A Broker for Cost-efficient QoS aware Resource Allocation in EC2.

Kurt Vermeersch

Coordinator: Kurt Vanmechelen

Universiteit Antwerpen

Thesis Goal

Broker	Seller <-> Buyer
Cost-efficient	Minimize cost
QoS aware	Constraints
Resource Allocation	Scheduler
EC2	Amazon

Mapping of QoS constrained workloads to Amazon pricing models, while trying to **minimize** the total **cost**.

Cloud Computing [1/2]

“Cloud computing is a large-scale **distributed** computing paradigm that is driven by **economies of scale**, in which a pool of abstracted, **virtualized**, **dynamically-scalable**, managed computing power, storage, platforms, and services are delivered **on demand** to external customers over the Internet.”

Ian Foster, Cloud Computing and Grid Computing 360-Degree Compared

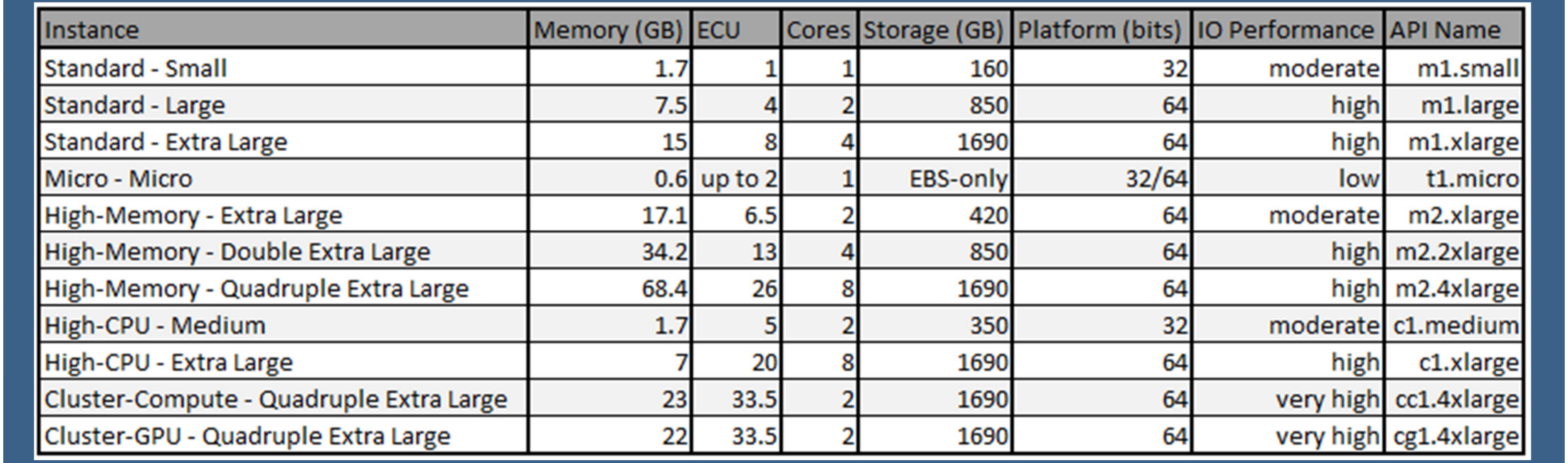
Cloud Computing [2/2]

- Distributed
  - location and device independence
- Economies of Scale
  - less expensive resources
- Virtualized
  - server consolidation
- Dynamically-scalable
  - no over or under provisioning
  - illusion of infinite amount
  - capEx to opEx

Amazon Cloud Computing

- Why?
  - Knowledge
  - Diversification
- What?
  - Public : an off-site third-party cloud provider
  - IaaS: hardware resources are provided
  - Product Portfolio: EC2, S3, etc.
  - ECU: equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron

Amazon Instance Types



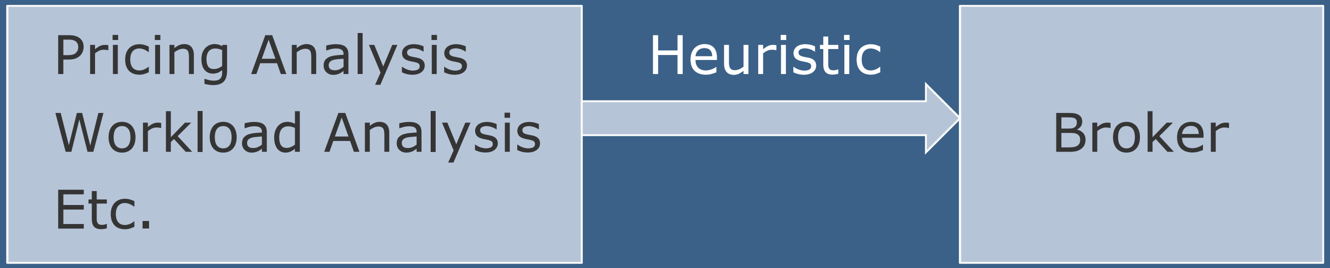
Instance	Memory (GB)	ECU	Cores	Storage (GB)	Platform (bits)	IO Performance	API Name
Standard - Small	1.7	1	1	160	32	moderate	m1.small
Standard - Large	7.5	4	2	850	64	high	m1.large
Standard - Extra Large	15	8	4	1690	64	high	m1.xlarge
Micro - Micro	0.6	up to 2	1	EBS-only	32/64	low	t1.micro
High-Memory - Extra Large	17.1	6.5	2	420	64	moderate	m2.xlarge
High-Memory - Double Extra Large	34.2	13	4	850	64	high	m2.2xlarge
High-Memory - Quadruple Extra Large	68.4	26	8	1690	64	high	m2.4xlarge
High-CPU - Medium	1.7	5	2	350	32	moderate	c1.medium
High-CPU - Extra Large	7	20	8	1690	64	high	c1.xlarge
Cluster-Compute - Quadruple Extra Large	23	33.5	2	1690	64	very high	cc1.4xlarge
Cluster-GPU - Quadruple Extra Large	22	33.5	2	1690	64	very high	cg1.4xlarge

Amazon Pricing

- Four Regions
  - US East, US West, EU and Asia Pacific
- On-Demand
  - Fixed hourly charging rate
  - Guaranteed to stay alive
- Reserved
  - Upfront payment (1y/3y), lower fixed hourly rate
  - Guaranteed to be available for launch
- Spot
  - Varying hourly rate
  - Instance can be terminated

Environmental Analysis

Variation in products, instances and pricing models complicates the mapping from workload to optimal resource division

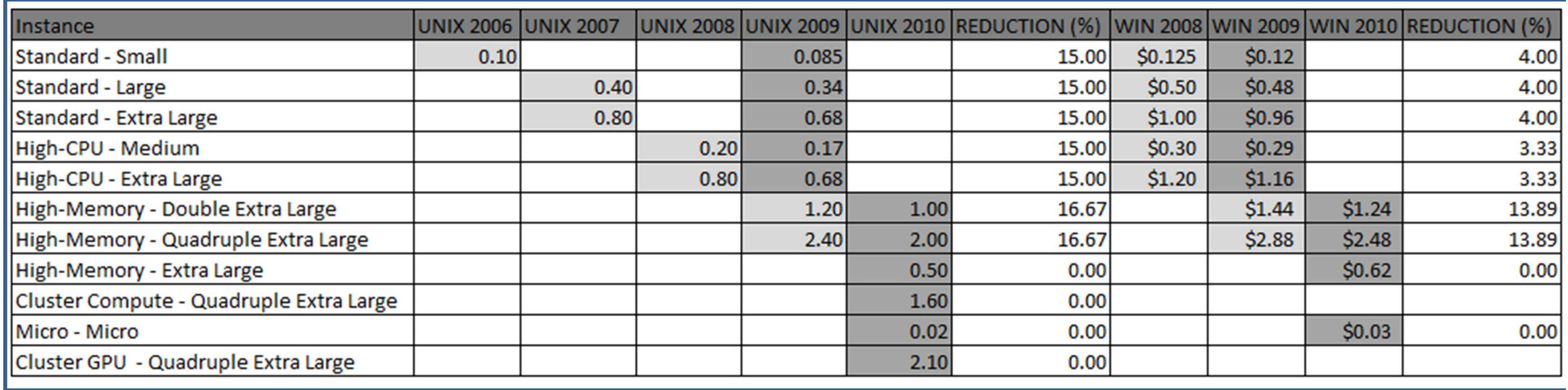


Pricing Analysis
Workload Analysis
Etc.

Heuristic

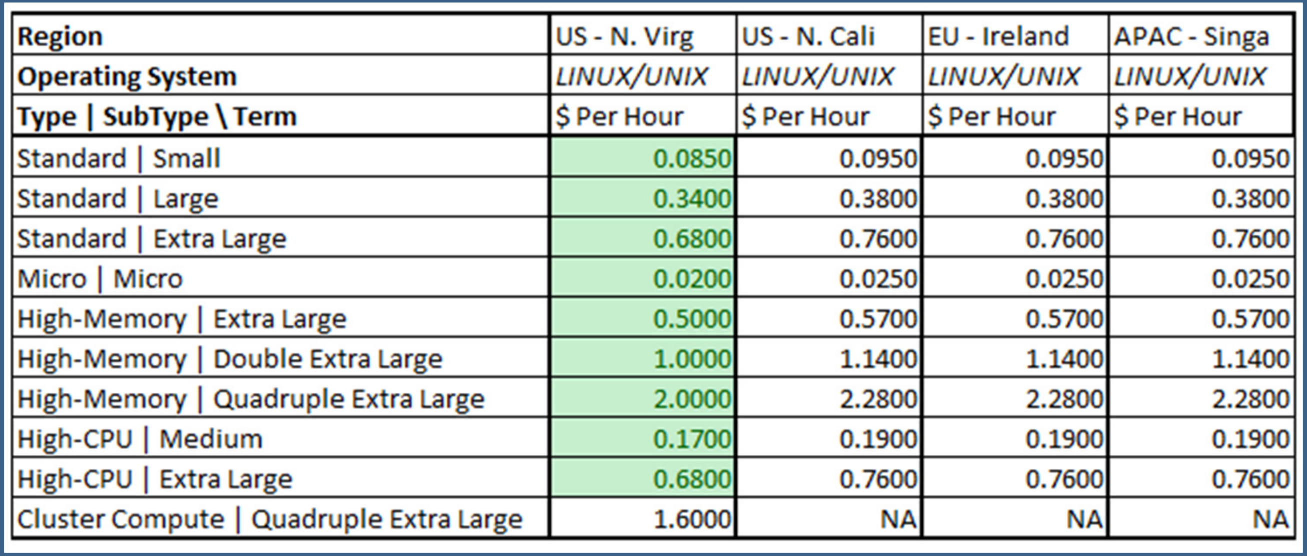
Broker

On-Demand Pricing Evolution



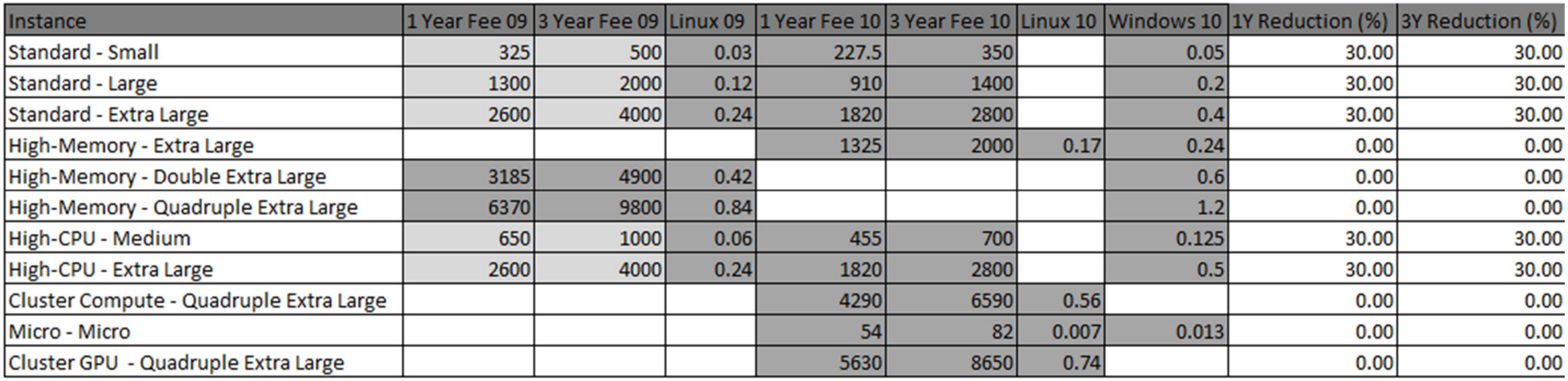
Instance	UNIX 2006	UNIX 2007	UNIX 2008	UNIX 2009	UNIX 2010	REDUCTION (%)	WIN 2008	WIN 2009	WIN 2010	REDUCTION (%)
Standard - Small	0.10			0.085		15.00	$0.125	$0.12		4.00
Standard - Large		0.40		0.34		15.00	$0.50	$0.48		4.00
Standard - Extra Large		0.80		0.68		15.00	$1.00	$0.96		4.00
High-CPU - Medium			0.20	0.17		15.00	$0.30	$0.29		3.33
High-CPU - Extra Large			0.80	0.68		15.00	$1.20	$1.16		3.33
High-Memory - Double Extra Large				1.20	1.00	16.67		$1.44	$1.24	13.89
High-Memory - Quadruple Extra Large				2.40	2.00	16.67		$2.88	$2.48	13.89
High-Memory - Extra Large					0.50	0.00			$0.62	0.00
Cluster Compute - Quadruple Extra Large					1.60	0.00				
Micro - Micro					0.02	0.00			$0.03	0.00
Cluster GPU - Quadruple Extra Large					2.10	0.00				

On-Demand Region Comparison



Region	US - N. Virg	US - N. Cali	EU - Ireland	APAC - Singa
Operating System	*LINUX/UNIX*	*LINUX/UNIX*	*LINUX/UNIX*	*LINUX/UNIX*
Type | SubType \ Term	$ Per Hour	$ Per Hour	$ Per Hour	$ Per Hour
Standard | Small	0.0850	0.0950	0.0950	0.0950
Standard | Large	0.3400	0.3800	0.3800	0.3800
Standard | Extra Large	0.6800	0.7600	0.7600	0.7600
Micro | Micro	0.0200	0.0250	0.0250	0.0250
High-Memory | Extra Large	0.5000	0.5700	0.5700	0.5700
High-Memory | Double Extra Large	1.0000	1.1400	1.1400	1.1400
High-Memory | Quadruple Extra Large	2.0000	2.2800	2.2800	2.2800
High-CPU | Medium	0.1700	0.1900	0.1900	0.1900
High-CPU | Extra Large	0.6800	0.7600	0.7600	0.7600
Cluster Compute | Quadruple Extra Large	1.6000	NA	NA	NA

Reserved Pricing Evolution



Instance	1 Year Fee 09	3 Year Fee 09	Linux 09	1 Year Fee 10	3 Year Fee 10	Linux 10	Windows 10	1Y Reduction (%)	3Y Reduction (%)
Standard - Small	325	500	0.03	227.5	350		0.05	30.00	30.00
Standard - Large	1300	2000	0.12	910	1400		0.2	30.00	30.00
Standard - Extra Large	2600	4000	0.24	1820	2800		0.4	30.00	30.00
High-Memory - Extra Large				1325	2000	0.17	0.24	0.00	0.00
High-Memory - Double Extra Large	3185	4900	0.42				0.6	0.00	0.00
High-Memory - Quadruple Extra Large	6370	9800	0.84				1.2	0.00	0.00
High-CPU - Medium	650	1000	0.06	455	700		0.125	30.00	30.00
High-CPU - Extra Large	2600	4000	0.24	1820	2800		0.5	30.00	30.00
Cluster Compute - Quadruple Extra Large				4290	6590	0.56		0.00	0.00
Micro - Micro				54	82	0.007	0.013	0.00	0.00
Cluster GPU - Quadruple Extra Large				5630	8650	0.74		0.00	0.00

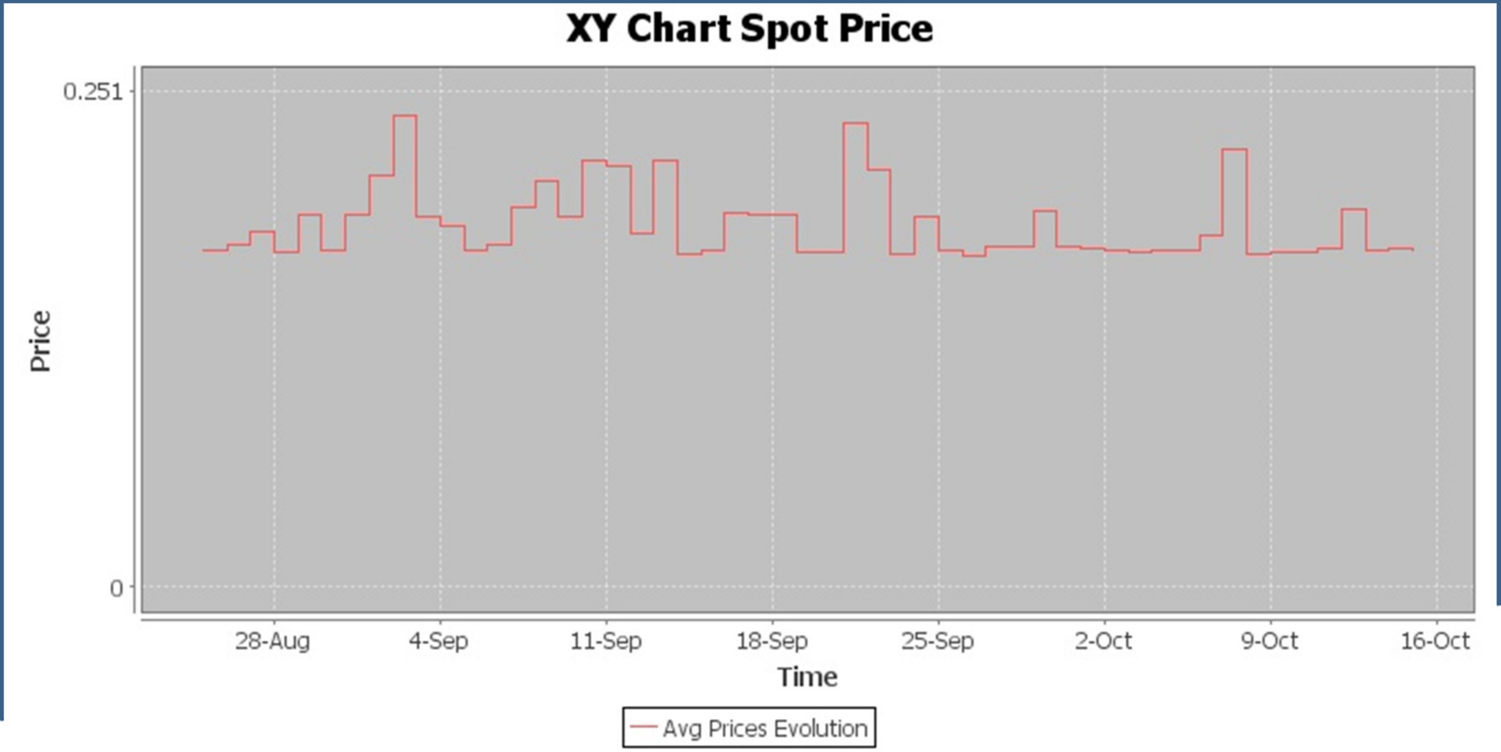
Spot Analysis

- Spot price history => CSV files
  - Cloudexchange.org
  - EC2 API

- Statistical analysis of Spot price evolution
  - High Memory Extra Large in US-East
- Analysis using average Spot prices

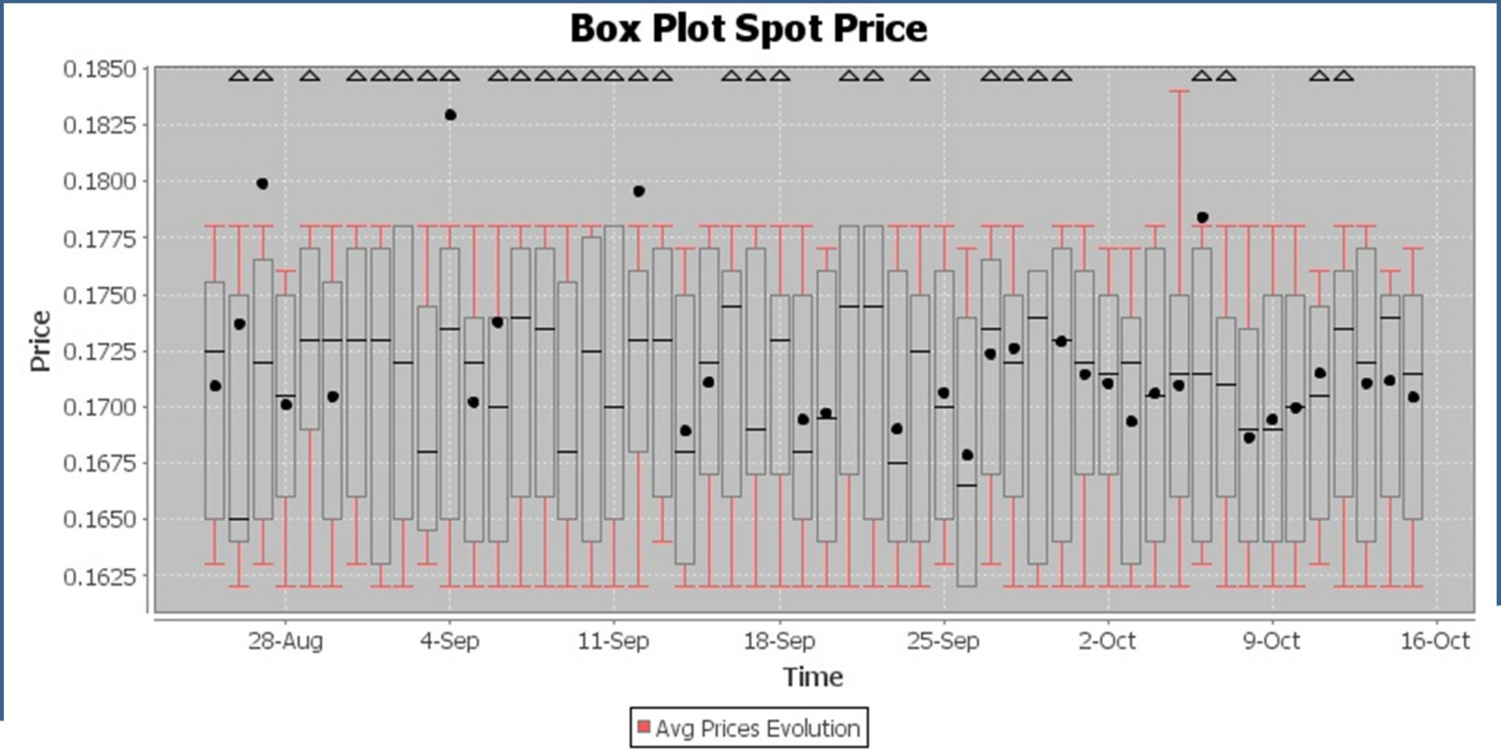
Spot History [1/4]

- fluctuating average price



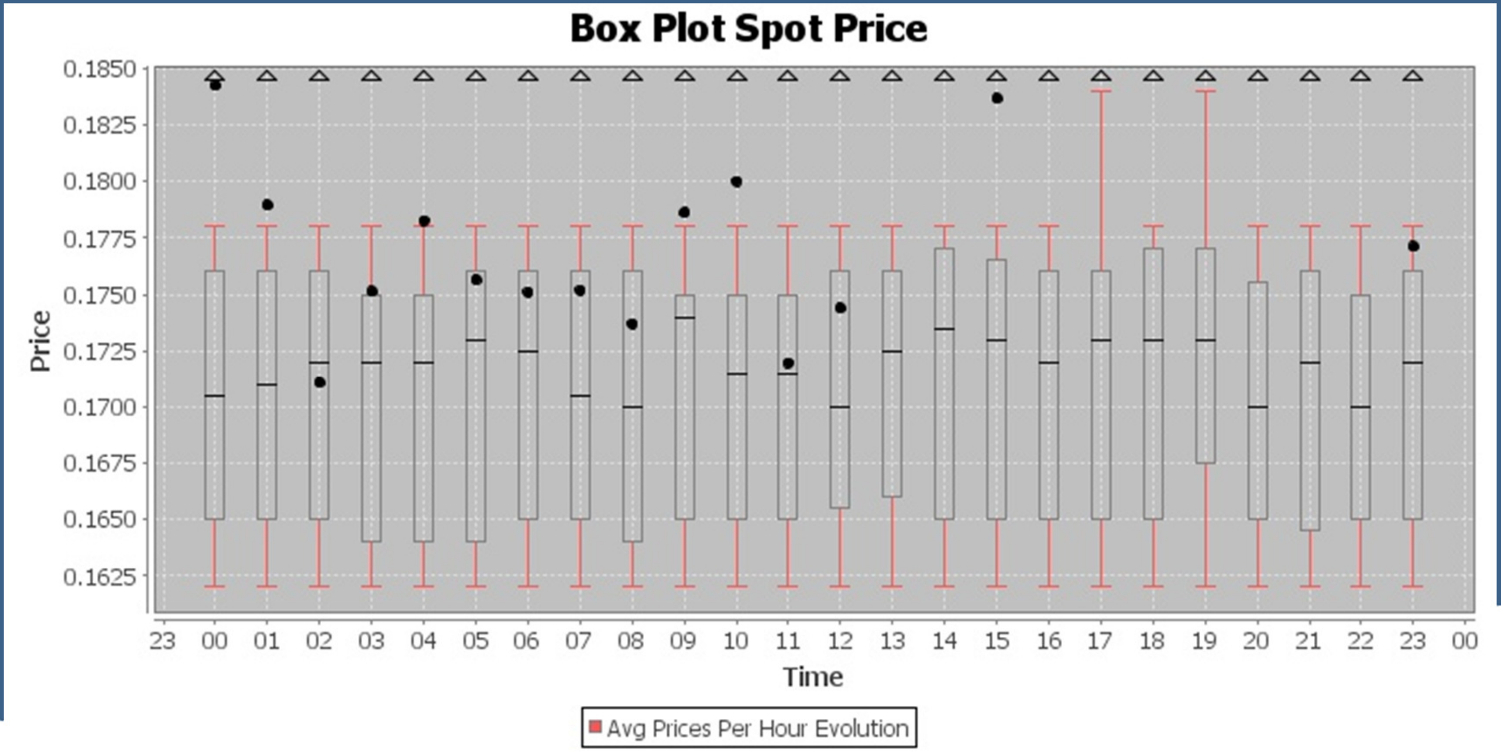
Spot History [2/4]

- Aligned boxplot percentiles => outliers



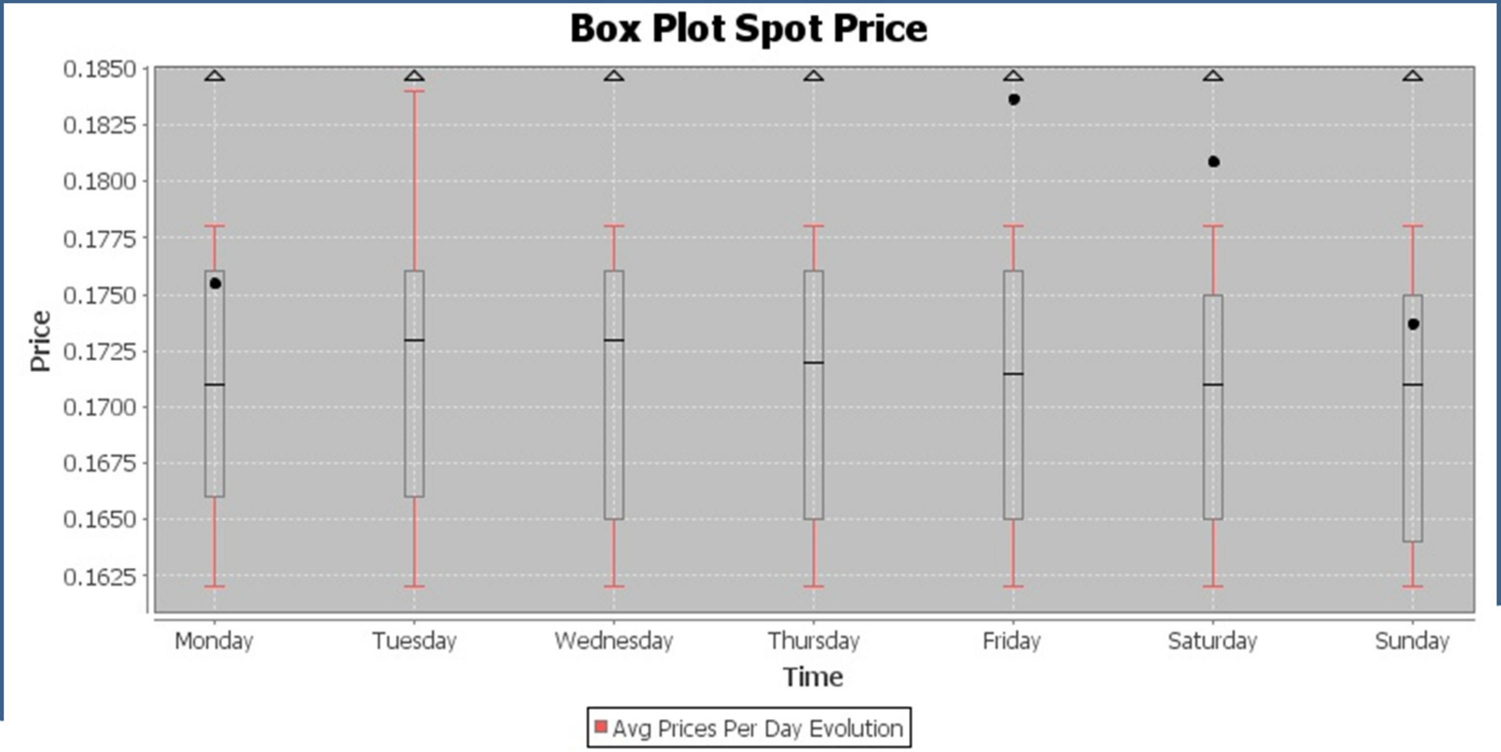
Spot History [3/4]

- Only small differences during the day

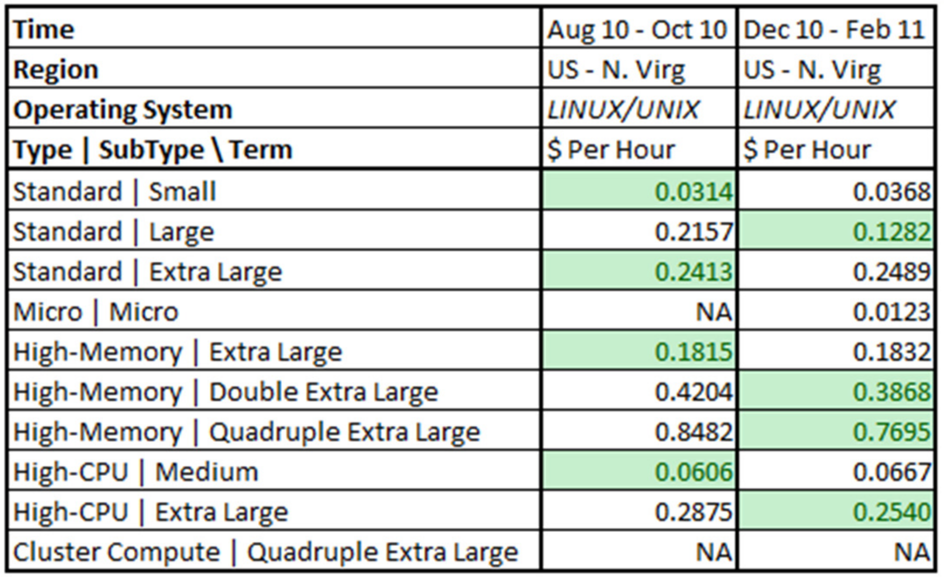


Spot History [4/4]

- Only a little cheaper during the weekend

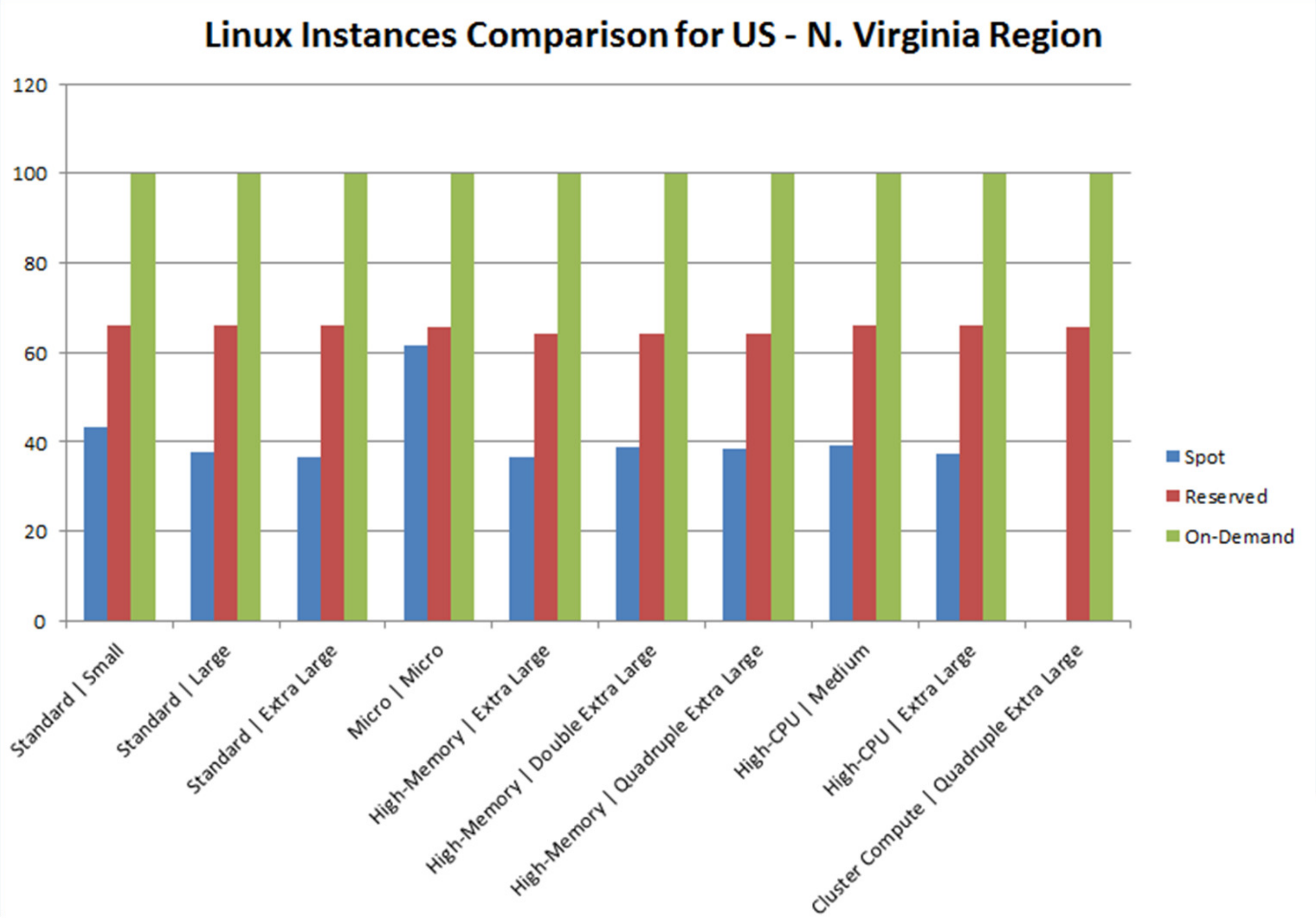


Spot Average [1/3]

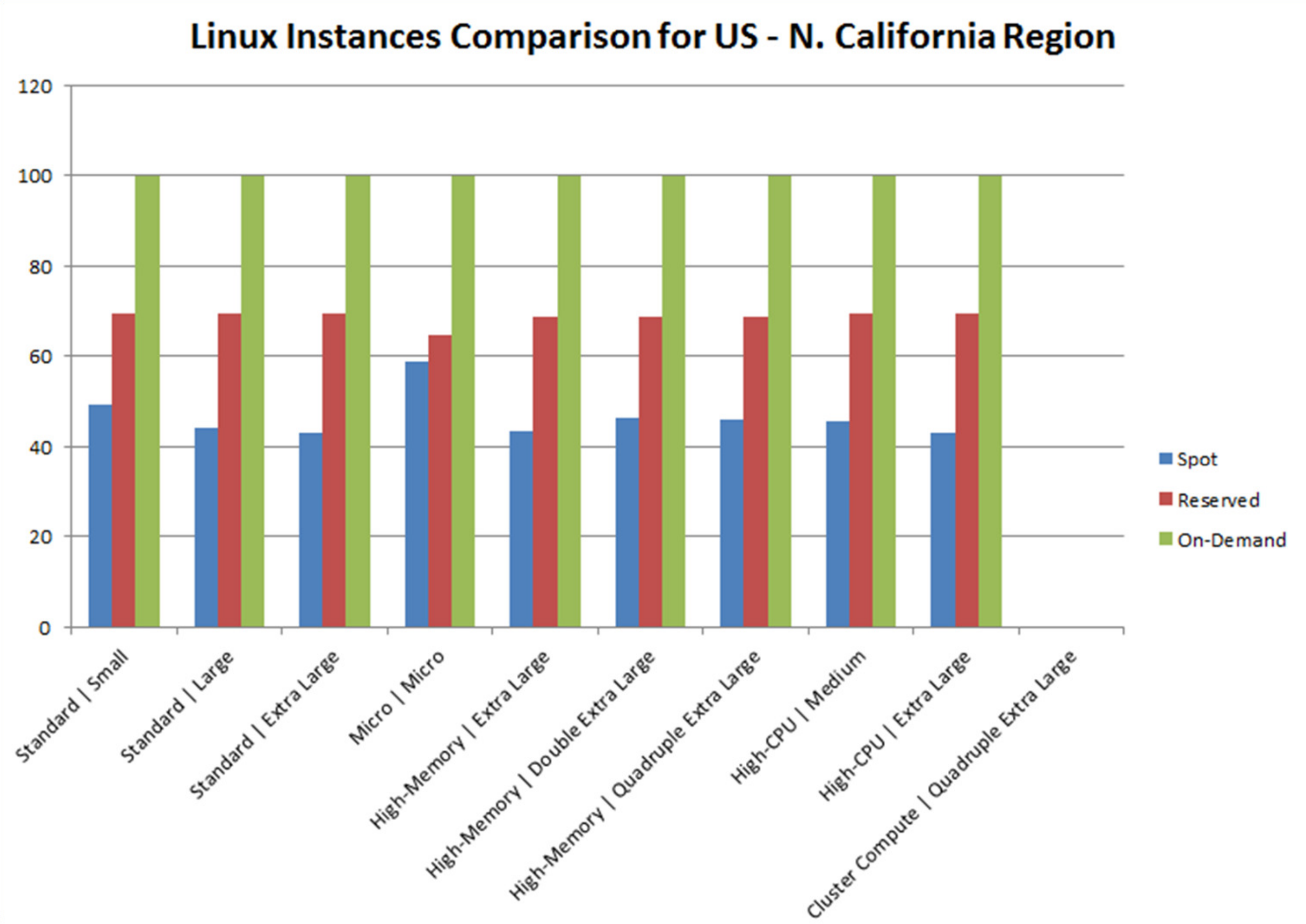


Time	Aug 10 - Oct 10	Dec 10 - Feb 11
Region	US - N. Virg	US - N. Virg
Operating System	*LINUX/UNIX*	*LINUX/UNIX*
Type | SubType \ Term	$ Per Hour	$ Per Hour
Standard | Small	0.0314	0.0368
Standard | Large	0.2157	0.1282
Standard | Extra Large	0.2413	0.2489
Micro | Micro	NA	0.0123
High-Memory | Extra Large	0.1815	0.1832
High-Memory | Double Extra Large	0.4204	0.3868
High-Memory | Quadruple Extra Large	0.8482	0.7695
High-CPU | Medium	0.0606	0.0667
High-CPU | Extra Large	0.2875	0.2540
Cluster Compute | Quadruple Extra Large	NA	NA

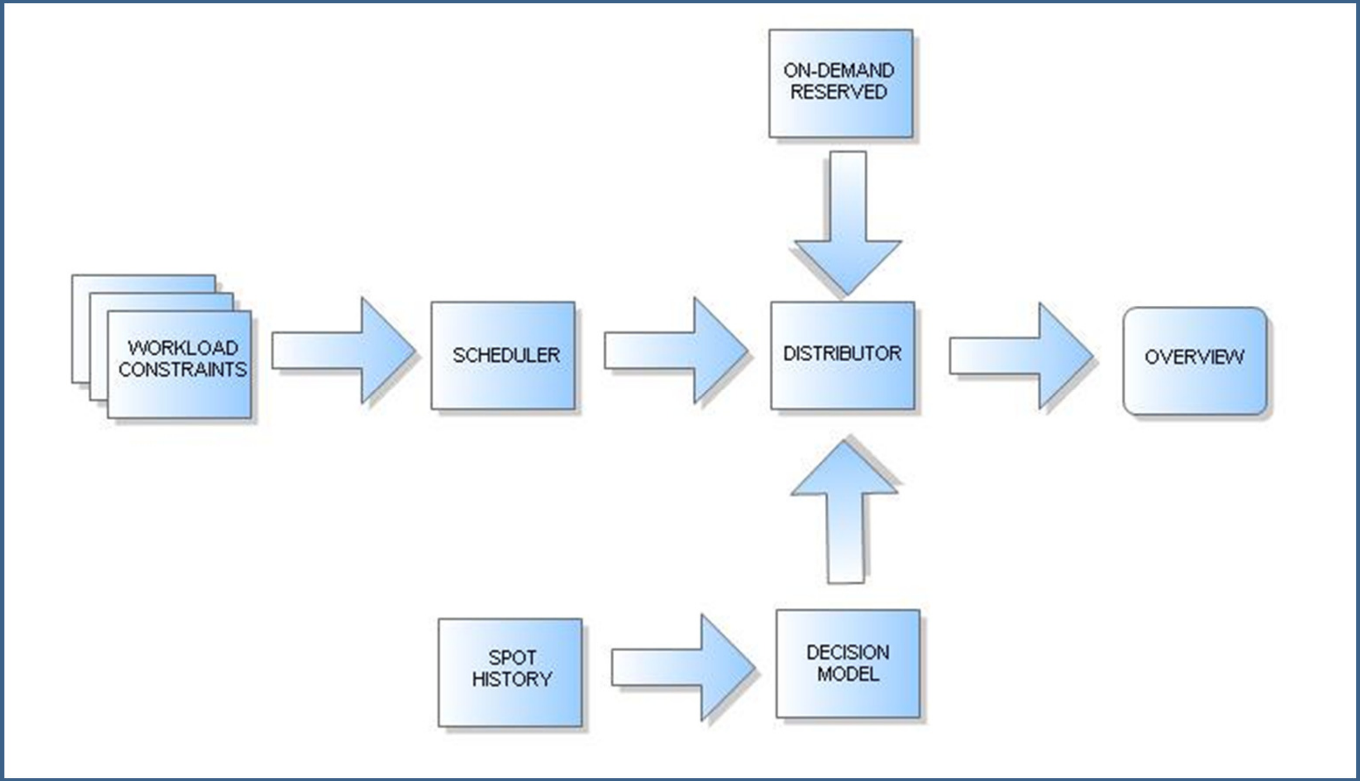
Spot Average - Regions [2/3]



Spot Average – Regions [3/3]

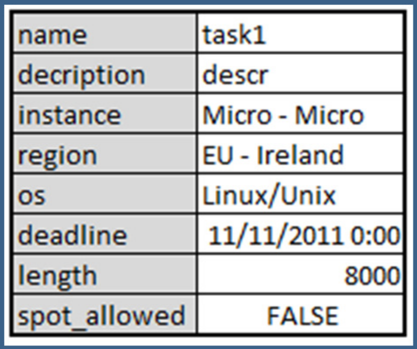


Broker Design



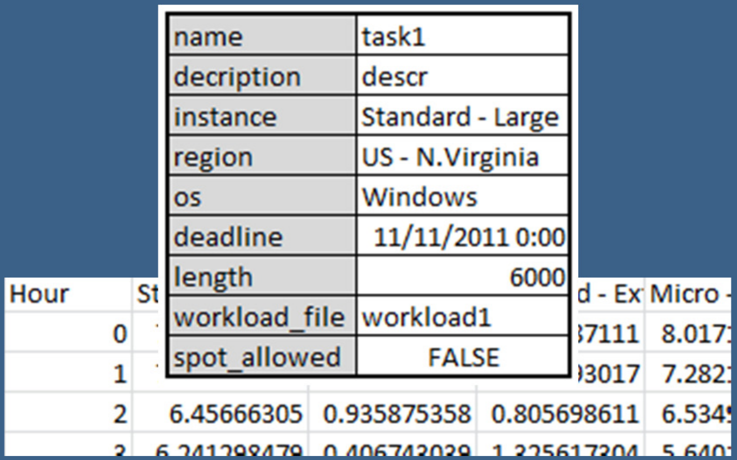
Workload & Constraints

W1: total VM hours specified



name	task1
decription	descr
instance	Micro - Micro
region	EU - Ireland
os	Linux/Unix
deadline	11/11/2011 0:00
length	8000
spot_allowed	FALSE

W2: Every hour # VMs specified



name	task1
decription	descr
instance	Standard - Large
region	US - N.Virginia
os	Windows
deadline	11/11/2011 0:00
length	6000
workload_file	workload1
spot_allowed	FALSE

Hour	St		d - Ex	Micro -
0			37111	8.0171
1			93017	7.2821
2	6.45666305	0.935875358	0.805698611	6.5349
3	6.241298479	0.406743039	1.325617304	5.6401

Spot Decision Model

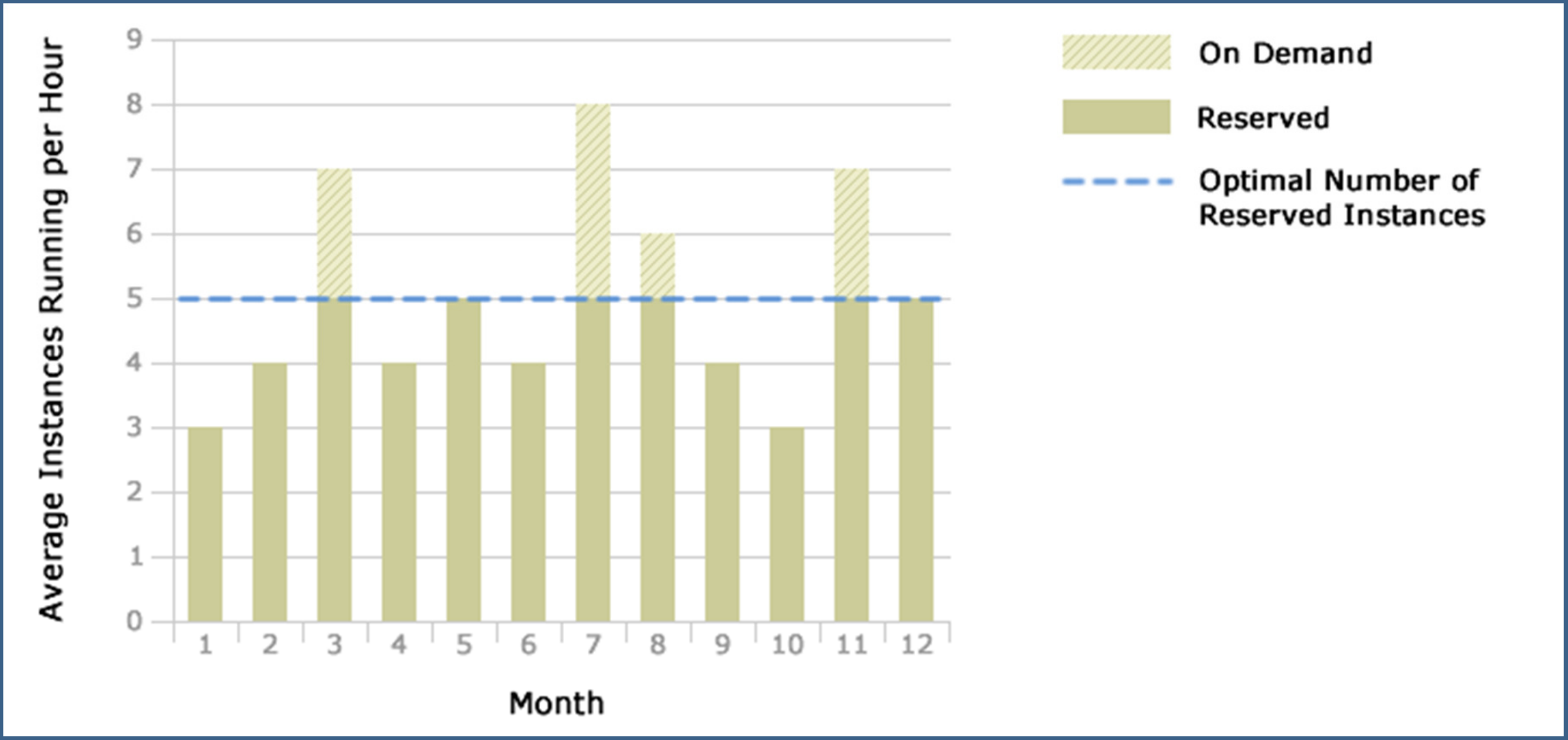
- Based on empirical data
- Checkpointing schemes
  - Hourly
  - Optimal
- Decision model determines spot bid that minimizes cost but ensures successful execution in terms of workload constraints
  - Java port: memory problem fix

Distributor

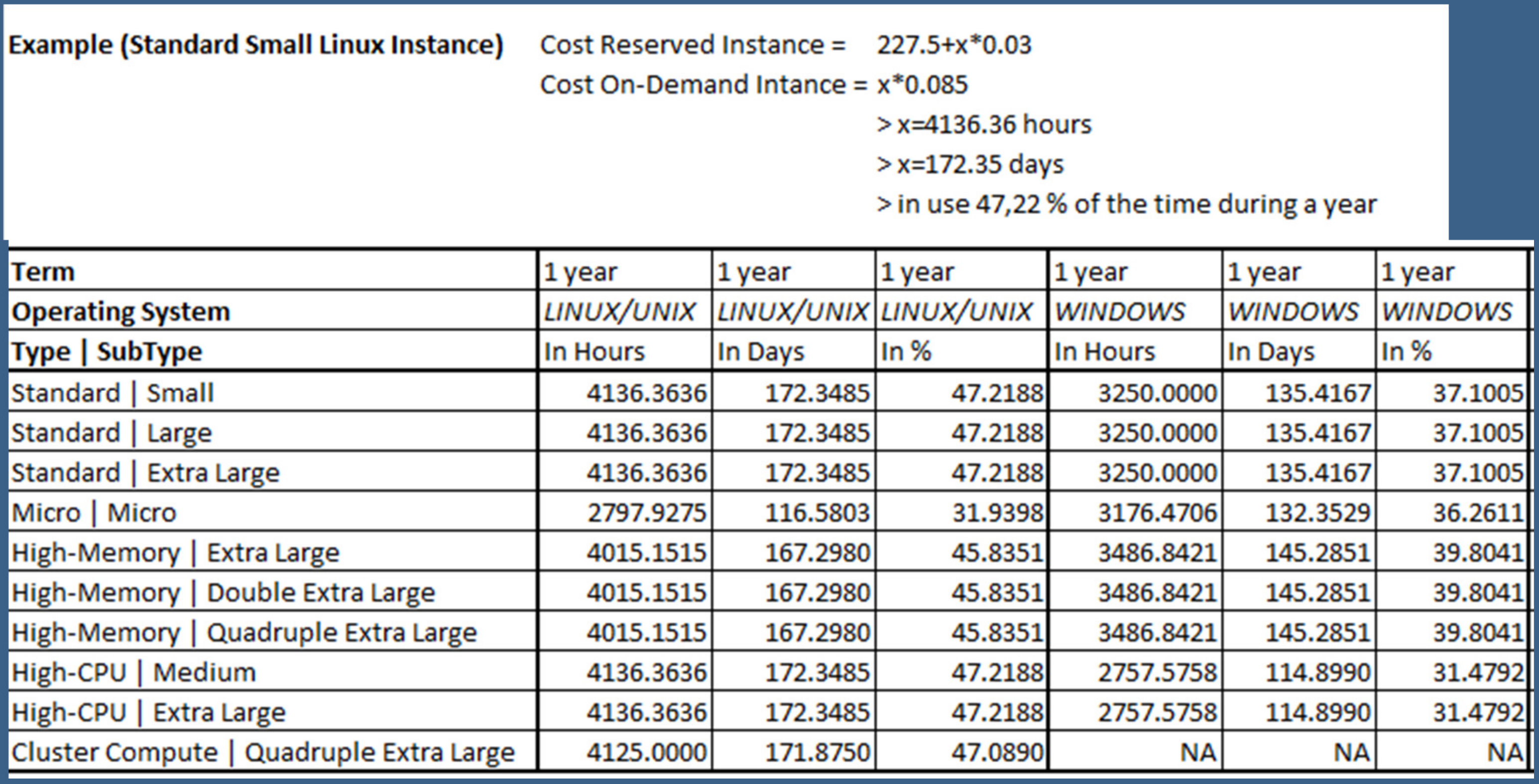
- Make division between different pricing models after scheduling has occurred
  - On-Demand vs Reserved -> optimal division possible
  - Spot vs Reserved -> optimal choice spot

- Spot is not always the best choice cfr. constraints.

On-Demand vs Reserved [1/3]



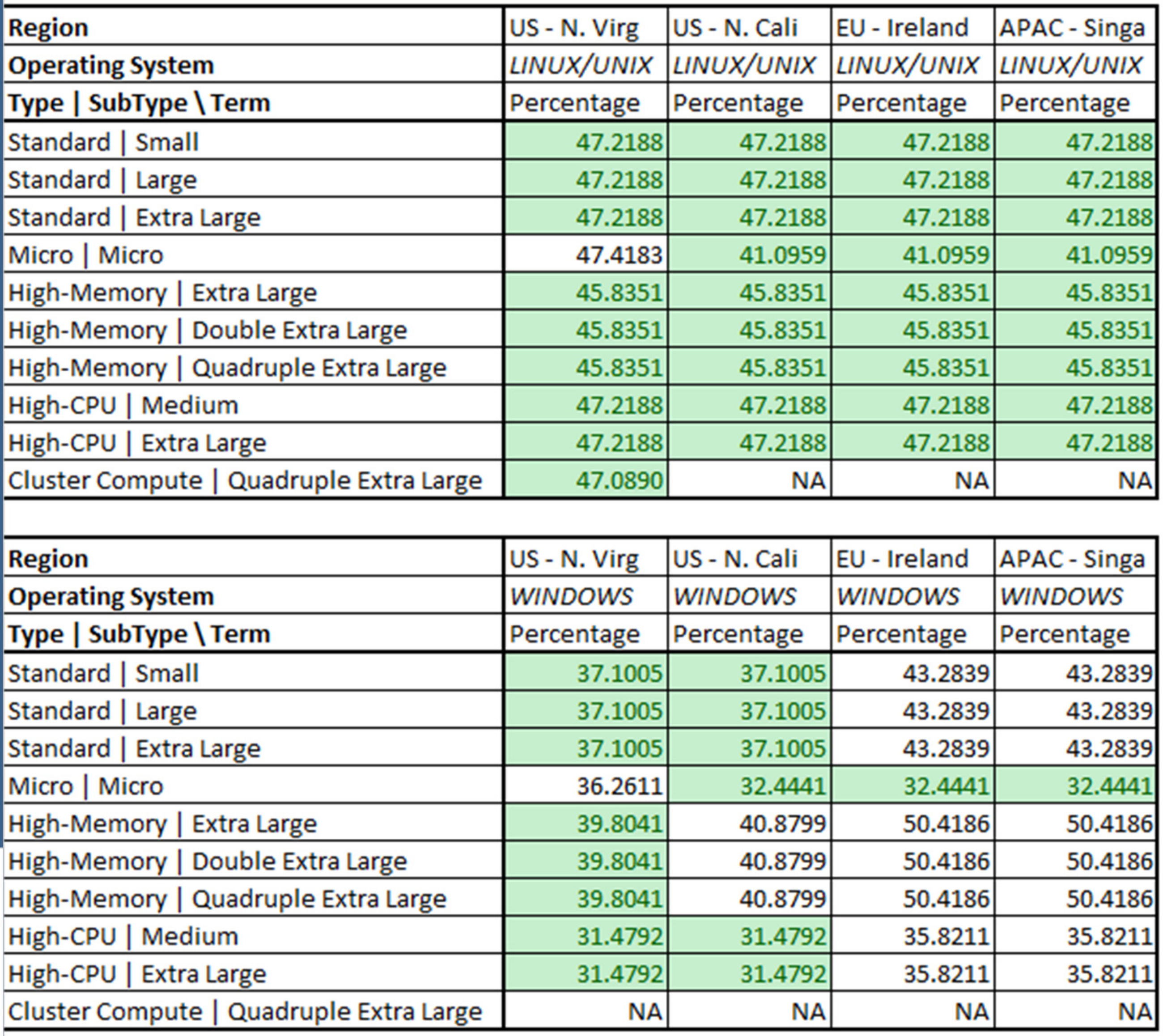
On-Demand vs Reserved [2/3]



Example (Standard Small Linux Instance) Cost Reserved Instance = 227.5+x*0.03
Cost On-Demand Intance = x*0.085
> x=4136.36 hours
> x=172.35 days
> in use 47,22 % of the time during a year

Term	1 year	1 year	1 year	1 year	1 year	1 year
Operating System	*LINUX/UNIX*	*LINUX/UNIX*	*LINUX/UNIX*	*WINDOWS*	*WINDOWS*	*WINDOWS*
Type | SubType	In Hours	In Days	In %	In Hours	In Days	In %
Standard | Small	4136.3636	172.3485	47.2188	3250.0000	135.4167	37.1005
Standard | Large	4136.3636	172.3485	47.2188	3250.0000	135.4167	37.1005
Standard | Extra Large	4136.3636	172.3485	47.2188	3250.0000	135.4167	37.1005
Micro | Micro	2797.9275	116.5803	31.9398	3176.4706	132.3529	36.2611
High-Memory | Extra Large	4015.1515	167.2980	45.8351	3486.8421	145.2851	39.8041
High-Memory | Double Extra Large	4015.1515	167.2980	45.8351	3486.8421	145.2851	39.8041
High-Memory | Quadruple Extra Large	4015.1515	167.2980	45.8351	3486.8421	145.2851	39.8041
High-CPU | Medium	4136.3636	172.3485	47.2188	2757.5758	114.8990	31.4792
High-CPU | Extra Large	4136.3636	172.3485	47.2188	2757.5758	114.8990	31.4792
Cluster Compute | Quadruple Extra Large	4125.0000	171.8750	47.0890	NA	NA	NA

On-Demand vs Reserved [3/3]

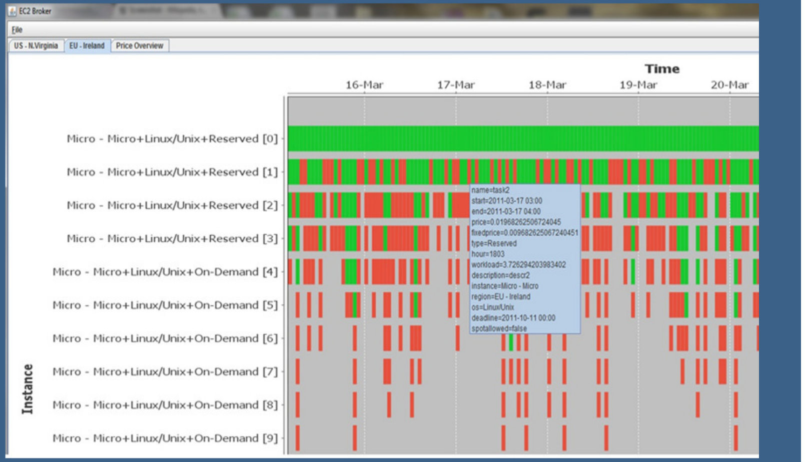


Region	US - N. Virg	US - N. Cali	EU - Ireland	APAC - Singa
Operating System	*LINUX/UNIX*	*LINUX/UNIX*	*LINUX/UNIX*	*LINUX/UNIX*
Type | SubType \ Term	Percentage	Percentage	Percentage	Percentage
Standard | Small	47.2188	47.2188	47.2188	47.2188
Standard | Large	47.2188	47.2188	47.2188	47.2188
Standard | Extra Large	47.2188	47.2188	47.2188	47.2188
Micro | Micro	47.4183	41.0959	41.0959	41.0959
High-Memory | Extra Large	45.8351	45.8351	45.8351	45.8351
High-Memory | Double Extra Large	45.8351	45.8351	45.8351	45.8351
High-Memory | Quadruple Extra Large	45.8351	45.8351	45.8351	45.8351
High-CPU | Medium	47.2188	47.2188	47.2188	47.2188
High-CPU | Extra Large	47.2188	47.2188	47.2188	47.2188
Cluster Compute | Quadruple Extra Large	47.0890	NA	NA	NA

Region	US - N. Virg	US - N. Cali	EU - Ireland	APAC - Singa
Operating System	*WINDOWS*	*WINDOWS*	*WINDOWS*	*WINDOWS*
Type | SubType \ Term	Percentage	Percentage	Percentage	Percentage
Standard | Small	37.1005	37.1005	43.2839	43.2839
Standard | Large	37.1005	37.1005	43.2839	43.2839
Standard | Extra Large	37.1005	37.1005	43.2839	43.2839
Micro | Micro	36.2611	32.4441	32.4441	32.4441
High-Memory | Extra Large	39.8041	40.8799	50.4186	50.4186
High-Memory | Double Extra Large	39.8041	40.8799	50.4186	50.4186
High-Memory | Quadruple Extra Large	39.8041	40.8799	50.4186	50.4186
High-CPU | Medium	31.4792	31.4792	35.8211	35.8211
High-CPU | Extra Large	31.4792	31.4792	35.8211	35.8211
Cluster Compute | Quadruple Extra Large	NA	NA	NA	NA

Scheduler

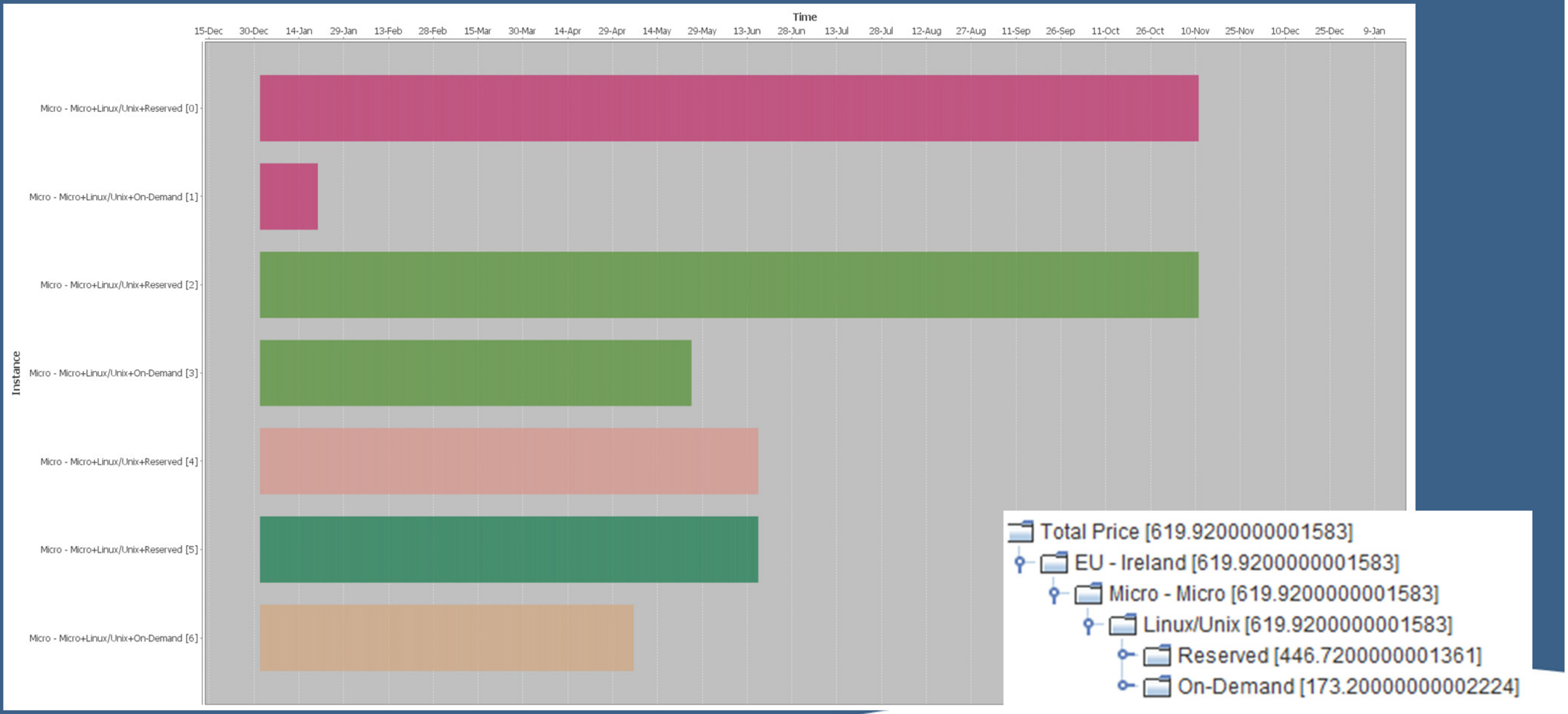
- Capacity fragmentation
- Based on workload model
- Computation intensive


EC2 Broker
File
US - N.Virginia
EU - Ireland
Price Overview
Time
16-Mar
17-Mar
18-Mar
19-Mar
20-Mar
Instance
Micro - Micro+Linux/Unix+Reserved [0]
Micro - Micro+Linux/Unix+Reserved [1]
Micro - Micro+Linux/Unix+Reserved [2]
Micro - Micro+Linux/Unix+Reserved [3]
Micro - Micro+Linux/Unix+On-Demand [4]
Micro - Micro+Linux/Unix+On-Demand [5]
Micro - Micro+Linux/Unix+On-Demand [6]
Micro - Micro+Linux/Unix+On-Demand [7]
Micro - Micro+Linux/Unix+On-Demand [8]
Micro - Micro+Linux/Unix+On-Demand [9]

```
name=task2
start=2011-03-17 03:00
end=2011-03-17 04:00
price=0.01968262506724045
fixedprice=0.009682625067240451
type=Reserved
hour=1803
workload=3.726294203983402
description=descr2
instance=Micro - Micro
region=EU - Ireland
os=Linux/Unix
deadline=2011-10-11 00:00
spotallowed=false
```

- Basic scheduling vs Optimized scheduling

Basic Scheduling (w1)

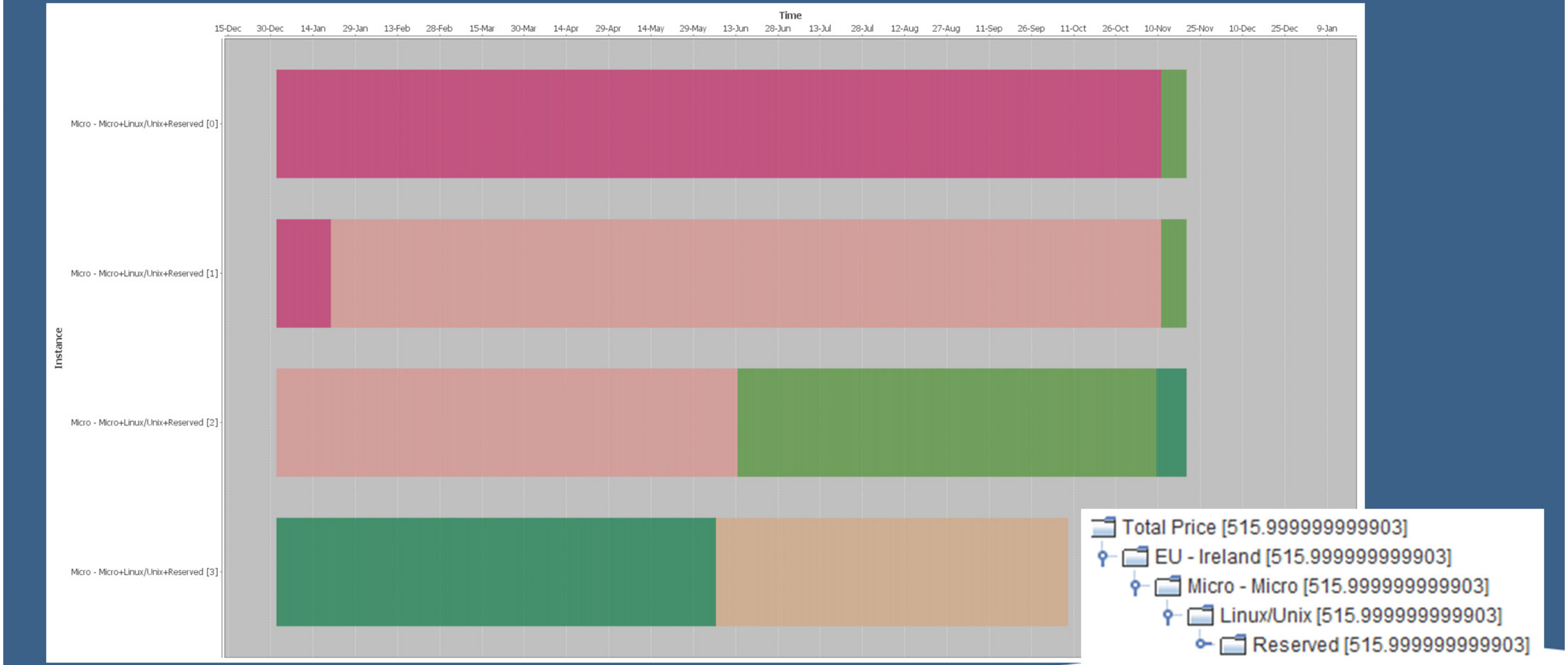


- Total Price [619.9200000001583]
  - EU - Ireland [619.9200000001583]
    - Micro - Micro [619.9200000001583]
      - Linux/Unix [619.9200000001583]
        - Reserved [446.7200000001361]
        - On-Demand [173.20000000002224]

Optimized Scheduling (w1) [1/2]

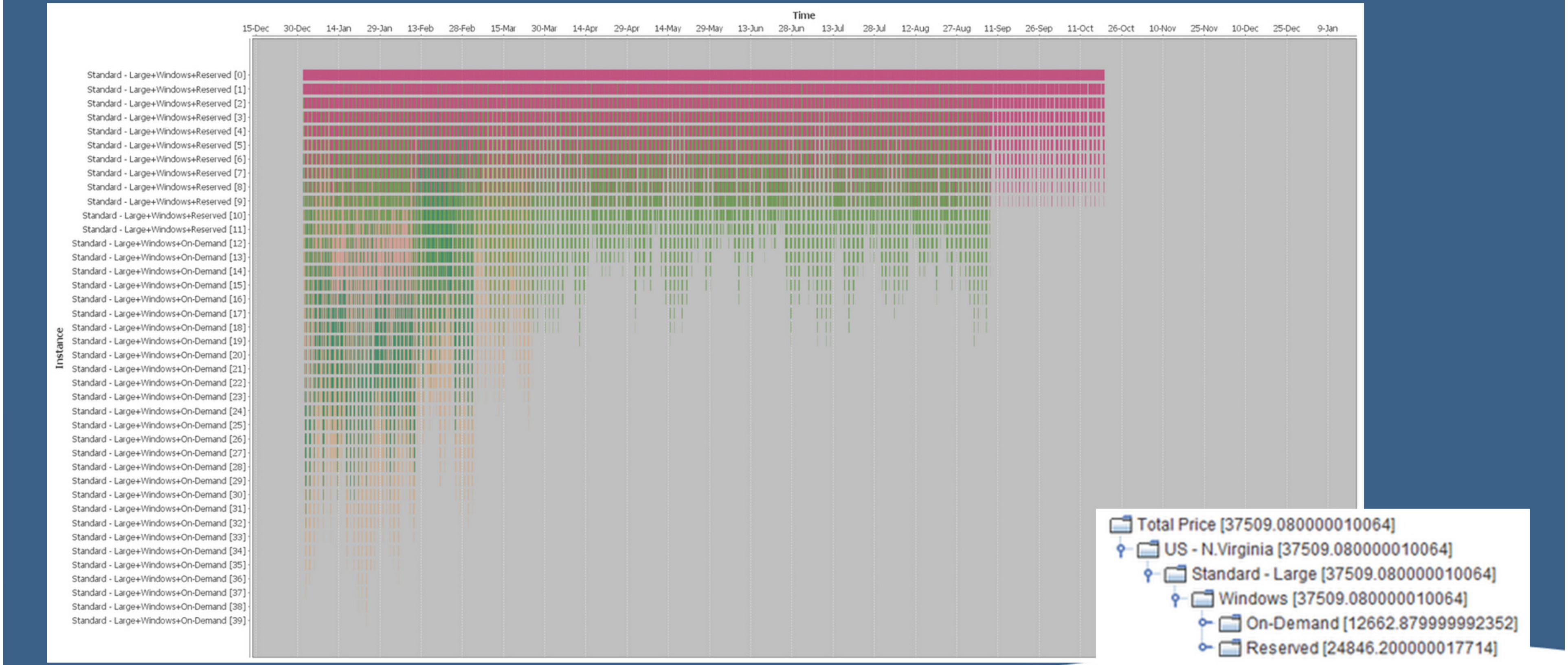
```
edfSort(tasks);
for(task t : tasks){
        for(instance i : instances){
                i.addPartTillDeadlineOrEnd(t);
                if(t.isDistributed()){ break; }
                if(i.isLast()) { instances.addNew(); }
        }
}
```

Optimized Scheduling (w1) [2/2]



Total Price [515.999999999903]
- EU - Ireland [515.999999999903]
  - Micro - Micro [515.999999999903]
    - Linux/Unix [515.999999999903]
      - Reserved [515.999999999903]

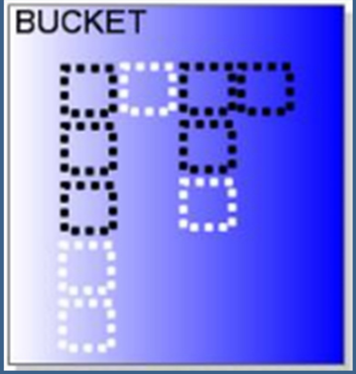
Basic Scheduling (w2)

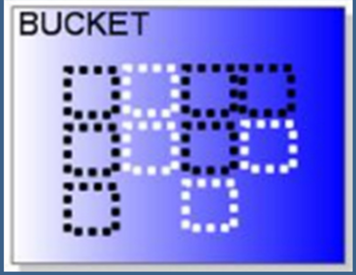


- Total Price [37509.080000010064]
  - US - N.Virginia [37509.080000010064]
    - Standard - Large [37509.080000010064]
      - Windows [37509.080000010064]
        - On-Demand [12662.879999992352]
        - Reserved [24846.200000017714]

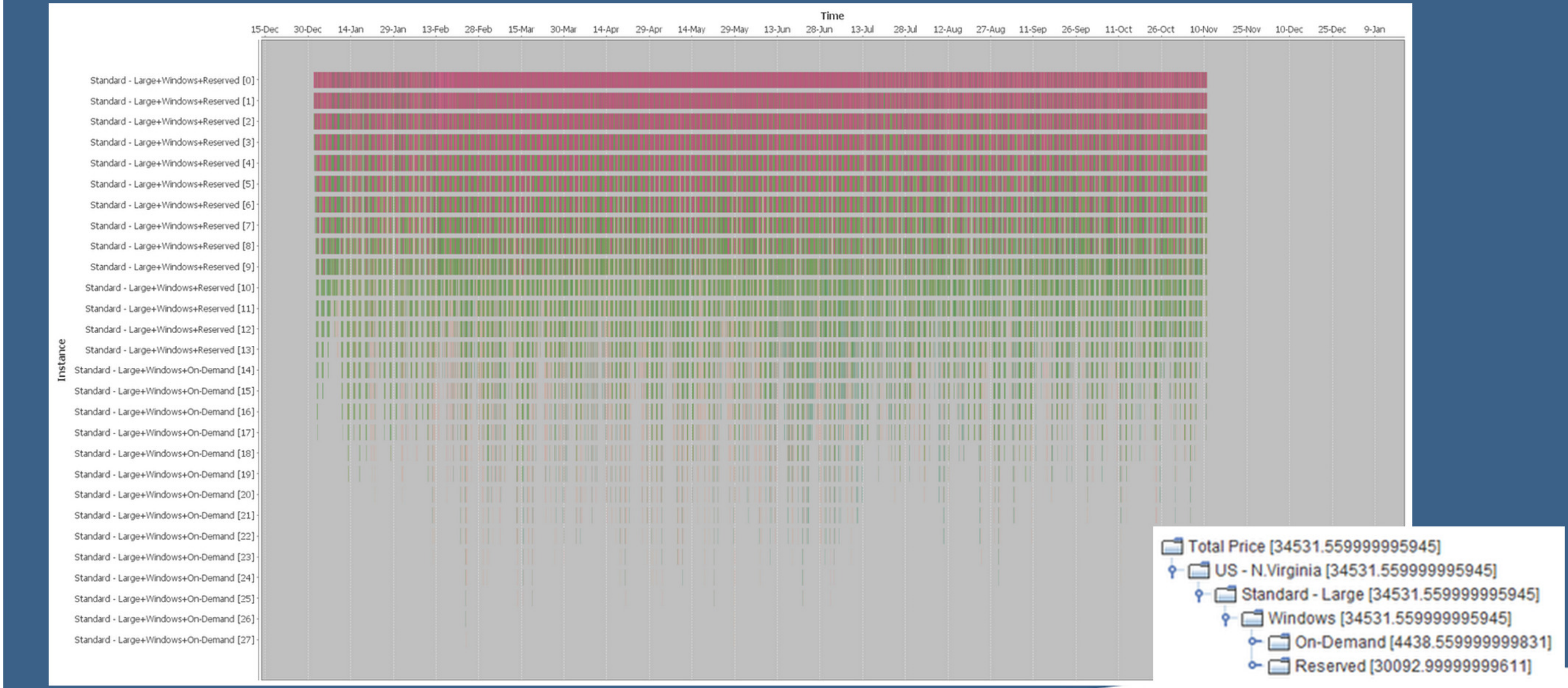
Optimized Scheduling (w2) [1/2]

```
for(Task t : tasks){
        buckets.divideEqually(t);
}
for(Bucket b : buckets){
        //try all combinations, choose the one
        //that minimizes the number
        //of needed instances
        b.makePlanning();
}
```





Optimized Scheduling (w2) [2/2]


Time
15-Dec
30-Dec
14-Jan
29-Jan
13-Feb
28-Feb
15-Mar
30-Mar
14-Apr
29-Apr
14-May
29-May
13-Jun
28-Jun
13-Jul
28-Jul
12-Aug
27-Aug
11-Sep
26-Sep
11-Oct
26-Oct
10-Nov
25-Nov
10-Dec
25-Dec
9-Jan
Instance
Standard - Large+Windows+Reserved [0]
Standard - Large+Windows+Reserved [1]
Standard - Large+Windows+Reserved [2]
Standard - Large+Windows+Reserved [3]
Standard - Large+Windows+Reserved [4]
Standard - Large+Windows+Reserved [5]
Standard - Large+Windows+Reserved [6]
Standard - Large+Windows+Reserved [7]
Standard - Large+Windows+Reserved [8]
Standard - Large+Windows+Reserved [9]
Standard - Large+Windows+Reserved [10]
Standard - Large+Windows+Reserved [11]
Standard - Large+Windows+Reserved [12]
Standard - Large+Windows+Reserved [13]
Standard - Large+Windows+On-Demand [14]
Standard - Large+Windows+On-Demand [15]
Standard - Large+Windows+On-Demand [16]
Standard - Large+Windows+On-Demand [17]
Standard - Large+Windows+On-Demand [18]
Standard - Large+Windows+On-Demand [19]
Standard - Large+Windows+On-Demand [20]
Standard - Large+Windows+On-Demand [21]
Standard - Large+Windows+On-Demand [22]
Standard - Large+Windows+On-Demand [23]
Standard - Large+Windows+On-Demand [24]
Standard - Large+Windows+On-Demand [25]
Standard - Large+Windows+On-Demand [26]
Standard - Large+Windows+On-Demand [27]

- Total Price [34531.559999995945]
  - US - N.Virginia [34531.559999995945]
    - Standard - Large [34531.559999995945]
      - Windows [34531.559999995945]
        - On-Demand [4438.559999999831]
        - Reserved [30092.99999999611]

What’s next?

- Extend scheduler with a checkpointing cost
- Extend scheduler/broker with spot instances
  - Using findings from spot analysis
  - Using decision model software
- Evaluate cost cuttings achieved by broker
- A lot of writing!

Thank You!

Questions?

Check out
http://www.thesis.kurtvermeersch.com