February 2011

DEFINING CHARACTERISTICS OF QFABRIC

By Pradeep Sindhu

The purpose of this note is to introduce a set of defining characteristics that set **QFabric** apart from other technologies that have been used or proposed for building the internal networks of large scale data centers. It is important to state that these characteristics complement each other and work synergistically to help QFabric deliver its benefits. Thus we will insist that for any technology to be labeled "QFabric" it must have *all* of the defining characteristics and not just some of them.

The note begins with a set of definitions and a description of the general benefits of QFabric. It then lists the defining characteristics and says why they are important to building efficient, cost effective, dynamic, and easy to manage data centers.

Definitions and General Benefits

QFabric is the name of a packet-switched networking technology purpose-built to enable the construction of highly efficient, cost effective, dynamic, and easy to manage data centers over a wide range of scales using standard off-the-shelf computing, storage, and services elements. The elements connect to QFabric via open standard network interfaces such as Ethernet and Fiber Channel.

QFabric will enable the performance of data centers to be improved at a much faster rate than is possible by relying on performance improvements of the infrastructure elements alone (e.g., Moore's Law for microprocessors). We call this enhanced ability of QFabric to improve the performance of data centers over time "exponential scaling".

Overall, QFabric embodies two quintessential capabilities:

- The ability to treat data center computing, storage, services, and network resources as *fully fungible* pools that can be dynamically and rapidly partitioned *without the infrastructure or the applications knowing details* about each other[1]. This is the key to simplicity, efficiency, and security in the data center.
- The ability to connect the resources to each other at very high speeds with apparently no limitations in the interconnect except fixed interface bandwidth and a small transit latency. This is the key to high performance as well as further efficiency improvements.

A **QFabric** is represented as a circular object with **N** identical interfaces to which server, storage, services, or network devices can attach themselves. The circular symbol is chosen to suggest the property that all interfaces of **QFabric** are equidistant to each other in terms of latency and equivalent in terms of bandwidth they can sink or source. This means that the internal network of QFabric appears to its users to be flat, not hierarchical. The number of interfaces **N** is called the scale of the QFabric.

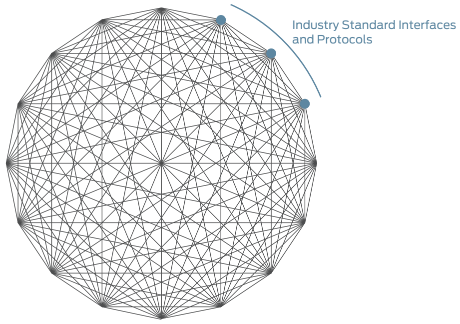

Figure 1: A QFabric with 16 Interfaces

Comparison to the Ideal Fabric

We define the **Ideal Fabric** as one that has fixed interface bandwidth, but infinite internal bandwidth and zero interface-to-interface latency. As we will see from the following defining characteristics, QFabric comes very close to the Ideal Fabric. In other words, QFabric is about as well as one can do given the constraints of physical implementation.

[1] In a multi-tenant data center it is completely impractical to assume that the infrastructure knows about the details of any application or that applications know about details of the infrastructure.

Defining Characteristics

A **QFabric** has the fundamental property of **scalability** which is defined as the ability to maintain a set of defining characteristics as the implementation grows in size from small values of **N** to large values of **N**. These defining characteristics are:

1. **Any-to-any connectivity with fairness and full non-blocking:** This is the ability of a set of interfaces to send and receive packets to and from any other set with no restrictions or pre-planning. Specifically, this includes the ability to absorb rapid changes to the rate of transmission, the number of active senders, and the number of active receivers. In all cases the full bandwidth of a target interface(s) is shared equally by all contending interfaces instantaneously and continuously, including in the special case of one interface sending to just one other interface; this equal sharing is referred to as "fairness". Finally, the only apparent congestion is due to the limited bandwidth of ingress and egress interfaces and any congestion of egress interfaces does not affect ingress interfaces sending to non-congested interfaces; this non-interference is referred to as "non-blocking".
2. **Low latency and jitter:** QFabric has interface-to-interface latency on the order of 2 microseconds at small scale growing slowly to about 10 microseconds at the largest scale. The latency also grows slowly with offered traffic load. QFabric also provides very low jitter.
3. **No packet drops under congestion:** When the instantaneous rate of packets coming in exceeds the instantaneous rate at which they are going out, QFabric signals the source (e.g. servers or VMs) causing congestion to slow down by the appropriate amount so that the arrival rate matches the departure rate. This throttling happens rapidly and continuously to match input rate to output rate in a smooth manner.
4. **Linear cost and power scaling:** The cost and power consumption of QFabric increases linearly with the number of interfaces **N**. This is in sharp contrast to traditional approaches where both cost and power increase much faster than linearly.
5. **Support of virtual networks and services:** QFabric implements **virtual** Layer 2 and Layer 3 networks to support multiple tenants each running multi-tier applications. Complex security and services requirements are supported by the insertion of Layer 4-7 processing at any point in an application's workflow. Full mobility of virtual machines from any interface to any other interface is supported. Support of virtual networks does not compromise any of the other properties.
6. **Modular distributed implementation that is highly reliable and scalable:** QFabric is built using a set of modular hardware and software components that are distributed and federated to provide high levels of redundancy. The modular implementation is designed to permit increasing or decreasing the number of interfaces while the system is running, a property we call "dynamic scalability".
7. **Single logical device:** Despite its distributed implementation, QFabric acts as a *single* logical packet switching device—the complexity of its distributed implementation is hidden without removing any of the desirable properties such as high reliability or dynamic scalability.

Why these characteristics are important

These defining characteristics represent a quantum jump in the networking technology used to build data centers. They are the principal way in which we will provide exponential scaling for data centers for the next decade.

Importance of Scalability

Scalability is the fundamental characteristic of QFabric. It is important for two reasons:

- **Economics:** Scale is a pre-requisite for achieving better economics in data centers. Small scale data centers simply cannot be made as efficient as large scale ones because of the *pooling principle*, which states that when P equal size partitions of a resource are combined into a single pool, we need $\sqrt{P}$ fewer units of the resource to provide the same level of service. Of course, to reap the benefits of pooling we need to be able to make the resources fully fungible.
- **Performance:** For most applications the raw performance of a set of tightly coupled computing elements in a single large data center is significantly better than the collective performance of these same elements distributed over a number of smaller data centers. This performance difference has to do with the inherently lower latency and higher bandwidth of inter-processor communication in a single data center.

Importance of Any-to-Any Fair, Non-Blocking Connectivity

This type of connectivity is critical to pooling the computing and storage resources in a data center. Without these capabilities resources in a large data center would remain stranded and not efficiently usable. This "flat" connectivity also goes a long way towards simplifying the writing of applications since there is no need to worry about the performance hierarchy of communication paths inside the data center. It also relieves the operations staff from having to worry about the "affinity" of application components in order to provide good performance.

Importance of Low Latency

Low latency in QFabric is critical to high performance, especially for modern applications where the ratio of communication to computation is relatively high compared to legacy applications. For example financial applications are especially sensitive to latency. High latency translates directly to lower performance because applications stall when they are waiting for a response over the interconnect—the more frequent and longer the stall, the lower the performance.

Importance of No Drop under Congestion

The ability to not drop packets when congestion occurs is critical to efficiently transporting "bursty" server to disk traffic. Most applications assume that reads and writes to disk will succeed with very high probability. Packet drops due to congestion breaks this assumption and forces the application to handle packet loss as an error, resulting in a drastic reduction in performance given the relative frequency of congestion events. As noted earlier, in QFabric the no drop property is provided by not harming "innocent bystanders" interfaces whose traffic is not destined to congested interfaces. This latter property of the implementation is key to supporting multi-tenancy and multiple applications within a tenant.

Importance of Linear Cost and Power Scaling

Since each interface requires some forwarding hardware, it is easy to see that the cost and power of a fabric must increase at least linearly with the number of interfaces N. QFabric is able to achieve the ideal of linear scaling, which is highly desirable from both a CAPEX and an OPEX standpoint.

Importance of Support for Virtual Networks and Services

Support for Virtual Networks is important because data centers are increasingly multi-tenant and have complex requirements for application security, performance, and reliability. Virtual networks provide the basic tools to allow resources to be partitioned and yet allow them to communicate securely when necessary. These abstractions are essential to decoupling the applications from the infrastructure.

Importance of a Distributed, Modular Implementation

A modular implementation where the modules are kept independent through the use of physical or logical separation has two important benefits. The first is that failures in either the hardware or software are unable to compromise the entire system. The second is that it becomes possible to increase or decrease the size of the system while it is running.

Importance of Acting as a Single Logical Device

This characteristic is important because it provides the simplest possible model for the network from a data center administrator's standpoint. It permits orchestration and management applications to be as simple as possible.

Conclusion

QFabric is a revolutionary new network technology purpose-built to allow the construction of large scale data centers using off-the-shelf computing and storage elements connected using industry standard network interfaces. It allows the performance of data centers to be "scaled exponentially", or improved at a rate much faster than is possible by relying on improvements of the computing and storage elements alone. This exponential scaling property is fundamental to making data centers not only much more powerful but also more economically efficient, dynamic, and easier to manage.

At the highest level, QFabric embodies two quintessential capabilities. The first is its ability to treat data center computing, storage, services, and network resources as fully fungible pools that can be dynamically and rapidly partitioned without the infrastructure or the applications knowing details about each other; this is the key to simplicity, efficiency, and security in data centers. The second is the ability to connect the resources to each other at very high speeds and with very low latency while ensuring fairness; this is the key to high performance and further efficiency improvements.

This paper has laid out seven defining characteristics of QFabric that set it apart from all other networking technologies currently in use or proposed for the data center. These defining characteristics are orthogonal to each other and encompass performance, cost, power, virtualization, security, reliability, and modularity dimensions that work together to provide the two quintessential capabilities mentioned above.

QFabric represent a quantum jump in the networking technology used to build data centers and will be the key to exponentially scaling high performance computing and storage over the next decade.

Biography

Pradeep Sindhu

Vice Chairman, Chief Technical Officer and Founder

Pradeep Sindhu founded Juniper Networks in February 1996, and is now the company's Vice Chairman and CTO. His founding vision was and still is to ***Connect Everything and Empower Everyone***.

Before founding Juniper Networks, Dr. Sindhu was a Principal Scientist and Distinguished Engineer at the Computer Science Lab at Xerox's Palo Alto Research Center (PARC).

He holds a bachelor's degree in electrical engineering from the Indian Institute of Technology in Kanpur, as well as a master's degree in the same discipline from the University of Hawaii. In addition, Dr. Sindhu holds both a master's and a doctorate degree in computer science from Carnegie Mellon University.

Corporate and Sales Headquarters
Juniper Networks, Inc.
1194 North Mathilda Avenue
Sunnyvale, CA 94089 USA
Phone: 888.JUNIPER (888.586.4737)
or 408.745.2000
Fax: 408.745.2100
www.juniper.net

APAC Headquarters
Juniper Networks (Hong Kong)
26/F, Cityplaza One
1111 King's Road
Taikoo Shing, Hong Kong
Phone: 852.2332.3636
Fax: 852.2574.7803

EMEA Headquarters
Juniper Networks Ireland
Airside Business Park
Swords, County Dublin, Ireland
Phone: 35.31.8903.600
EMEA Sales: 00800.4586.4737
Fax: 35.31.8903.601

Copyright 2011 Juniper Networks, Inc. All rights reserved. Juniper Networks, the Juniper Networks logo, Junos, NetScreen, and ScreenOS are registered trademarks of Juniper Networks, Inc. in the United States and other countries. All other trademarks, service marks, registered marks, or registered service marks are the property of their respective owners. Juniper Networks assumes no responsibility for any inaccuracies in this document. Juniper Networks reserves the right to change, modify, transfer, or otherwise revise this publication without notice.

2000384-001-EN Feb 2011

Printed on recycled paper