



Open. Together.



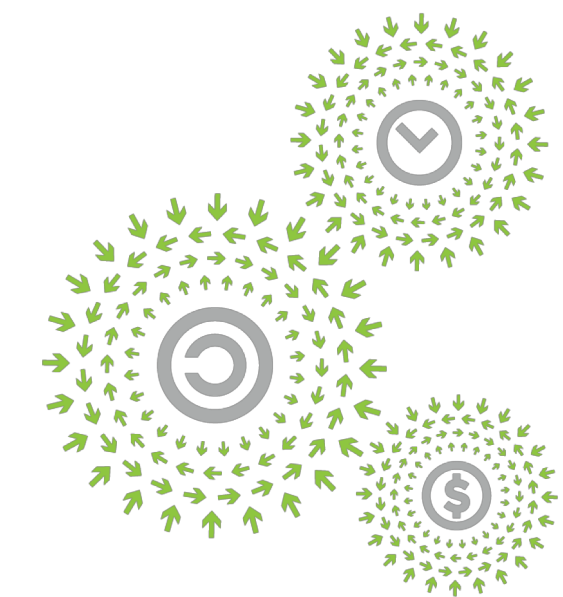
OCP
SUMMIT



NETWORKING

Scaled Troubleshoot Process for Optical Interconnects at Facebook Data Centers Using Network Switches: BER Illusion Methodology (BIM)

Abhijit Chakravarty, Manufacturing & Quality Engineer - Network, Facebook
Eddie Galley, Manufacturing Quality Manager, Facebook
Ashwin Poojary, Systems Engineer - Network, Facebook



OPEN
PLATINUM™



Open. Together.

Overview

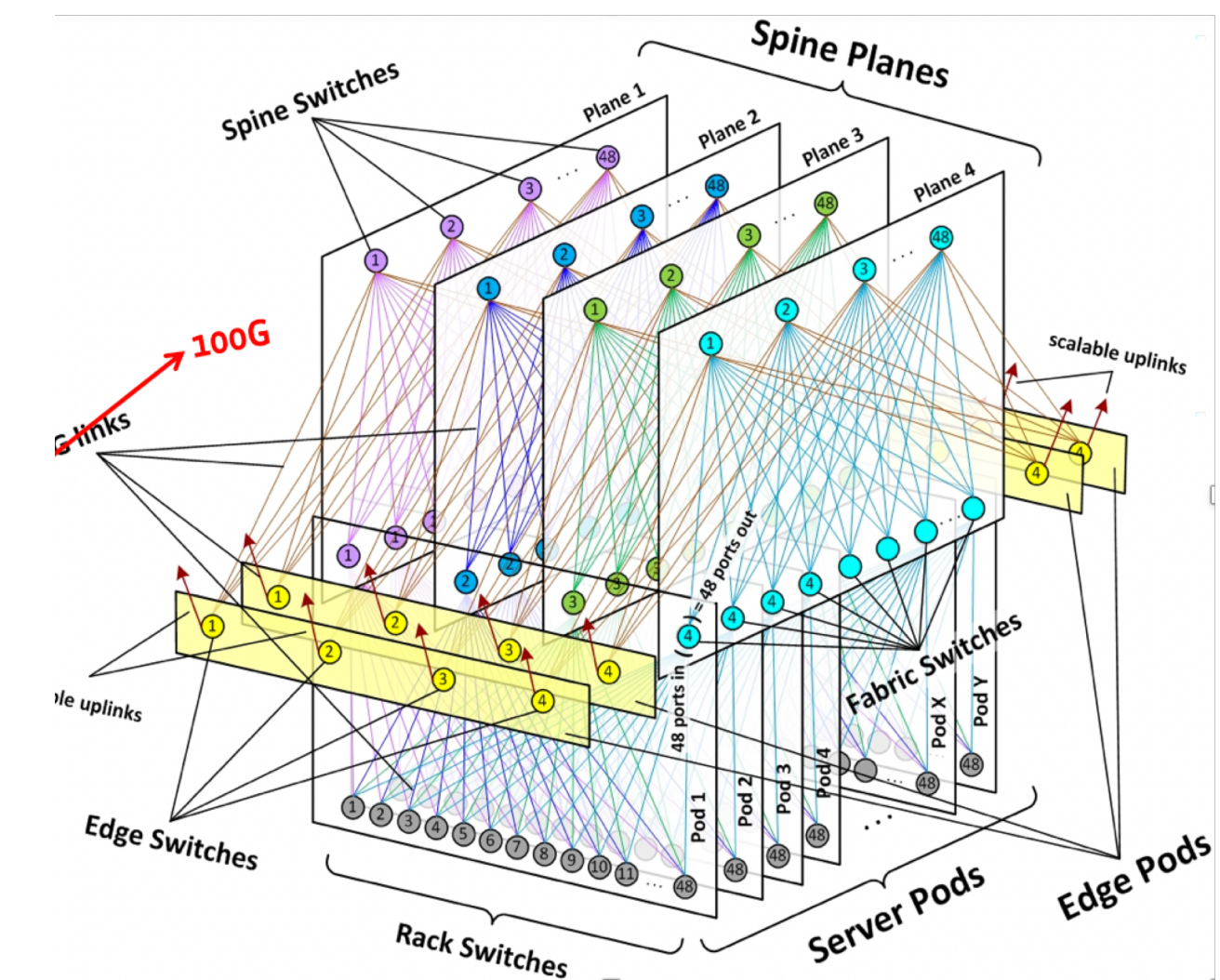
- 100G CWDM4 & Troubleshoot Challenges @ Data Centers
- BIM Concept Overview
- BIM Architecture & Advantages
- Tool Development & Troubleshoot Steps
- Conclusion & Future Development (Long-Term Reliability)



NETWORKING

Challenges in FB Infra: Troubleshooting Optics @Scale

- TOR to FSW to Spine connectivity
- More than ~1M 100G CWDM4-OCP lite in production across all FB data centers
- Need to support high volume
- Existing troubleshooting tools are not sufficient to support high volume
- Tedious process (one part at a time) to troubleshoot with the existing tools
- ~50% No Trouble Found (NTF) rate after RMA
- Identifying good optics (multiple parts) at data centers before RMA is very important to meet high volume requirement



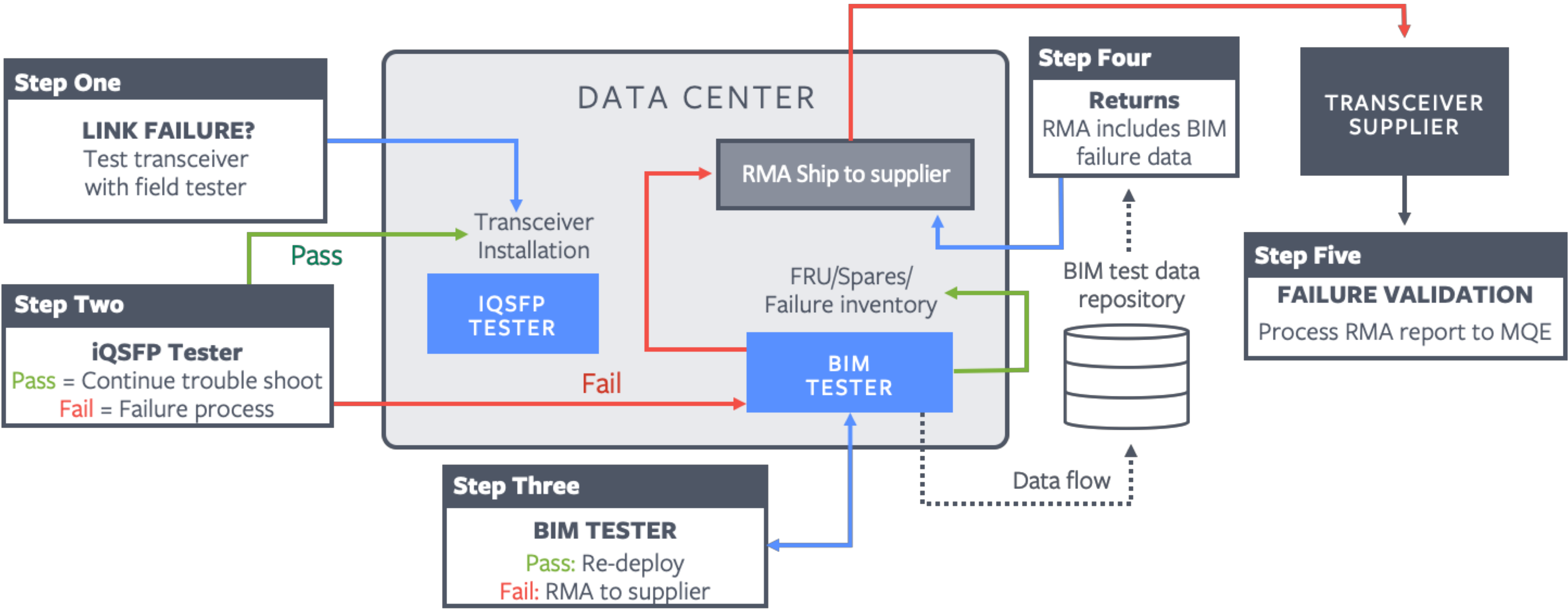
BIM Concept Overview



NETWORKING

- OCP Wedge 100 switch to run BER test; No external BERT
- 4 switches connected together
- Customized script to pull Digital Optical Monitoring (DOM) info (Tx and RX) including supplier info
- 128 optical interconnect tests simultaneously (low cost)
- ~5-10 minutes to troubleshoot all 128 optics
- Long-term test capabilities as required
- Goal to bring down the NTF number significantly and save time and effort for the engineers
- More stringent test (higher confidence in test result), capture of results in repository
- RMA to supplier now includes failure information – faster failure confirmation

Overview of Optical Troubleshoot Process @ FB Data Centers



- 1 When link failure occurs during deployment or post deployment, ability to test a single transceiver at the point of failure dramatically speeds the trouble-shooting process
- 2 Reduction of false failures flowing to “fail” inventory equals a reduction of false failures returned to supplier
- 3 BIM test provides second line of defense; more stringent test (higher confidence in test results), capture of results in repository
- 4 RMA to supplier now includes failure information for faster failure confirmation
- 5 Reduction in No Fault Found (order of magnitude expected); Quality data TAT improvement

Advantages of BIM Tool

- Low cost troubleshooting tool
- Reduce troubleshooting downtime
- Decrease in false failures
- Speed up supplier RMA process
- Improve failure inventory tracking
- Faster quality response time from suppliers



NETWORKING

Steps to troubleshoot (3 Steps)



NETWORKING

Step 1: Load the link failed optical transceivers on the OCP Wedge network switch with loop back cables

Step 2: Run the tool remotely

Step 3: Get the report, which includes status (pass/fail) and detailed information such as:

- Vendor information
- Parametric data by channel
- BER information by channel
- Pass/Fail identification

Software environment

- Worked with Edgecore to come up with a custom kernel and diagnostics script
- CLI tool for intuitive and accessible interfacing
- RPM, RedHat Package Manager for packaging the software
- Used chef, a system configuration tool
- The open source GlusterFS network attached storage system



NETWORKING

What the CLI tool does

- Connect to the OCP Wedge Network switch test station
- Optics initialization
- Check for all ports to come up
- Set pre-emphasis values
- Run the PRBS tests for user defined time
- Probe the PRBS on every channel for errors
- Query the DOM
- Parse the results
- Create a report with PASS/FAIL and all required information



NETWORKING

Help Menu



NETWORKING

```
Usage: ber_illusion_cli [OPTIONS]
```

```
ber_illusion_cli is a tool to troubleshoot Optics in datacenter test-  
stations. Wiki: https://fburl.com/ber\_illusion
```

Options:

```
-ip, --ip_address TEXT      IP/IPV6 address of the BER test station  
                             [required]  
-l, --log_directory TEXT   Log directory, (default: /tmp)  
-p, --password TEXT        Password to access the BER test station (default:  
                             facebook)  
-t, --time INTEGER         Time in seconds to run the BER tests (default:  
                             180)  
-u, --username TEXT        Username to access the BER test station (default:  
                             root)  
-h, --help                 Show this message and exit.
```

Sample results



NETWORKING

	A	B	C	D	E	F
1	0	1	2	3		31
2	Parameters	Port1	Port2	Port3	...	Port32
			Supplier A	Supplier B	...	Supplier M
3	Vendor Name	NA			...	
4	Vendor SN	NA	AAAAAAAAAAAA	BBBBBBBBBBBB	...	MMMMMMMMMMMM
5	Technology	NA	100G_CWDM_LITE	CWDM_MSA_100G	...	LR4_100G
6	Vendor PN	NA	XXXXXXXXXXXX	YYYYYYYYYY	...	ZZZZZZZZZZ
			1.0	2.3	...	p.q
7	Vendor rev	NA			...	
8	TX1 Power(mW)	NA	0.7800	0.6219	...	1.2424
9	TX2 Power(mW)	NA	0.7542	0.6487	...	1.1653
10	TX3 Power(mW)	NA	0.5819	0.6559	...	1.1752
11	TX4 Power(mW)	NA	0.7731	0.6253	...	1.2201
12	RX1 Power(mW)	NA	0.8325	0.0313	...	0.1173
13	RX2 Power(mW)	NA	0.8684	0.0312	...	0.9813
14	RX3 Power(mW)	NA	0.6840	0.0298	...	0.9408
15	RX4 Power(mW)	NA	0.8507	0.0320	...	0.7805
16	TX1 Bias(mA)	NA	60.950	60.950	...	35.210
17	TX2 Bias(mA)	NA	60.950	60.950	...	34.598
18	Tx3 Bias(mA)	NA	60.950	60.420	...	35.210
19	Tx4 Bias(mA)	NA	60.950	60.420	...	34.598
20	Temperature(degC)	NA	35.33984375	41.11718750	...	35.43359375
21	Lane1_BER	NA	better than5E-5 (PASS)	>5E-5(FAIL)	...	>5E-5(FAIL)
22	Lane2_BER	NA	better than5E-5 (PASS)	>5E-5(FAIL)	...	>5E-5(FAIL)
23	Lane3_BER	NA	better than5E-5 (PASS)	>5E-5(FAIL)	...	>5E-5(FAIL)
24	Lane4_BER	NA	better than5E-5 (PASS)	>5E-5(FAIL)	...	>5E-5(FAIL)
25	Test Results	NA	PASS	FAIL	...	FAIL
26						
27						
28	Test station IP:	2001:DB8::4				
29	Test start time:	2018-10-29 17:11:10.381871				
30	BER test time (s):	180				
31						

Conclusion & Future Development (Long-Term Reliability)

- BIM is highly scalable tool; 128 parts troubleshoot in ~5-10mins; can be increased just by adding OCP network switches
- Significant operational troubleshoot time reduction in data center environment
- Reduction in No Fault Found (order of magnitude expected)
- Quality data TAT improvement; Improve failure inventory tracking
- RMA to supplier now includes failure information – faster failure confirmation
- Can be used for long-term reliability and performance tests to understand the laser related issues both at supplier sites and customer data centers
- For more details, please contact FB and/or OCP community; Tools are available through OCP



NETWORKING



Open. Together.

OCP Global Summit | March 14–15, 2019

