Horizon Card Sort Findings

January 2015

Objectives

Why bother with a card sort?

The goal of this research was to understand how users grouped the top of the tree items in Horizon, as well as how they labeled these groupings.

This data can then be used to *help inform* future IA decisions.

An open card sort was used to explore the category types users generated, in addition to allowing them to create the groupings that were most meaningful to them.

Method

Participants

Data from 65 participants was used in the final analysis.

Data from 90 participants with incomplete/ abandoned surveys were deleted.

- Data from 4 participants who sorted into 13 or more categories were also removed from the analysis.
- Data from 3 participants who sorted in 2 or fewer categories was also removed from the analysis.

This resulted in 45 participants with recent (past 3 months) Horizon experience, and 20 without.

Participant were recruited through a variety of sources, including:

- OpenStack mailing lists,
- IRC channels,
- LinkedIn groups, and
- University mailing lists

An effort was made to recruit a mix of participants in terms of their recent experience with Horizon.

Participants had the opportunity to enter their name in a drawing for a tablet PC after completing their card sort.

Method

Procedures & Materials

Participants completed a brief demographic survey, and then the open card sort.

- The sort was provided using OpenSort.
- After completing the sort, participants were then asked to enter their contact information for an entry into the tablet drawing.

18 cards were used in the sort.

These cards used in the study are currently top-of-tree for Horizon or recommended by the Horizon PTL, David Lyle.

- Initial pilot data indicated that the terms themselves were ambiguous, and not all participants reviewed the definitions.
- Therefore, the cards were presented to participants as a term definition, followed by the term itself.

The cards definitions were written by Tiffany Bockman and reviewed by David Lyle.

Findings
Visualizations

Similarity Matrix

Combined Model

Similarity matrix

Network Topology																
61	API Access															
60	60	Routers														
60	60	61	Networks													
53	53	54	54	Load Balancers												
47	48	46	46	41	Key Pairs											
46	47	45	45	40	60	Security Groups										
52	53	51	51	45	55	53	Floating IPs									
34	34	35	35	41	31	31	32	Data Processing (Hadoop clusters)								
35	35	36	36	42	31	31	32	56	Database							
39	39	40	40	47	33	32	36	49	51	VPNs						
39	39	40	40	47	33	32	36	49	51	61	Firewalls					
39	39	40	40	47	33	32	36	49	51	61	61	Stacks				
39	39	40	40	47	34	33	36	50	53	57	57	57	Object Storage			
34	34	33	33	29	44	45	38	24	24	25	25	25	26	Images		
32	32	31	31	29	40	41	35	25	24	25	25	25	26	54	Volumes	
28	28	27	27	25	34	35	31	22	21	22	22	22	23	43	48	Instances
22	22	22	22	21	24	24	23	27	25	23	23	23	24	22	23	28	Overview

Indicates how often each card is paired with the other.

Maximum score is 65.

Take-aways

Most of the cards show semi-consistent pairings (70% agreement; SM score of 45 or>).

Data Processing and Overview have the lowest consistency for pairing.

- Overview is often placed in a group by itself.

Similarity Matrix

Combined Model – Balanced Samples

Similarity matrix

Key Pairs														
35	Security Groups													
31	29	Floating IPs												
26	25	30	Routers											
26	25	30	36	Networks										
26	25	30	36	36	Network Topology									
27	26	31	36	36	36	API Access								
23	22	27	33	33	33	33	Load Balancers							
16	16	18	19	19	19	19	21	Data Processing (Hadoop clusters)						
16	16	18	20	20	20	20	22	32	Database					
18	17	22	24	24	24	24	27	28	30	VPNs				
18	17	22	24	24	24	24	27	28	30	36	Firewalls			
18	17	22	24	24	24	24	27	28	30	36	36	Stacks		
18	17	21	23	23	23	23	26	29	31	34	34	34	Object Storage	
25	26	19	17	17	17	17	14	12	12	13	13	13	13	Images
22	23	17	15	15	15	15	14	12	12	13	13	13	13	33	Volumes
18	19	15	13	13	13	13	12	10	10	11	11	11	11	24	26	Instances
11	11	10	9	9	9	9	9	13	13	11	11	11	11	10	10	15	Overview

Indicates how often each card is paired with the other.

Maximum score is 40.

Take-aways

Images and Instances have lower consistency than before (65%).

- Once again, Data Processing and Overview have the lowest consistency for pairing.
- Overview is still an oddball.
  - It is still most frequently placed in a group by itself.

Findings

Preliminary Analysis

Explanations for Groupings

53/65 participants provided meaningful explanations for their groupings.

The most frequent explanation was functional proximity.

- *"I grouped them largely by the functionality... that is, if it had to do with virtual instances, it went into virtualization, if it had to do with allocation of storage it went into resource allocation."*
- *"I find it easier when using a GUI for related things to be grouped together. Even if you might not use all things in that group at the same time, it creates logical pairing so you don't have to hunt something down."*

A closely related explanation was to group items based on what are commonly used together

	Horizon User	Non-Horizon User	Total
Functional Proximity	22	9	31
Commonly Used Together	10	2	12
Experience/ Familiarity	0	4	4
Perceived Importance	1	1	2
Frequency of Use	2	0	2
Contribution Interest Areas	1	0	1
Business Value	1	0	1

PCA

Full Sample

"Top-down from Concepts via Functionality to Services."

PCA (Participant-Centric Analysis) compares each individual's sort to determine which individual sort has the greatest number of participants' sorts that are consistent with it.

PCA indicated that the IA supported by the greatest number of sorts (37/65) utilized 6 categories.

- A second analysis was conducted that excluded sorts with unnamed and/or nonsense categories.
  - Again, PCA results in a best fit with the same 6 category model (28/47).

However, Horizon users were over-represented in this analysis.

When corrected to give equal weight to both populations, 50% of sorts agreed with this model.

Group 1 *Networking* *Network*	**Group 2** *Storage* *Persistent Storage* *Cloud Storage*	**Group 3** *Overview* *Horizon* *UI* *End User*	**Group 4** *Data Management* *Clusters* *Data Processing* *PaaS*	**Group 5** *Concepts* *VMs* *Instances* *Shared Services*	**Group 6** *Security*
Security Groups	Volumes	Overview	Database	Instance	Key Pairs
Floating IPs	Object Storage	API Access	Database Processing	Image	
Network Topology				Stacks	
Networks					
Routers					
Load Balancers					
Firewalls					
VPNs					

PCA

Corrected Sample

The best fit (60%) PCA with sample sizes corrected for Horizon experience is with a 5 category IA.

This model was similar to the 6 category model resultant from the full sample PCA, except:

- Database & Data Processing were merged into the Storage category.
- Stacks was moved from the Concept category to Overview.
- Security Groups moved from Networking to Security.
- Load Balancers moved to VMs, from Networking.

"I grouped them by what I have used in the past and how I associated them with each other."

Group 1 *Networking*	**Group 2** *Storage* *Data*	**Group 3** *Overview* *User Interface Management*	**Group 4** *VMs*	**Group 5** *Security*
Floating IPs	Volumes	Overview	Instance	Key Pairs
Network Topology	Object Storage	API Access	Image	Security Groups
Networks	Database	Stacks	Load Balancers	
Routers	Data Processing			
Firewalls				
VPNs				

Category Analysis

Combined Sort , Corrected Sample

Category standardization was employed utilizing Donna Spencer's approach.

This resulted in 19 standardized categories.

Cards best-fit, based on correlational analysis, identified 6/19 categories for primary groupings.

- Generally consistent with PCA.

Both Firewalls and Database had 2 categories with close correlations.

Group 1 *Networking*	**Group 2** *Storage*	**Group 3** *Overview*	**Group 4** *Data*	**Group 5** *Instances Concepts*	**Group 6** *Security*
Floating IPs	Volumes	Overview	*Database (12.5%)*	Instance	Key Pairs
Network Topology	Object Storage	API Access	Database Processing	Images	Security Groups
Networks	*Database (12.5 %)*			*Stacks*	*Firewalls (25%)*
Routers					
Load Balancers					
Firewalls (35%)					
VPNs					

Cluster Analysis

6 Category Model

Group 1: 100%	Group 2: 81%	Group 3:77%	Group 4: 82%	Group 5: 83%	Group 6: 100%
Overview	Network Topology	Data Processing (Hadoop clusters)	Images	Key Pairs	Instances
	API Access	Database	Volumes	Security Groups	
	Routers	VPNs		Floating IPs	
	Networks	Firewalls			
	Load Balancers	Stacks			
		Object Storage			

Conclusion

Limitations with the open card sort methodology.

Why people group items the way they did is unknown.

- Self-report data is incomplete and sometimes ambiguous.

The groupings may not be task based.

- How people organize abstract concepts, versus how they organize them in the flow of task performance, often differ.
  - Therefore, this grouping may not result in the greatest usability.

Next Steps

1. Determine which of the potential model(s) to take forward.
   - Inclined to focus on a 5 category model
   - Standardize the category names
2. Conduct a closed card sort utilizing the most promising categories identified in this research.
   - This provides a mechanism for validating the labels generated.
3. Follow up with a usability study with proposed IA.
   - This is done to determine if the categories hold up in a task-based context.

Thank you

Categories Included

1. Overview provides a high-level view of the resources allocated to the current project. (Overview)
2. Instances are virtual machines that are started from an image. You can boot an instance directly from an image or from an existing volume. (Instances)
3. Block storage volumes can be attached to instances. These volumes provide persistent fixed storage. (Volumes)
4. An image contains the operating system and other software for launching a virtual machine. (Images)
5. Security groups allow administrators to create IP filter rules that determine network traffic for instances. Rules are project-specific. (Security Groups)
6. Key pairs can be injected into virtual machines at launch to grant users SSH access to instances hosted on the virtual machine. A single key pair can be used with multiple instances within the same project. (Key Pairs)
7. Cloud users can associate a Floating IP to an instance or project from a pre-allocated pool. A Floating IP provides access to an instance from a public network. (Floating IPs)
8. OpenStack services and service endpoints are displayed in a table in the Horizon User Portal for easy reference. (API Access)
9. Network Topology presents a visual representation of the cloud's network configuration. (Network Topology)
10. Networks provide the communication channels for instances in the cloud. (Networks)

Categories Included, cont.

11. Routers provide a connection between two isolated networks. (Routers)
12. Load balancers distribute network traffic across virtual machines. You can create pools to receive and process traffic. (Load Balancers)
13. Stacks are template-driven resource pools that are designed to run specific workloads or applications. (Stacks)
14. Firewalls provide protection by limiting the network traffic that enters and exits Private (tenant) networks. (Firewalls)
15. Virtual Private Networks (VPNs) extend Private (tenant) networks across unsecured public networks. (VPNs)
16. Object storage is managed using containers, where users can store, retrieve and delete objects using a REST API. (Object Storage)
17. One or more servers can be created and grouped together to manage database instances. (Database)
18. These special clusters (also known as Hadoop clusters) are used to store and analyze huge amounts of unstructured data. (Data Processing)

Cluster Analysis

6 Category Model – Combined Data, Corrected Sample

Group 1: 100%	Group 2: 83%	Group 3:73%	Group 4: 83%	Group 5: 83%	Group 6: 100%
Overview	Network Topology	Data Processing (Hadoop clusters)	Images	Key Pairs	Instances
	API Access	Database	Volumes	Security Groups	
	Routers	VPNs		Floating IPs	
	Networks	Firewalls			
	Load Balancers	Stacks			
		Object Storage			

Similarity Matrix

Horizon vs. Non-Horizon Users

Horizon users (N= 45)

Similarity matrix

Network Topology																
42	API Access															
41	41	Routers														
41	41	42	Networks													
34	34	35	35	Load Balancers												
31	31	30	30	25	Key Pairs											
30	30	29	29	24	40	Security Groups										
34	34	33	33	27	37	36	Floating IPs									
19	19	20	20	25	17	17	17	Data Processing (Hadoop clusters)								
19	19	20	20	25	16	16	17	38	Database							
23	23	24	24	30	19	18	21	33	33	VPNs						
23	23	24	24	30	19	18	21	33	33	41	Firewalls					
23	23	24	24	30	19	18	21	33	33	41	41	Stacks				
23	23	24	24	30	19	18	21	33	33	39	39	39	Object Storage			
21	21	20	20	17	29	29	25	13	12	14	14	14	14	Images		
19	19	18	18	17	25	25	22	13	12	14	14	14	14	34	Volumes	
16	16	15	15	14	22	22	19	11	10	12	12	12	12	29	33	Instances
12	12	12	12	12	14	14	12	18	17	15	15	15	15	13	13	14	Overview

Non-Horizon users (N= 20)

Similarity matrix

Key Pairs																
18	Security Groups															
12	13	Images														
12	13	19	Volumes													
9	10	12	12	Instances												
9	9	7	7	6	VPNs											
9	9	7	7	6	19	Firewalls										
9	9	7	7	6	19	19	Stacks									
9	9	7	7	6	18	18	18	Database								
9	9	7	7	6	18	18	18	19	Object Storage							
11	11	8	8	7	17	17	17	16	16	Routers						
11	11	8	8	7	17	17	17	16	16	19	Networks					
11	11	8	8	7	17	17	17	16	16	19	19	Network Topology				
10	10	8	8	7	17	17	17	16	16	18	18	18	Load Balancers			
9	9	7	7	6	16	16	16	17	17	15	15	15	15	Data Processing (Hadoop clusters)		
13	13	8	8	7	15	15	15	14	14	17	17	17	16	13	API Access	
15	14	9	9	8	13	13	13	12	12	15	15	15	14	12	17	Floating IPs
5	5	5	5	9	4	4	4	4	4	4	4	4	4	4	4	5	Overview

4 Category Model

Combined Data

Horizon users (N =45) are over-represented as compared to non-Horizon users (N=20)

Group 1: 100% *Overview* *Dashboard*	**Group 2: 68%** *Services* *Basic Concepts* *Basic Cloud Services*	**Group 3: 72%** *Advanced Concepts* *Secure Access* *Networking Services*	**Group 4: 77%** *Orchestration* *Clustering*
Overview	Images	Network Topology	Data Processing (Hadoop clusters)
	Volumes	API Access	Database
	Instances	Routers	VPNs
		Networks	Firewalls
		Load Balancers	Stacks
		Key Pairs	Object Storage
		Security Groups	
		Floating IP	

4 Category Model

Horizon vs. Non-Horizon Users

Horizon Users (N = 45)

Group 1 *Insight* *Tenant* *Overview* *General Status*	**Group 2** *Networking* *Networking* *Services*	**Group 3** *Concepts* *Basic Concepts* *Basic Cloud* *Services*	**Group 4** *Clustering* *Orchestration*
Overview	API Access	Images	Data Processing
	Network Topology	Volumes	Database
	Routers	Instances	VPNs
	Networks		Firewalls
	Load Balancers		Object Storage
	Floating IPs		Stacks
	Security groups		
	Key Pairs		

Non-Horizon Users (N = 20

Group1 *Overview* *Resources*	**Group2** *Network* *Cloud Config*	**Group3** *Storage* *Machines*	**Group4** *VMs* *Instances*
Overview	API Access	Key Pairs	Instances
	Network Topology	Security Groups	
	Routers	Images	
	Networks	Volumes	
	Load Balancers		
	Floating IPs		
	VPNs		
	Stacks		
	Database		
	Data Processing		
	Firewalls		
	Object Storage		