

OPEN

Compute Project

# System Test Plan for Wedge-100


Authors:

Hany Morsy, Network Engineer, Facebook




# 1. Revision History

| Date | Name | Description |
|---|---|---|
| 07/26/2016 | Hany Morsy | initial version |
| | | |
| | | |
| | | |
| | | |
| | | |
| | | |

© 2016 Facebook.

As of July 26, 2016, the following persons or entities have made this Specification available under the Open Compute Project Hardware License (Permissive) Version 1.0 (OCPHL-P), which is available at http://www.opencompute.org/community/get-involved/spec-submission-process/.

Facebook, Inc.

Your use of this Specification may be subject to other third party rights. THIS SPECIFICATION IS PROVIDED "AS IS." The contributors expressly disclaim any warranties (express, implied, or otherwise), including implied warranties of merchantability, non-infringement, fitness for a particular purpose, or title, related to the Specification. The Specification implementer and user assume the entire risk as to implementing or otherwise using the Specification. IN NO EVENT WILL ANY PARTY BE LIABLE TO ANY OTHER PARTY FOR LOST PROFITS OR ANY FORM OF INDIRECT, SPECIAL, INCIDENTAL, OR CONSEQUENTIAL DAMAGES OF ANY CHARACTER FROM ANY CAUSES OF ACTION OF ANY KIND WITH RESPECT TO THIS SPECIFICATION OR ITS GOVERNING AGREEMENT, WHETHER BASED ON BREACH OF CONTRACT, TORT (INCLUDING NEGLIGENCE), OR OTHERWISE, AND WHETHER OR NOT THE OTHER PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

## 2. Scope

This document provides the high level network test plan for facebook's 32x100G TOR Switch (a.k.a wedge-100).

## 3. Contents


1. Revision History ..... 2
2. Scope ..... 4
3. Contents ..... 4
4. Overview ..... 5
5. **Test Topology** ..... 6
6. Test Plan ..... 7
   - 6.1. Functionality Testing ..... 7
   - 6.2. Scale & Stress Testing ..... 7
   - 6.3. Integration with FB Tools ..... 7
   - 6.4. Failure Conditions and OIR ..... 7
   - 6.5. Performance Testing ..... 8
   - 6.6. Longevity Test ..... 8
7. Ixia Configuration ..... 9

# 4. Overview

Wedge-100 platform is facebook's 32x100G Top of Rack (TOR) switch based on Broadcom's Tomahawk Switch ASIC. It is intended to be used as the Rack Switch (RSW) in Facebook's data-center fabric networks.

The RSW attaches southbound to servers in the same physical rack. DAC cables with variety of breakout options are used to provide connectivity to various NIC speeds in the rack. On the northbound side, the RSW attaches to 4 Fabric Switches (FSWs) using 4 uplinks.

The purpose of this document is to provide a high level description of the network test plan of the wedge-100 platform. This test plan doesn't cover the server-side testing of the RSW. Facebook datacenter network engineering team uses this plan to qualify the functionality and performance of the switch before deploying it into service.

# 5. Test Topology

The following diagram shows the topology of facebook's network testing setup for wedge-100 RSW (Rack Switch) and sixpack-40 FSW (Fabric Switch):

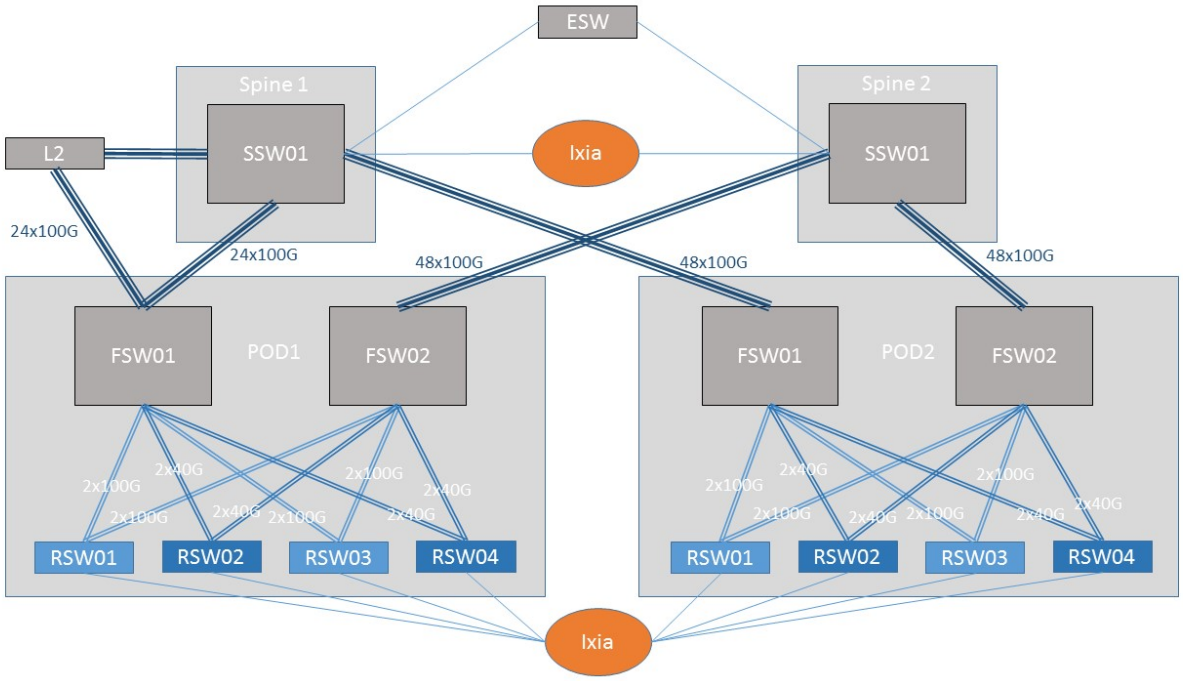



Figure 1: Wedge-100 Development test Topology

**Notes on Topology:**

- Wedge-100 is the Device-under-Test (DUT) in Rack Switch (RSW) Role.
- Test Topology consists of 2 Server Pods, 2 Spine Planes and 1 Edge Pod.
- Server Pods are 2-Post. Typical production deployments are 4-Post with one uplink to each FSW.
- To simulate 4 uplinks from each RSW, 2 parallel uplinks from a given RSW to each FSW have been used.
- A Spine Plane has only one Spine Switch (SSW). Typical production deployments can scale up to 48 SSWs per Plane.
- To simulate a full-capacity Fabric, 48 parallel uplinks have been run to the SSW.
- Wedge-40 and 3rd party vendors have been tested to establish a baseline and provide comparison with Wedge100.
- Traffic Generation and BGP Scale Route simulation from Ixia.

# 6. Test Plan

## 6.1. Functionality Testing

1. BGP Peerings with FSW
2. ECMP among 40G and 100G uplinks to FSW
3. BGP Policy Operation
4. Redundancy failover events and convergence speed
5. Graceful Restart by bgpd kill on RSW
6. Graceful Restart by clearing bgp neighbor on FSW
7. Recovery from wedge_agent bounce
8. Fbossdeploy bgp restart
9. Optical Transceiver EEPROM and DOM

## 6.2. Scale & Stress Testing

1. High Frequency interface shut/no-shut on the uplinks
2. High Frequency BGP neighbor shut/no-shut on the uplinks
3. Generation of 8K IPv4 + 8K IPv6 BGP Routes from Ixia
4. Running all traffic flows at once (Intra-Pod, Inter-Pod and External)
5. Flapping 8K IPv4 + 8K IPv6 BGP Routes every 60 seconds
6. Shifting routes from plane to plane
7. Checking for Memory Leaks and CPU Utilization

## 6.3. Integration with FB Tools

1. Config Generation
2. Auto-Provisioning
3. FBOSS Deployments
4. Monitoring via Dashboards
5. Auditing and alerting
6. Manageability
7. Drain/Undrain

## 6.4. Failure Conditions and OIR

1. Interface recovery with shut/no-shut in SW
2. Interface recovery with fiber pulls
3. Interface recovery with optic + fiber pulls
4. BGP session recovery with BGP neighbor shut/no-Shut




5. BGP session recovery with BGP Process Kill and Re-spawn

## 6.5. Performance Testing

1. Impact of one interface shut on FSW side towards RSW
2. Impact of one interface shut on RSW side towards FSW
3. Impact of interface unshut for the above
4. Impact of one BGP neighbor Shut on RSW side towards FSW
5. Impact of BGP neighbor unshut on the FSW side wards RSW
6. Impact of one BGP neighbor shut followed by interface shut on RSW side towards FSW
7. Impact of one interface unshut followed by BGP neighbor unshut for the above
8. Impact of drain/undrain
9. Impact of bgpd & agent updates

## 6.6. Longevity Test

1. A continuous Run of all Traffic flows with full scale of BGP Routes to confirm stability

# 7. Ixia Configuration

- BGP Route Scale Simulation

  Total Routes 8K V4 + 8K V6 routes in a prefix mix close to our production environments.

- Traffic Flows
  - 50 simulated servers behind each RSW
  - Traffic Flows are a mesh of 50 Src IPs, 50 Dst IPs with random TCP/UDP ports
  - Both Fixed 400B and Random (74B – 1500B)
  - Bi-directional traffic
  - Intra-Pod traffic (Ixia → RSW → 6-Pack → RSW → Ixia)
  - Inter-Pod traffic (Ixia → RSW → 6-Pack → SSW → 6-Pack → RSW → Ixia)
  - External traffic (Ixia → RSW → 6-Pack → SSW → Ixia)