OpenStack Swift Reference Designs

This document contains four OpenStack Swift reference designs: small, medium, large, and compute and object cloud.

The four designs have separate high level specifications and architecture diagrams but all re-use a common set of bill of materials, racking rules, and network plug suggestions.

Small	Medium	Large
Integrated Proxy. 24 object server limit.	Dedicated proxy nodes.	Dedicated proxy and dedicated meta-data nodes.

Guidelines for choosing between small and medium

Storage size:

Small is limited to a maximum of 24 object servers. If you need more storage than can fit in 24 object servers you should choose medium.

Background:

Swift small contains exactly 3 Swift proxies which run on the 3 controllers. There are no horizontal scaling guidelines going beyond 3 controllers. Given the horizontal scaling rule of thumb of 1 proxy server to 8 object servers you are limited to a maximum of 24 object servers.

Performance:

Depending on your object storage workload characteristics you may find that the proxy servers become the bottleneck due to either the workload or the sharing of controller server resources between the control plane services and the Swift proxy service. Additionally, depending on the workload you may need more than 3 proxies to handle 24 object servers. If either of these issues becomes a factor, moving to Swift medium with its dedicated Swift proxy nodes would alleviate the issue.

Guidelines for choosing between medium and large

Cost savings:

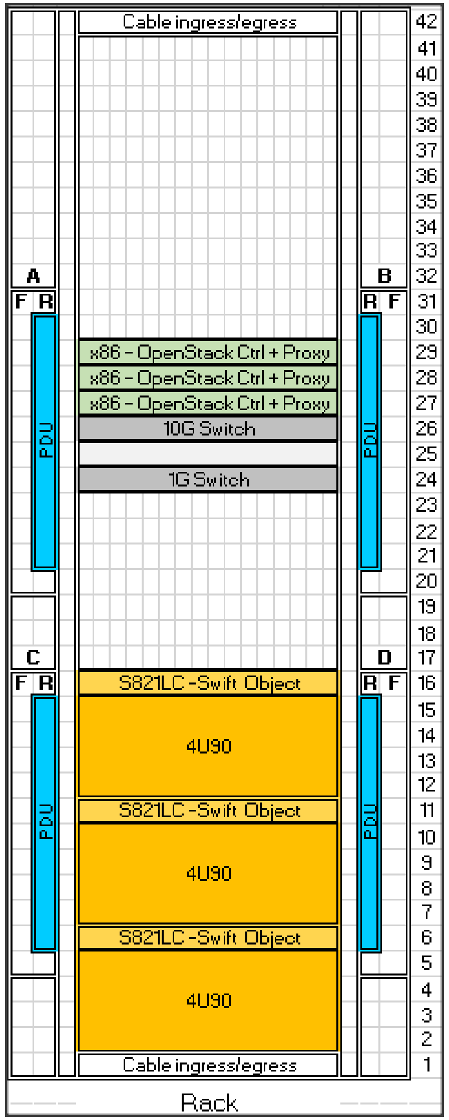
As you scale the medium architecture horizontally, given workload specifics you may begin to have under utilized SSDs which are used to hold the account and container Swift rings. At some point you hit a tipping point where it is more cost effective to host the account and container rings with their associated SSDs in dedicated metadata servers. You would then scale the metadata servers horizontally with a rule of thumb ratio of 1 metadata server to 6 object servers. The exact point you when you hit this cost savings threshold is dependent upon server and SSD pricing.

Performance:

The object storage workload specifics could favor large with its dedicated metadata servers before the cost savings threshold is hit. For example, if the workload has an extremely high number of users and containers but lower raw object storage needs, and the workload is doing a lot of account and container lookup, the large configuration with its dedicated metadata servers may be a better fit.

Small Swift Cluster

Swift Small – Starter Config– High Level Specification Sheet



OpenStack Software Stack:
Ubuntu 14.04 (all nodes)
..Openstack
..
..OpsPanel + Horizon DashBoard
-Nagios
- ELK Stack (Elasticsearch, Logstash, Kibana)

**Contact IBM for Redundant/Bonding Options

Network : (non HA) – no Bonding **
1 x Mellanox SX1410 (8831-S48)
1 x Lenovo G8052 (7120-48E)

Rack:
QTY: 1
SlimRack 7965-94Y
PDUs x 4

OpenStack Controller & Proxy: x86
QTY: **3**

Server Config: (Lenovo 3550-M5 (1U)
20 Cores (2.0Ghz), 256GB,
2 x 4TB SATA HDDs
1 x 2-Port 10G NIC (Intel 10G/Mellanox)

Swift Object /MetaData
QTY: **3**

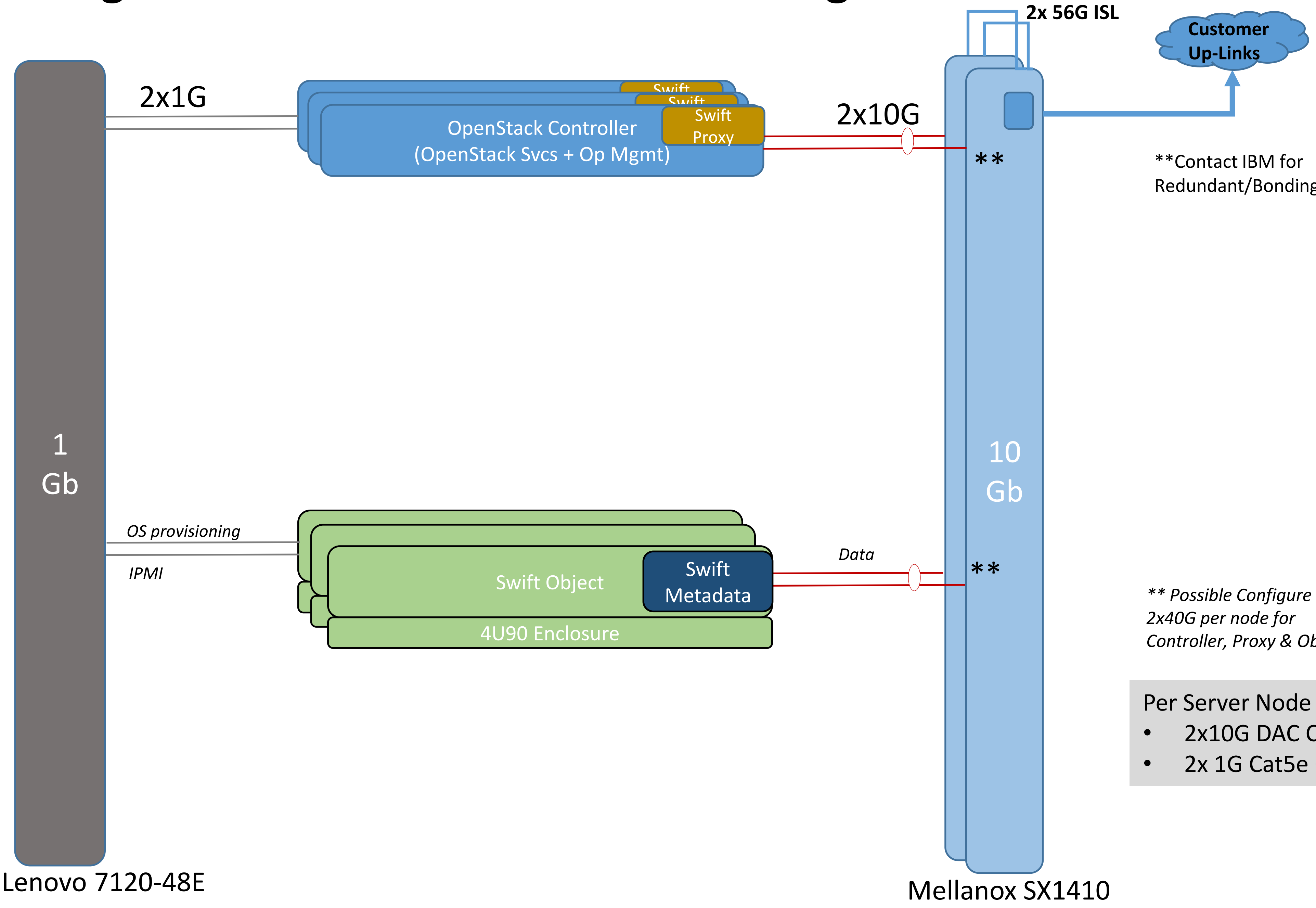
Per Server Config: (Stratton 8001-21C) (1U)
16 Cores (2.3Ghz), 256GB

- (OS) 2+ 128GB DOM + 4 x SSDs x 240GB
- 1 x 2-Port 10G NIC (Intel/Mellanox)
- 1 x LSI 3008 External SAS
- 1 x MegaRAID SAS controller

Expansion Drawer (4U) : Supermicro SC946ED - 4U90
90 LFF – 2TB SAS HDDs

****Notes:**
a) Proc + Memory config change is required based on actual performance requirement

Swift Small - High Level Network Architecture Diagram



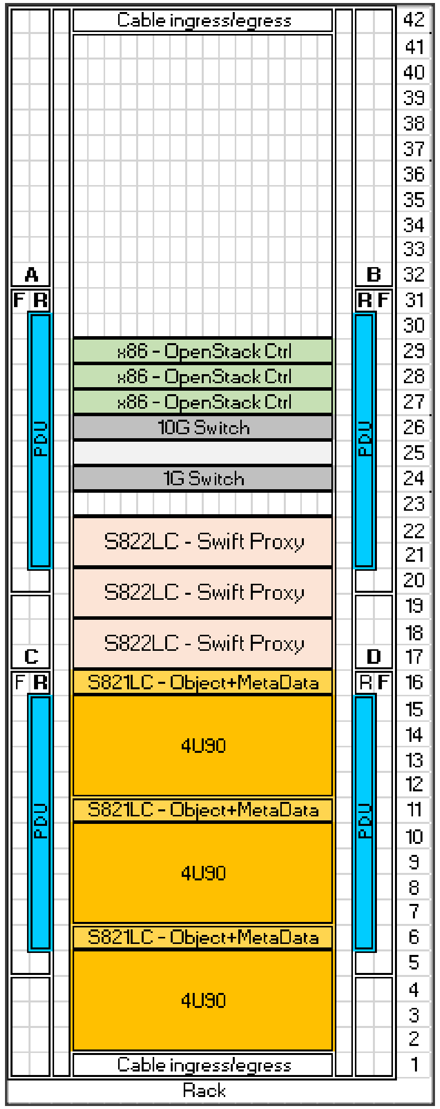
**Contact IBM for
Redundant/Bonding Options

*** Possible Configure
2x40G per node for
Controller, Proxy & Object*

Per Server Node
- 2x10G DAC Cables
- 2x 1G Cat5e Cables

Medium Swift Cluster

Swift Medium– Starter Config– High Level Specification Sheet



OpenStack Software Stack:
Ubuntu 14.04 (all nodes)
..Openstack
..
..OpsPanel + Horizon DashBoard
-Nagios
- ELK Stack (Elasticsearch, Logstash, Kibana)

**Contact IBM for Redundant/Bonding Options

Network : (non HA) – no Bonding **
1 x Mellanox SX1410 (8831-S48)
1 x Lenovo G8052 (7120-48E)

Rack:
QTY: 1
SlimRack 7965-94Y
PDUs x 4

OpenStack Controller
QTY: 3

Server Config: (Lenovo 3550-M5 (1U)
20 Cores (2.0Ghz), 256GB,
2 x 4TB SATA HDDs
1 x 2-Port 10G NIC (Intel 10G/Mellanox)

Swift Object /MetaData
QTY: 3

Per Server Config: (Stratton 8001-21C) (1U)
16 Cores (2.3Ghz), 256GB

- (OS) 2+ 128GB DOM + 4 x SSDs x 240GB
- 1 x 2-Port 10G NIC (Intel/Mellanox)
- 1 x LSI 3008 External SAS
- 1 x MegaRAID SAS controller

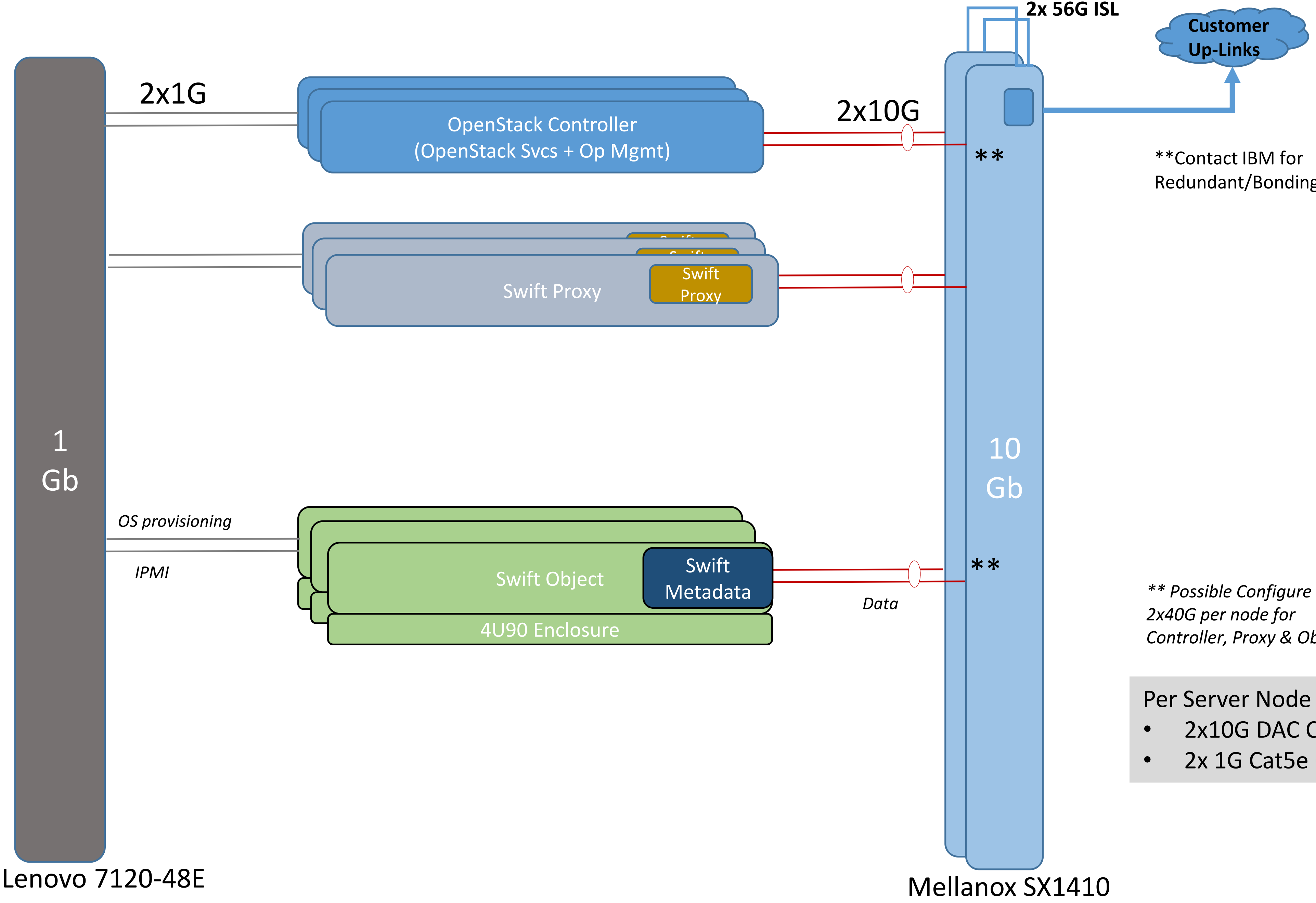
Expansion Drawer (4U) : Supermicro SC946ED - 4U90
90 LFF – 2TB SAS HDDs

Swift Proxy:
QTY: 3

Per Server Config: (Briggs 8001-22C) (2U)
20 Cores @2.92Ghz, 256GB
2 x 2 TB SATA HDDs
1 x 2-Port 10G NIC (Intel 10G/Mellanox)

****Notes:**
a) Proc + Memory config change is required based on actual performance requirement

Swift Medium - High Level Network Architecture Diagram



**Contact IBM for Redundant/Bonding Options

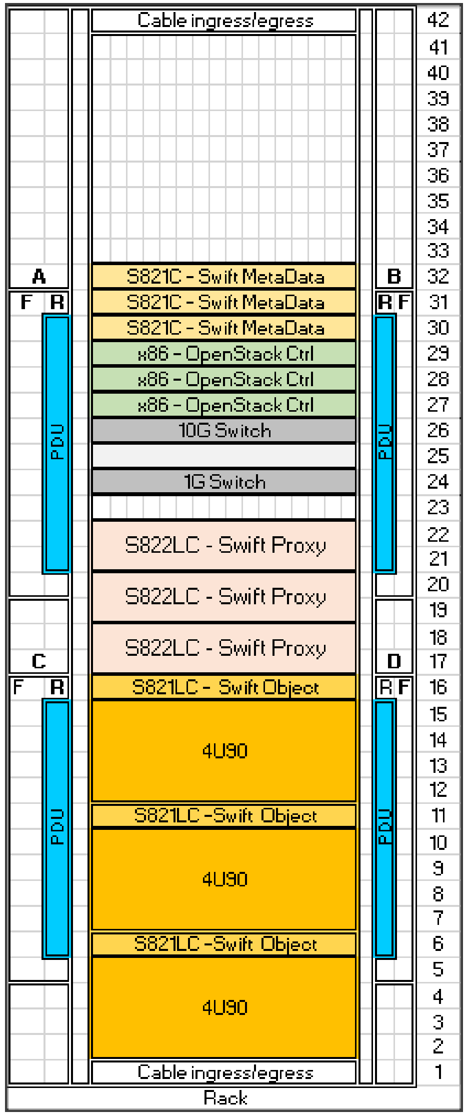
*** Possible Configure 2x40G per node for Controller, Proxy & Object*

Per Server Node

- 2x10G DAC Cables
- 2x 1G Cat5e Cables

Large Swift Cluster

Swift Large – Starter Config– High Level Specification Sheet



OpenStack Software Stack:
Ubuntu 14.04 (all nodes)
..Openstack
..
..OpsPanel + Horizon DashBoard
-Nagios
- ELK Stack (Elasticsearch, Logstash, Kibana)

**Contact IBM for
Redundant/Bonding Options

Network : (non HA) – no Bonding **
1 x Mellanox SX1410 (8831-S48)
1 x Lenovo G8052 (7120-48E)

Rack:
QTY: 1
SlimRack 7965-94Y
PDUs x 4

OpenStack Controller
QTY: 3

Server Config: (Lenovo 3550-M5 (1U)
20 Cores (2.0Ghz), 256GB,
2 x 4TB SATA HDDs
1 x 2-Port 10G NIC (Intel 10G/Mellanox)

Swift MetaData
QTY: 3

Per Server Config: (Stratton 8001-21C) (1U)
16 Cores (2.3Ghz), 256GB
- (OS) 2+ 128GB DOM + 4 x SSDs x 240GB
- 1 x 2-Port 10G NIC (Intel/Mellanox)

Swift Proxy:
QTY: 3

Per Server Config: (Briggs 8001-22C) (2U)
20 Cores @ 2.92Ghz, 256GB
2 x 2 TB SATA HDDs
1 x 2-Port 10G NIC (Intel 10G/Mellanox)

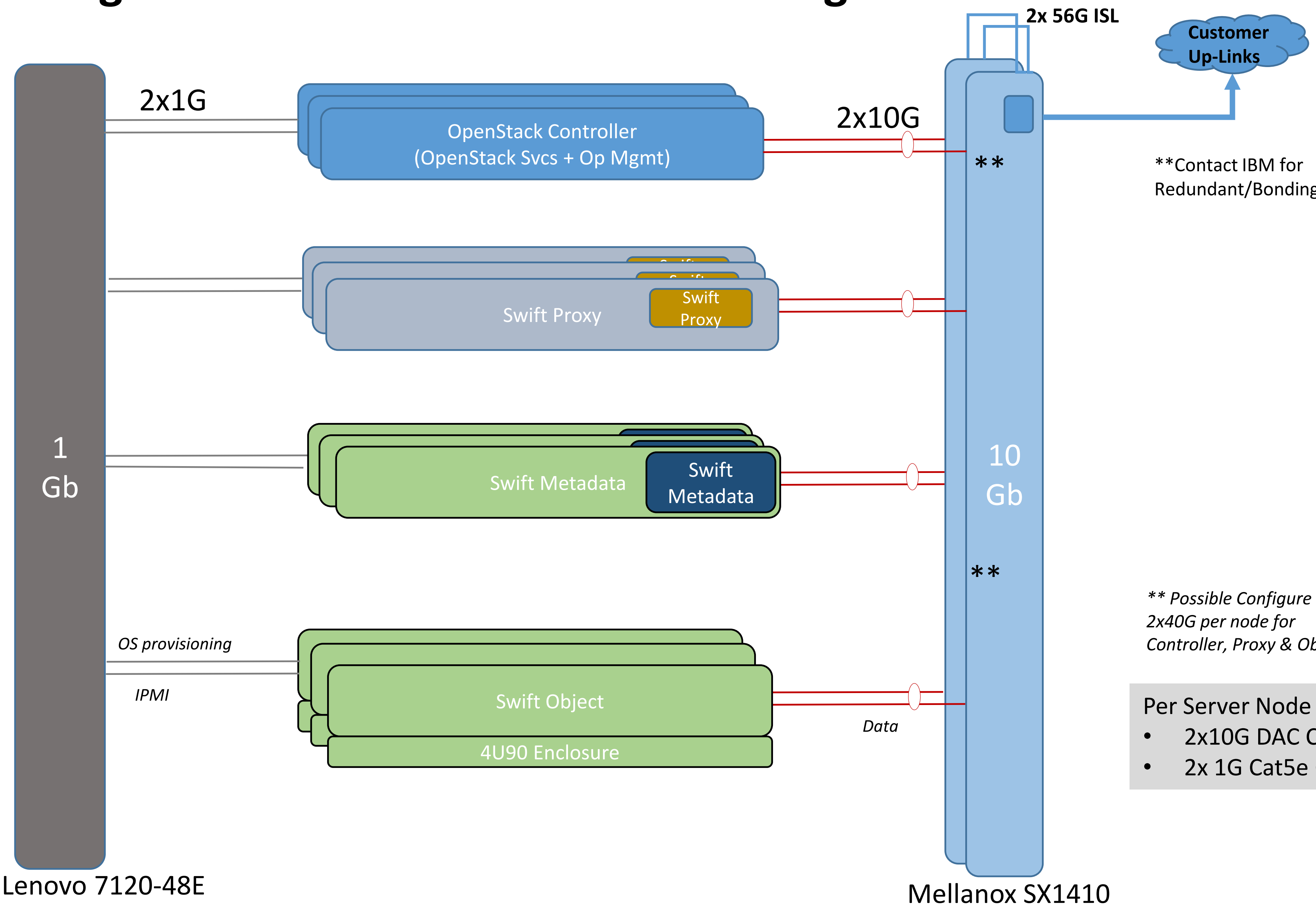
Swift Object
QTY: 3

Per Server Config: (Stratton 8001-21C) (1U)
16 Cores (2.3Ghz), 256GB
- (OS) 2+ 128GB DOM
- 1 x 2-Port 10G NIC (Intel/Mellanox)
- 1 x LSI 3008 External SAS
- 1 x MegaRAID SAS controller

Expansion Drawer (4U) : Supermicro SC946ED - 4U90
90 LFF – 2TB SAS HDDs

****Notes:**
a) Proc + Memory config change is required based on actual performance requirement

Swift Large - High Level Network Architecture Diagram



**Contact IBM for Redundant/Bonding Options

*** Possible Configure 2x40G per node for Controller, Proxy & Object*

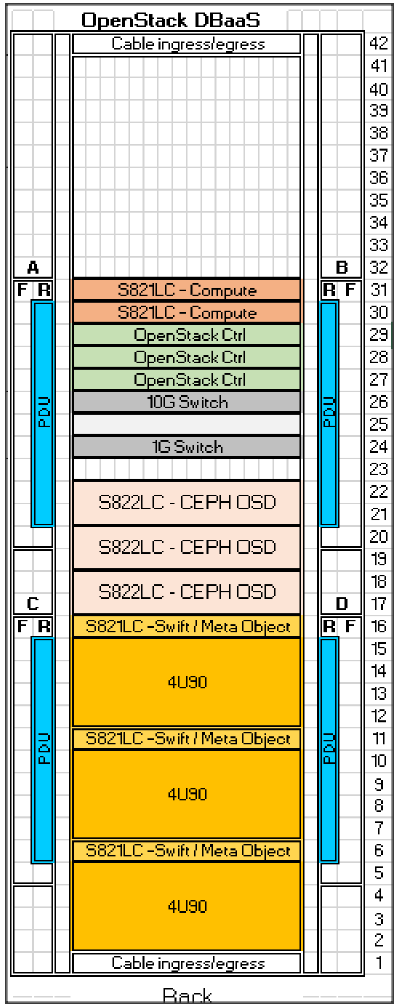
Per Server Node

- 2x10G DAC Cables
- 2x 1G Cat5e Cables

Private Cloud with Object Storage and Compute

Swift with Private Compute Cloud – Starter Config– High Level Specification Sheet

OpenStack Software Stack:
Ubuntu 14.04 (all nodes)
..Openstack
..
..OpsPanel + Horizon DashBoard
-Nagios
- ELK Stack (Elasticsearch, Logstash, Kibana)



****Notes:**
a) Openstack & Proxy Node can be combined (if requires lesser than 24 SWIFT Objects)
b) Compute qty + Memory config change is required based on actual performance requirement
c) Dedicated Swift Meta Data Server maybe required

OpenStack Controller & Proxy: x86
QTY: 3

Server Config: (Lenovo 3550-M5 (1U)
20 Cores (2.0Ghz), 256GB,
2 x 4TB SATA HDDs
1 x 2-Port 10G NIC (Intel 10G/Mellanox)

OpenStack Compute:
QTY: 2

Server Config: (Stratton 8001-12C) (1U)
16 Cores (2.3Ghz), 128GB ,
2 x 4TB SATA HDDs
1 x 2-Port 10G NIC (Intel 10G/Mellanox)

CEPH Config :
QTY: 3

Per Server Config: (Briggs 8001-22C) (2U)
20 Cores (2.93Ghz), 256GB
- (OS) 2+ 128GB DOM + (Journal) 2x SSD 240GB (**1.2** DWPD) + (Storage) 10 x 8TB SAS HDDs (~80TB)
- 1 x 2-Port 10G NIC (Intel/Mellanox)
- 1 x MegaRAID SAS controller

**Contact IBM for Redundant/Bonding Options

Network : (non HA) – no Bonding **
1 x Mellanox SX1410 (8831-S48)
1 x Lenovo G8052 (7120-48E)

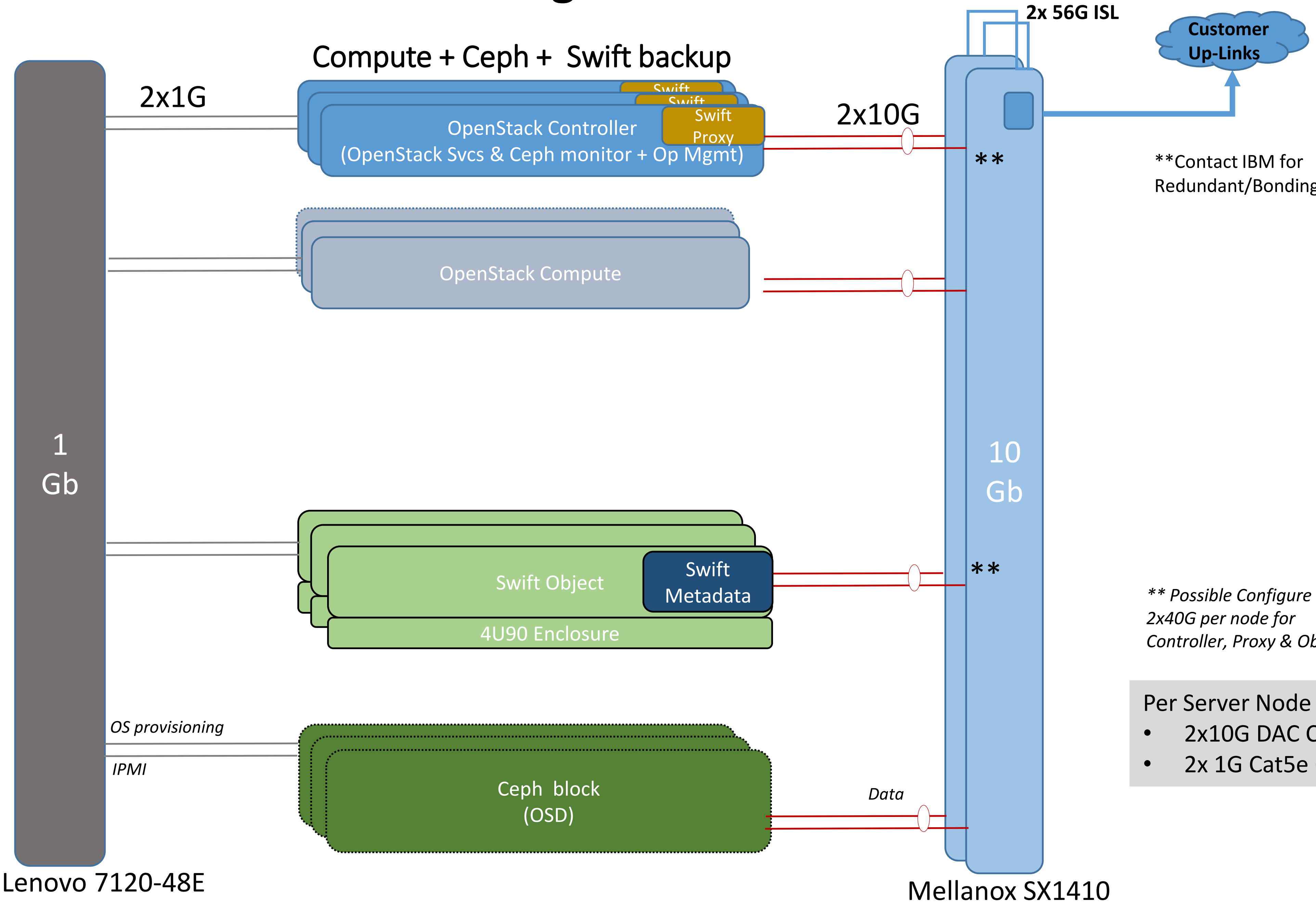
Rack:
QTY: 1
SlimRack 7965-94Y
PDUs x 4

Swift Object /MetaData
QTY: 3

Per Server Config: (Stratton 8001-21C) (1U)
16 Cores (2.3Ghz), 256GB
- (OS) 2+ 128GB DOM + 4 x SSDs x 240GB
- 1 x 2-Port 10G NIC (Intel/Mellanox)
- 1 x LSI 3008 External SAS
- 1 x MegaRAID SAS controller

Expansion Drawer (4U) : Supermicro SC946ED - 4U90
90 LFF – 2TB SAS HDDs

High Level Network Architecture Diagram



**Contact IBM for Redundant/Bonding Options

*** Possible Configure 2x40G per node for Controller, Proxy & Object*

Per Server Node
- 2x10G DAC Cables
- 2x 1G Cat5e Cables

Common Suggested Racking Rules,
Server Bill of Materials,
and Networking Diagrams

Suggested Racking Rule

Suggested Racking Rule

Reserved for accessibility

Reserved **U37-U41** for **Rack-Rack** Switches
if ordered

Any component placed above 32U should be accessed by the Compliance team for Tipping (if use MFG factory rack integration)

Reserved **U24-U26** for **In-rack** Network Switches

Start at U2and go UP

Common Rule:

--Place Pod / same MTM / Similar Funtion together

--Place Heavier MTM first starting U2

--Follow native MTM unique racking requirement

Racking Rules:

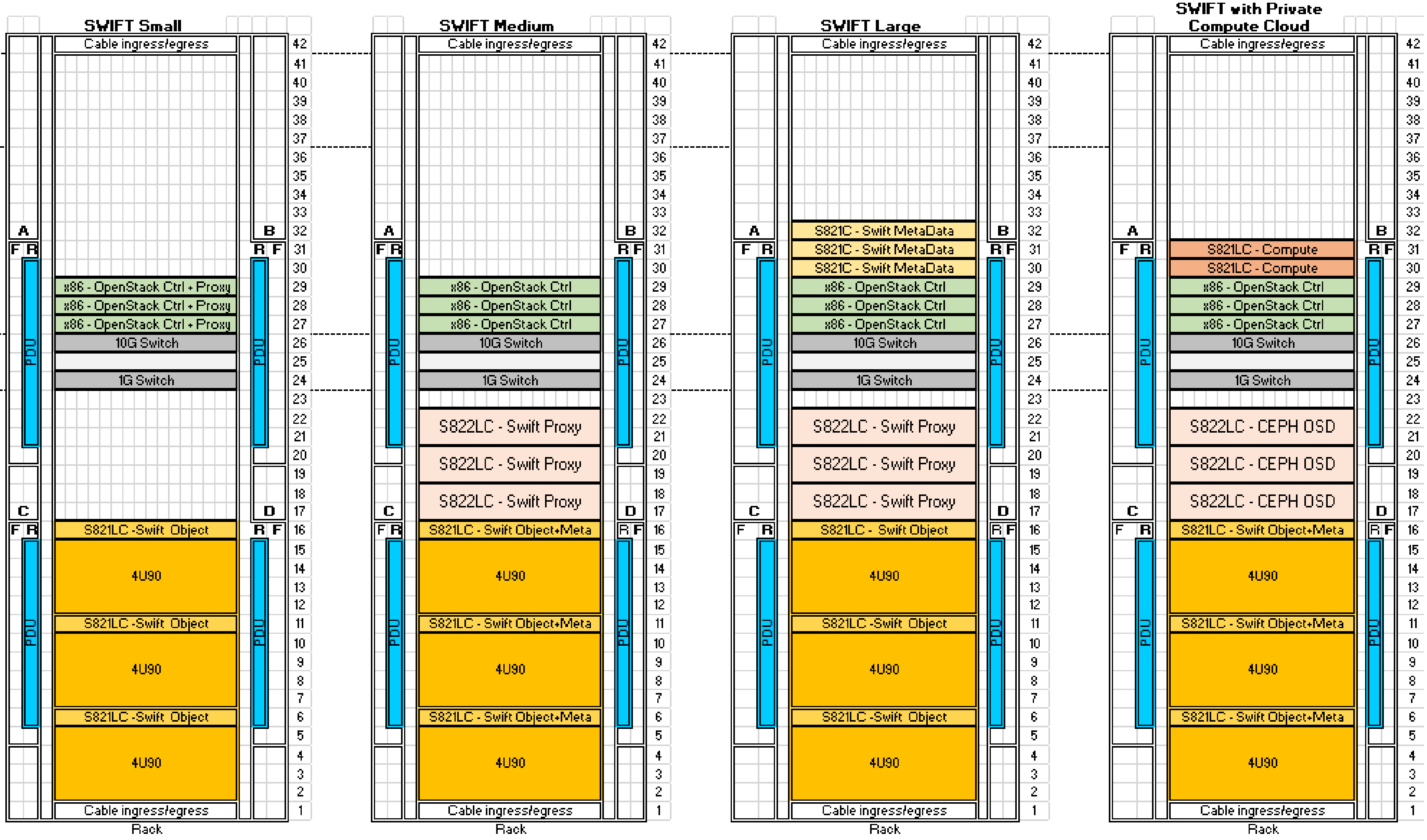
-- Recommend (0 -18) 2U Servers per Rack

-- Swift Object are placed in the lower part of the rack

-- Then Proxy 2U servers, follow by 1U MetaData and Controller servers

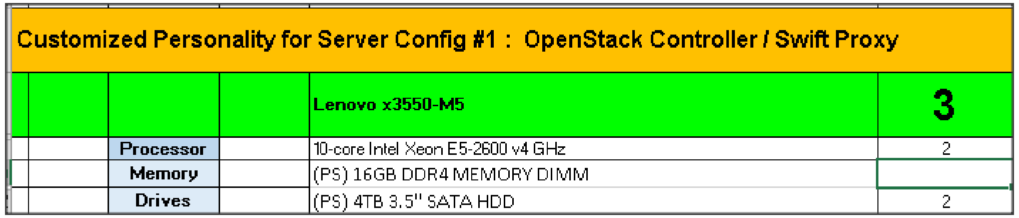
-- 2 racks per Set of network switch

Reserved for accessibility



U	SWIFT Small	SWIFT Medium	SWIFT Large	SWIFT with Private Compute Cloud
42	Cable ingress/egress	Cable ingress/egress	Cable ingress/egress	Cable ingress/egress
41				
40				
39				
38				
37				
36				
35				
34				
33				
32			S821C - Swift MetaData	
31			S821C - Swift MetaData	S821LC - Compute
30			S821C - Swift MetaData	S821LC - Compute
29	x86 - OpenStack Ctrl + Proxy	x86 - OpenStack Ctrl	x86 - OpenStack Ctrl	x86 - OpenStack Ctrl
28	x86 - OpenStack Ctrl + Proxy	x86 - OpenStack Ctrl	x86 - OpenStack Ctrl	x86 - OpenStack Ctrl
27	x86 - OpenStack Ctrl + Proxy	x86 - OpenStack Ctrl	x86 - OpenStack Ctrl	x86 - OpenStack Ctrl
26	10G Switch	10G Switch	10G Switch	10G Switch
25				
24	1G Switch	1G Switch	1G Switch	1G Switch
23				
22		S822LC - Swift Proxy	S822LC - Swift Proxy	S822LC - CEPH OSD
21				
20		S822LC - Swift Proxy	S822LC - Swift Proxy	S822LC - CEPH OSD
19				
18		S822LC - Swift Proxy	S822LC - Swift Proxy	S822LC - CEPH OSD
17				
16	S821LC -Swift Object	S821LC - Swift Object+Meta	S821LC - Swift Object	S821LC - Swift Object+Meta
15	4U90	4U90	4U90	4U90
14				
13				
12				
11	S821LC -Swift Object	S821LC - Swift Object+Meta	S821LC -Swift Object	S821LC - Swift Object+Meta
10	4U90	4U90	4U90	4U90
9				
8				
7				
6	S821LC -Swift Object	S821LC - Swift Object+Meta	S821LC -Swift Object	S821LC - Swift Object+Meta
5	4U90	4U90	4U90	4U90
4				
3				
2				
1	Cable ingress/egress	Cable ingress/egress	Cable ingress/egress	Cable ingress/egress

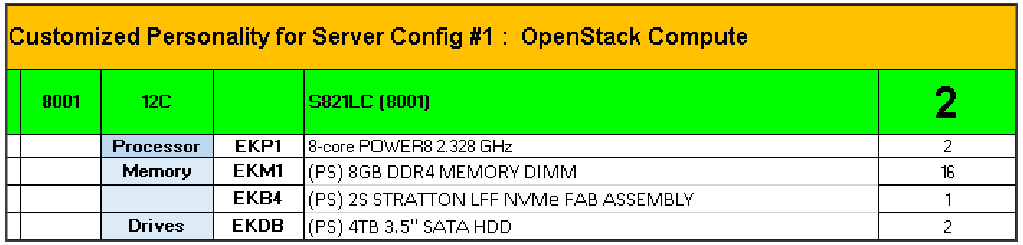
Server BOMs- Please Select the appropriate BOM for each Node Personality



Customized Personality for Server Config #1 : OpenStack Controller / Swift Proxy

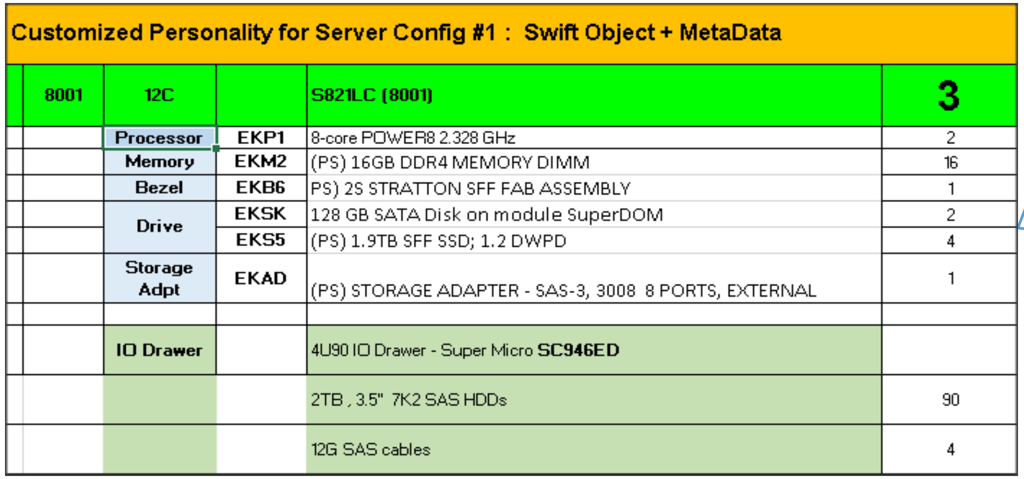
			Lenovo x3550-M5	3
	Processor		10-core Intel Xeon E5-2600 v4 GHz	2
	Memory		(PS) 16GB DDR4 MEMORY DIMM	
	Drives		(PS) 4TB 3.5" SATA HDD	2

This Server is offered by external supplier. Customer can configure similar server from other supplier as need



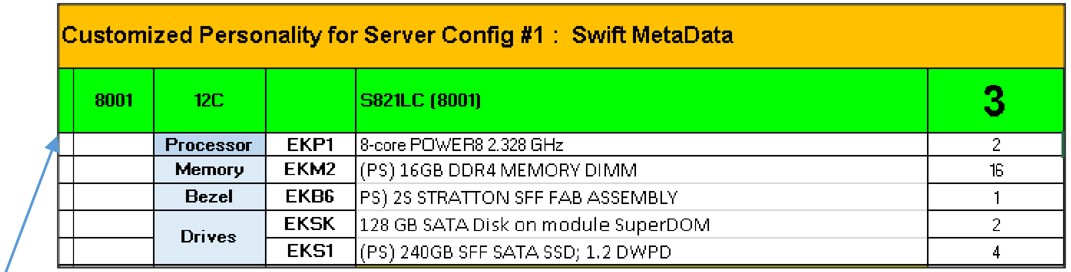
Customized Personality for Server Config #1 : OpenStack Compute

8001	12C		S821LC (8001)	2
	Processor	EKP1	8-core POWER8 2.328 GHz	2
	Memory	EKM1	(PS) 8GB DDR4 MEMORY DIMM	16
		EKB4	(PS) 2S STRATTON LFF NVMe FAB ASSEMBLY	1
	Drives	EKDB	(PS) 4TB 3.5" SATA HDD	2



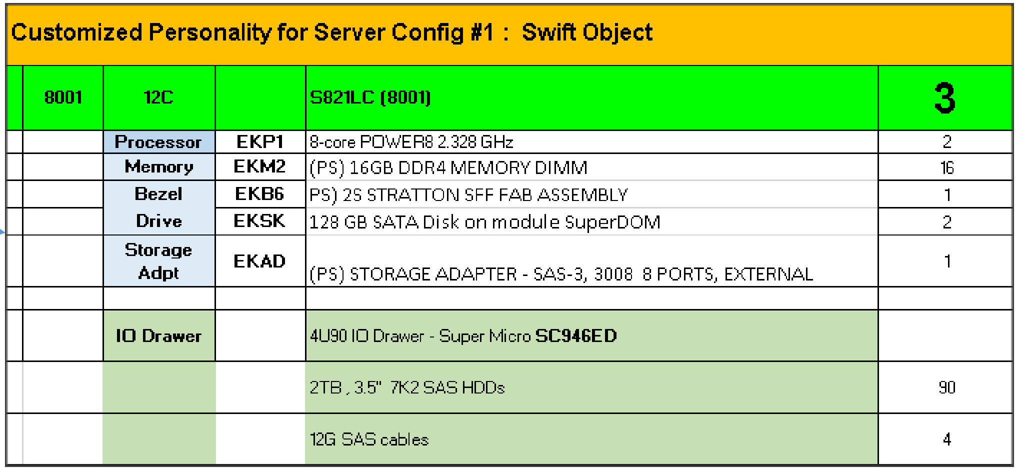
Customized Personality for Server Config #1 : Swift Object + MetaData

8001	12C		S821LC (8001)	3
	Processor	EKP1	8-core POWER8 2.328 GHz	2
	Memory	EKM2	(PS) 16GB DDR4 MEMORY DIMM	16
	Bezel	EKB6	PS) 2S STRATTON SFF FAB ASSEMBLY	1
	Drive	EKSK	128 GB SATA Disk on module SuperDOM	2
		EKS5	(PS) 1.9TB SFF SSD; 1.2 DWPD	4
	Storage Adpt	EKAD	(PS) STORAGE ADAPTER - SAS-3, 3008 8 PORTS, EXTERNAL	1
	IO Drawer		4U90 IO Drawer - Super Micro **SC946ED**	
			2TB , 3.5" 7K2 SAS HDDs	90
			12G SAS cables	4



Customized Personality for Server Config #1 : Swift MetaData

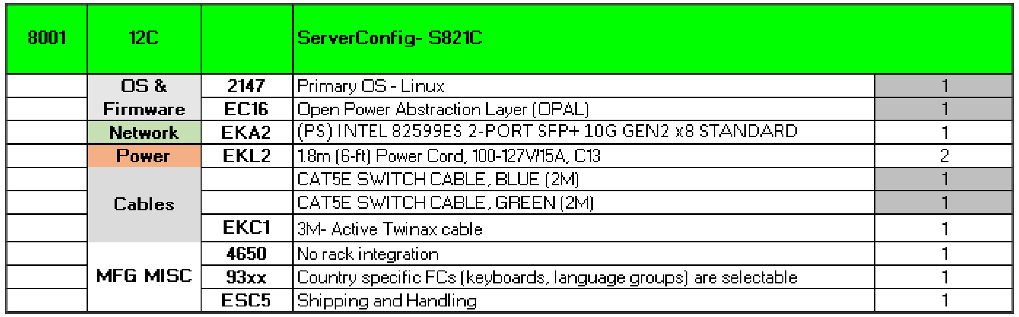
8001	12C		S821LC (8001)	3
	Processor	EKP1	8-core POWER8 2.328 GHz	2
	Memory	EKM2	(PS) 16GB DDR4 MEMORY DIMM	16
	Bezel	EKB6	PS) 2S STRATTON SFF FAB ASSEMBLY	1
	Drives	EKSK	128 GB SATA Disk on module SuperDOM	2
		EKS1	(PS) 240GB SFF SATA SSD; 1.2 DWPD	4



Customized Personality for Server Config #1 : Swift Object

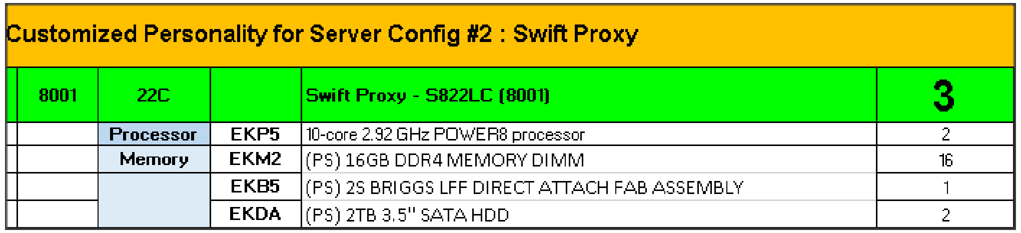
8001	12C		S821LC (8001)	3
	Processor	EKP1	8-core POWER8 2.328 GHz	2
	Memory	EKM2	(PS) 16GB DDR4 MEMORY DIMM	16
	Bezel	EKB6	PS) 2S STRATTON SFF FAB ASSEMBLY	1
	Drive	EKSK	128 GB SATA Disk on module SuperDOM	2
	Storage Adpt	EKAD	(PS) STORAGE ADAPTER - SAS-3, 3008 8 PORTS, EXTERNAL	1
	IO Drawer		4U90 IO Drawer - Super Micro **SC946ED**	
			2TB , 3.5" 7K2 SAS HDDs	90
			12G SAS cables	4

Based Server Config for 8001-12C: (For All Server Type above)

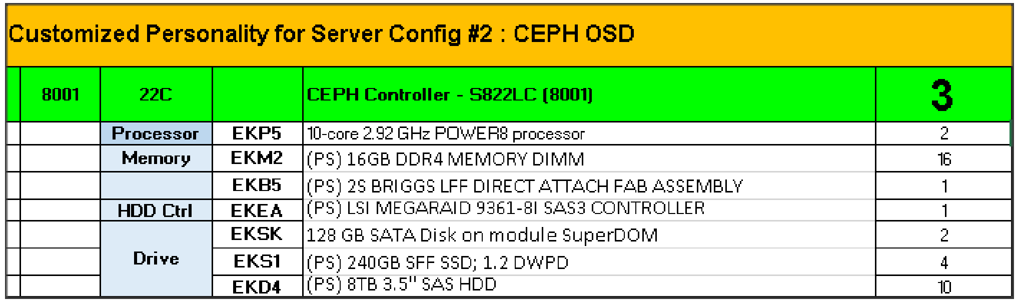


8001	12C		ServerConfig- S821C	
	OS & Firmware	2147	Primary OS - Linux	1
		EC16	Open Power Abstraction Layer (OPAL)	1
	Network	EKA2	(PS) INTEL 82599ES 2-PORT SFP+ 10G GEN2 x8 STANDARD	1
	Power	EKL2	1.8m (6-ft) Power Cord, 100-127V/15A, C13	2
	Cables		CAT5E SWITCH CABLE, BLUE (2M)	1
			CAT5E SWITCH CABLE, GREEN (2M)	1
		EKC1	3M- Active Twinax cable	1
	MFG MISC	4650	No rack integration	1
		93xx	Country specific FCs (keyboards, language groups) are selectable	1
		ESC5	Shipping and Handling	1

Server BOMs- Please Select the appropriate BOM for each Node Personality

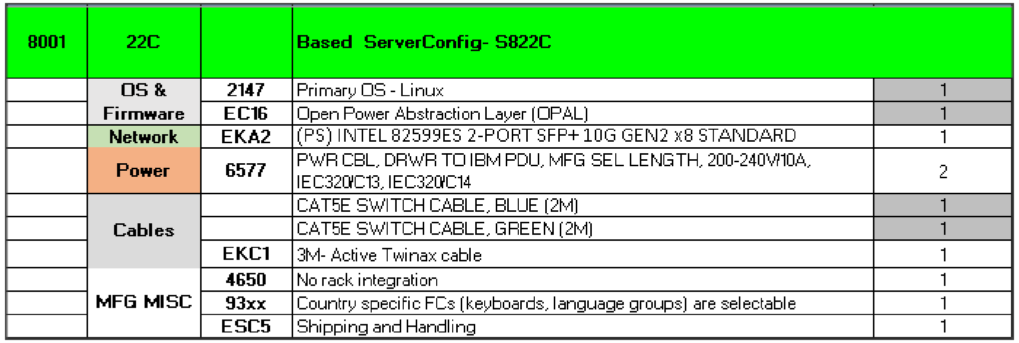


Customized Personality for Server Config #2 : Swift Proxy				
8001	22C		Swift Proxy - S822LC (8001)	3
	Processor	EKP5	10-core 2.92 GHz POWER8 processor	2
	Memory	EKM2	(PS) 16GB DDR4 MEMORY DIMM	16
		EKB5	(PS) 2S BRIGGS LFF DIRECT ATTACH FAB ASSEMBLY	1
		EKDA	(PS) 2TB 3.5" SATA HDD	2



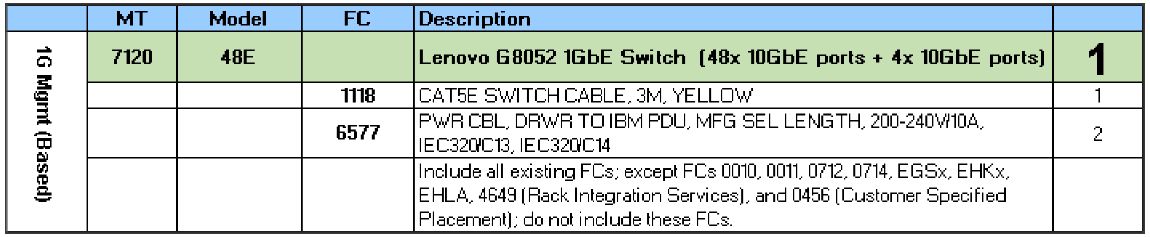
Customized Personality for Server Config #2 : CEPH OSD				
8001	22C		CEPH Controller - S822LC (8001)	3
	Processor	EKP5	10-core 2.92 GHz POWER8 processor	2
	Memory	EKM2	(PS) 16GB DDR4 MEMORY DIMM	16
		EKB5	(PS) 2S BRIGGS LFF DIRECT ATTACH FAB ASSEMBLY	1
	HDD Ctrl	EKEA	(PS) LSI MEGARAID 9361-8I SAS3 CONTROLLER	1
	Drive	EKSK	128 GB SATA Disk on module SuperDOM	2
		EKS1	(PS) 240GB SFF SSD; 1.2 DWPD	4
		EKD4	(PS) 8TB 3.5" SAS HDD	10

Based Server Config for 8001-22C: (For All Server Type above)

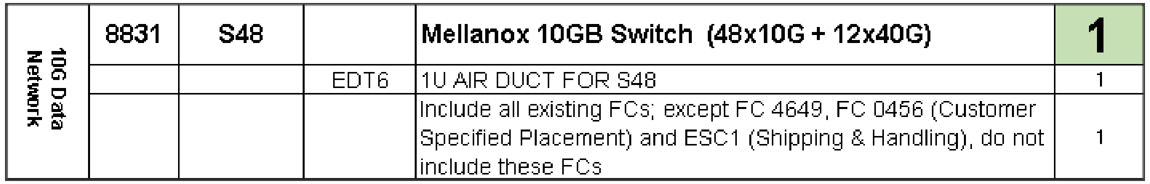


8001	22C		Based ServerConfig- S822C	
	OS & Firmware	2147	Primary OS - Linux	1
		EC16	Open Power Abstraction Layer (OPAL)	1
	Network	EKA2	(PS) INTEL 82599ES 2-PORT SFP+ 10G GEN2 x8 STANDARD	1
	Power	6577	PWR CBL, DRWR TO IBM PDU, MFG SEL LENGTH, 200-240V/10A, IEC320/C13, IEC320/C14	2
	Cables		CAT5E SWITCH CABLE, BLUE (2M)	1
			CAT5E SWITCH CABLE, GREEN (2M)	1
		EKC1	3M- Active Twinax cable	1
	MFG MISC	4650	No rack integration	1
		93xx	Country specific FCs (keyboards, language groups) are selectable	1
		ESC5	Shipping and Handling	1

Network Switch BOMs

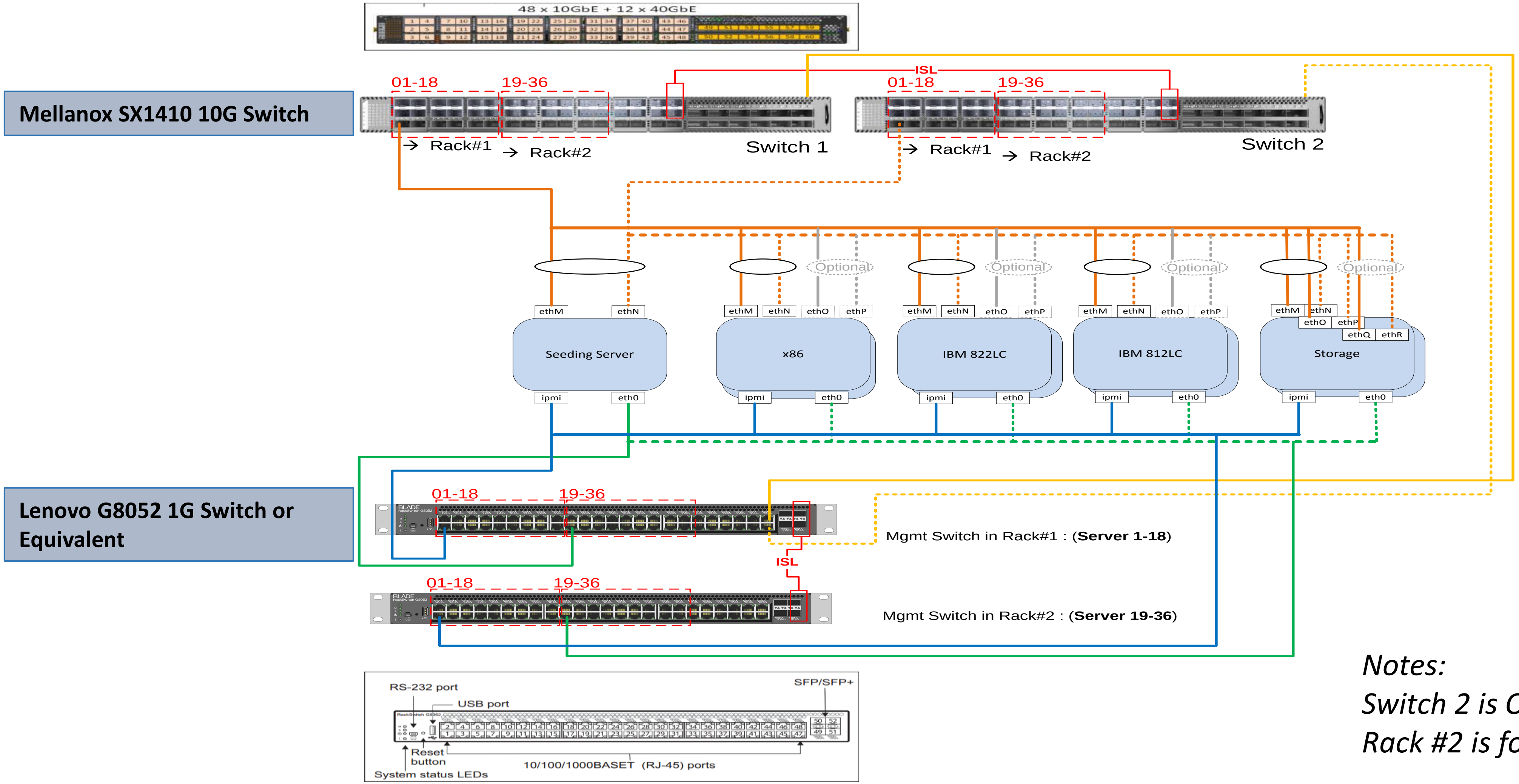


	MT	Model	FC	Description	
1G Mgmt (Based)	7120	48E		**Lenovo G8052 1GbE Switch (48x 10GbE ports + 4x 10GbE ports)**	**1**
			1118	CAT5E SWITCH CABLE, 3M, YELLOW	1
			6577	PWR CBL, DRWR TO IBM PDU, MFG SEL LENGTH, 200-240V/10A, IEC320/C13, IEC320/C14	2
				Include all existing FCs; except FCs 0010, 0011, 0712, 0714, EGSx, EHKx, EHLA, 4649 (Rack Integration Services), and 0456 (Customer Specified Placement); do not include these FCs.	



10G Data Network	8831	S48		**Mellanox 10GB Switch (48x10G + 12x40G)**	**1**
			EDT6	1U AIR DUCT FOR S48	1
				Include all existing FCs; except FC 4649, FC 0456 (Customer Specified Placement) and ESC1 (Shipping & Handling), do not include these FCs	1

Network Plug Rule - Sample



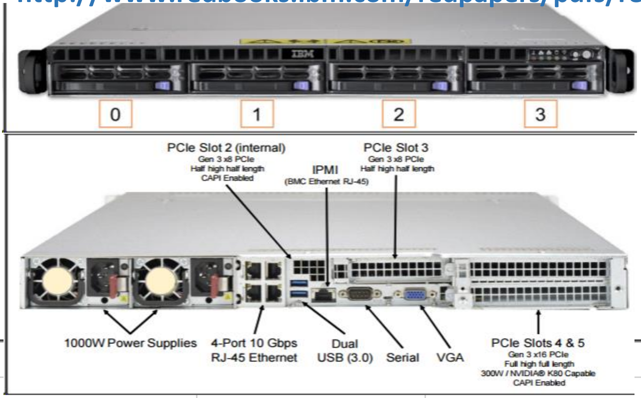
Notes:

Switch 2 is Optional.

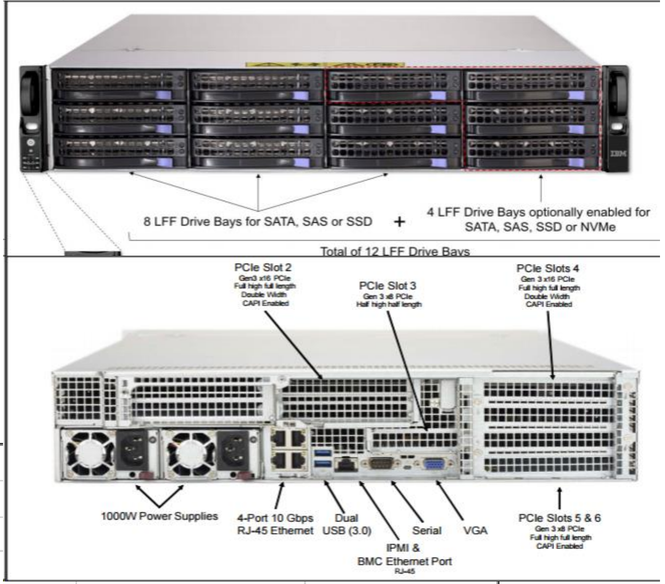
Rack #2 is for future expansion

Network Plug P2P Label -- Sample

MTM: 8001-12C
http://www.redbooks.ibm.com/redpapers/pdfs/redp5406.pdf



MTM: 8001-22C
http://www.redbooks.ibm.com/redpapers/pdfs/redp5407.pdf



Server PCI Slot Placement

8001-12C/22C Statton/Briggs

	adapter	PCI slot	Port	Cabling
Primary NIC	10GbE	slot 3	T1	yes
			T2	
Optional NIC	10GbE		T1	
			T2	
Mgmt-OS	1GbE	LOM	T1	yes
BMC	1GbE	LOM	impi	yes

Cable P2P Label for H_TOR : capable of 36 Downlink-36 Uplink (ie Mellanox SX1410) ~1:1 Network Subscriptions

		10GbE	10GbE	1GbE	1GbE
		H_TOR_1	H_TOR_2	M_TOR_1	M_TOR_1
Server #	Name <opt>	P2P Data network Cable Label	P2P Data network Cable Label	P2P Mgmt RJ4-5 Cable Label	P2P IPMI RJ-45 Cable Label
1		1A/SVR1/slot 3/T1 <> H_TOR_1/Port1		1A/SVR1/LOM/T1 <> M_TOR_1/Port1	1A/SVR1/LOM/impi <> M_TOR_1/Port19
2		1A/SVR2/slot 3/T1 <> H_TOR_1/Port2		1A/SVR2/LOM/T1 <> M_TOR_1/Port2	1A/SVR2/LOM/impi <> M_TOR_1/Port20
3		1A/SVR3/slot 3/T1 <> H_TOR_1/Port3		1A/SVR3/LOM/T1 <> M_TOR_1/Port3	1A/SVR3/LOM/impi <> M_TOR_1/Port21
4		1A/SVR4/slot 3/T1 <> H_TOR_1/Port4		1A/SVR4/LOM/T1 <> M_TOR_1/Port4	1A/SVR4/LOM/impi <> M_TOR_1/Port22
5		1A/SVR5/slot 3/T1 <> H_TOR_1/Port5		1A/SVR5/LOM/T1 <> M_TOR_1/Port5	1A/SVR5/LOM/impi <> M_TOR_1/Port23
6		1A/SVR6/slot 3/T1 <> H_TOR_1/Port6		1A/SVR6/LOM/T1 <> M_TOR_1/Port6	1A/SVR6/LOM/impi <> M_TOR_1/Port24
7		1A/SVR7/slot 3/T1 <> H_TOR_1/Port7		1A/SVR7/LOM/T1 <> M_TOR_1/Port7	1A/SVR7/LOM/impi <> M_TOR_1/Port25
8		1A/SVR8/slot 3/T1 <> H_TOR_1/Port8		1A/SVR8/LOM/T1 <> M_TOR_1/Port8	1A/SVR8/LOM/impi <> M_TOR_1/Port26
9		1A/SVR9/slot 3/T1 <> H_TOR_1/Port9		1A/SVR9/LOM/T1 <> M_TOR_1/Port9	1A/SVR9/LOM/impi <> M_TOR_1/Port27
10		1A/SVR10/slot 3/T1 <> H_TOR_1/Port10		1A/SVR10/LOM/T1 <> M_TOR_1/Port10	1A/SVR10/LOM/impi <> M_TOR_1/Port28
11		1A/SVR11/slot 3/T1 <> H_TOR_1/Port11		1A/SVR11/LOM/T1 <> M_TOR_1/Port11	1A/SVR11/LOM/impi <> M_TOR_1/Port29