PRICE®

ICEAA 2019

Using Predictive Analytics and Open Source Data to Estimate IT and Cloud Related Costs for Government IT Systems

Presented by:
Richard Mabe, PRICE Systems, LLC
Dan Harper, The MITRE Corporation

Estimate with Confidence™

Overview

- Purpose
- Background/Issues
- Research Focus/Scope
- Approach
  - Part I: Data
  - Part II: CERs
- Other Findings from Open Source Data
- Summary/Next Steps

Purpose/BLUF

- Mine publicly available (open-source) federal and commercial databases for usable cloud and IT data (cost, schedule, technical)
- Based on this data, develop methods to predict provider-agnostic cloud-related costs for Government IT systems
- **Bottom line:** using public data, we were able to create both single-variable and multi-variable CERs to predict costs for Compute Environment (servers, RAM) and Storage Environment (capacity)

Background

- Federal agencies are adapting their IT system budgets and funding strategies to accommodate commercial cloud services
- The White House's "Cloud Smart" strategy encourages agencies to consider the life-cycle costs for cloud support
- Defensible life-cycle cost estimates supporting IT budgets need to:
  - Include modifications costs for SW and data being moved to cloud
  - Incorporate commercial cloud rates and fees to ensure adequate funding is provided in the budget
- Multiple agencies have developed ad-hoc tools to estimate all or part of the cloud-related costs for specific providers
  - Incorporate rates and fees from commercial provider web sites
  - Complex effort to include multiple providers and multiple available instances for hosting Government IT systems

Issues

- The cloud transition process is not instantaneous
- Estimating costs for the budget requires a two-year lead time:
  - Exact vendor and cloud requirements may not be finalized when budgets are being developed
  - But management requires significant “real-time” analysis for economic trade-offs
- Data on actual cloud costs for multiple providers is not readily available to Government cost estimators
- Estimating methodologies based on predictive analytic methods may provide acceptable, vendor-agnostic estimates to support IT budgets
  - Estimates, not “Exactimates”
  - Use open-source commercial and Government databases

Research Focus

- Identify/develop data, methods, and tools, which can be used to:
  - Estimate defensible, verifiable cloud life cycle costs (plan, modify, migrate, re-host, go live, sustain)
  - Help establish a realistic budget for current systems transitioning to the cloud
  - Provide estimates two-to-five years ahead of when the funds will be spent

Scope

- On-going, detailed research effort
  - This paper focuses primarily on *cloud infrastructure* costs
- Business IT systems supporting finance, HR, medical, logistics
  - Federal government only
  - Commercial cloud only (no DISA)
- Open-source infrastructure data (non-proprietary)
  - Summary-level costs from commercial cloud providers (*this paper*)
  - Expenditure data from the Federal IT Dashboard (*future effort*)
- Used PRICE® TruePlanning® tools as a modeling platform
  - Results transferrable to other platforms

Approach
Part I: Data

Sources

- Open-source data describing IaaS costs
  - Non-proprietary
  - Financial (Planned budgets; Annual expenditures)
  - Technical (Server quantity and capacity, Storage capacity, Data transport bandwidth)
- Source for government IT system financial data:
  - DOD IT Dashboard (Budget and Limited Technical) (https://itdashboard.gov/)
- Source for commercial cloud data:
  - Individual Cloud Provider web sites and calculators
  - Summary sites of multiple providers
    - *Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/)*
    - *Banzai Cloud web site (https://github.com/banzaicloud)*

Normalization

- Commercial cloud data
  - Standard, High Memory and High CPU cloud instances
  - Linux OS
  - Compute Environment HW (2vCPUs per VM; GB of RAM)
  - Storage Environment HW (Separate Storage Devices; GB of storage capacity)
- Government IT Dashboard data
  - Reported infrastructure budgets/expenditures by PY, BY and CY (P/IaaS; SaaS)
  - Sorted by:
    - *Federal Agency and IT Program (both DoD and Civilian)*
    - *Federal Enterprise Architecture (FEA), Business Reference Model (BRM) codes - Primary Service Area*
    - *Project Status (Complete; In-Work)*

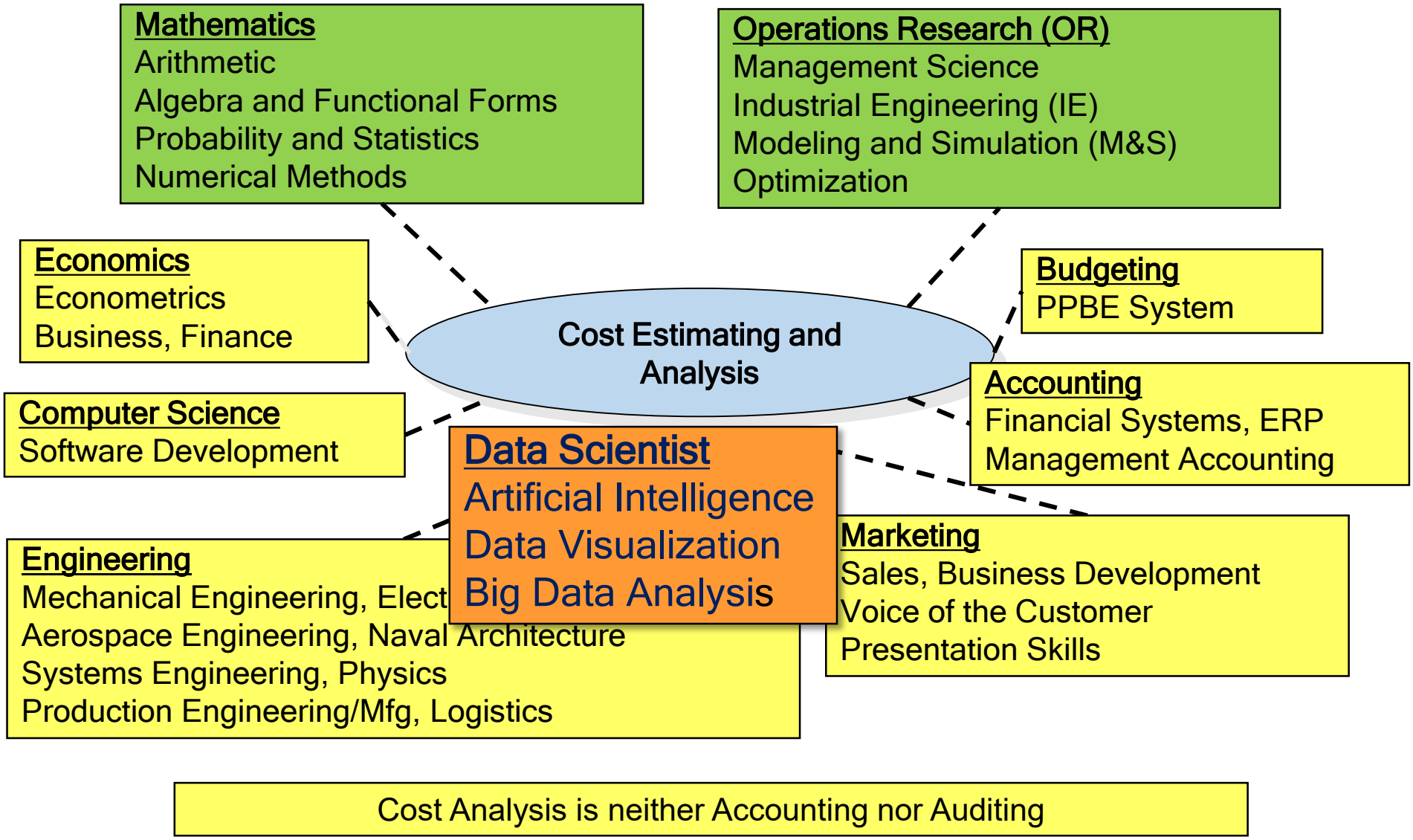
Challenge of Big Data and Cost Analysis

- There is a lot of data available, but the challenge lies in making it usable and standardizing it for cost analysis, i.e., normalization
- Data preparation is the foundation of impactful analysis. Data scientists spend 80 percent of their time on data preparation*
- ICEAA members are exploring this problem, e.g., March 2019 presentation "Practical Approach to Data Science and Data Analytics" from Tom Sullivan (Herren)
- Most of us are most comfortable in Excel, so let's keep it there for now
  - Pivot tables, data visualization, pivot charts, power pivot etc.

DH

*"How to Cut Data Preparation Time for Visualization Tools" https://tdwi.org/articles/2019/02/15/diq-all-how-to-cut-data-preparation-time-for-visualization-tools.aspx

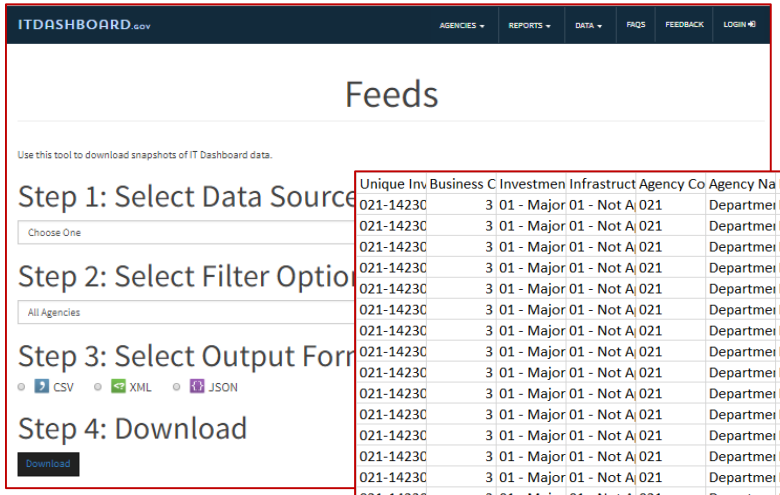
Cost Estimating Disciplines-Need 1 More!?



Cost Analysis is neither Accounting nor Auditing

Normalizing and Organizing Data (1)

- OMB IT Dashboard ".csv" data can be downloaded into Excel but must be normalized and manipulated (https://itdashboard.gov/)
  - Over 7000 Federal IT programs
  - Over 26,000 records (rows) ; 42 pieces of data (columns) per record
  - over 1.1 million total cells; includes text, dates, categories, budget phases, investment titles, ID numbers, dollars planned, dollars spent



Unique Inv	Business C	Investmen	Infrastruct	Agency Co	Agency Na	Investmen	Agency Ac	Project ID	Activity Na	Activity De	Structure I	Start Date	Start Date	Start Date	Completio	Completio	Completio	Total Cost	Total Cost	Total Cost
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10283	7143	Colorado \	Colorado \	689.1	########	########	########	########		########	2.166123	2.166123	1.723838
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10284	7143	Colorado \	Colorado \	689.2	8/1/2011	########	########	########		########	5.094964	5.094964	6.807892
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10285	7143	Colorado \	Colorado \	689.3	6/1/2012	6/1/2012	6/1/2012	########		########	0.690332	0.690332	0.528947
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10304	7143	Initial Ope	Initial Ope	689.4	########	########	########	########		########	1.483872	1.483872	1.288615
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10300	7145	Draft Targ	3 Nautical	690.1	7/1/2011	7/1/2011	7/1/2011	########		########	1.24195	1.24195	0.851305
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10301	7145	Final Targe	3 Nautical	690.2	########	########	########	########		########	1.883112	1.883112	1.58271
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10287	7145	Complete	3 Nautical	690.3	4/1/2012	4/1/2012	4/1/2012	########		########	0.364811	0.364811	0.658373
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10303	7147	Initial Ope	Common /	691.1	########	########	########	########		########	11.98827	11.98827	12.71118
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10280	7147	ATOP - Re	ATOP Req	691.2	########	########	########	########		2/9/2012	0.058355	0.058355	0.120536
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10307	7147	IOC - ERA	Initial Ope	691.3	########	########	########	########		########	38.8009	38.8009	40.01979
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10306	7147	IOC - ADS	ADS-B cap	691.4	########	########	########	########		########	12.5286	12.5286	10.17553
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10279	7147	ATOP - 1st	ATOP Soft	691.5	1/1/2012	1/1/2012	1/1/2012	########		########	9.065	9.065	1.90905
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10313	7147	MEARTS Fi	Micro-En f	691.6	5/1/2013	5/1/2013	5/1/2013	########		########	4.308091	4.308091	3.288085
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10311	7147	MEARTS Fi	Micro-En f	691.8	########	########	########	########		8/6/2015	1.255896	1.255896	1.2559
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10263	7149	(IOC) - Sur	ATC Surfac	692.1	########	########	########	########		########	15.56592	15.56592	13.16
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10265	7149	(IOC) - Sur	ATC Surfac	692.2	########	########	########	########		########	9.38773	9.38773	3.64222
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10262	7149	(IOC) - Sur	ATC Surfac	692.3	4/1/2012	4/1/2012	4/1/2012	########		########	2.98908	2.98908	4.653845
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10272	7149	Achieve Su	ATC Surfac	692.4	########	########	########	########		########	4.490989	4.490989	4.646
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10264	7149	(IOC) - Sur	ATC Surfac	692.5	########	########	########	########		########	3.06784	3.06784	3.051
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10305	7149	Install Surf	Advisory S	692.6	########	########	########	########		########	32.92754	32.92754	44.114
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10302	7151	Ground Ba	GIM Requi	693.1	########	########	########	########		########	1.550406	1.550406	1.263338
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10315	7151	TBFM Syst	GIM Time	693.2	########	########	########	########		########	10.82456	10.82456	11.60742
021-14230	3	01 - Major	01 - Not A	021	Departmei	FAAXX704	10314	7151	SSS and VF	Sub Systen	693.3	4/1/2012	4/1/2012	4/1/2012	########		########	5.332676	5.332676	6.94192

Normalizing and Organizing Data (2)

- Use Lookups, Pivot Tables, Charts/visualizations, Excel BI, or whatever is most useful to you and conducive to analysis, e.g., tabular layout
  - Data needs to make intuitive sense

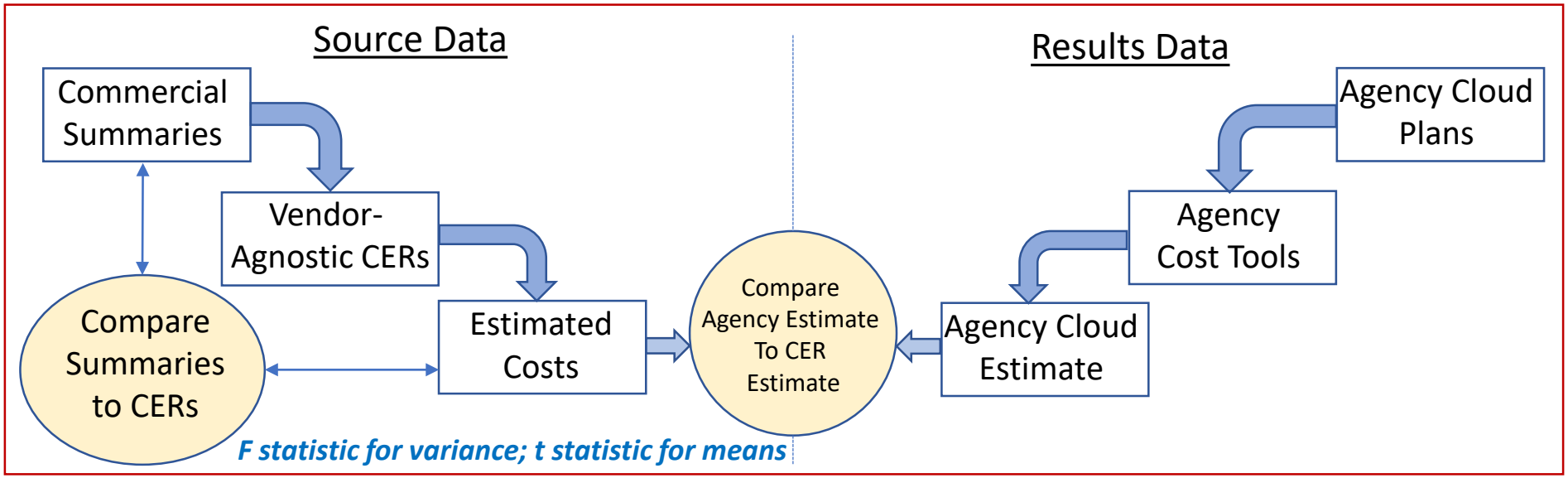
Ref Col; Restore Sort	Part I: Project Level Costs	Unique Investment Identifier	Agency Code	Agency Name	Investment Title	Project ID	Project Name	FEA BRM Services - Primar	FEA BRM Services - Primary service area (Possible App Type)	Start Date	Completion Date	Days to Complet	Production Deployment Date	Project Life Cycle Cos ($A*1
1	X	005-000000038	005	Department of Agriculture	APHIS-Enterprise Infrastructure (AEI)					10/1/2010	8/8/2014	1407		$ 113.93
2		005-000000038	005	Department of Agriculture	APHIS-Enterprise Infrastructu	665	Enterprise Management			10/1/2010	9/30/2011	364		$ 46.96
3		005-000000038	005	Department of Agriculture	APHIS-Enterprise Infrastructu	666	VoIP Riverdale			8/15/2011	4/30/2012	259		$ 0.66
4		005-000000038	005	Department of Agriculture	APHIS-Enterprise Infrastructu	667	Enterprise Security			9/6/2011	12/31/2012	482		$ 3.16
5		005-000000038	005	Department of Agriculture	APHIS-Enterprise Infrastructu	669	Enterprise Software license optimization			10/3/2011	9/7/2012	340		$ 10.06
6		005-000000038	005	Department of Agriculture	APHIS-Enterprise Infrastructu	670	Hardware Maintenance			11/1/2011	5/31/2012	212		$ 4.27
7		005-000000038	005	Department of Agriculture	APHIS-Enterprise Infrastructu	671	Enterprise Life Cycle Management			3/1/2012	8/24/2012	176		$ 4.01
8		005-000000038	005	Department of Ag										
9		005-000000038	005	Department of Ag										
10		005-000000038	005	Department of Ag										
11		005-000000038	005	Department of Ag										
12		005-000000038	005	Department of Ag										
13		005-000000038	005	Department of Ag										
14	X	005-000000039	005	Department of A										

Sum	Years													
A	Ir A Activity Description2	<2/28/2001	2001	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	Grand Total
Department of Agriculture		**0.165**			**$35.249**				**$24.406**	**$68.147**	**$67.678**	**$ 163.869**	**$86.168**	**$64.202**	**$ 1.297**	**$ 511.182**
	APHIS-Animal Disease Traceability Information System (ADTIS)									**$ 0.379**		**$ 0.187**	**$ 0.187**			**$ 0.753**
	Key Deliverable															
	Adapt OAS applications to new framework called weblogic									$ 0.379						$ 0.379
	Adapt OAS applications to new framework called Weblogic.															
	Provides capability to enter non ISO standard tags.															
	N/A															
	Help Desk for Tracibility											$ 0.102				$ 0.102
	Licenses and support											$ 0.015				$ 0.015
	SA&A RMF Step 4											$ 0.016				$ 0.016
	Steady State IT Support											$ 0.054				$ 0.054
	Usable Functionality															
	Help Desk for Tracibility															
	NITC Hosting Costs															
	POA&M Remediation															
	SA&A RMF Steps 1-3 & POA&M Remediation															
	Server Hosting at NITC															
	Steady State IT Support															
	(blank)															
	Help Desk for Tracibility												$ 0.102			$ 0.102
	Licenses and support												$ 0.015			$ 0.015
	SA&A RMF Step 4												$ 0.016			$ 0.016
	Steady State IT Support												$ 0.054			$ 0.054
	APHIS-Certification, Accreditation, Registration, Permitting, and Other Licensing (CARPOL)												**$ 2.823**	**$ 2.312**		**$ 5.135**
	15 - This is not a software development related activity															
	Support CARPOL PMO activities.															
	Support CARPOL PMO FY2016 first and second quarter activities.															
	Support CARPOL PMO FY2016 third and fourth quarter activities.															
	5 - Development															
	Develop permitting capability.												$ 2.823	$ 2.312		$ 5.135
	Milestone activity noting the completion of APHIS eFile development															
	Review requirements and begin framework development.															

Approach Part II: Develop CERs

Experimental Design

- Create vendor-agnostic CERs from rates of 4 cloud providers
  - Validate by comparing estimates to the commercial actuals
- Examine costs estimated by models from two federal agencies
- Compare the CER estimates to the federal agency costs
- If the CER estimates are statistically equivalent to the agency costs (F, t), then the vendor-agnostic models are just as valid



Infrastructure Source Data Analysis

1. Evaluate correlated pairs within the commercial data summaries

DEPENDENT	INDEPENDENT	R-Value
Linux Hrly On Demand Rate per Instance	GB RAM per Instance	0.7839
Linux Hrly On Demand Rate per Instance	GB Storage per Instance	0.3392
DEPENDENT	**INDEPENDENT**	**R-Value**
Linux Annual On Demand Rate per Instance	GB RAM per Instance	0.8007
Linux Annual On Demand Rate per Instance	Cloud Type Number	0.2898
DEPENDENT	**INDEPENDENT**	**R-Value**
Linux Annual Discount Rate per Instance	GB RAM per Instance	0.7030
Linux Annual Discount Rate per Instance	GB Storage per Instance	0.5402

- Rates for infrastructure correlated most strongly with the amount of RAM required for Compute Environment vCPUs
- Lesser correlation occurred between the Rates and Storage Environment requirements

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/)
Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

PRICE

Infrastructure Source Data Analysis

2. Develop CERs based on the correlated pairs:

- Single variable and multi-variable regression models; Vendor-agnostic
- Includes statistical measures of quality (compare vendor data with CER estimates; n=24; alpha=0.05)
- P-values > (0.05) indicate actual and estimate sample sets have equal variance and means

Rate Basis	Component*	CER (Cost per vCPU)	R^2	F for Variance	P-value	t for Means	P-Value
Hourly	Compute	0.077*e^(0.040 * GB RAM)	0.649	1.331	0.249	-1.021	0.317
Hourly	Compute + Store	(0.005*GB RAM) + (0.0000981*GB Storage)+0.065	0.766	1.161	0.361	-0.326	0.746
Annual	Compute	43.390 * GB RAM+ 558.998	0.641	1.559	0.147	-0.00006	0.999
Annual	Compute + Store	(43.901*GB RAM) + (0.354*GB Storage) + 529.714	0.663	1.507	0.166	0.0004	0.997
Annual (Discount)	Compute	430.087 * e^(0.041 * GB RAM)	0.536	1.936	0.060	0.234	0.815
Annual (Discount)	Compute + Store	(27.045 * GB RAM) + (0.936 * GB Storage) + 338.483	0.837	1.194	0.337	-0.001	0.998

No valid CER for just storage

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/)
Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

PRICE

Infrastructure Source Data Analysis

3. Verify accuracy of the CERs as compared to costs from two Government agencies for similar IaaS solutions

- Government costs derived from commercial provider web sites
- Compared the CER results to costs estimated by the Government tools
- Government value = point value within a range of commercial values
- Evaluate to see if CER estimated value falls in same range

- Process steps:
  - Compute cost per vCPU with each CER
  - Multiply this value times the number of vCPUs in the agency-defined range (Low, Optimal, High)
  - Identify a “number-of-vCPUs” multiplier and CER value that best matches the Government “point” value
  - Identify additional limitations or assumptions with the CER estimate

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/) Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

PRICE

Results: Compute Environment (Hourly Rate)

Agency 1: $ = 0.077*e^(0.04*RAM) *vCPU Multiplier

HW Resource	vCPU Range	MB RAM (Memory)	Gov't Value ($/Hour)	CER Value ($/ Hour) (Best Fit)	vCPU Multiplier for Best Fit	% Error (Best Fit)	Predicted Low CER Value	Predicted High CER Value
X-Large Level 1	16 – 32	128 – 256	2.829	2.793	36	1%	1.232	2.489
X-Large Level 2	8 – 32	64 – 128	1.663	1.701	22	- 2%	0.618	2.470
Large Std level1	8 – 32	32- 64	1.328	1.389	18	- 5%	0.616	2.467
Large Std level2	4 – 16	16 – 32	0.558	0.540	7	3%	0.308	1.232
Med Std level1	4 – 16	8 – 16	0.436	0.462	6	- 6%	0.308	1.232
Med Std level2	4 – 8	4 – 8	0.335	0.308	4	8%	0.308	0.616
Small	2 - 4	2 – 4	0.253	0.231	3	9%	0.154	0.308

Notes:

- **X-Large Level 1 required a multiplier above the range to yield an error of < 10%**
- **All other HW resource hourly costs can be predicted within 10% error**
- **Conclusion: Vendor-agnostic CER as good as tailored Gov't agency model**

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/) Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

PRICE

Results: Compute Environment (Annual Rate)

Agency 2: $ =((43.39*RAM)+558.998)*vCPU Multiplier

HW Resource	vCPU Range	MB RAM (Memory)	Actual Annual $	Predicted / Annual (Best Fit)	Multiplier (vCPUs in Range)	% Error (Best Fit)	Predicted (Low vCPU Value)	Predicted (High vCPU Value)
Xlarge 1	2 – 4	15	1,505	1,572	1.3	-6%	2,419	4,839
Xlarge 2.2	2 – 4	34	3,528	3,254	1.6	2%	4,068	8,137
Xlarge 2.4	4 – 8	68	7,132	6,668	1.9	2%	14,038	28,076
Xlarge 3.2	4 – 8	30	5,954	5,396	2.9	1%	7,442	14,885

Notes:

- **Multipliers yielding <10% error are not within the vCPU range for each instance**
- **Government values include only 1 cloud vendor; CERs were developed using 4 vendors**
- **Not a perfect fit for this vendor; work-in-process to verify with additional federal model results**

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/)
Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

PRICE

Results: Compute Env. (Discount Annual Rate)

Agency 2: $ =(430.087*e^(0.041*GB RAM))*vCPU Multiplier

HW Resource	vCPU Range	MB RAM (Memory)	Actual Annual $	Predicted / Annual (Best Fit)	Multiplier (vCPUs in Range)	% Error (Best Fit)	Predicted (Low vCPU Value)	Predicted (High vCPU Value)
Xlarge 1	2 – 4	15	1,505	1,591	2	7%	1,591	3,182
Xlarge 2.2	2 – 4	34	3,528	3,467	2	10%	3,467	6,934
Xlarge 2.4	4 – 8	68	7,132	6,988	1	9%	27,953	55,906
Xlarge 3.2	4 – 8	30	5,954	5,885	4	9%	5,885	11,771

Notes:

- Multipliers yielding <10% error are within the vCPU range, except for one instance (68GB RAM); note that the Government actual cost is also low
- Government values include only 1 cloud vendor; CERs were developed using 4 vendors
- Discount rates generally provided better fit to Government estimates

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/) Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

Results: Compute + Store (Hrly/Monthly Rate)

Agency 1: Extended data range for GB Storage (up to 5 TB)

Source	Storage Range	GB RAM Range	0 – 1 TB (Average)	2 – X TB (Average)	5 TB Required per Month	% Delta
Government Value (Agency 1)	0 – 4 TB	4 – 16 GB	0.030/GB	0.450/GB	$601	
Vendor Data Estimate	0 – 1 TB	4 – 16 GB	0.124/GB	0.377/GB	$572	4.8%

Notes:

Created 2 hourly curves:

- **Summary Data: $=(0.005*GB RAM)+(0.0009815*GB Storage) + 0.065**
- **Government Data: $ = ((0.1189*GB Storage)+6.7786)/720**

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/) Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

PRICE

Results: Compute + Store (Annual Rate)

Agency 2: $=43.901 * GB RAM + 0.354 * GB Storage + 529.714

HW Resource	vCPU Range	MB RAM (Memory)	GB Storage	Actual Annual $	Predicted/ Annual (Best Fit)	Multiplier (vCPUs in Range)	% Error (Best Fit)	Predicted (Low vCPU Value)	Predicted (High vCPU Value)
Xlarge 1	2 – 4	15	1,680	1,693	1,783	1	-5%	3,565	7,131
Xlarge 2.2	2 – 4	34	850	3,623	3,485	1.7	-9%	4,646	9,2924
Xlarge 2.4	4 – 8	68	1,680	7,320	7,808	1.9	-7%	8,219	16,438
Xlarge 3.2	4 – 8	30	30	5,957	5,572	3.3	-7%	3,714	7,429

Notes:

- **Multipliers yielding <10% error are not within the vCPU range for each instance**
- **Government values include only 1 cloud vendor; CERs were developed using 4 vendors**
- **Not a perfect fit for this vendor; work-in-process to verify with additional federal model results**

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/) Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

PRICE

Results: Compute + Store (Disc. Annual Rate)

Agency 2: $=27.045 * GB RAM + 0.936 * GB Storage + 338.483

HW Resource	vCPU Range	MB RAM (Memory)	GB Storage	Actual $ / Annual	Predicted / Annual (Best Fit)	Multiplier (vCPUs in Range)	% Error (Best Fit)	Predicted (Low vCPU Value)	Predicted (High vCPU Value)
Xlarge 1	2 – 4	15	1,680	1,693	1,853	0.8	-9%	4,633	9,267
Xlarge 2.2	2 – 4	34	850	3,623	3,902	1.9	-8%	4,107	8,214
Xlarge 2.4	4 – 8	68	1,680	7,320	7,875	2.1	-6%	7,500	15,000
Xlarge 3.2	4 – 8	30	30	5,957	5,890	5	1%	2,356	4,712

Notes:

- **Government values include only 1 cloud vendor; CERs were developed using 4 vendors**
- **Marginally better fit; improves as vCPUs increase**

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/) Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

Conclusions

- Met our Research Objective
  - CERs developed from multiple commercial provider rates able to predict costs for Compute Environment (servers, RAM) and Storage Environment (capacity)
  - Predicted costs statistically equivalent to costs estimated by two government agency models
  - Costs for Compute Environment were better match than costs for Storage environment
  - Estimates using the CERs for discounted commercial costs (Compute and Store) were better match than CERs for non-discounted costs

Next Steps

- Developing predictive infrastructure costs still a work-in-process
  - More vCPUs per VM
  - Extended range for Storage
  - Look for “Storage Only” instances
- Expand verification to other Federal agencies
  - Two more agency models still to evaluate
  - 35 additional cloud programs
- Expand research to predict other cloud-related costs
  - Refactor SW for re-hosting in cloud
  - Modify/migrate data

PRICE

Revised CER: Compute (Hourly Rate)

Agency 1: $ = ((0.036 * vCPUs) + (0.007 * GB RAM)) - 0.010

HW Resource	vCPU Range	MB RAM (Memory)	Gov't Value ($/Hour)	CER Value ($/ Hour) (Best Fit)	vCPU Multiplier for Best Fit	% Error (Best Fit)
X-Large Level 1	16 – 32	128 – 256	2.829	2.583	36	9%
X-Large Level 2	8 – 32	64 – 128	1.663	1.574	22	5%
Large Std level1	8 – 32	32- 64	1.328	1.286	18	3%
Large Std level2	4 – 16	16 – 32	0.558	0.566	8	-1%
Med Std level1	4 – 16	8 – 16	0.436	0.422	6	3%
Med Std level2	4 – 8	4 – 8	0.335	0.350	5	-5%
Small	2 - 4	2 – 4	0.253	0.278	4	-9%

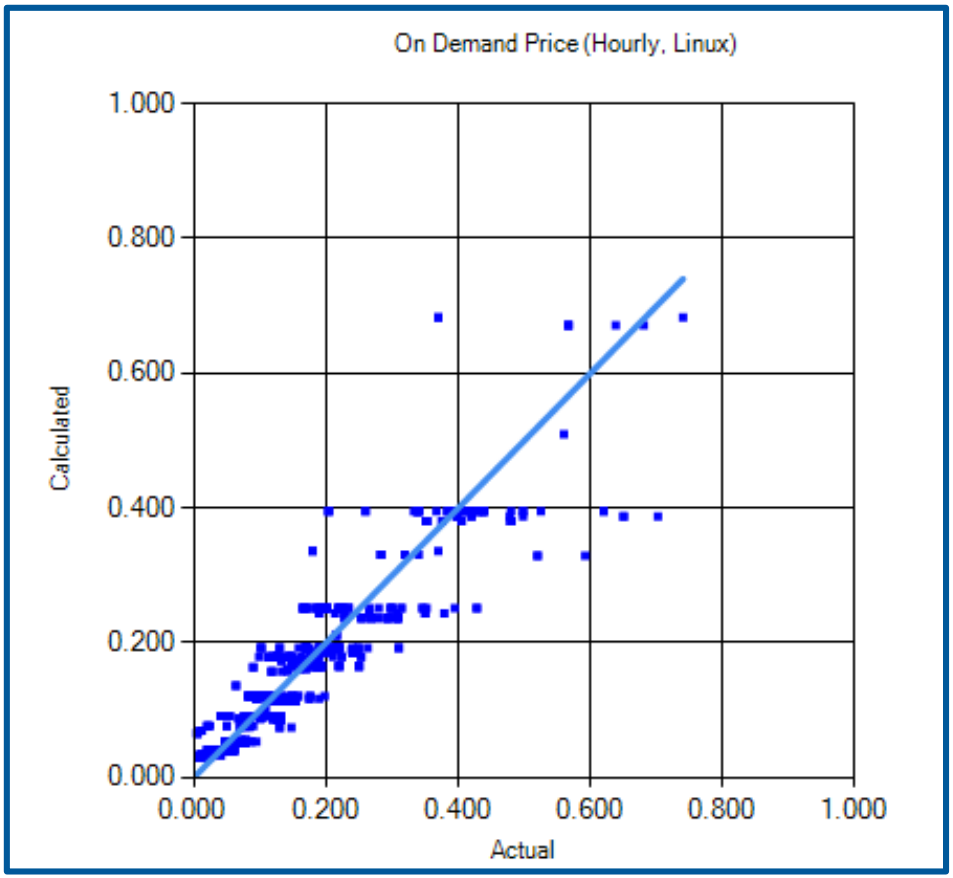
Notes:

- Minor adjustment to vCPUs; same general results
- Predicted still fits to within 10% error
- Conclusion: Revised Vendor-agnostic CER as good as tailored Gov't agency model

Data Sources: Right Scale/Flexera web site (https://www.rightscale.com/; https://www.flexera.com/)
Banzai Cloud web site (https://github.com/banzaicloud); OMB ITDashboard (https://itdashboard.gov/)

Revised Hourly CER- Includes VMs

$ = 0.036 * (CPUs per VM) + 0.007 * (GB RAM) - 0.010



Regression Table	
Statistic	Value
Multiple R	0.878
R Square	0.771
Adjusted R Square	0.771
Standard Error	0.062
Observations	771

ANOVA Table					
Source	SSE	DF	MSE	P-value	F-Statistic
Regression	9.874	1	9.874	1.49E-248	2593.072
Error	2.928	769	0.004		
Total	12.802	770	0.017		

Coefficients Table						
Name	Value	SE	t-Test	P-value	Upper CL	Lower CL
CPUs Per VM	0.045	8.85E-04	50.922	1.49E-248	0.047	0.043
Intercept	0.026	0.004	7.172	1.74E-12	0.033	0.019

Other Findings from Open Source Data (Data Points, Analogies)

Cost Analogies

- One of four recommended approaches to cost analysis in the GAO cost guide is to use analogous systems
- The IT Dashboard categorizes programs using the Federal Enterprise Architecture Framework
  - Combine data for "like" systems together to be analyzed and compared
- For example: estimate the cost to develop a Business Intelligence capability; VOIP, Help Desk, etc.

Software Licenses

- There are dozens of data points enterprise licenses for agency level applications such as Windows and Oracle
  - DHS "Purchase and install 25 Riverbeds and Licenses" @$847K (FY13$K), or $34K each
  - DHS CIO had the following ELA payments in FY 14:

Adobe Annual Payment Due	$	6.187
Oracle Annual Payment Due	$	45.543
Symantec Annual Payment Due	$	13.404
McAfee Annual Payment Due	$	4.129
IBM Annual Payment Due	$	0.250

PRICE

IT Metrics

- Includes thousands of metrics collected and reported in the OMB Business Case 300. Just for Help Desk:
  - # of help desk complaints compared to total number of help desk customers
  - % Customers Satisfied with WMA 3rd tier speed of response
  - % Helpdesk Tickets Closed
  - % of help desk tickets related to system issues resolved.
  - % of incident tickets resolved and closed within 3 days
  - % of incident tickets resolved and closed within 30 days.
  - % of Severity 1 tickets at Tier 3 closed in under 90 days
  - % of tickets escalated to Tier III Customer Support

Business Case Detail **UII 024-000005358**

Table II.C.1A/B Performance Metrics and Actuals

Metric ID	Metric Description	Unit of Measure	Performance Measurement Category Mapping	Agency Baseline Capability	2018 Target	2019 Target	Measurement Condition	Reporting Frequency	Agency Strategic Objective / Agency Priority Goal	Is Metric Retired?
					25321	186527	99.820000	03/31/2017		
25320	Active Users: Support up to 10,000 active users that process a transaction in the ICM system	Number of Users	2 - Strategic and Business Results		10000.000000	10000.000000	Under target	Monthly	024SO18227 : Not Available	No

Source: https://itdashboard.gov/drupal/summary/024/024-000005358

Back-Up Slides

Predictive Analytics

- "Encompasses a variety of statistical techniques from predictive modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future or otherwise unknown events."
  - *(https://en.wikipedia.org/wiki/Predictive_analytics)*

- Techniques used in this study:
  - Data mining
  - Correlation
  - Regression
  - Analysis of Variance

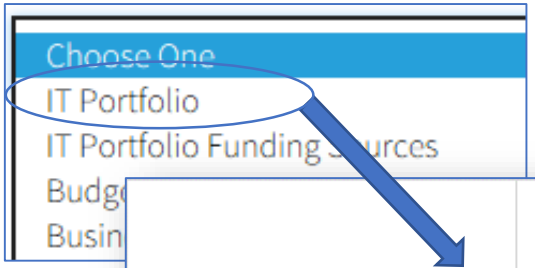
Example: Summary of Commercial Data

- Right Scale Website (https://www.rightscale.com/)
- Reference to cloud data by web site locations for multiple providers
- Summary data sorted by commercial provider for comparison

VM Type (Linux OS)					Rate Basis (US $)		
GP Instance	Case	vCPU per VM	GB RAM	GB Storage	On Demand Hourly	On Demand Annual	Discount Annual
m5d.large	Std	2	8	75	0.113	990	631
m5.large	Std	2	8	0	0.096	841	534
r5d.large	High_Mem	2	16	75	0.114	126	797
r5.large	High_Mem	2	16	0	0.126	1104	692
c5d.large	High_CPU	2	4	50	0.096	841	534
c5.large	High_CPU	2	4	0	0.085	745	473

Example: Summary of IT Dashboard

- From OMB IT Dashboard (https://www.itdashboard.gov/)
- Web portal for general public to view details of 7000+ Federal IT investments
- Launched in 2009, to view details online and track their progress over time
- Displays data received from agency IT Portfolio and Business



Unique Investment Identifier	Start Date	Completion Date	Total IT Spending FY2017 (PY) ($ M)	Total IT Spending FY2018 (CY) ($ M)	Total Cloud PaaS/IaaS Cost CY	Total Cloud SaaS Cost CY	Total Other Managed Services Cost CY
005-000000039	10/3/2011	3/30/2012	$14.14	$13.06	$3.60		
005-000000098	1/1/2013	9/30/2017	$172.59	$162.42		$5.34	
005-000000185	10/1/2014	9/30/2017	$12.13	$13.63		$3.60	
005-000001425	6/1/2012	12/31/2012	$99.96	$88.79	$5.35		$18.16

Transition Process

As Is System
(User Owned)
- Operate
- Sustain

Recurring Costs:
- Labor
- Materials
- Overhead
- ODCs
- Facilities
- PM/SE

Plan for Transition:
- *Business Case*
- *Change Mgmt*
- *Svc Level Agreement*

- What
- When
- Where To
- Security
- Access

Transition:
- Software
- Data
- Interfaces

Non-Recurring Costs:
- Modify/Refactor SW apps
- Prep data for migration
- Develop new middleware Interfaces
- Adapt to Cloud OS and Middleware Services
- PM/SE

Execute Plan:
- SW Porting
- *Data Migration*
- *User Training*

- Migrate
- Instantiate
- Test/Verify
- Parallel Ops
- Changeover
- Go Live

To Be System
(Cloud Host)
- IaaS
- PaaS
- SaaS

Recurring Costs:
- Fees
- Licenses
- Subscriptions

For:
- Infrastructure
- Run Time Env
- SW Services
- Access
- Cybersecurity
- PM/SE

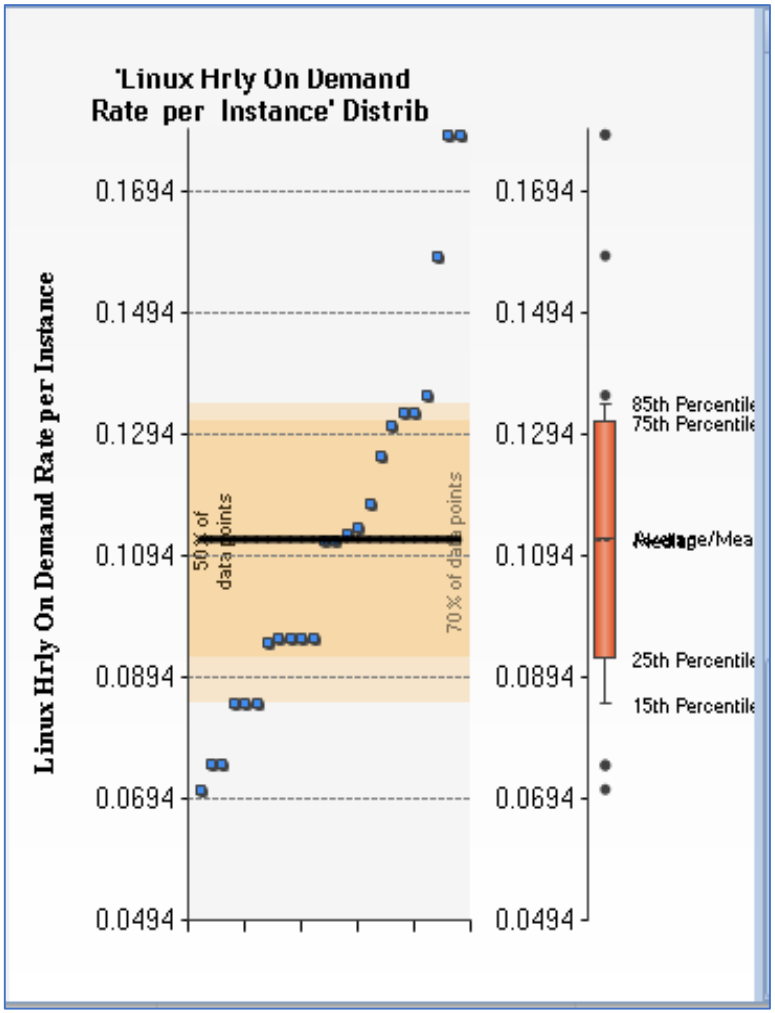
Linus On-Demand Rate per Hour

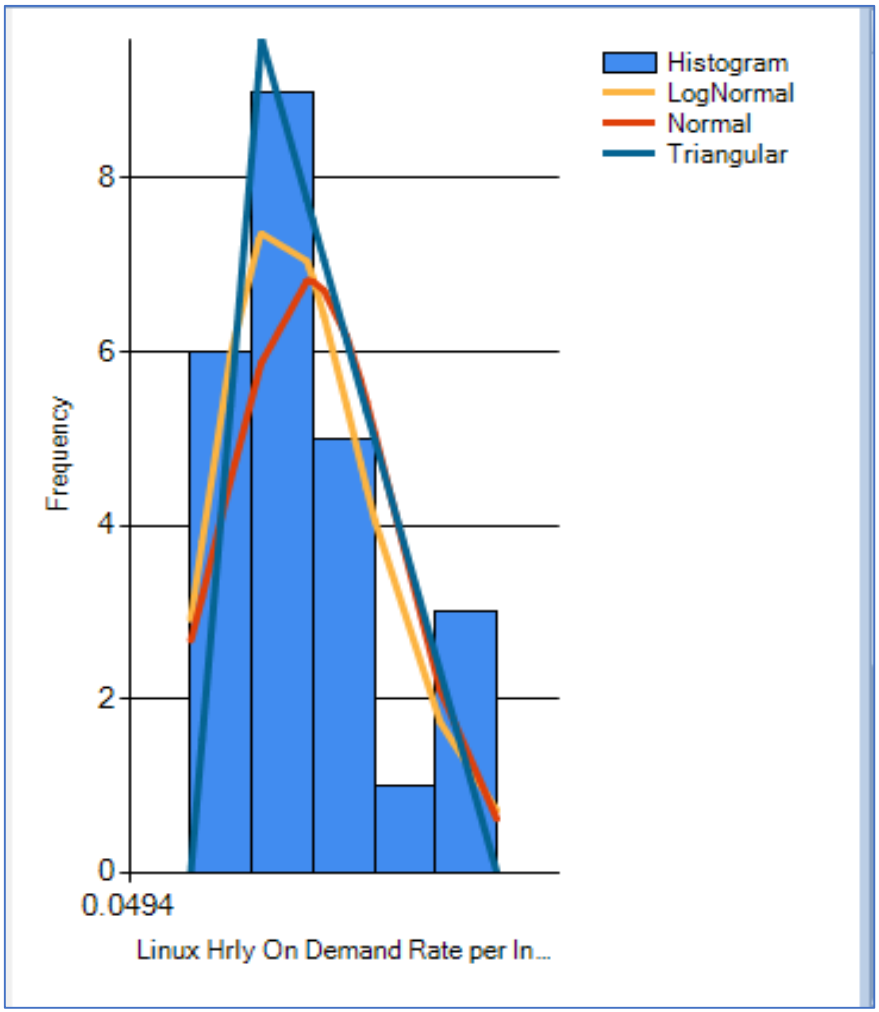
	Data	LN	Normal	TRI
Min	0.071			
Max	0.179			
25%	0.093			
75%	0.132			
Mean	0.113	0.113	0.113	0.115
Median	0.112	0.109	0.113	0.112
Mode	0.096	0.102	0.113	0.096
Standard Deviation	0.03	0.03	0.03	0.022
Coefficient of Variation	0.27	0.265	0.27	0.191

DEPENDENT	INDEPENDENT	R-Value
Linux Hrly On Demand Rate per Instance	**GB RAM per Instance**	**0.7839**
Linux Hrly On Demand Rate per Instance	**GB Storage per Instance**	**0.3392**
Linux Hrly On Demand Rate per Instance	Cloud Instance Number	0.3354
Linux Hrly On Demand Rate per Instance	Cloud Type Number	0.3283
Linux Hrly On Demand Rate per Instance	CPU Plus SSD Number	0.2217

PREDICTIVE CER	R^2-Value	Fit
= 0.077 * e^(0.040 * [GB RAM per Instance])	**R^2 = 0.649**	**Exponential**
= 0.064 * ([GB RAM per Instance]^0.272)	R^2 = 0.617	Power
= -2.226E-005 * ([GB RAM per Instance]^2) + 0.005 * [GB RAM per Instance] + 0.072	R^2 = 0.615	Polynomial
= 0.005 * [GB RAM per Instance] + 0.073	R^2 = 0.615	Linear
= 0.005 * [GB RAM per Instance] + 9.815E-005 * [GB Storage per Instance] + 0.065	**R^2 = 0.766**	**Multiple**

Linus On-Demand Rate per Hour





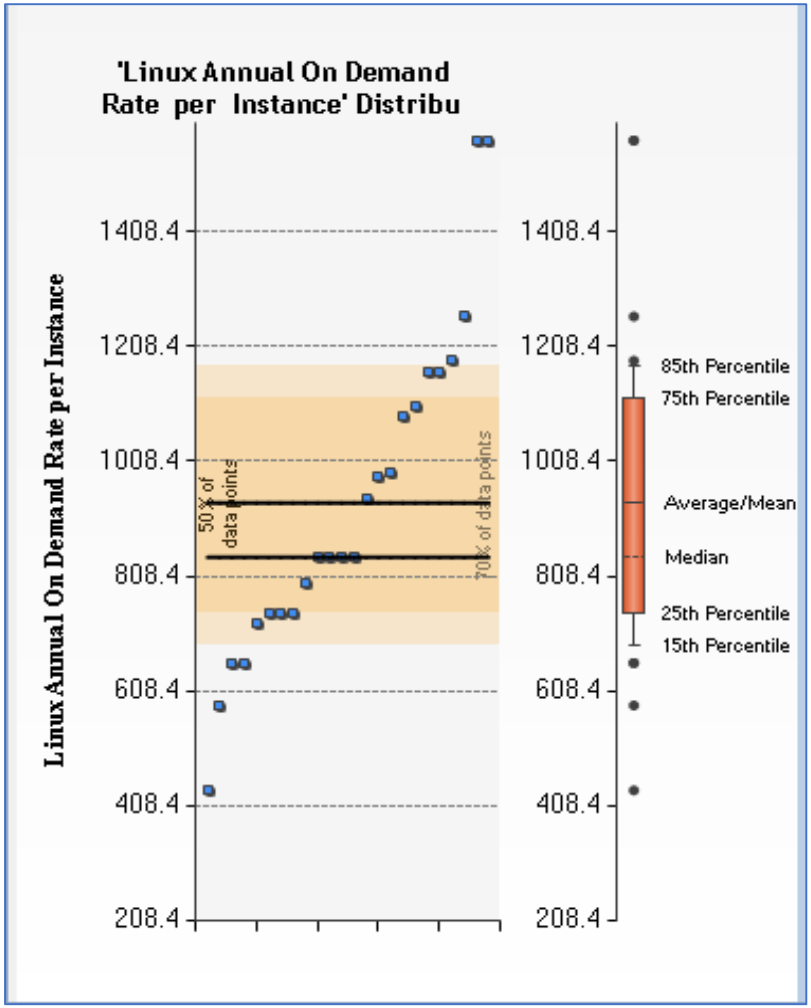
Linux On-Demand Annual Rate ($)

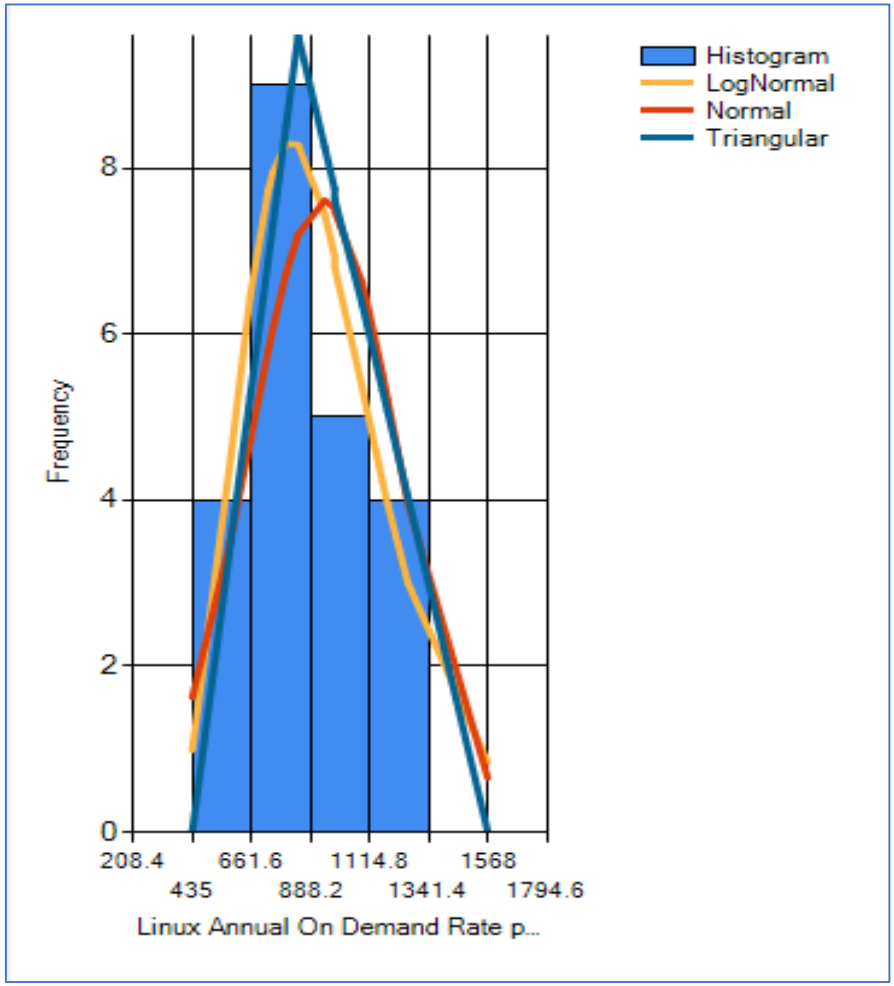
	Data	LN	Normal	TRI
Min	435			
Max	1568			
25%	745			
75%	1119.25			
Mean	936.125	938.489	936.125	948
Median	841	895.904	936.125	926.248
Mode	841	816.443	936.125	841
Standard Deviation	284.94	292.78	284.94	231.273
Coefficient of Variation	0.304	0.312	0.304	0.244

DEPENDENT	INDEPENDENT	R-Value
Linux Annual On Demand Rate per Instance	GB RAM per Instance	0.8007
Linux Annual On Demand Rate per Instance	Cloud Type Number	0.2898
Linux Annual On Demand Rate per Instance	Cloud Instance Number	0.2927
Linux Annual On Demand Rate per Instance	**GB Storage per Instance**	**0.0990**
Linux Annual On Demand Rate per Instance	CPU Plus SSD Number	0.2517

PREDICTIVE CER	R^2-Value	Fit
= 1.013 * ([GB RAM per Instance]^2) + 24.464 * [GB RAM per Instance] + 620.154	R^2 = 0.647	Polynomial
= 43.390 * [GB RAM per Instance] + 558.998	**R^2 = 0.641**	**Linear**
= 602.129 * e^(0.046 * [GB RAM per Instance])	R^2 = 0.622	Exponential
= 489.797 * ([GB RAM per Instance]^0.312)	R^2 = 0.597	Power
= 43.901 * [GB RAM per Instance] + 0.354 * [GB Storage per Instance] + 529.714	**R^2 = 0.663**	**Multiple**

Linux On-Demand Annual Rate ($)



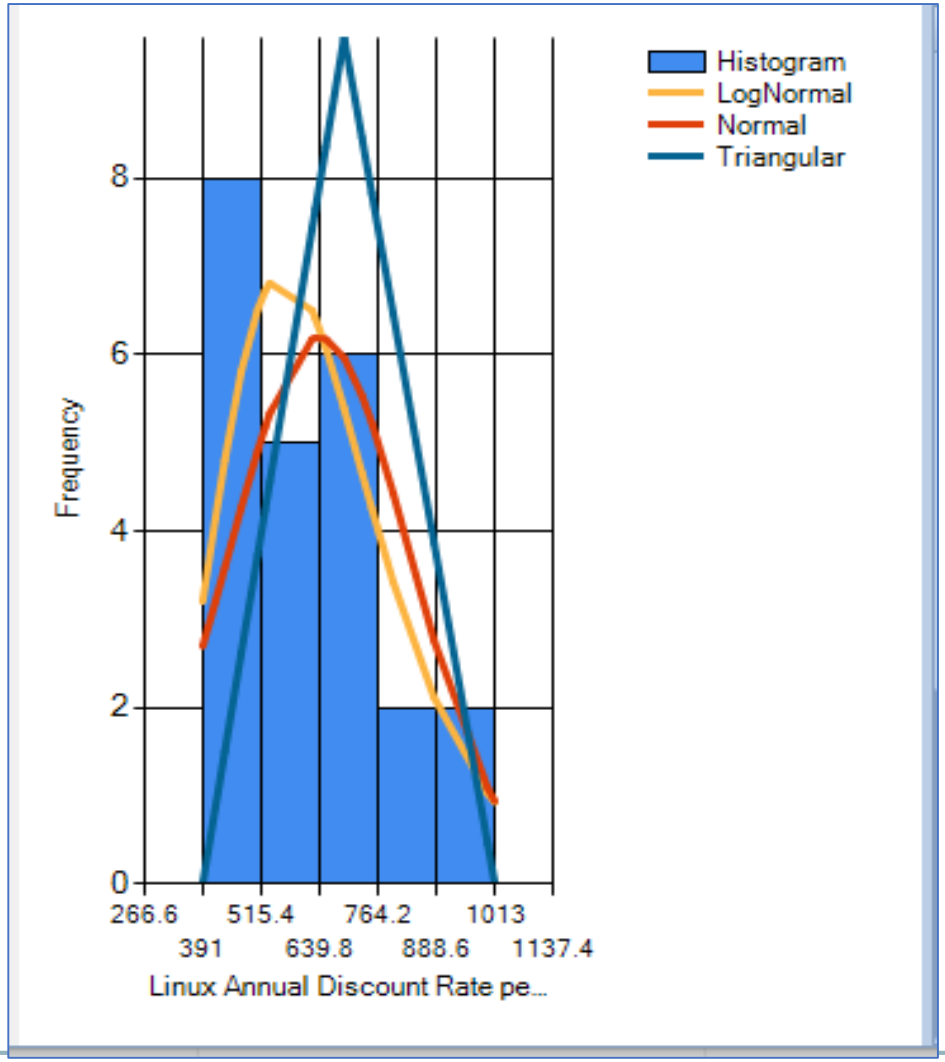


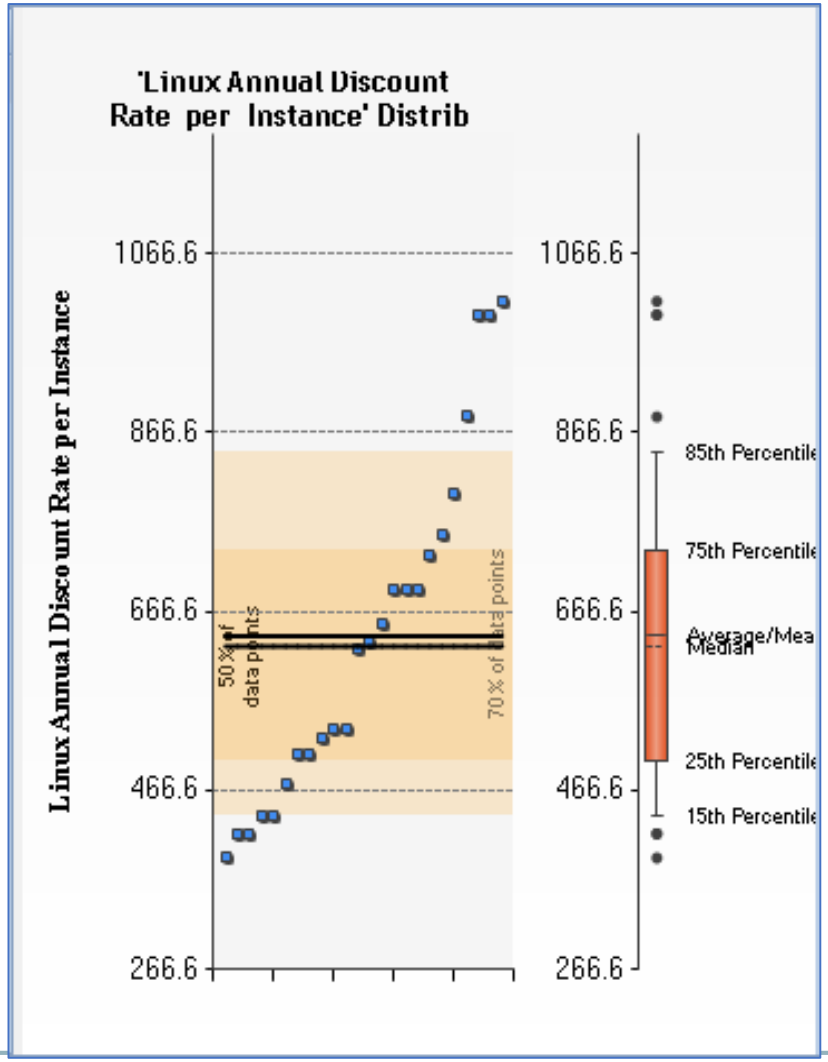
Linux On-Demand Annual Discount Rate

	Data	LN	Normal	TRI
Min	391			
Max	1013			
25%	499.25			
75%	735.25			
Mean	639.125	639.905	639.125	698.667
Median	627.5	613.036	639.125	697.04
Mode	692	562.636	639.125	692
Standard Deviation	192.185	191.521	192.185	126.965
Coefficient of Variation	0.301	0.299	0.301	0.182

DEPENDENT	INDEPENDENT	R-Value
Linux Annual Discount Rate per Instance	GB RAM per Instance	0.7030
Linux Annual Discount Rate per Instance	GB Storage per Instance	0.5402
Linux Annual Discount Rate per Instance	CPU Plus SSD Number	0.3209
Linux Annual Discount Rate per Instance	Cloud Instance Number	0.3014
Linux Annual Discount Rate per Instance	Cloud Type Number	0.2935
PREDICTIVE CER	**R^2-Value**	**Fit**
= 430.087 * e^(0.041 * [GB RAM per Instance])	R^2 = 0.536	Exponential
(All other recommended Regression CERs had R^2 of less than 0.50.)		
= 27.045 * [GB RAM per Instance] + 0.936 * [GB Storage per Instance] + 338.483	R^2 = 0.837	Multiple

Linux On-Demand Annual Discount Rate





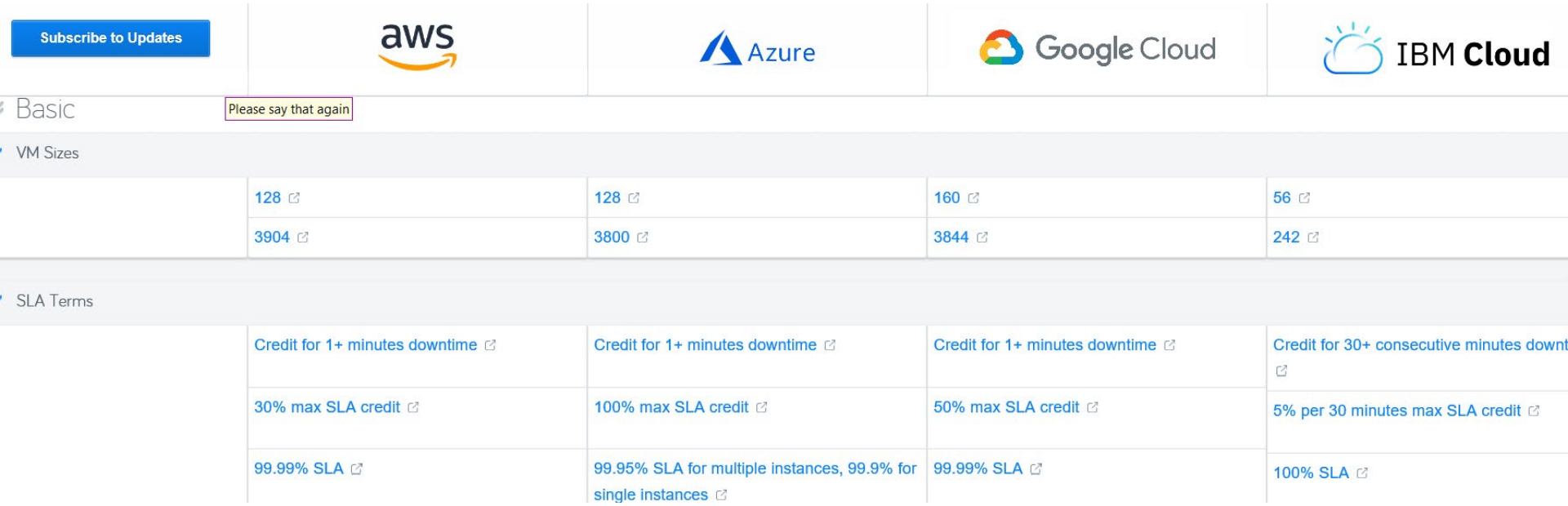
IT Dashboard Sources

Exhibit 53A	An overview of each Exhibit 53 submission
Exhibit 53C	An overview of each agencies' Cloud Computing Portfolio.
Exhibit 53 Funding Sources	Funding source information includes Total IT Spending for the PY, CY and BY, and a breakdown of DME and O&M costs.
Activities	Activity data includes cost, schedule and variance information
Investment Baseline History	Provides a history of investment rebaselines, replans and corrections.
CIO Evaluation History	Provides a history of the evaluation by CIO ratings and comments.
Contracts	Overview of contract information including information matched with USAspending.gov
Exhibit 300A	Exhibit 300A submission information, as well as Evaluation (by Agency CIO) information, and dates of last change to activities, contracts, and performance metrics.
Investment Trends	Provides a history of investment CIO Evaluations over time, as well as legacy investment ratings and variances
Performance Metrics	Overview of an investment's operational performance in regards to "Results-Specific" and "Activities or Technology-Specific" metrics
Performance Metric Actuals	Overview of the actual results of an investment's performance metrics in comparison to their targets
Projects	Project data includes cost, schedule and variance information
Life Cycle Costs	Life Cycle Cost information broken down into planning costs, DME, O&M, and Government FTE costs.
Techincal Solutions Requirements	Indicates whether an investment in whole or in part specifically addresses various technical solution requirements
Techincal Solutions Requirements URLs	Provides a list of URLs which correspond to the technical solution requirements
Investments Eliminated or Reduced	An agency-provided list of investments that will result in the elimination or reduction of other investments.
Investment Related URLs	Provides URLs related to the investment and indicates how they are related
Bureau	Provides a list of Bureaus based on the OMB Circular A-11 Appendix C
Agency	Provides a list of Agencies based on the OMB Circular A-11 Appendix C

Key Deliverables

- There are dozens of data points for acquisition documents, for example:
  - AoA
  - TEMP
  - CONOPS

Right scale cloud comparison tool



Subscribe to Updates

	aws	Azure	Google Cloud	IBM Cloud
Basic	Please say that again			
VM Sizes				
	128	128	160	56
	3904	3800	3844	242
SLA Terms				
	Credit for 1+ minutes downtime	Credit for 1+ minutes downtime	Credit for 1+ minutes downtime	Credit for 30+ consecutive minutes downt
	30% max SLA credit	100% max SLA credit	50% max SLA credit	5% per 30 minutes max SLA credit
	99.99% SLA	99.95% SLA for multiple instances, 99.9% for single instances	99.99% SLA	100% SLA

https://www.rightscale.com/cloud-comparison-tool/