White Paper

Enhancing Database Performance with Cisco Nexus 9000 and Application Centric Infrastructure

Contents

Overview **3**

Test Setup **4**

Switch Configuration **5**

Test Results **6**

Test Completion Times for Separate Flows 6

Test Completion Times for Mixed Flows 7

Test Conclusions **8**

For More Information **8**

Overview

In a typical data center fabric, all packets flowing through a switch port that belong to the same quality-of-service (QoS) class are treated equally: that is, all packets experience the same average latencies and the same probability of being dropped when congestion occurs on the switch port. However, the performance experienced by a user is usually the average time required to complete a transaction. Therefore, the switch fabric's behavior of treating all packets equally translates to a greater performance penalty on smaller flows because the percentage of delay due to packet drops and buffer latency is greater for smaller flows.

Both Cisco Nexus® 9000 Series Switches and Cisco Application Control Infrastructure offer features that allow data centers to better manage the priority of packets based on flow size.

- Dynamic Packet Prioritization (DPP): DPP prioritizes small flows over large flows, helping ensure that small flows are not affected by larger flows due to excessive queuing.
- Approximate Fair Drop (AFD): AFD introduces flow-size awareness and fairness to early-drop congestion-avoidance mechanism.

Both DPP and AFD ensure small flows are detected and prioritized and not dropped to avoid timeouts, while large flows are given early congestion notification through TCP to prevent over use of buffer space. As a result, smart buffers allow large and small flows to share the switch buffers in a much more fair and efficient manner. This provides buffer space for small flows to burst and large flows to fully utilize the link capacity with much lower latency times than a simple large buffer approach implemented in most merchant silicon switches.

This document analyzes the performance improvement that can be achieved with DPP for database workloads. Online transaction processing (OLTP) workloads are the core workloads for most enterprise applications and represent the transactions performed by users. These workloads usually have very small transaction sizes, ranging from a few kilobytes to a few tens of kilobytes, and are very sensitive to performance. Any performance degradation for these transactions directly affects the user experience.

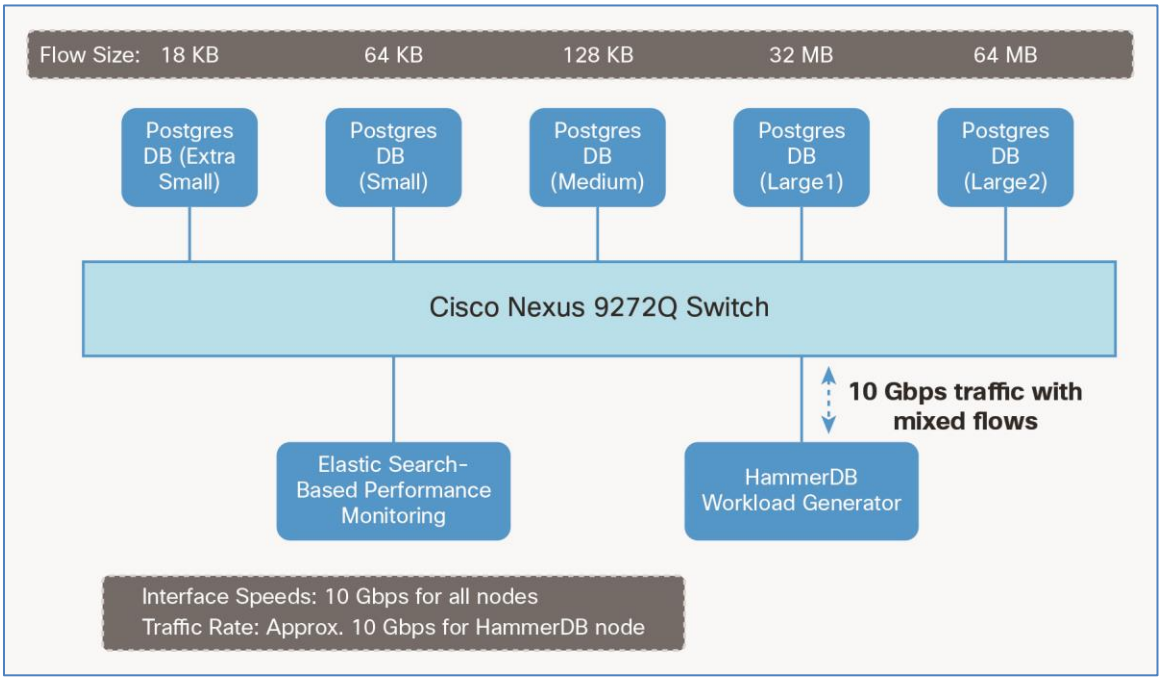
Today's virtualized data centers as well as evolving complex application topologies compel the traffic from OLTP workloads to be mixed with other workloads, which can create conditions for congestion and performance degradation. The mixing of traffic from various flows on a single switch port can occur in a variety of scenarios: for example, if multiple workload types are run on a single hypervisor, if a single application sends multiple OLTP or sequential online analytical processing (OLAP) queries to a variety of data sources, or if a variety of flows are aggregated in uplink ports.

The Cisco Nexus 9000 Series DPP & AFD features are extremely useful for enhancing performance of OLTP workloads by giving higher priority to smaller flows. This document describes the database tests conducted with Cisco Nexus 9272Q Switches with the DPP feature enabled. The tests demonstrated that the DPP feature significantly boosts the performance of smaller database flows in a scenario in which multiple flow sizes share a congested switch port. While the tests described in this document were conducted using Cisco Nexus 9272Q switches, the behavior would be similar for Application Centric Infrastructure switches as well.

Test Setup

Figure 1 shows the test setup.

Figure 1. Topology of Test Setup



The tests were performed as follows:

- Five database nodes were set up, with each on a separate bare-metal server.
- HammerDB was used as the load generator.
  - Five instances of the HammerDB workload generator were run: one for each database.
  - All five workload generator instances were run on the same bare-metal server to help ensure that traffic from all the databases was concentrated on the application node, leading to bandwidth congestion.
  - Modified TPC-C scripts were used to create the schema and generate queries. TPC-C workloads are used to simulate an OLTP environment. Multiple users concurrently perform short transactions, and the performance is measured by the time taken to complete the transactions. The scale of the IT infrastructure is determined by the number of warehouses, and the load is determined by the number of concurrent users. For the purposes of these tests, TPC-C scripts were modified. The field sizes of the databases were modified according to the test requirements, and queries were modified to reach the fields with the required payload sizes.
  - Each user in the HammerDB workload used a different port, resulting in a unique flow tuple.
- To allow comparison of tests for different workloads, the total number of transactions for each database workload was adjusted so that each workload was completed in 300 seconds when run individually.

- When all the database workloads were run simultaneously, the switch port on the HammerDB node was saturated with approximately 10-Gbps traffic flow. The following additional checks were performed:
  - To verify that the conditions for congestion indeed were being created, packet discards were looked for on the switch port.
  - To help ensure that no packet drops were occurring on the server side, the TCP buffer size was increased to 4 MB.
- The maximum transmission unit (MTU) size was set to 9000 for all servers and switch ports.

Switch Configuration

The following configurations were used to set up DPP on the switch:

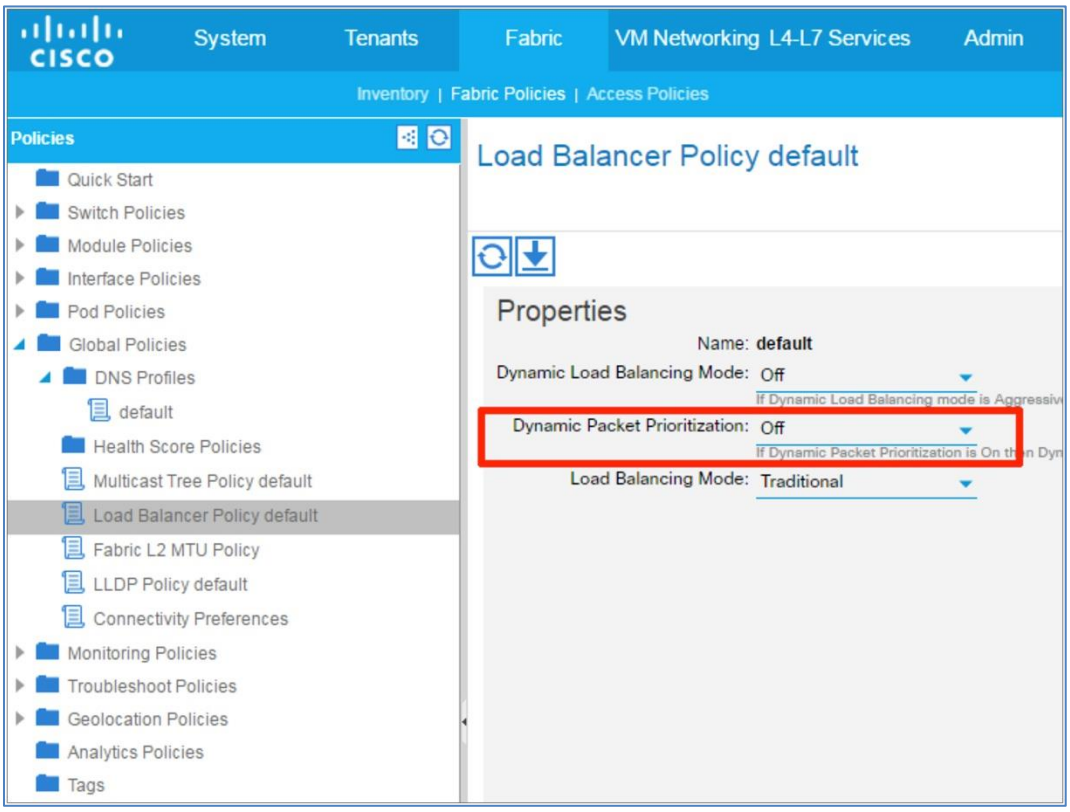
```
switchname Lacrosse-1
class-map type network-qos c-8q
class-map type network-qos c-8q-nq
policy-map type network-qos c-8q
 class type network-qos c-8q-nq-default
 dpp set-qos-group 7
 mtu 1500
policy-map type network-qos dpp-nwqos
 class type network-qos c-8q-nq-default
 dpp set-qos-group 7
 mtu 1500

hardware qos dynamic-packet-prioritization max-num-pkts 20
```

The last command, **hardware qos dynamic-packet-prioritization max-num-pkts 20**, is critical for tuning the behavior of DPP. It specifies the number of packets in a flow that must be prioritized. The setting **20** means that the first 20 packets of any flow will be prioritized in the no-drop queue, and subsequent packets will be sent to the drop queue. If a flow is shorter than 20 packets, the entire flow will use the priority queue.

The DPP definition of a flow tuple is (Src_IP, Src_Port, Dest_IP, Dest_Port). If two consecutive transactions in the same tuple occur less than 5 milliseconds (ms) apart, the two transactions will be considered part of the same flow. Multiple small flows that occur less than 5 ms apart thus coalesce into a single flow from the point of view of the switch. Therefore, delays between HammerDB transactions were adjusted to be approximately 8 ms, which is a reasonable assumption for OLTP transactions.

DPP configuration for ACI can be performed through the ACI's user interface. Figure 2 shows the configuration screen in ACI.

Figure 2. DPP Configuration in ACI



Test Results

Test completion times were assessed for both separate and mixed flows.

Test Completion Times for Separate Flows

Table 1 shows the results for separate flows.

Table 1. Completion Time for Separate Flows

	Flow Type	Transaction Size	Number of Transactions	Flow Completion Time (Seconds)
1	Extra Small	18 KB	41805	300
2	Small	64 KB	36585	300
3	Medium	128 KB	30612	300
4	Large 1	32 MB	507	300
5	Large 2	64 MB	240	300

The number of transactions for each test was adjusted until processing for each flow type was completed in 300 seconds when run individually. The small and medium flows by themselves do not saturate the network bandwidth on the switch port. However, when large flows are added to the mix, the bandwidth is saturated on the switch port of the HammerDB node.

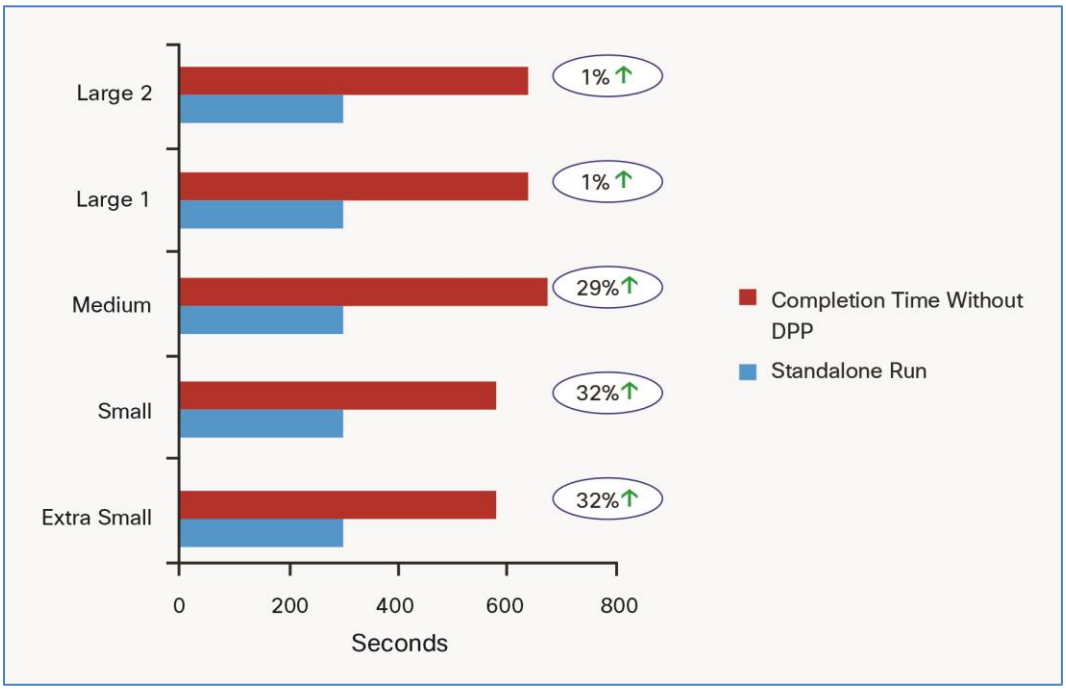
Test Completion Times for Mixed Flows

All five database workloads were run simultaneously, and the completion times for the workloads were measured. Five trials of the test were run, with and without DPP. If there was no congestion, each workload would finish in approximately 300 seconds. However, as a result of congestion, the flows contend for bandwidth and take longer to complete. Table 1 and Figure 2 show the results.

Table 2. Completion Time for Mixed Flows

	Flow Type	Transaction Size	Number of Transactions	Flow Completion Time Without DPP (Seconds)	Flow Completion Time with DPP (Seconds)
1	Extra Small	18 KB	41805	572	390
2	Small	64 KB	36585	585	396
3	Medium	128 KB	30612	672	479
4	Large 1	32 MB	507	646	641
5	Large 2	64 MB	240	651	642

Figure 3. Comparison of Completion Times With and Without DPP



DPP accelerated the small flows by approximately 32 percent, whereas the impact on large flows was negligible. Medium flows were accelerated by 29 percent because only the first 20 packets in the flow received priority, and last few packets of each flow had to use the drop queue.

Test Conclusions

The objective of the tests was to evaluate Cisco Nexus 9000 Series Switch performance with the DPP feature for small flows associated with OLTP workloads. The test results clearly demonstrated that when flows of multiple sizes contend for bandwidth on the same switch port, all the flows are delayed, but the effect on small flows is the worst. DPP is extremely effective in correcting this performance distortion by assigning higher priority to the smaller flows, making this feature very useful for running virtualized OLTP workloads.

While the tests were conducted for Cisco Nexus 9000 Series Switches, the same conclusions would hold true for Cisco Application Centric Infrastructure as well that provide DPP and AFD features.

For More Information

http://miercom.com/pdf/reports/20160210.pdf

http://www.cisco.com/c/dam/en/us/products/collateral/switches/nexus-9000-series-switches/at-a-glance-c45-737330.pdf

Americas Headquarters
Cisco Systems, Inc.
San Jose, CA

Asia Pacific Headquarters
Cisco Systems (USA) Pte. Ltd.
Singapore

Europe Headquarters
Cisco Systems International BV Amsterdam,
The Netherlands

Cisco has more than 200 offices worldwide. Addresses, phone numbers, and fax numbers are listed on the Cisco Website at www.cisco.com/go/offices.

Printed in USA C11-738425-00 01/17