OpenStack Swift Reference Designs

This document contains three OpenStack Swift reference designs:

- Small
- Medium
- Large

The three 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. The three designs vary based on the location of the proxy service and the account/container services and their associated SSDs.

	Small	Medium	Large
Proxy service location	Container on control plane	Dedicated server	Dedicated server
Account/Container SSD location	Object server	Object server	Dedicated server

The small and medium sizes are the recommended sizes because the account and container SSDs are in the server which is connected to the JBOD object storage drawer. Separating the SSDs into dedicated servers is not economically beneficial when using high capacity JBOD drawers for the object storage.

The choice between small and medium is normally driven by workload performance needs. The proxy service becomes CPU intensive under load. In the small configuration the proxy service competes with the rest of the control plane for CPU cycles.

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

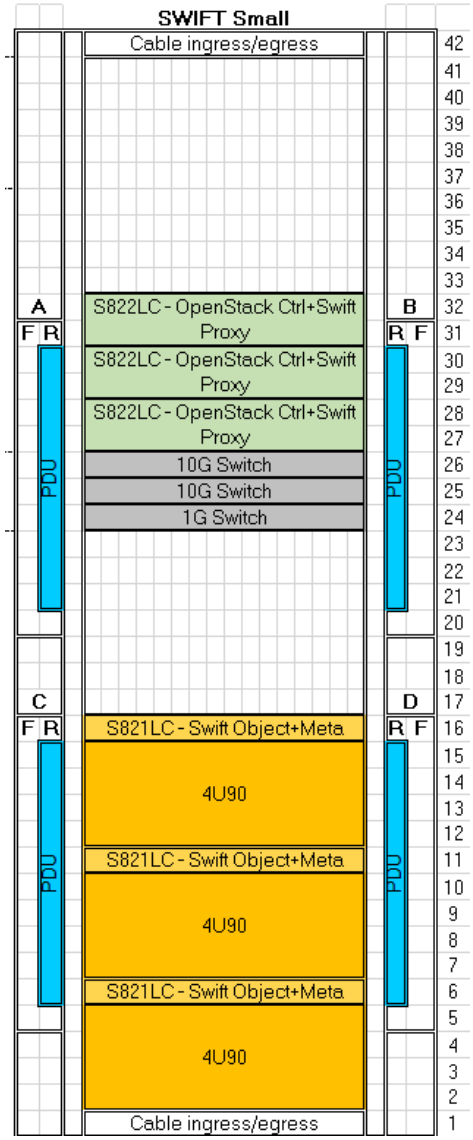
Cost savings:

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.

High Level Component Architecture Diagram – Swift Small



Ubuntu 16.04 (all nodes)
Software:
OpenStack Newton
- Horizon, Keystone, Swift
- RabbitMQ, Galera, etc.
Operational Management
- Nagios Core
- ELK Stack (Elasticsearch, Logstash, Kibana)

Network : (HA) – Bonding
2 x Mellanox SX1410 (8831-S48)
1 x Lenovo G8052 (7120-48E)

Rack:
QTY: 1
SlimRack 7965-94Y (Standard 19" rack)
PDUs x 4: Each node should have 2 power cords cabled to two different PDUs

OpenStack Controller, Swift Proxy & OpsMgr
QTY: 3

Server Config: (Briggs S822LC MTM 8001-22C) (2U)
20 Cores (2.92 Ghz), 128 GB,
1 x 4TB SATA HDD
2 x 2-Port 10G NIC (Intel 10G / Mellanox)

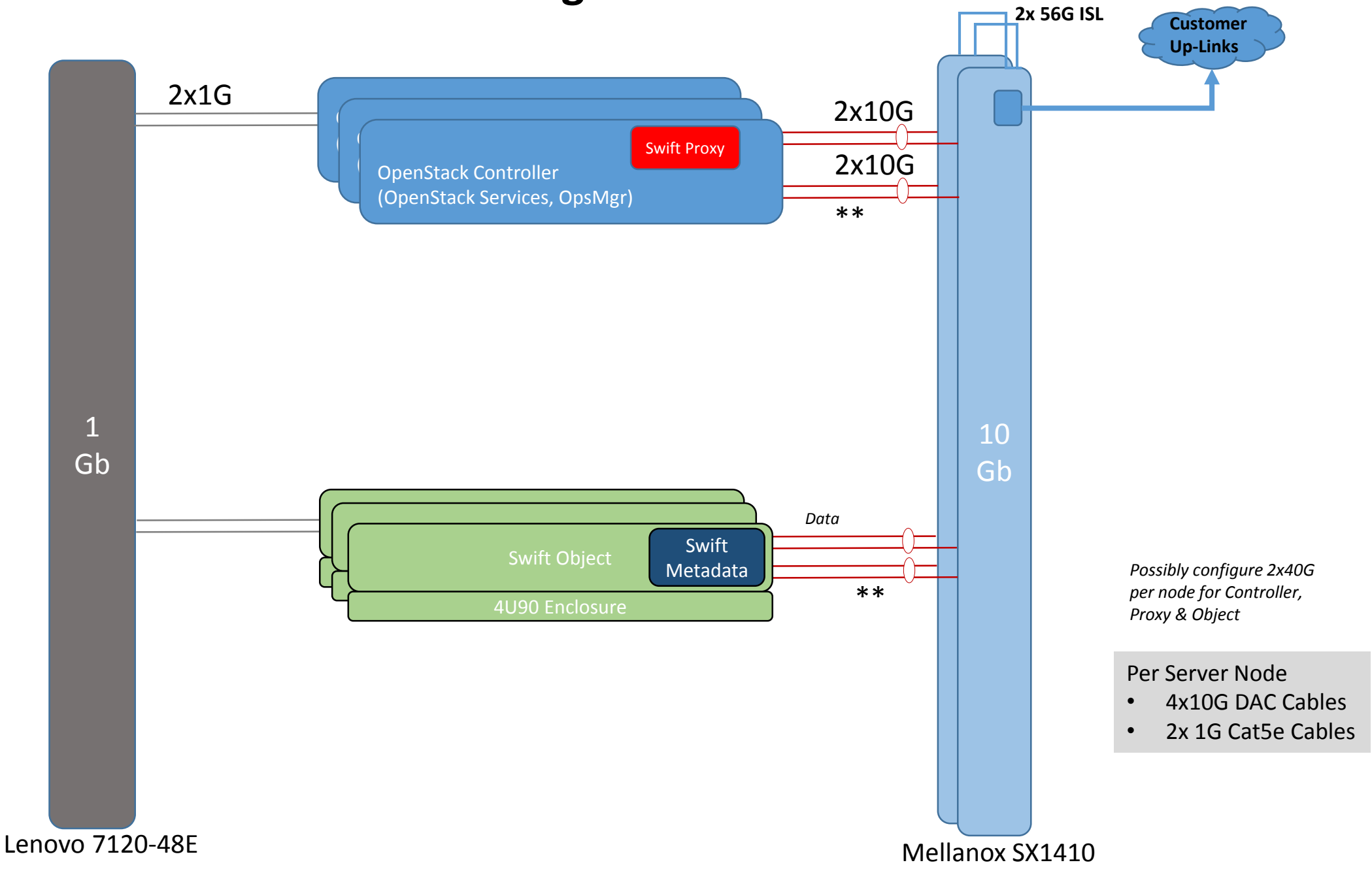
Swift Object / Metadata
QTY: 3

Server Config: (Stratton S821LC MTM 8001-12C) (1U)
16 Cores (2.3Ghz), 128GB
- (OS) 1 x 4TB SATA HDD + 4 x 240 GB SSDs
- 2 x 2-Port 10G NIC (Intel / Mellanox)
- 1 x External SAS (8 port SAS3) LSI 3008 based

Expansion Drawer (4U) :
90 LFF JBOD Storage SMC PN SE-946ED-R2KJBOD
90 LFF – 2TB SAS HDDs

****Notes:**
a) Proc + Memory config may need to be altered based on actual performance requirements

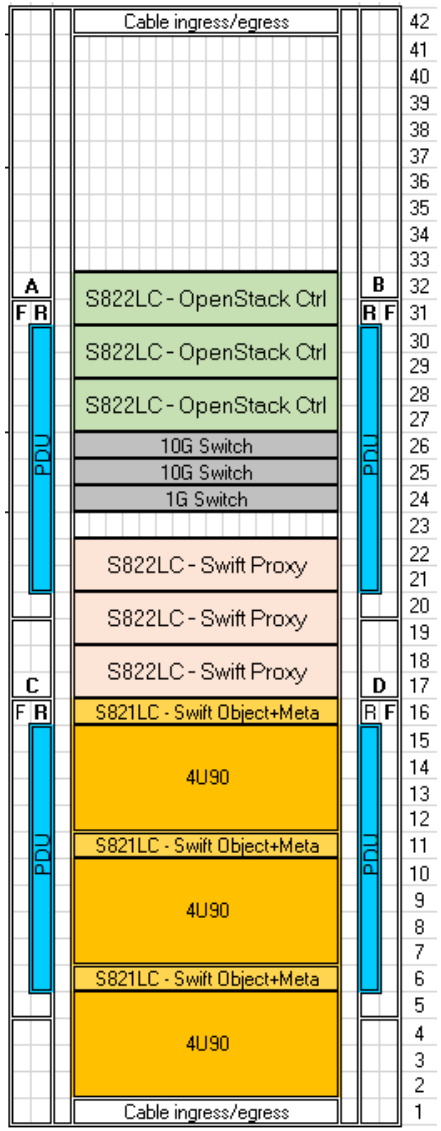
High Level Network Architecture Diagram – Swift Small



Possibly configure 2x40G per node for Controller, Proxy & Object

Per Server Node
- 4x10G DAC Cables
- 2x 1G Cat5e Cables

High Level Component Architecture Diagram – Swift Medium



Ubuntu 16.04 (all nodes)
Software:
OpenStack Newton
- Horizon, Keystone, Swift
- RabbitMQ, Galera, etc.
Operational Management
- Nagios Core
- ELK Stack (Elasticsearch, Logstash, Kibana)

Network : (HA) – Bonding
2 x Mellanox SX1410 (8831-S48)
1 x Lenovo G8052 (7120-48E)

Rack:
QTY: 1
SlimRack 7965-94Y (Standard 19" rack)
PDUs x 4: Each node should have 2 power cords cabled to two different PDUs

OpenStack Controller & OpsMgr
QTY: 3

Server Config: (Briggs S822LC MTM 8001-22C)
(2U)
20 Cores (2.92 Ghz), 128 GB,
1 x 4TB SATA HDD
2 x 2-Port 10G NIC (Intel 10G / Mellanox)

Swift Proxy
QTY: 3

Server Config: (Briggs S822LC MTM 8001-22C)
(2U)
20 Cores (2.92 Ghz), 128 GB,
1 x 4TB SATA HDD
2 x 2-Port 10G NIC (Intel 10G / Mellanox)

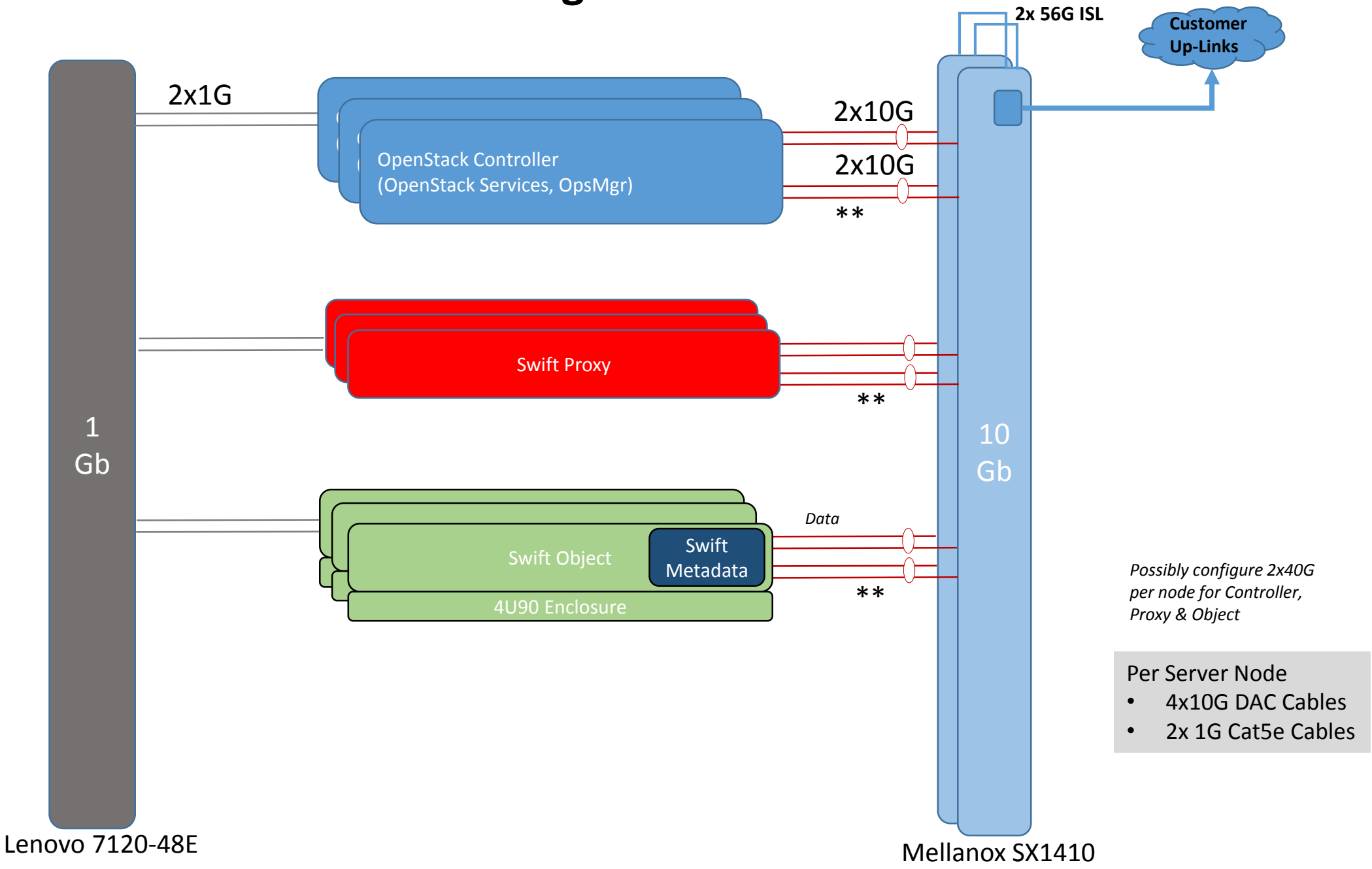
Swift Object / Metadata
QTY: 3

Server Config: (Stratton S821LC MTM 8001-12C)
(1U)
16 Cores (2.3Ghz), 128GB
- (OS) 1 x 4TB SATA HDD + 4 x 240 GB SSDs
- 2 x 2-Port 10G NIC (Intel / Mellanox)
- 1 x External SAS (8 port SAS3) LSI 3008 based

Expansion Drawer (4U) :
90 LFF JBOD Storage SMC PN SE-946ED-R2KJBOD
90 LFF – 2TB SAS HDDs

****Notes:**
a) Proc + Memory config may need to be altered based on actual performance requirements

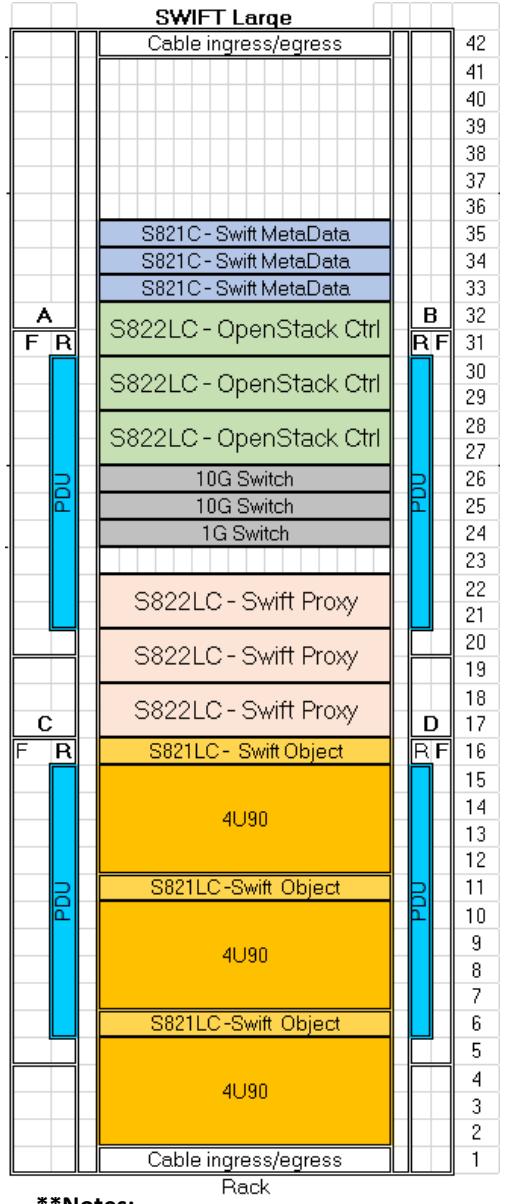
High Level Network Architecture Diagram – Swift Medium



Possibly configure 2x40G per node for Controller, Proxy & Object

Per Server Node
- 4x10G DAC Cables
- 2x 1G Cat5e Cables

High Level Component Architecture Diagram – Swift Large



Ubuntu 16.04 (all nodes)
Software:
OpenStack Newton
- Horizon, Keystone, Swift
- RabbitMQ, Galera, etc.
Operational Management
- Nagios Core
- ELK Stack (Elasticsearch, Logstash, Kibana)

Network : (HA) – Bonding
2 x Mellanox SX1410 (8831-S48)
1 x Lenovo G8052 (7120-48E)

Rack:
QTY: 1
SlimRack 7965-94Y (Standard 19" rack)
PDUs x 4: Each node should have 2 power cords cabled to two different PDUs

OpenStack Controller & OpsMgr
QTY: 3

Server Config: (Briggs S822LC MTM 8001-22C)
(2U)
20 Cores (2.92 Ghz), 128 GB,
1 x 4TB SATA HDD
2 x 2-Port 10G NIC (Intel 10G / Mellanox)

Swift Metadata
QTY: 3

Server Config: (Stratton 8001-12C)
(1U)
16 Cores (2.3Ghz), 128GB
- (OS) 1 x 4TB SATA HDD + 4 x 240 GB SSDs
- 2 x 2-Port 10G NIC (Intel / Mellanox)

Swift Proxy
QTY: 3

Server Config: (Briggs S822LC MTM 8001-22C)
(2U)
20 Cores (2.92 Ghz), 128 GB,
1 x 4TB SATA HDD
2 x 2-Port 10G NIC (Intel 10G / Mellanox)

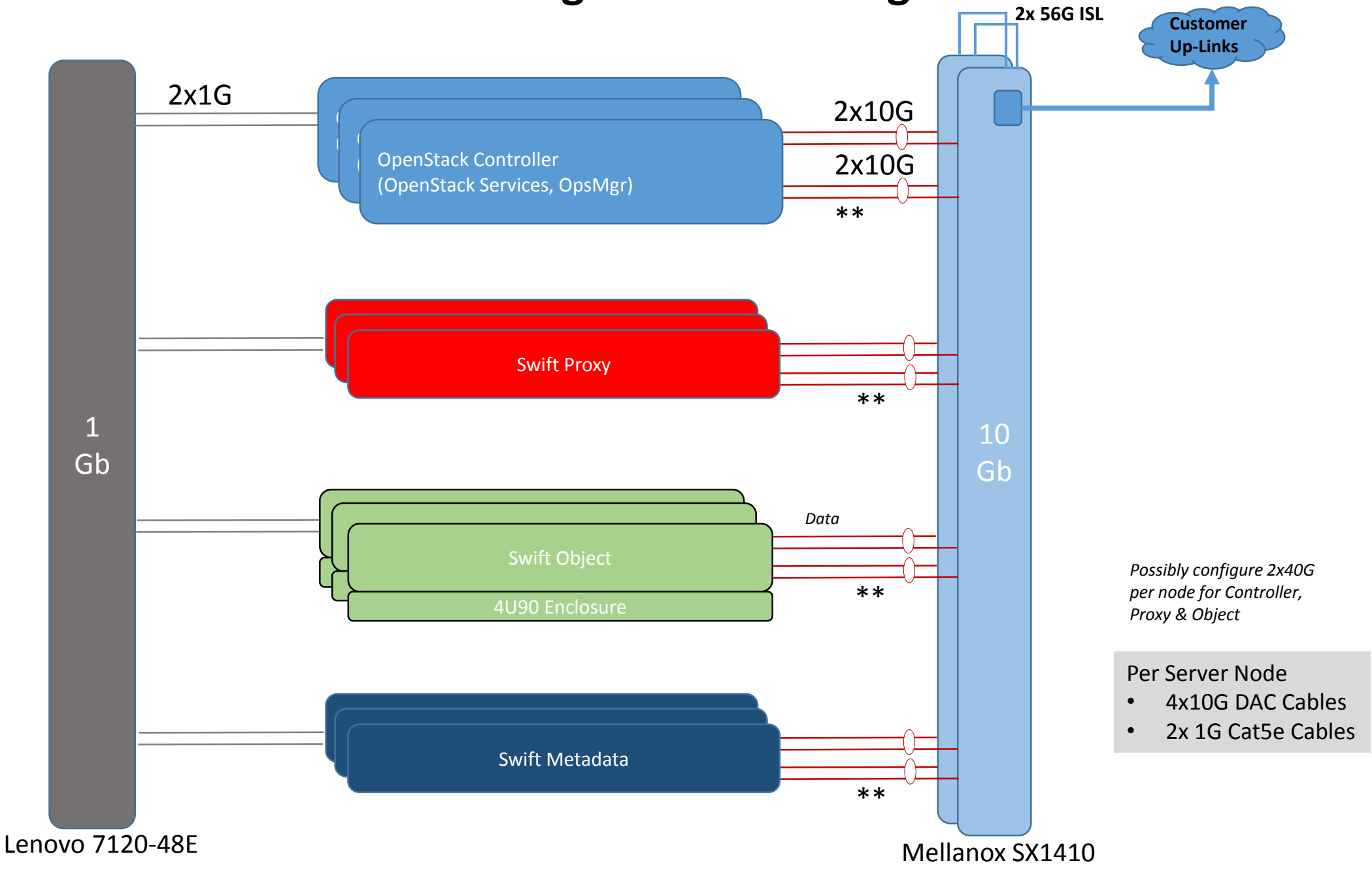
Swift Object
QTY: 3

Server Config: (Stratton S821LC MTM 8001-12C)
(1U)
16 Cores (2.3Ghz), 128GB
- (OS) 1 x 4TB SATA HDD + 4 x 240 GB SSDs
- 2 x 2-Port 10G NIC (Intel / Mellanox)
- 1 x External SAS (8 port SAS3) LSI 3008 based

Expansion Drawer (4U) :
90 LFF JBOD Storage SMC PN SE-946ED-R2KJBOD
90 LFF – 2TB SAS HDDs

****Notes:**
a) Proc + Memory config may need to be altered based on actual performance requirements

High Level Network Architecture Diagram – Swift Large



Possibly configure 2x40G per node for Controller, Proxy & Object

Per Server Node
- 4x10G DAC Cables
- 2x 1G Cat5e Cables