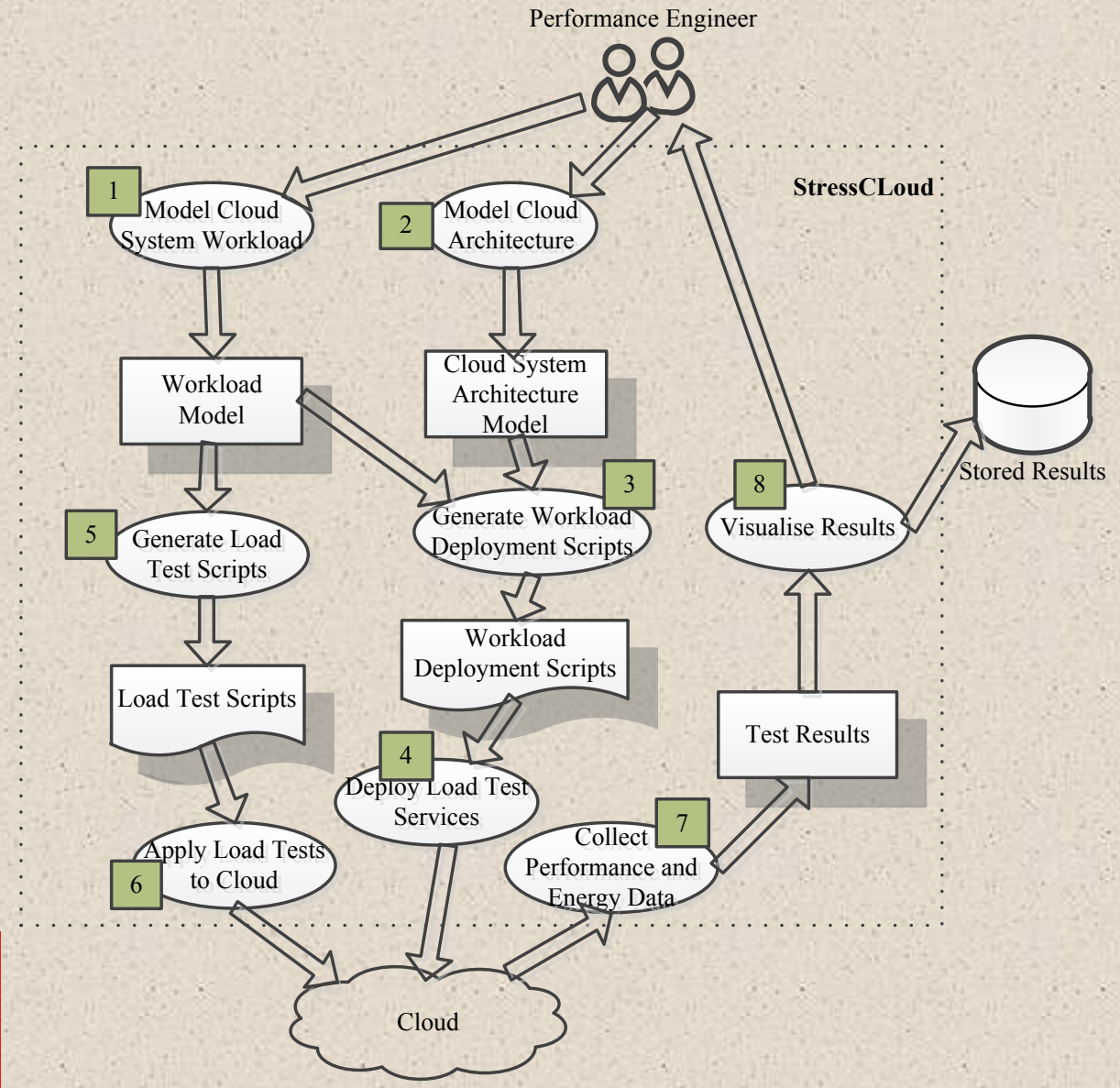
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Understand both system performance and energy consumption pattern:

- running extensive experiments with heterogeneous parameters/metrics and workloads;
- collecting appropriate cloud and application energy/ performance measurements;
- performing energy/performance trade-off analysis.

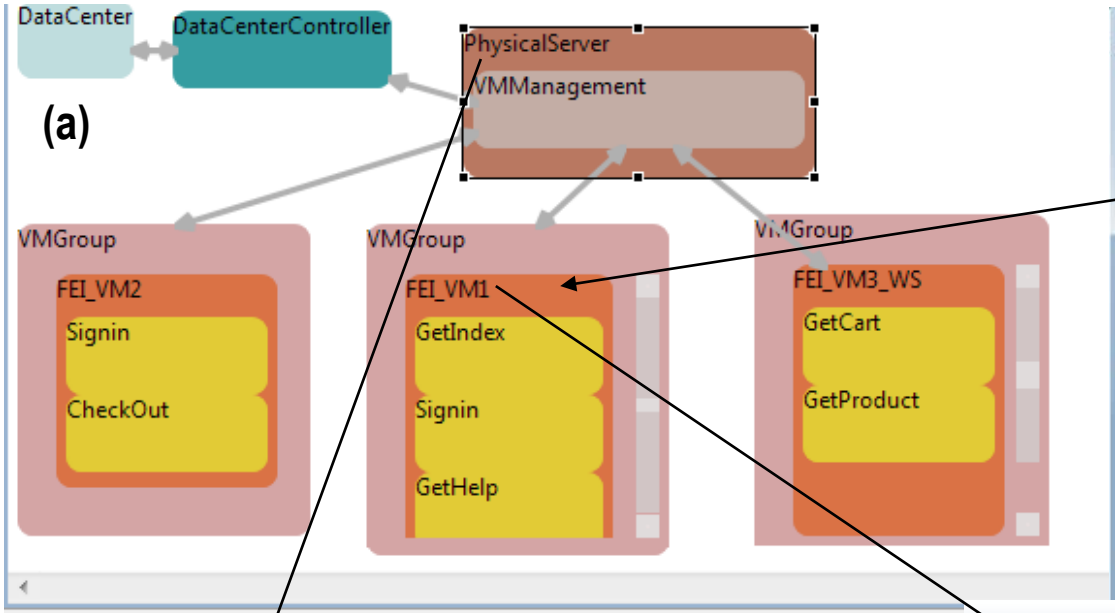
SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

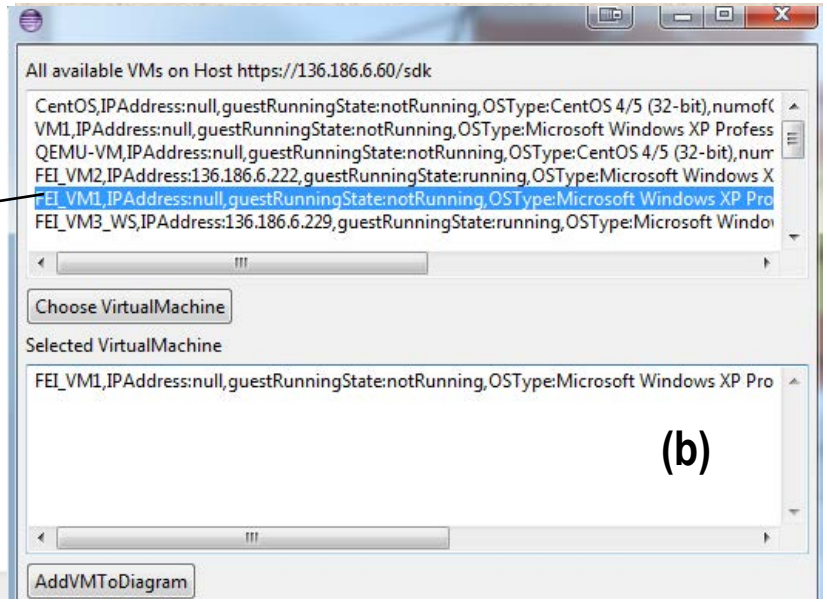
Performance and energy analysis tool - StressCloud



SWIN BUR NE

CENTRE FOR INFORMATION TECHNOLOGY RESEARCH





Tasks | Properties

PhysicalServer

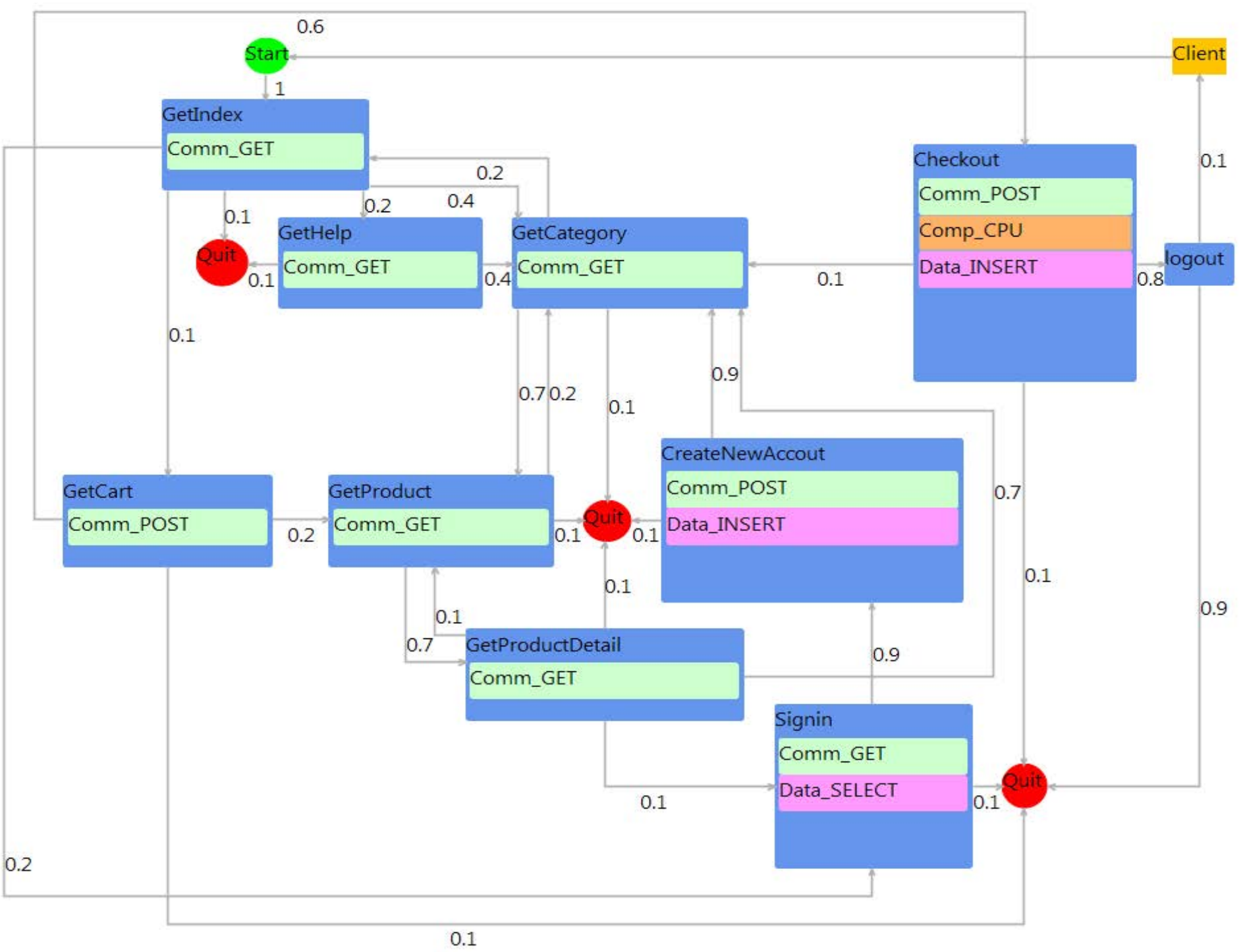
Core | Appearance

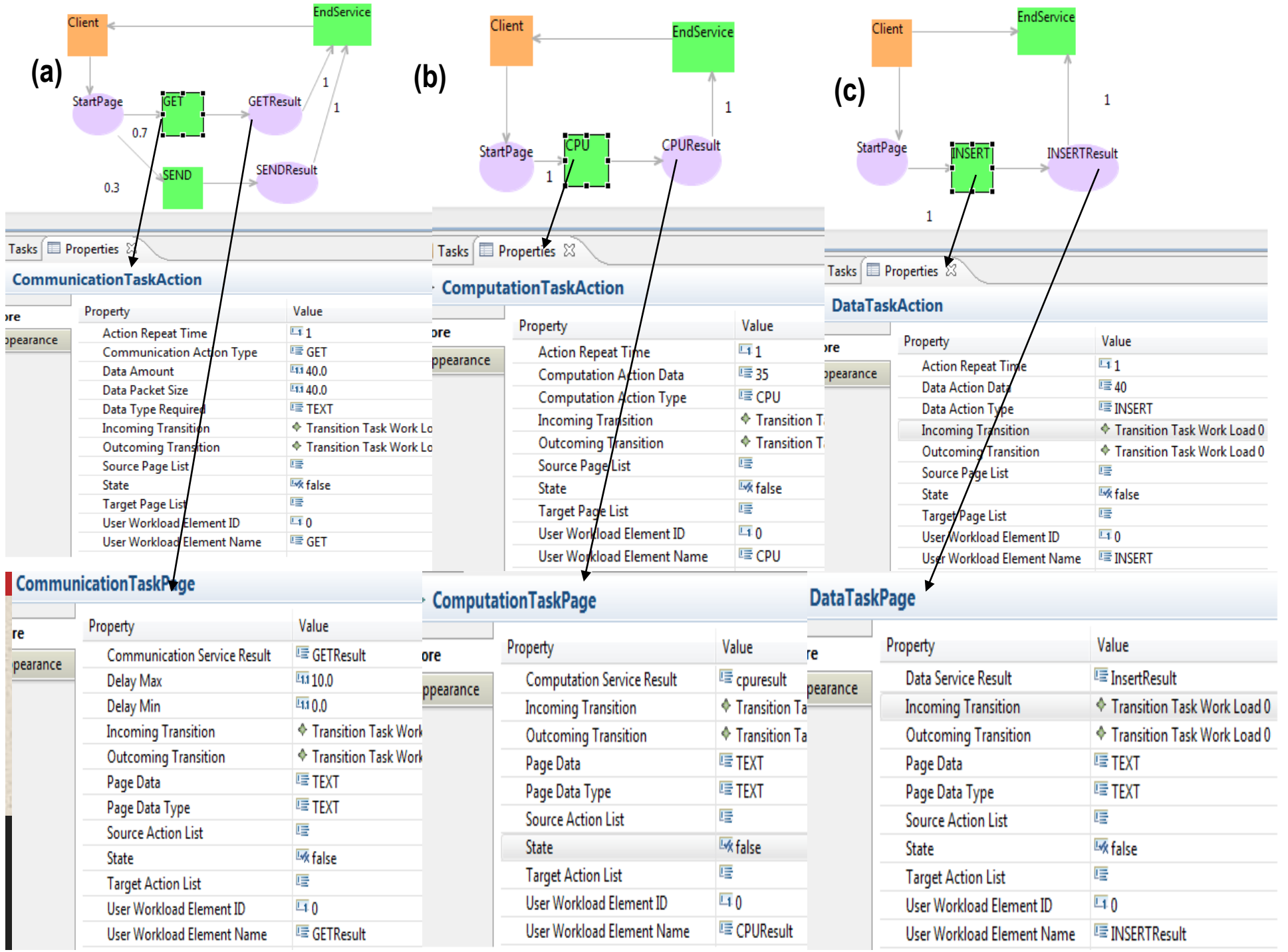
Property	Value
Access Password	amani123
Access User Name	root
Cache Of Physical Processor	N/A
Cloud Element ID	0
Cloud Element Name	PhysicalServer
Frequency Of Physical Processor	2.799999615E9
Frequency Of RAM	0.0
Hyper Threading	true
Incoming Connection	
Number Of Physical Processor	4
Number Of Virtual Machine	14
Number Of VM Group	2
Outcomming Connection	
Processor Additional Description	null
RAM Additional Description	null
Server Service Address	https://136.186.6.60/sdk
Size Of RAM	1.0718543872E10
Size Of Storage	9.98848331776E11
Type Of Physical Processor	Intel

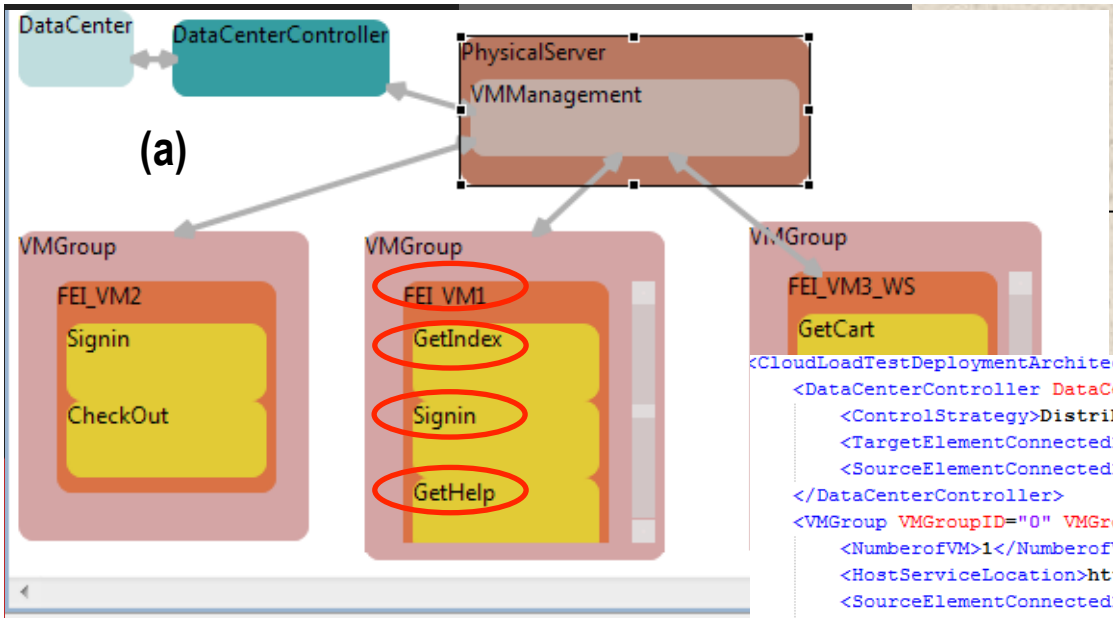
VirtualMachine

Core | Appearance

Property	Value
Cache Of Virtual Processor	N/A
Cloud Element ID	0
Cloud Element Name	FEI_VM1
Communicationl Task Allocated List	
Compositional Task Allocated List	
Computation Task Allocated List	
Datal Task Allocated List	
Frequency Of RAM	2133.0
Frequency Of Virtuall Processor	2.8
Host Access Password	amani123
Host Access Username	root
Host Service Location	https://136.186.6.60/sdk
Incoming Connection	
Number Of Task Allocated	3
Number Of Virtual Processor	3
OS Type	Microsoft Windows XP Professional (32-bit)
Outcomming Connection	
Processor Additional Description	null
RAM Additional Description	null
Size Of RAM	6144.0
Size Of Storage	8.388608E7
Type Of Virtual Processor	Intel
VM Group ID	0
VM Running Status	running
VM Service Location	http://136.186.6.235:8080


Start
Client
GetIndex
Comm_GET
Checkout
Comm_POST
Comp_CPU
Data_INSERT
logout
GetHelp
Comm_GET
Quit
GetCategory
Comm_GET
GetCart
Comm_POST
GetProduct
Comm_GET
Quit
CreateNewAccout
Comm_POST
Data_INSERT
GetProductDetail
Comm_GET
Signin
Comm_GET
Data_SELECT
Quit
0.6
1
0.1
0.2
0.4
0.1
0.1
0.4
0.1
0.8
0.1
0.1
0.7
0.2
0.1
0.9
0.7
0.2
0.1
0.1
0.1
0.1
0.7
0.9
0.1
0.9
0.1
0.2
0.1


(a)
Client
EndService
StartPage
GET
GETResult
SEND
SENDResult
0.7
0.3
1
1
Tasks
Properties
CommunicationTaskAction
Property
Value
Action Repeat Time
1
Communication Action Type
GET
Data Amount
40.0
Data Packet Size
40.0
Data Type Required
TEXT
Incoming Transition
Transition Task Work Lo
Outcoming Transition
Transition Task Work Lo
Source Page List
State
false
Target Page List
User Workload Element ID
0
User Workload Element Name
GET
CommunicationTaskPage
Property
Value
Communication Service Result
GETResult
Delay Max
10.0
Delay Min
0.0
Incoming Transition
Transition Task Work
Outcoming Transition
Transition Task Work
Page Data
TEXT
Page Data Type
TEXT
Source Action List
State
false
Target Action List
User Workload Element ID
0
User Workload Element Name
GETResult
(b)
Client
EndService
StartPage
CPU
CPUResult
1
1
Tasks
Properties
ComputationTaskAction
Property
Value
Action Repeat Time
1
Computation Action Data
35
Computation Action Type
CPU
Incoming Transition
Transition T
Outcoming Transition
Transition T
Source Page List
State
false
Target Page List
User Workload Element ID
0
User Workload Element Name
CPU
ComputationTaskPage
Property
Value
Computation Service Result
cpuresult
Incoming Transition
Transition Ta
Outcoming Transition
Transition Ta
Page Data
TEXT
Page Data Type
TEXT
Source Action List
State
false
Target Action List
User Workload Element ID
0
User Workload Element Name
CPUResult
(c)
Client
EndService
StartPage
INSERT
INSERTResult
1
1
Tasks
Properties
DataTaskAction
Property
Value
Action Repeat Time
1
Data Action Data
40
Data Action Type
INSERT
Incoming Transition
Transition Task Work Load 0
Outcoming Transition
Transition Task Work Load 0
Source Page List
State
false
Target Page List
User Workload Element ID
0
User Workload Element Name
INSERT
DataTaskPage
Property
Value
Data Service Result
InsertResult
Incoming Transition
Transition Task Work Load 0
Outcoming Transition
Transition Task Work Load 0
Page Data
TEXT
Page Data Type
TEXT
Source Action List
State
false
Target Action List
User Workload Element ID
0
User Workload Element Name
INSERTResult



Tasks | Properties

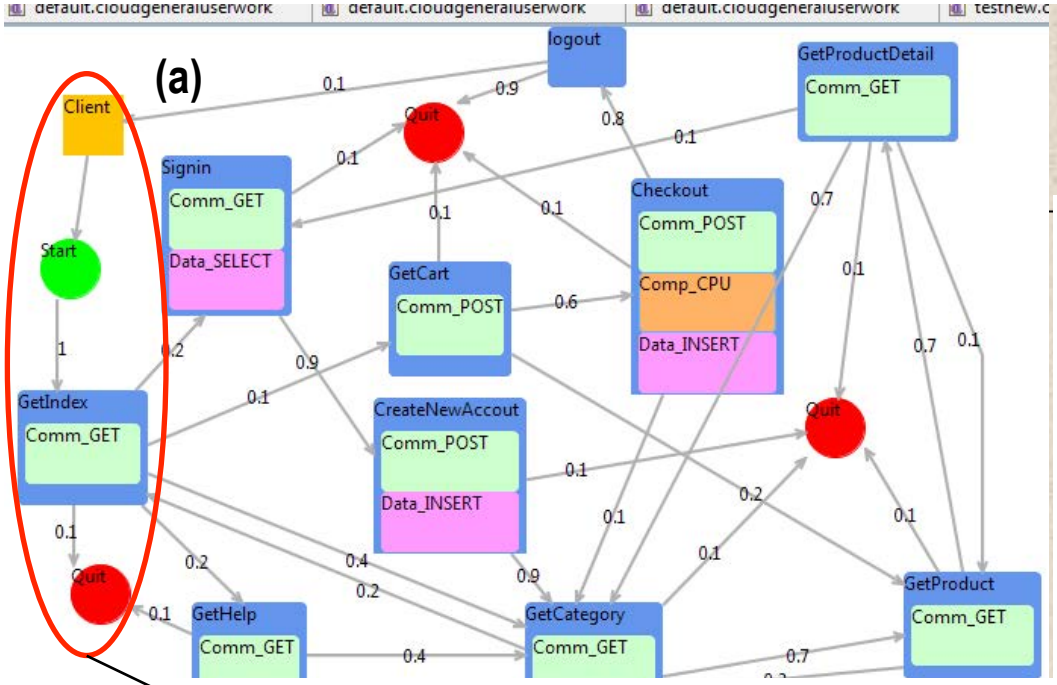
PhysicalServer

Core | Appearance

Property	Value
Access Password	amani123
Access User Name	root
Cache Of Physical Processor	N/A
Cloud Element ID	0
Cloud Element Name	PhysicalSe
Frequency Of Physical Processor	2.7999996
Frequency Of RAM	0.0
Hyper Threading	true
Incoming Connection	
Number Of Physical Processor	4
Number Of Virtual Machine	14
Number Of VM Group	2
Outcomming Connection	
Processor Additional Description	null
RAM Additional Description	null
Server Service Address	https://13
Size Of RAM	1.0718543
Size Of Storage	9.9884833
Type Of Physical Processor	Intel

(b)

```
<CloudLoadTestDeploymentArchitecture>
    <DataCenterController DataCenterControlerID="1" DataCenterControlerName="DataCenterController">
        <ControlStrategy>Distribute</ControlStrategy>
        <TargetElementConnected>VMManagement</TargetElementConnected>
        <SourceElementConnected>DataCenter</SourceElementConnected>
    </DataCenterController>
    <VMGroup VMGroupID="0" VMGroupName="VMGroup">
        <NumberofVM>1</NumberofVM>
        <HostServiceLocation>https://136.186.6.60/sdk</HostServiceLocation>
        <SourceElementConnected>VMManagement</SourceElementConnected>
        <VirtualMachine VMID="0" VMName="FEI_VM1">
            <CacheofVirtualProcessor>N/A</CacheofVirtualProcessor>
            <ProcessorDescription>null</ProcessorDescription>
            <RAMDescription>null</RAMDescription>
            <FrequencyofVirtualProcessor>2.8</FrequencyofVirtualProcessor>
            <FrequencyofRAM>2133.0</FrequencyofRAM>
            <TypeofVirtualProcessor>Intel</TypeofVirtualProcessor>
            <SizeofRAM>6144.0</SizeofRAM>
            <SizeofStorage>8.388608E7</SizeofStorage>
            <NumberofVirtualProcessor>3</NumberofVirtualProcessor>
            <NumberofTaskAllocated>3</NumberofTaskAllocated>
            <VMGroupID>0</VMGroupID>
            <OSType>Microsoft Windows XP Professional (32-bit)</OSType>
            <VMServiceLocation>http://136.186.6.235:8080</VMServiceLocation>
            <HostServiceAddress>https://136.186.6.60/sdk</HostServiceAddress>
            <HostUsername>root</HostUsername>
            <HostPassword>amani123</HostPassword>
            <Task TaskID="0" TaskName="GetIndex">
                <MappedTaskID>0</MappedTaskID>
                <MappedTaskName>GetIndex</MappedTaskName>
            </Task>
            <Task TaskID="0" TaskName="Signin">
                <MappedTaskID>0</MappedTaskID>
                <MappedTaskName>Signin</MappedTaskName>
            </Task>
            <Task TaskID="0" TaskName="GetHelp">
                <MappedTaskID>0</MappedTaskID>
                <MappedTaskName>GetHelp</MappedTaskName>
            </Task>
        </VirtualMachine>
    </VMGroup>
```

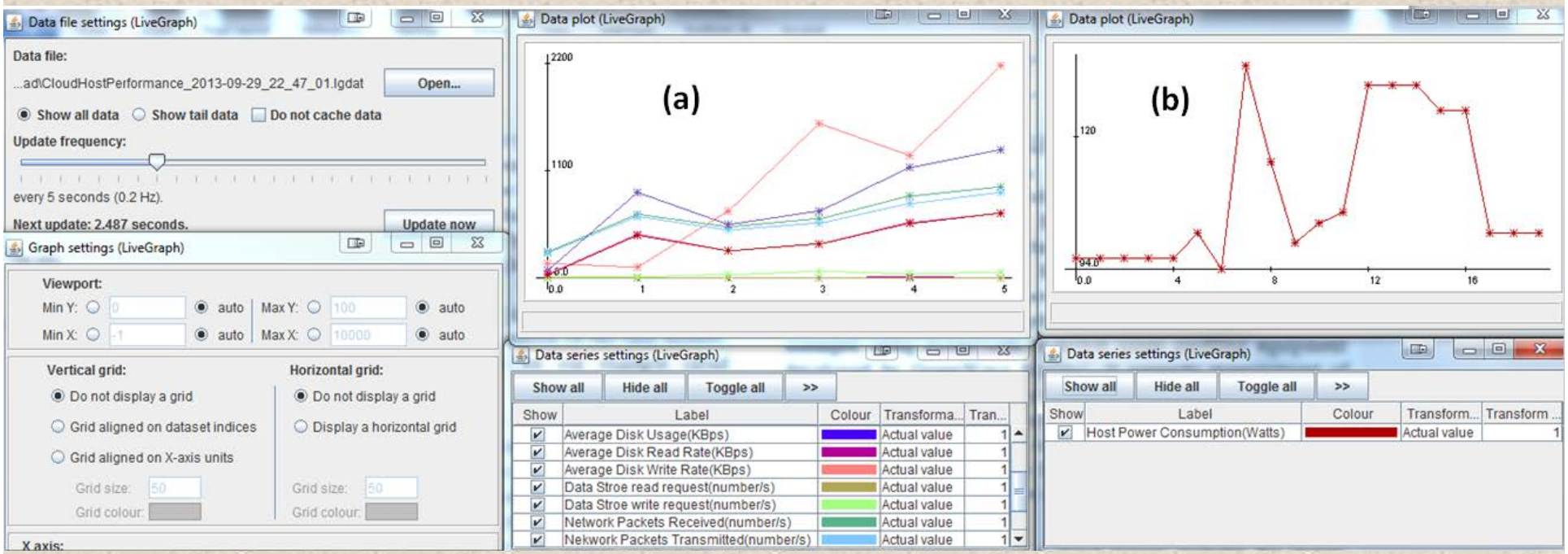



```
<Path PathID="0">
    <GeneralUserClient UserClientID="0" UserClientName="Client">
        <ClientState>false</ClientState>
    </GeneralUserClient>
    <StartPoint StartPointID="0" StartPoint="Start">
        <IsStartState>false</IsStartState>
        <IsTerminateState>false</IsTerminateState>
        <State>false</State>
    </StartPoint>
    <CompositionalTask TaskID="0" TaskName="GetIndex">
        <TransitProbabilityFromPriorPage>1.0</TransitProbabilityFromPriorPage>
        <IsStartState>false</IsStartState>
        <IsTerminateState>false</IsTerminateState>
        <State>false</State>
        <Sub_CommunciationIntensiveTask TaskID="0" TaskName="Comm_GET">
            <WorkloadModelFile>C:\test\decisionmodel\JPetStore\GET.xml</WorkloadModelFile>
            <RepeatTime>1.0</RepeatTime>
            <ResourceLocation>http://136.186.6.235:8080</ResourceLocation>
            <NumberOfVM>1</NumberOfVM>
            <VMTypeRequired>Large</VMTypeRequired>
            <IsStartState>false</IsStartState>
            <IsTerminateState>false</IsTerminateState>
            <State>false</State>
        </Sub_CommunciationIntensiveTask>
    </CompositionalTask>
    <EndPoint EndPointID="0" EndPoint="End">
        <IsStartState>false</IsStartState>
        <IsTerminateState>false</IsTerminateState>
        <State>false</State>
    </EndPoint>
```

(b)

(c)

```
<ParentTask TaskID="0" TaskName="Comm_GET">
    <ResourceLocation>http://136.186.6.235:8080</ResourceLocation>
</ParentTask>
<Path PathID="0">
    <CommunicationServiceClient ServiceClientID="0" ServiceClientName="Client">
        <isMultiThread>false</isMultiThread>
        <ServiceClientNumber>10</ServiceClientNumber>
        <ClientState>false</ClientState>
    </CommunicationServiceClient>
    <CommunicationTaskPage PageID="0" PageName="StartPage">
        <ServiceResult>start</ServiceResult>
        <DelayMax>0.0</DelayMax>
        <DelayMin>0.0</DelayMin>
        <PageData>Start</PageData>
        <PageDataType>Start</PageDataType>
        <PageState>false</PageState>
    </CommunicationTaskPage>
    <CommunicationTaskAction ActionID="0" ActionName="GET">
        <RepeatTime>1</RepeatTime>
        <ActionType>GET</ActionType>
        <DataAmount>40.0</DataAmount>
        <PacketSize>40.0</PacketSize>
        <DataType>TEXT</DataType>
        <ActionState>false</ActionState>
        <TransitProbabilityFromPriorPage>1.0</TransitProbabilityFromPriorPage>
    </CommunicationTaskAction>
    <CommunicationTaskPage PageID="0" PageName="GETResult">
        <ServiceResult>GETResult</ServiceResult>
        <DelayMax>10.0</DelayMax>
        <DelayMin>0.0</DelayMin>
        <PageData>TEXT</PageData>
        <PageDataType>TEXT</PageDataType>
        <PageState>false</PageState>
    </CommunicationTaskPage>
    <CommunicationTaskAction ActionID="0" ActionName="EndService">
        <RepeatTime>1</RepeatTime>
        <ActionType>END</ActionType>
        <DataAmount>40.0</DataAmount>
        <PacketSize>40.0</PacketSize>
        <DataType>TEXT</DataType>
        <ActionState>false</ActionState>
        <TransitProbabilityFromPriorPage>1.0</TransitProbabilityFromPriorPage>
    </CommunicationTaskAction>
```



SWINBURNE

CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Performance and energy analysis tool - StressCloud

Workload Model:

A set of **Tasks** modelling the target cloud application behaviour

- ✓ Computation-Intensive
  - ✓ CPU-Intensive
  - ✓ Memory-Intensive
- ✓ Data-Intensive
- ✓ Communication-Intensive

Task:

A stochastic form chart specifying the detailed user requests and required responses from the cloud system

Task Type	Service Type in StressCloud
CPU-intensive	Fibonacci sequence calculating
Memory-intensive	File processing
Data-intensive	Rational database operating
Communication-intensive	HTTP request/response

SWINBURNE CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Performance and energy analysis tool - StressCloud

Cloud Architecture Model:

All available resources in the target cloud system and their detailed configurations

SWIN BUR NE

CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Agenda

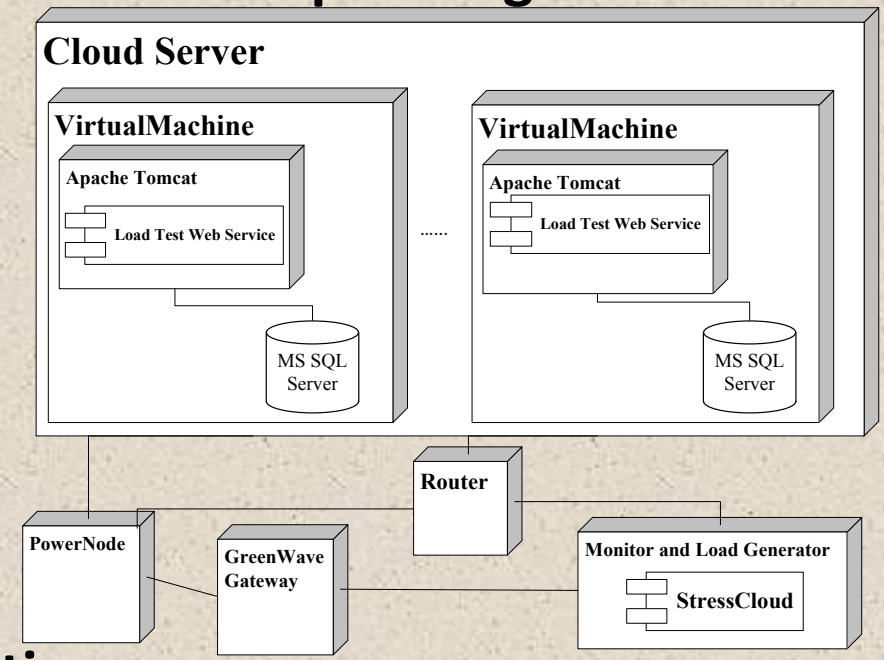
- Introduction
- Approach
- **Experiment Setup**
- Experimental Results
- Conclusion

SWINBURNE

CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Experiment Setup

- Energy and performance profiling framework



- VM configuration

Virtual Machine	Number of Cores	RAM	Hard Disk
Small	1	2GB	80GB
Medium	2	4GB	80GB
Large	3	6GB	80GB
XLarge	4	8GB	80GB

SWINBURNE CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Test cases

Workload Type	Test set Number	Test Point
CPU-intensive	1	Resource & Workload
CPU-intensive	2	Resource Allocation Strategy
Memory-intensive	1	Resource & Workload
Data-intensive	1	Workload
Data-intensive	2	Workload
Communication-intensive	1	Workload
Communication-intensive	2	Resource Allocation Strategy
Mix Computation- and Data-intensive	1	Workload
Mix Computation- and Data-intensive	2	Resource Allocation Strategy
Mix Computation- , Data- and Communication-intensive	1	Workload
Mix Computation- , Data- and Communication-intensive	2	Resource Allocation Strategy

SWINBURNE INFORMATION TECHNOLOGY RESEARCH

Agenda

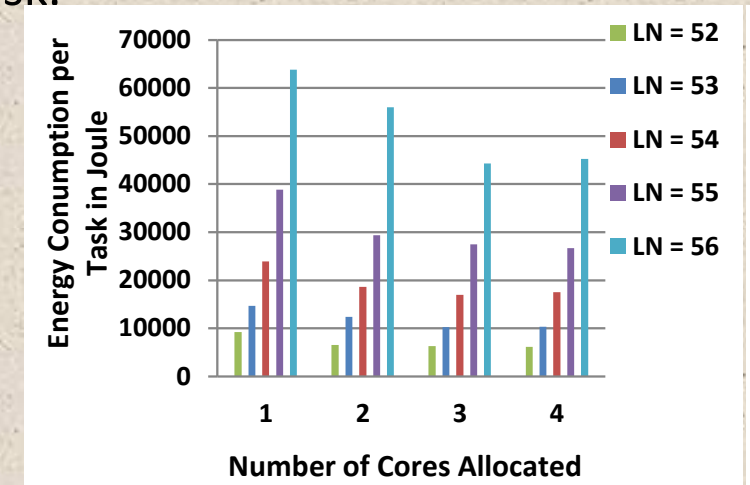
- Introduction
- Approach
- Experiment Setup
- **Experimental Results**
- Conclusion

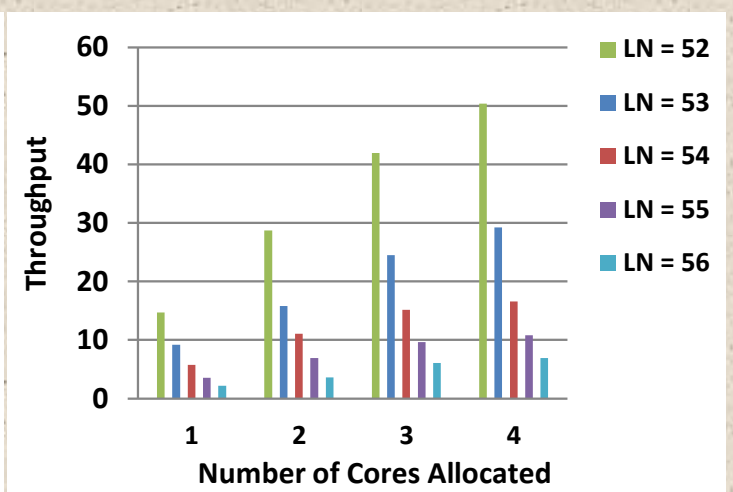
SWINBURNE

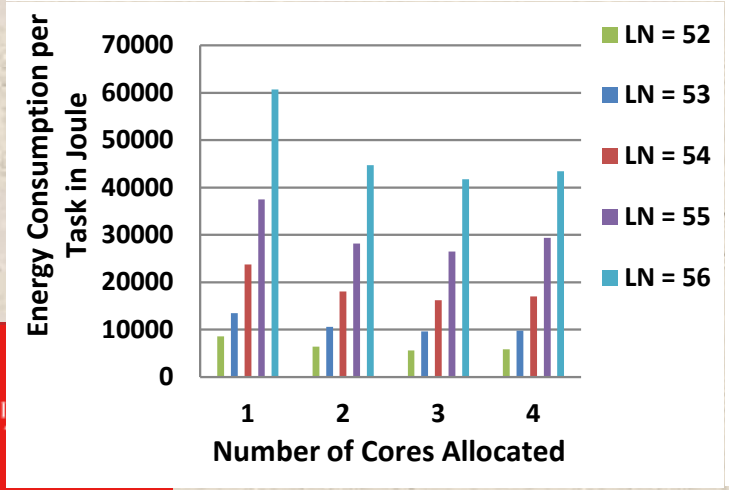
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

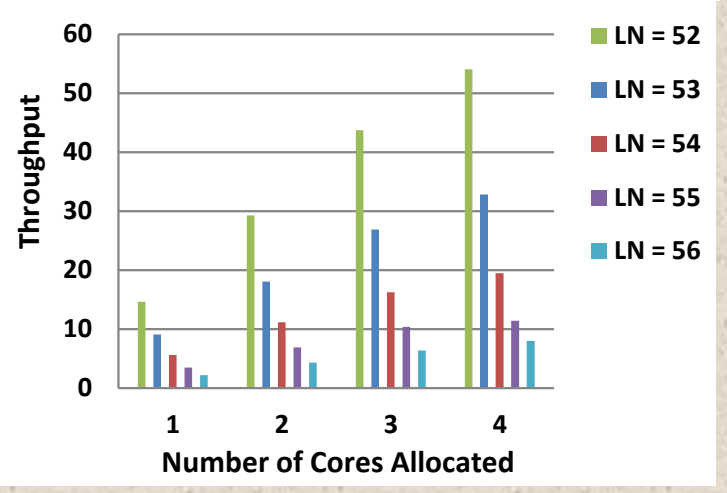
CPU-Intensive Workload

- **Test set #1**: Keeping the number of tasks constant, while gradually increasing the CPU cores allocated to the task, and the workload of the task.



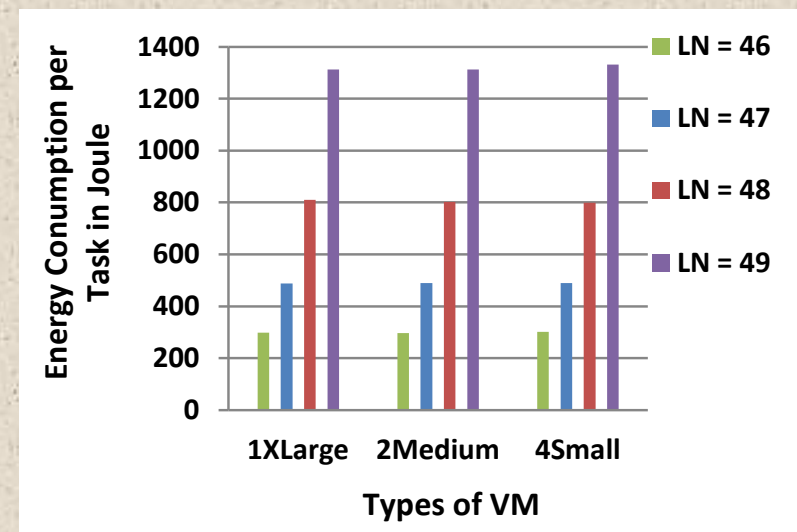


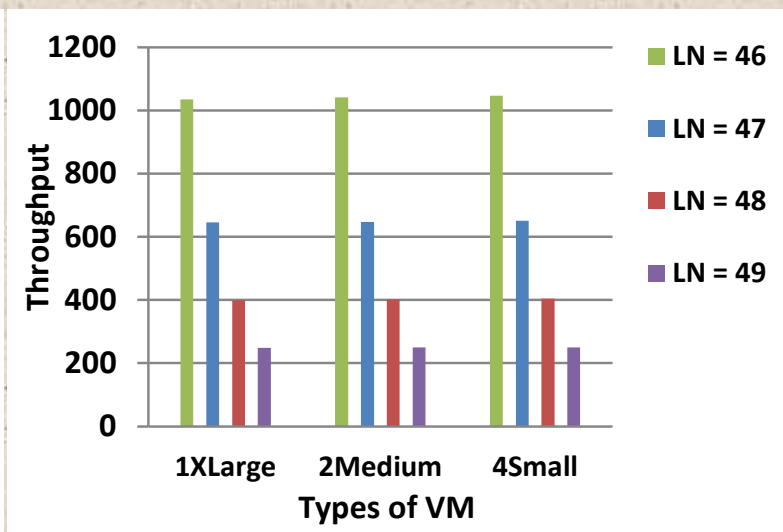




CPU-Intensive Workload

- **Test set #2**: Keeping the number of tasks and resource allocated to the tasks constant, while changing the resource allocation strategy.

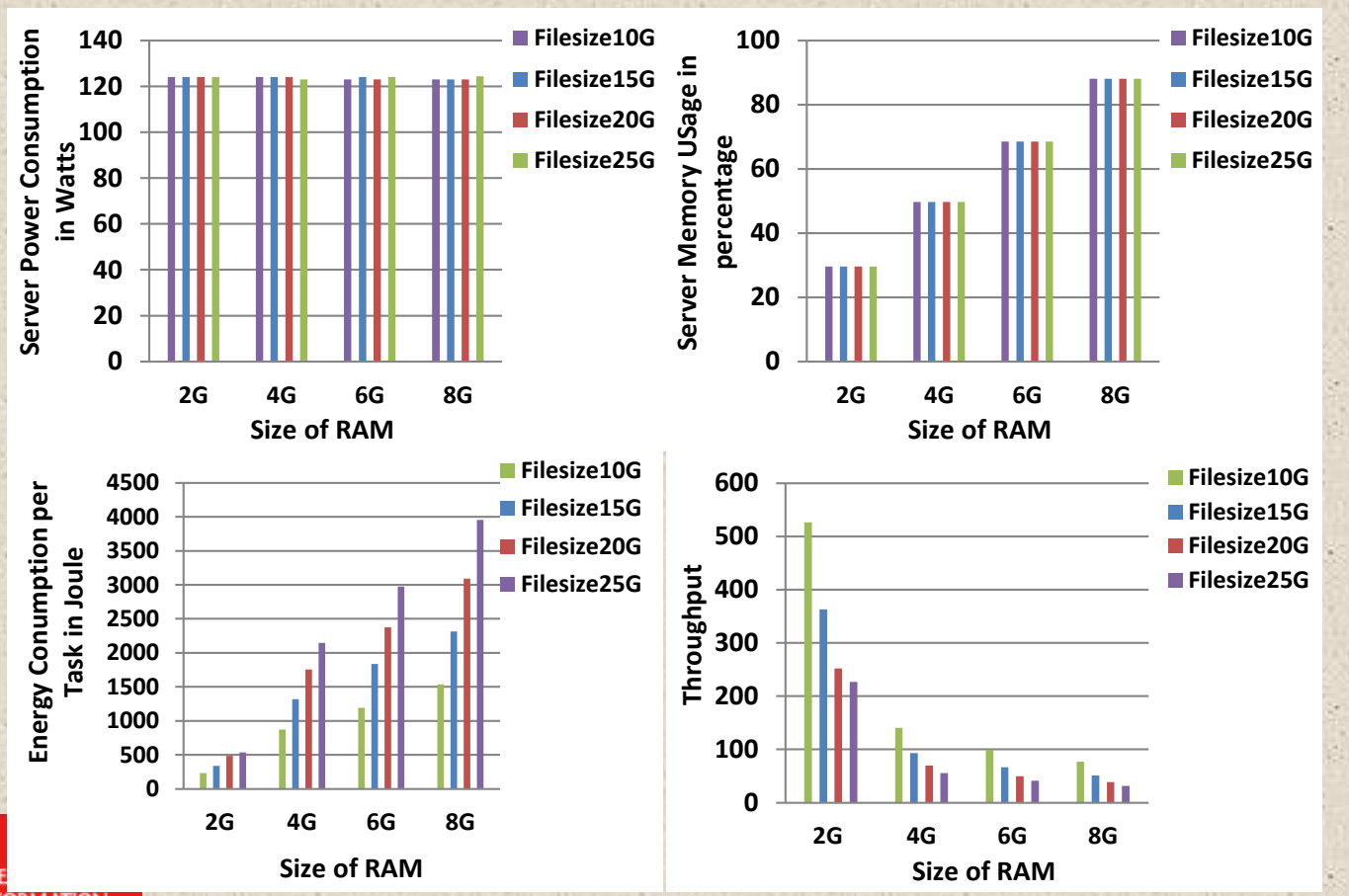




SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Memory-Intensive Workload

- **Test set #1**: Keeping the number of tasks constant, while gradually increasing the size of RAM allocated to the task, and the workload of the task

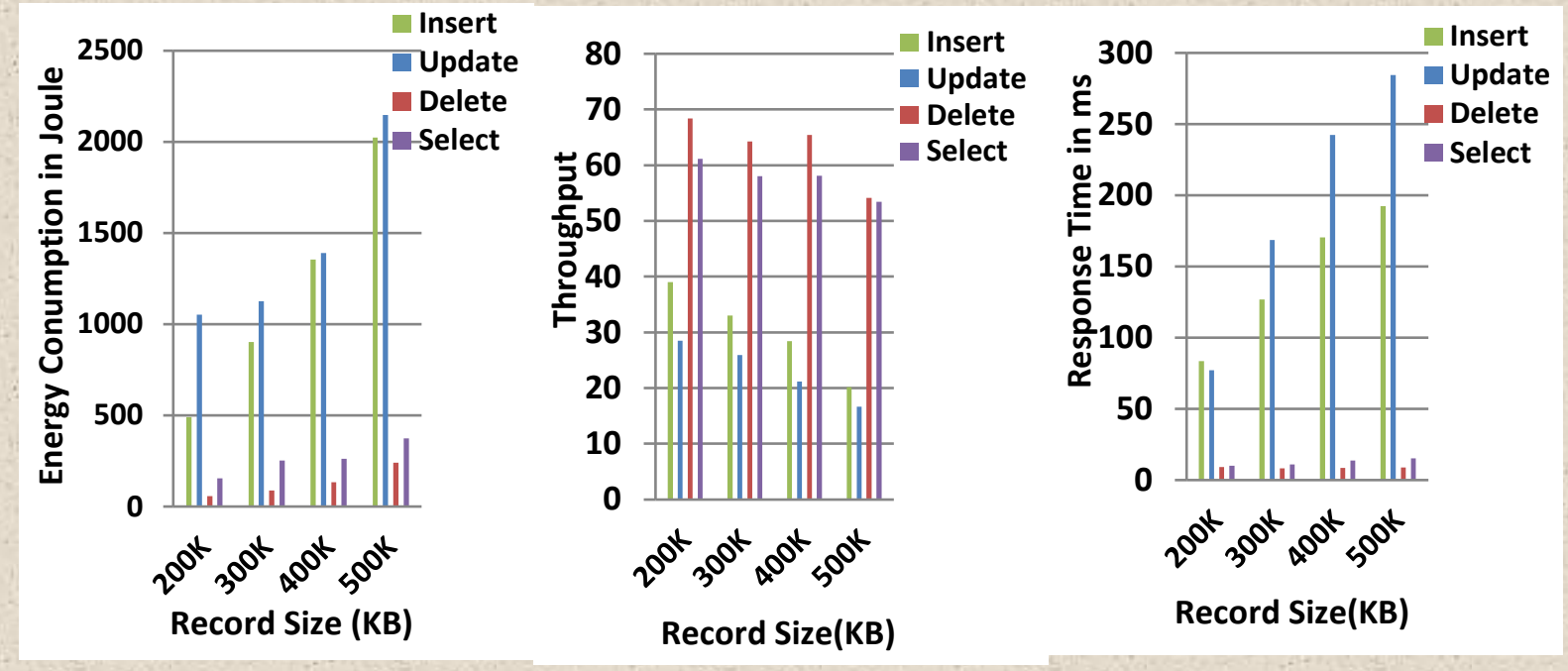


SWIN BURNE

CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

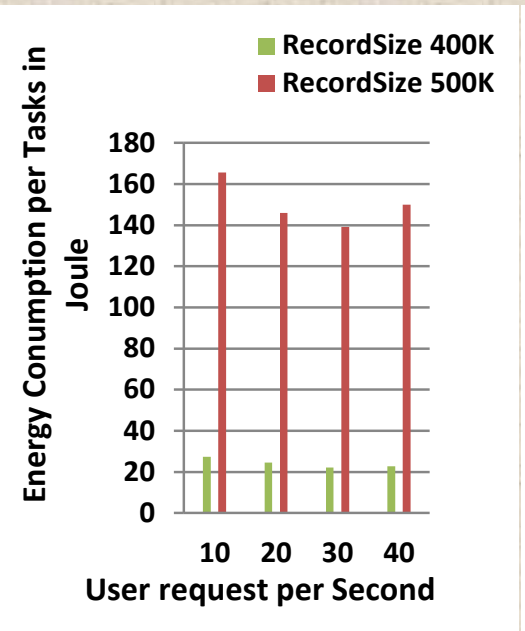
Data-Intensive Workload

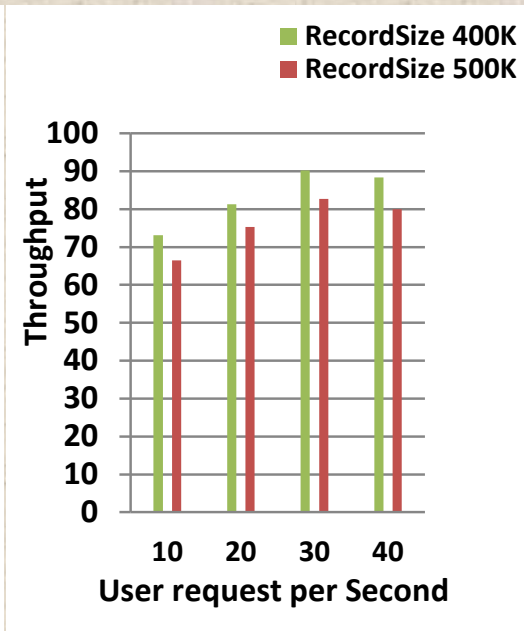
- **Test set #1**: Keeping the number of tasks constant, while gradually increasing the workload of each task

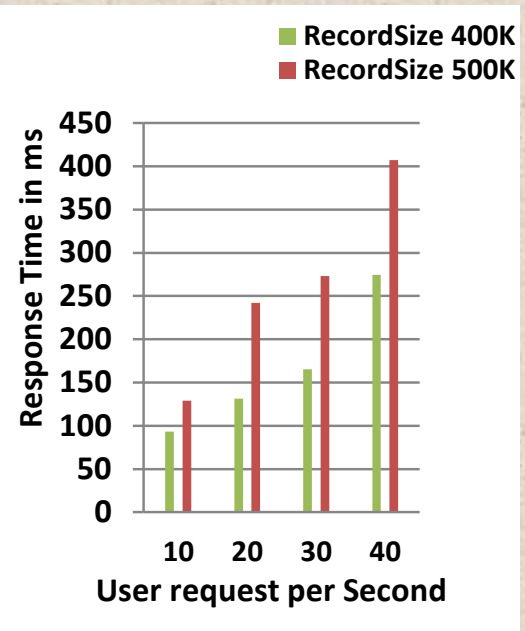


Data-Intensive Workload

- **Test set #2**: Keeping the ratio of each type of operation and total number of tasks constant while changing record size of database requests and user request number per second



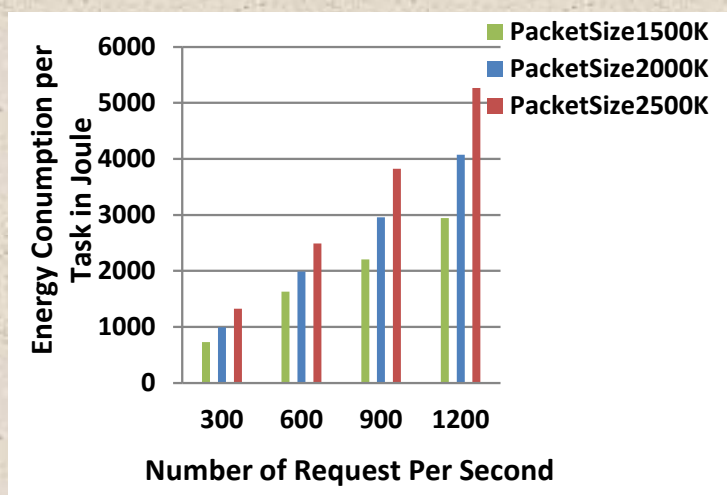


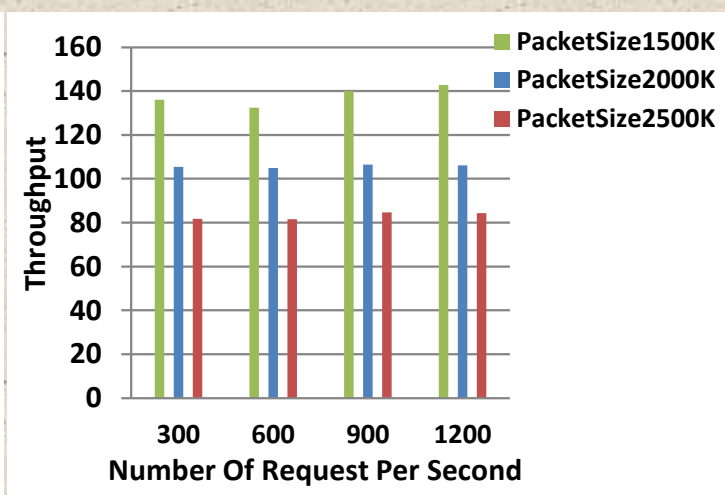


SWINBURNE CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Communication-Intensive Workload

- **Test set #1**: Keep the resource allocation strategy constant while increasing the number of user requests and the packet size of each request

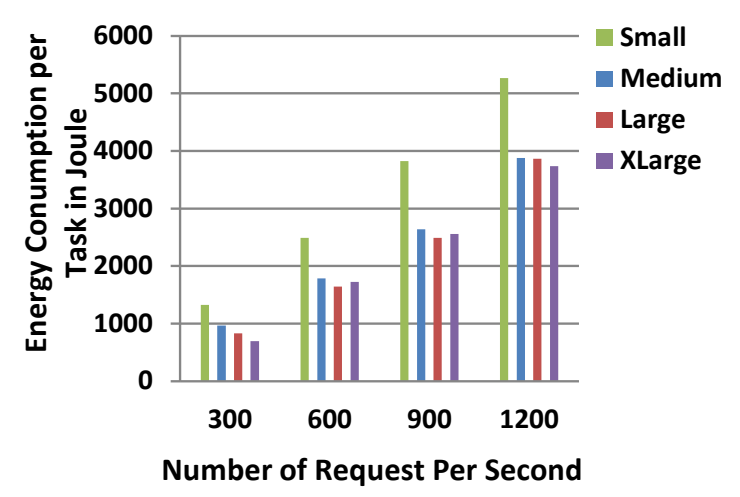


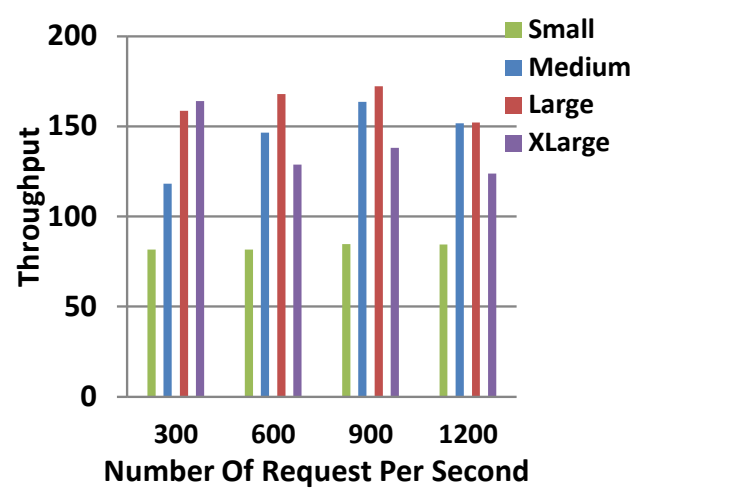


SWINBURNE CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Communication-Intensive Workload

- **Test set #2**: Keep the packet size of each request constant while changing the number of user request per second and resource allocation strategy

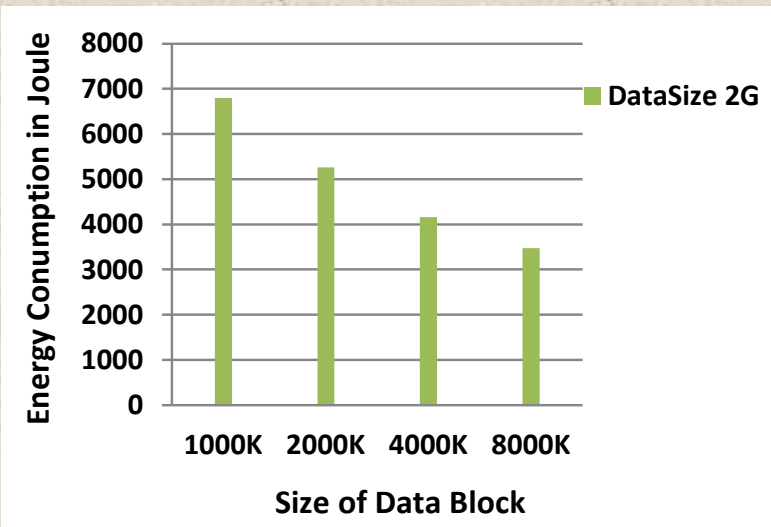


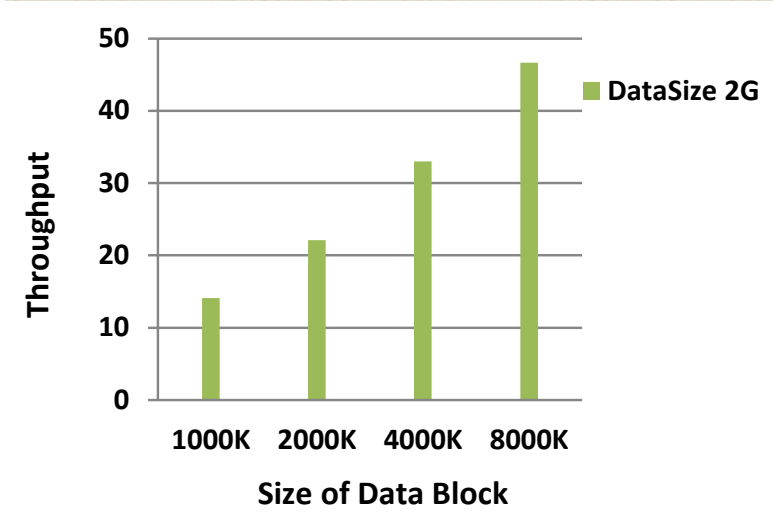


SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Mix Computation- and Data-Intensive Workload

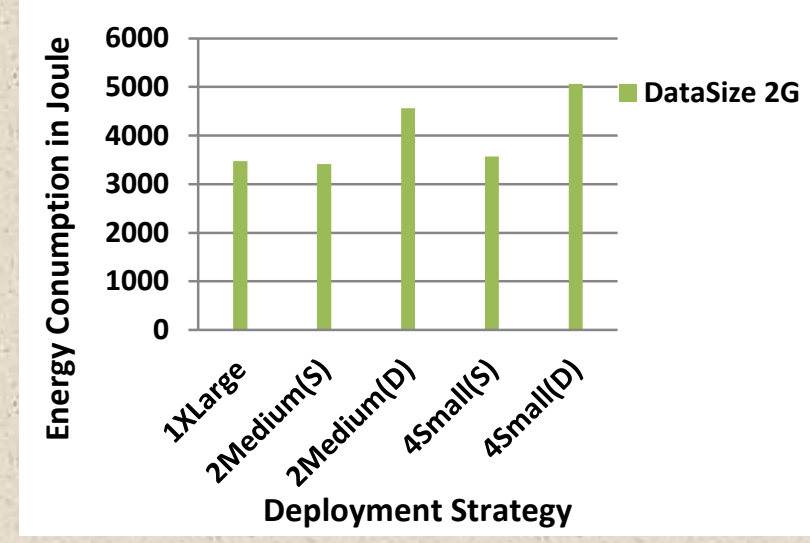
- **Test set #1**: Keep the resource allocation strategy and total amount of data processed constant, while changing the size of each data set.

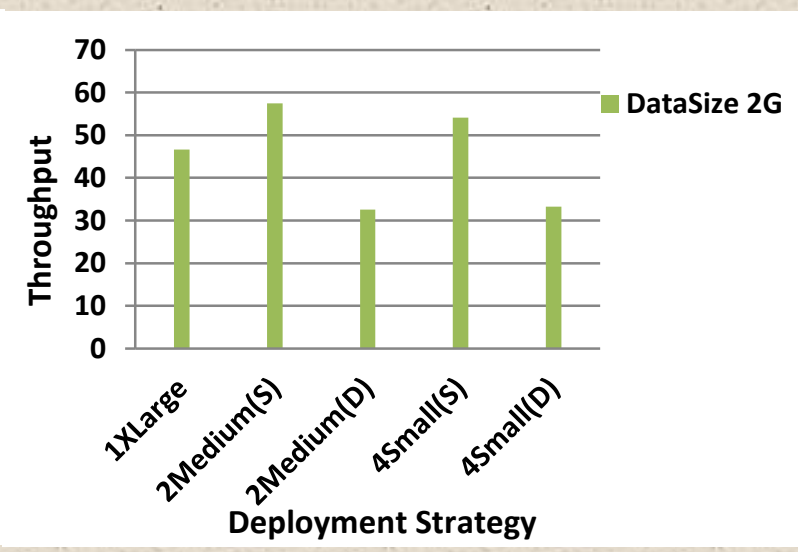




Mix Computation- and Data-Intensive Workload

- **Test set #2**: Keep the total amount of data processed and size of each data set constant, while changing the resource allocation strategy

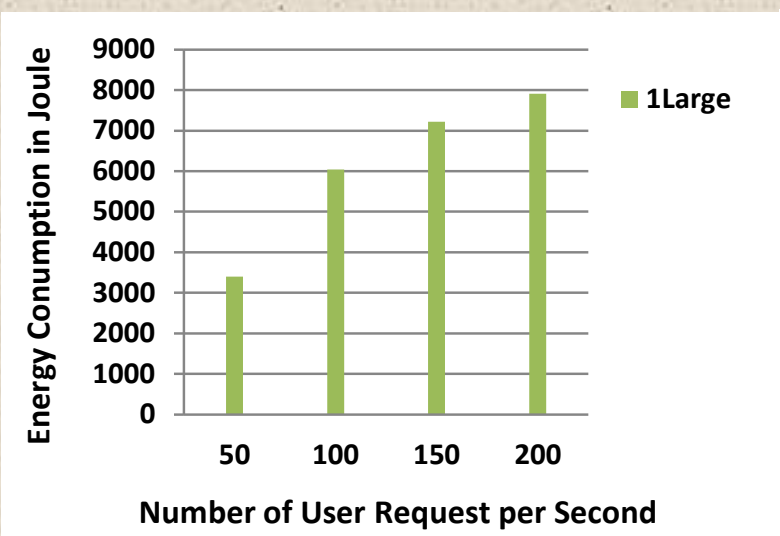


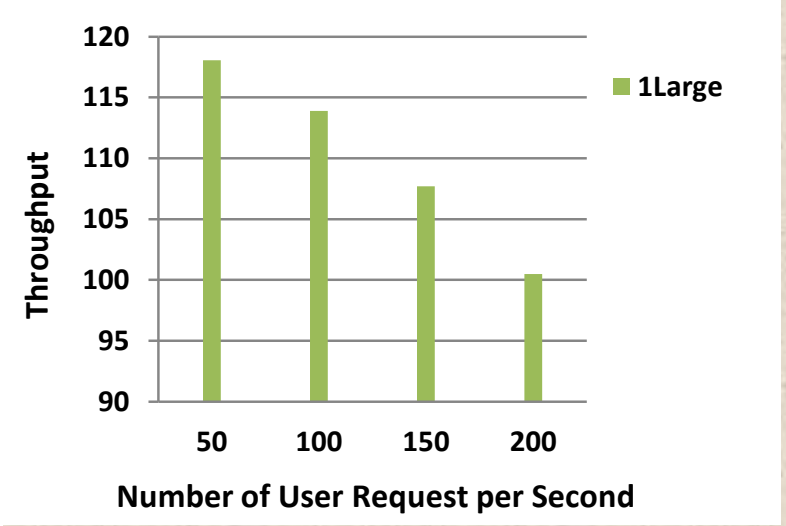


SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Mix Computation-, Data- and Communication-Intensive Workload

- **Test set #1**: : Keep the resource allocation strategy constant while changing workload

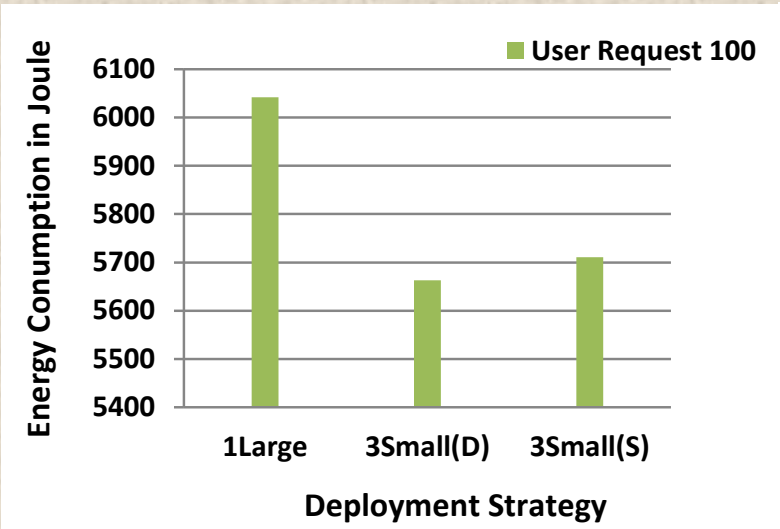


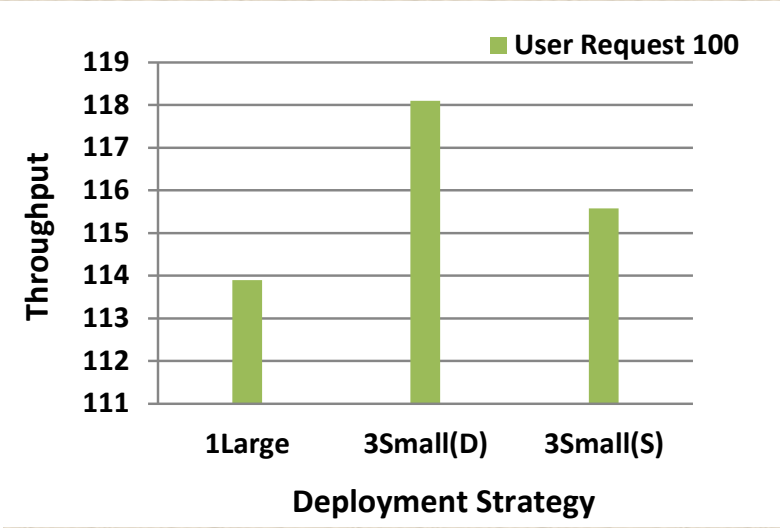


SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Mix Computation-, Data- and Communication-Intensive Workload

- **Test set #2**: Keep the workload constant while changing the resource allocation strategy





SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Agenda

- Introduction
- Approach
- Experiment Setup
- Experimental Results
- Conclusion

SWINBURNE
CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Conclusion

- StressCloud
- Experiment results show energy consumption and system performance are related to:
  - Organisation of Cloud application workload.
  - Type of Cloud application workload.
  - Resource allocation strategy.

SWIN BUR NE

CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

Thank You!

SWINBURNE

CENTRE FOR INFORMATION TECHNOLOGY RESEARCH

References

- Chen, F., Grundy, J.C., Schneider, J-G, Yang, Y. and He, Q. Automated Analysis of Performance and Energy Consumption for Cloud Applications, 2014 ACM/SPEC International Confernce on Performance Engineering, Dublin, Ireland, Mar 23-26 2014, ACM Press.
- Chen, F., Grundy, J.C., Yang, Y., Schneider, J-G., He, Q. Experimental Analysis of Task-based Energy Consumption in Cloud Computing Systems, 4th ACM/SPEC International Conference on Performance Engineering (ICPE2013), April 21-24 2013, Praque, Czech Republic, ACM Press
- Chen, F., Schneider, J-G., Yang, Y., Grundy, J.C. and He, Q. An Energy Consumption Model and Analysis Tool for Cloud Computing Environments, 1st ICSE Workshop on Green and Sustainable Software Engineering, Zurich, Switzerland, 3rd June 2012.
- Cai, Y., Grundy, J.C. and Hosking, J.G. Synthesizing Client Load Models for Performance Engineering via Web Crawling, In Proceedings of the 2007 IEEE/ACM International Conference on Automated Software Engineering, Atlanta, Nov 5-9 2007, IEEE CS Press.
- Draheim, D., Grundy, J.C., Hosking, J.G., Lutteroth, C. and Weber, G. Realistic Load Testing ofWeb Applications, In Proceedings of the 10th European Conference on Software Maintenance and Re-engineering, Berlin, 22-24 March 2006.
- Grundy, J.C., Hosking, J.G., Li, L. And Liu, N. Performance engineering of service compositions, ICSE 2006 Workshop on Service-oriented Software Engineering, Shanghai, May 2006.
- Grundy, J.C., Cai, Y. and Liu, A. SoftArch/MTE: Generating Distributed System Test-beds from High-level Software Architecture Descriptions, Automated Software Engineering, Kluwer Academic Publishers, vol. 12, no. 1, January 2005, pp. 5-39.