FIBRE CHANNEL WRITE ACCELERATION

The Cisco® MDS 9000 Family of Multilayer Directors and Fabric Switches provide a complete set of intelligent storage services that include Fibre Channel Write Acceleration (FC-WA), which can dramatically improve the number of Small Computer System Interface (SCSI) I/O operations per second over long-distance in a Fibre Channel-based storage area network (SAN), by reducing the effect of transport latency on the processing of each operation.

SUMMARY

Modern data centers are built around SANs that provide secure, reliable, and scalable centralized data-storage facilities. An effective business-continuance and disaster-recovery plan mandates the deployment of multiple data centers located at optimal distances to protect against regional power failures or disasters, yet close enough for synchronous data replication without affecting the application performance. Achieving this balance for successful business continuance poses a significant challenge for businesses.

The time in which the business needs to recover during a disaster determines the type of data replication required. Synchronous data replication provides the least amount of downtime but requires the data centers to be close enough so that the application performance is not affected by the latencies introduced for every I/O operation. Asynchronous or semi-synchronous replication allows for greater distance but the data centers are in lock step with each other, with the secondary data center lagging behind the primary by a fixed amount of time. This in turn implies that there is a loss of data for that period of time. An optimal solution is to increase the distance between the data centers without introducing additional latency for I/O operations.

The Cisco MDS 9000 Family provides FC-WA, capable of dramatically reducing the effect of transport latency during the time needed to complete an SCSI operation over distance. Consequently FC-WA allows longer-distance between the primary and secondary data centers, increasing the resilience of the solution and the flexibility in selecting the locations with the least impact on the application performances, and improving application performance for a given configuration.

The FC-WA feature is supported on the Cisco MDS 9000 Family Storage Services Module, and may be used on any Cisco MDS 9200 Series of Multilayer Fabric Switches and Cisco MDS 9500 Series of Multilayer Directors running the Cisco SAN-OS Release 2.0(2b) or later, with the Enterprise Licensing package.

CHALLENGE

The term "business continuity" implies the availability of the entire set of physical, logical, and human resources needed to run the business in the event of an unforeseen stoppage. The physical infrastructure includes the location and its associated resources, with respect to the IT area; these resources include the physical servers, storage, and network equipment. The logical infrastructure is the data or information needed to continue with the business. It is imperative that a finite set of procedures and processes be put in place along with the adequate technology to maintain business continuity.

Figure 1. Relationship Between Data Center Distances and Application Availability
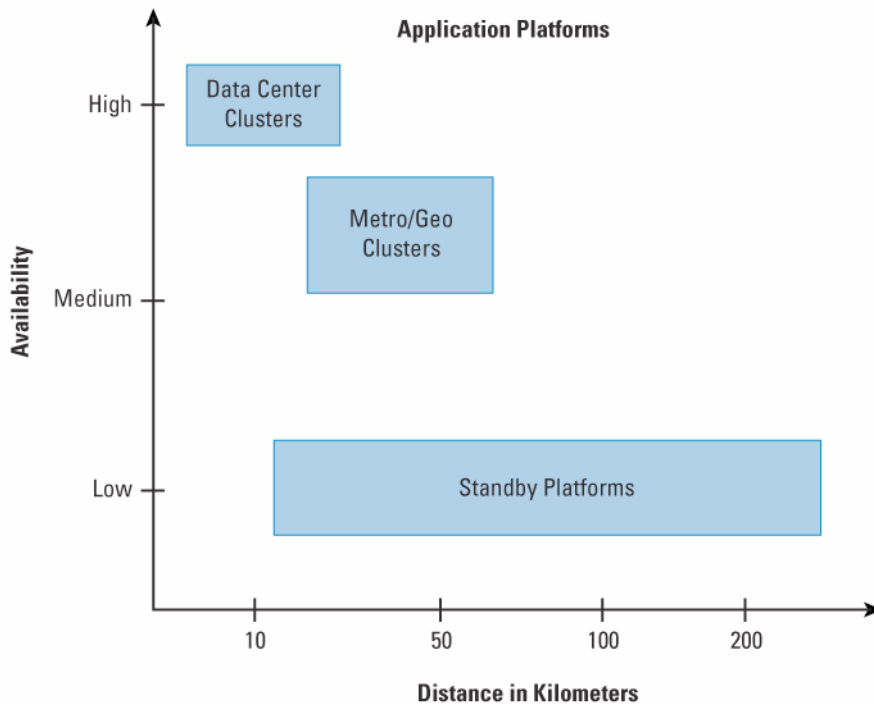
Technology plays a critical role in determining how, when, and where the business recovery starts after an outage. This technology includes the protection or duplication of the physical IT infrastructure, a one-time operation and replication of the data from the primary site to the secondary site, and ongoing daily operations. Applications platforms, including their supporting hardware and software, can be protected by the following technologies:

- Data center clusters for application availability within a given data center
- Metro or geo clusters for metropolitan distances, to keep the applications continuously available
- Standby platforms and applications to help restore business operations in time to meet business objectives

Figure 1 shows the relationship of application-platform availability and distance between data centers.

- Data replication is a complex task and the method of replication is determined by the criticality of the application using the data. For mission-critical applications it is mandatory that no data be lost and that downtime is kept to an absolute minimum, whereas other applications can withstand significantly longer interruptions. Data-replication technologies that are available are; include coarse wavelength-division multiplexing (CWDM), dense wavelength-division multiplexing (DWDM), and Fibre Channel over IP (FCIP). CWDM combines up to 16 wavelengths onto a single fiber. CWDM technology uses an ITU-standard 20-nanometer (nm) spacing between the wavelengths, from 1310 nm to 1610 nm. This technology does not amplify or regenerate the light signals, or in other words is passive repeater, and therefore is limited to shorter distances of up to 80 km.
- DWDM combines up to 64 wavelengths onto a single fiber. DWDM technology uses an ITU-standard 100-GHz or 200-GHz spacing between the wavelengths, arranged in several bands at ~1500 to 1600 nm. In this technology the wavelengths are closer together with respect to CWDM, and therefore more accurate. The advantage is a much higher density of wavelengths, and the fact that the light signals can be amplified and regenerated allows for longer distances, with regeneration for approximately every 150 km.

The data transport medium in wavelength division multiplexing is an optical fiber cable, often referred to as "dark fiber."

- FCIP is a technology where the Fibre Channel frames are encapsulated into an IP packet and sent over to the other site on a dedicated TCP/IP tunnel created for this purpose. This technology can very effectively deliver high throughput over long distances, but the latency of the I/O depends on the distance.

These replication technologies can use either synchronous replication or asynchronous replication methods.

- Synchronous replication, also referred to as real-time replication, implies that an I/O is not complete until both sides acknowledge its completion. To maintain consistency of data, every write to the primary storage is written over to the secondary storage in the same sequence before the application writing is considered successful. This method helps ensure that data stored at both sites is always consistent, and the only data that could be lost in the event of a failure is the data that was not committed or was being transmitted at the time of the failure. The disadvantage of performing synchronous replication is that Fibre Channel protocol (FCP) defines four steps, or two round trips, before every I/O can complete. The speed of light becomes the limiting factor because a latency of 1 millisecond (ms) is introduced for every 50 km of distance between the data centers, even with the use of dark fiber.
- Asynchronous replication, often referred to as "store-and-forward" replication, stores changes to the data at the primary storage site for a predetermined amount of time and then forwards the changes to the secondary storage system for updates. This way the secondary storage system lags behind the primary storage system by a fixed amount of time. This technology allows the data at the secondary storage system to be synchronized with the primary storage system in a given amount of time. Some data changes may not be recoverable if those changes were still stored on the primary storage system and were not yet forwarded at the time of the outage.

Given the different technologies and methods of replication, the challenge for data-center managers is to ensure that the data is replicated in a timely manner without affecting the performance of the applications using that data. The best possible solution for a data-center manager is to implement synchronous replication on a wavelength division multiplexing solution. However the distance between the data centers, which, with the existing technologies, cannot exceed 80 to 100 km without affecting the applications, limits this solution.

To mitigate this issue, Cisco Systems® introduced Fibre Channel Write Acceleration (FC-WA), an intelligent feature in the Cisco MDS 9000 Family switches. FC-WA accelerates the writing process over long distances, thus helping perform synchronous replication over significantly longer distances than what was previously possible.

SOLUTION

Cisco MDS 9000 Family Storage Services Module

The Cisco MDS 9000 Family Storage Services Module provide the intelligent service of identifying the SCSI I/O flow for a given initiator-target pair. This information is used to provide the FC-WA feature and the feature to gather advanced I/O statistics for a given initiator-target pair. The FC-WA feature decreases the latency of an I/O over long distances and the advanced I/O statistics collected can be used to evaluate the storage performance for the initiator-target pair.

The Storage Services Module incorporates all the capabilities of the Cisco MDS 9000 Family 32-Port Fibre Channel Switching Module and also provide scalable and intelligent storage services. The Storage Services Module features a highly distributed processing architecture and hardware-based SCSI processing, capable of optimizing storage access. Data is processed in band by an application-specific processor called Data Path Processor (DPP). The control commands and information are processed by a dedicated Control Path Processor (CPP). To deploy FC-WA in the SAN, both the initiator and target devices must be directly attached to a Storage Services Module.

Fibre Channel Write Acceleration

FC-WA minimizes storage latency and improves the number of application transactions per second over long distances. It increases the distance of replication or reduces effective latency to improve performance during synchronous replication.

The improved performance results from a coordinated effort performed by the Storage Services Module local to the initiator and the Storage Services Module local to the target. The initiator Storage Services Module, bearing the host-connected intelligent port (HI-port), allows the initiator

to send the data to be written well before the write command has been processed by the remote target, and an SCSI Transfer Ready message has had the time to travel back to start the data transfer in the traditional way. The exchange of information between the HI-port and the disk-connected intelligent port (DI-port) allows the transfer to begin earlier than in a traditional transfer. The procedure makes use of a set of buffers for temporarily storing the data as near to the DI-port as possible. The information between the HI-port and DI-port is piggybacked on the SCSI command and the SCSI Transfer Ready command, so there are no additional FC-WA-specific frames traveling on the SAN. Data integrity is maintained by the fact that the original message that states the correct execution disk side of the write operation (SCSI Status Good) is transferred from the disk to the host.

Figure 2 shows the effect of latency in the communication channel to the time taken to complete the I/O operation during a SCSI write operation. The time added to the net execution time of the operation is at least four times the trip delay between the host and the disk, because of the transfer of the command, the Transfer Ready message, the data, and the status. Figure 3 shows how FC-WA allows the data to be sent on the line without waiting for the disk Transfer Ready message to be transferred all the way back to the host. To preserve data integrity, the status message is not emulated. Depending on the timing, the latency added by the communication time may be as low as two times the trip delays, transfer of the command, and transfer of status. Therefore the expected distance between the host and the disk can now be increased by up to 50 percent.

Figure 2. Normal SCSI Write Operation

Normal SCSI Write

Figure 3.   SCSI Write Operation with Storage Services Module Cisco Solution Performance
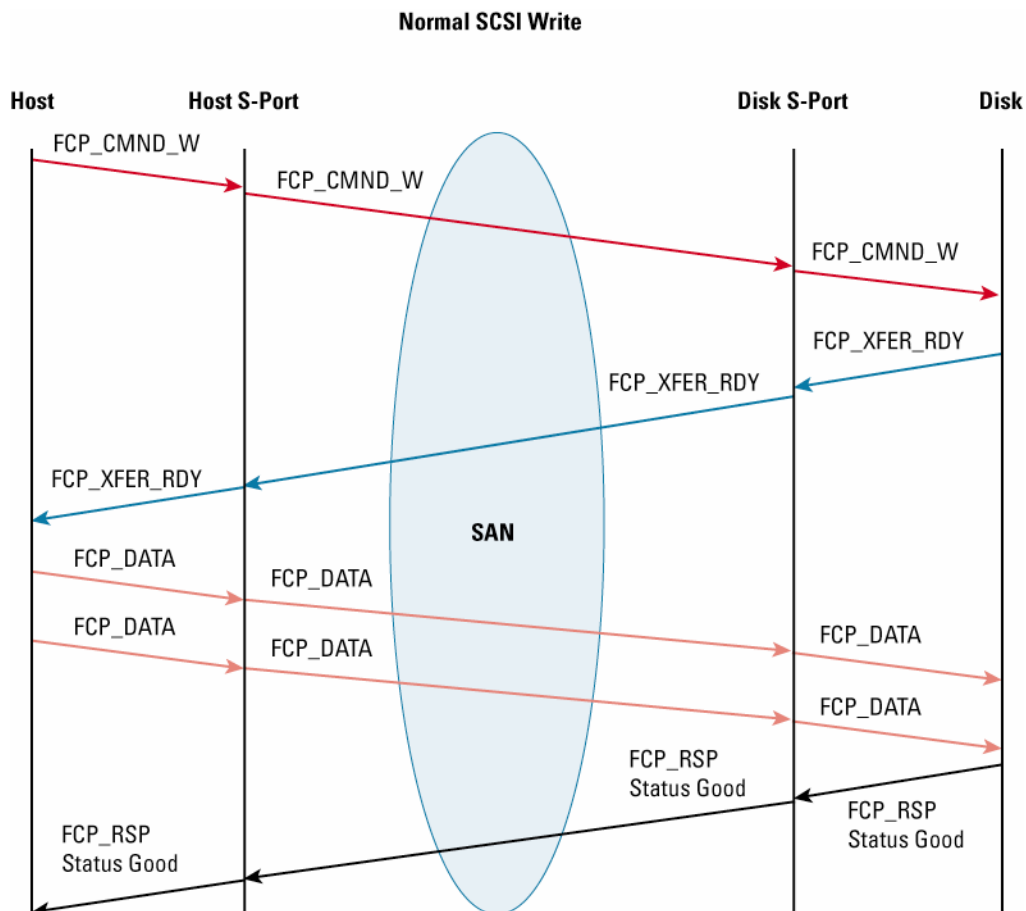
SCSI Write with FC-WA
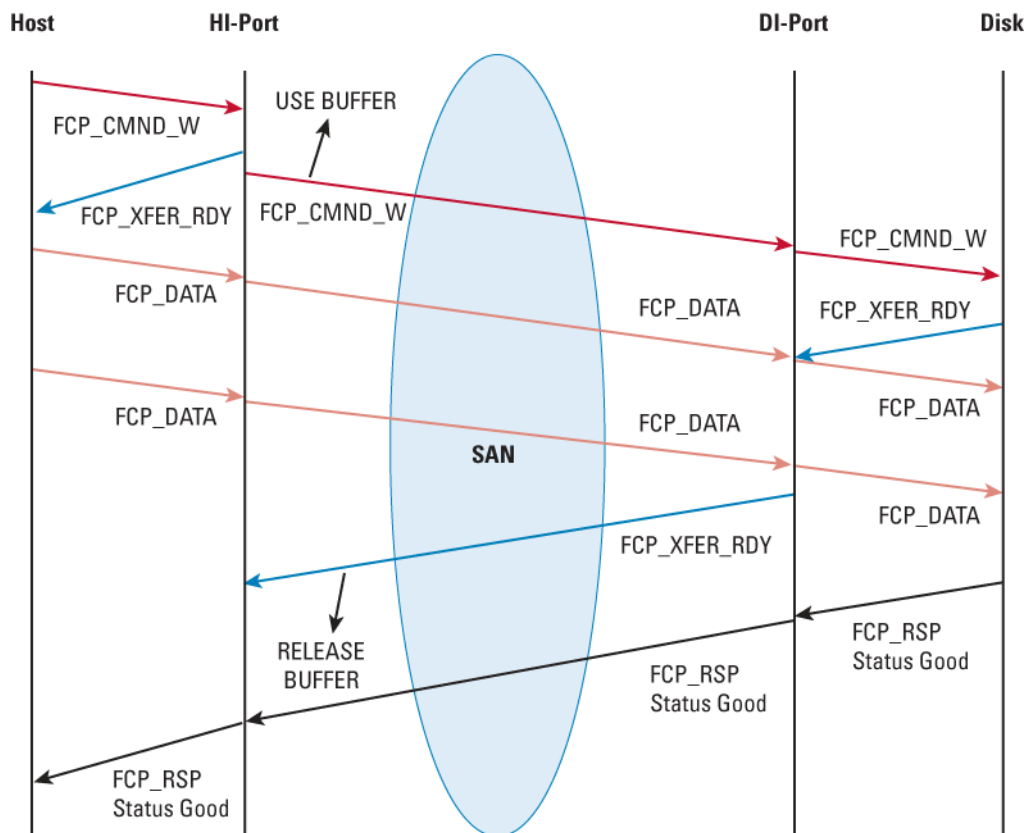
Cisco MDS 9000 Family Storage Services Module performance depends heavily on the traffic profile of the SCSI operation performed by the specific application, on the performances of the storage system, and on the distance between the host and the target disk.

The following factors can enhance the benefits of the Storage Services Module:

- Larger transfer latency, or greater distances
- Small SCSI writes
- Storage with low write latencies
- Larger percentage of writes than reads

It is clear from the preceding criteria that online transaction processing (OLTP) traffic with short and fast operations over long distances will benefit the most, but the fact that the read-to-write percentage is not favorable in OLTP applications may restrict the total application throughput gain, whereas a write-intensive application can benefit more from FC-WA—in the range of 20 percent. To highlight the performance characteristics of FC-WA, two sets of tests have been performed:

- Emulated SCSI traffic write only, optimized for high performances from the disk. The performance increase has been measured in terms of relative increment in I/Os per second (IOPS) and latency over a distance of 100 km.
- OLTP traffic from an Oracle application; the performance increase has been measured in terms of relative increment of transactions per second.

Both tests were performed accelerating the traffic flow between two EMC storage arrays running Symmetrix Remote Data Facility (SRDF) synchronous replication for multiple Remote Data Facility (RDF) pairs. Figure 4 shows the gain in IOPS and reduction in latency for emulated SCSI write traffic at 100-km distances. Figure 5 shows the transactions per second (TPS) gained for an Oracle application using FC-WA.

Figure 4.   Percentage Gain of IOPS and Reduction of Latency
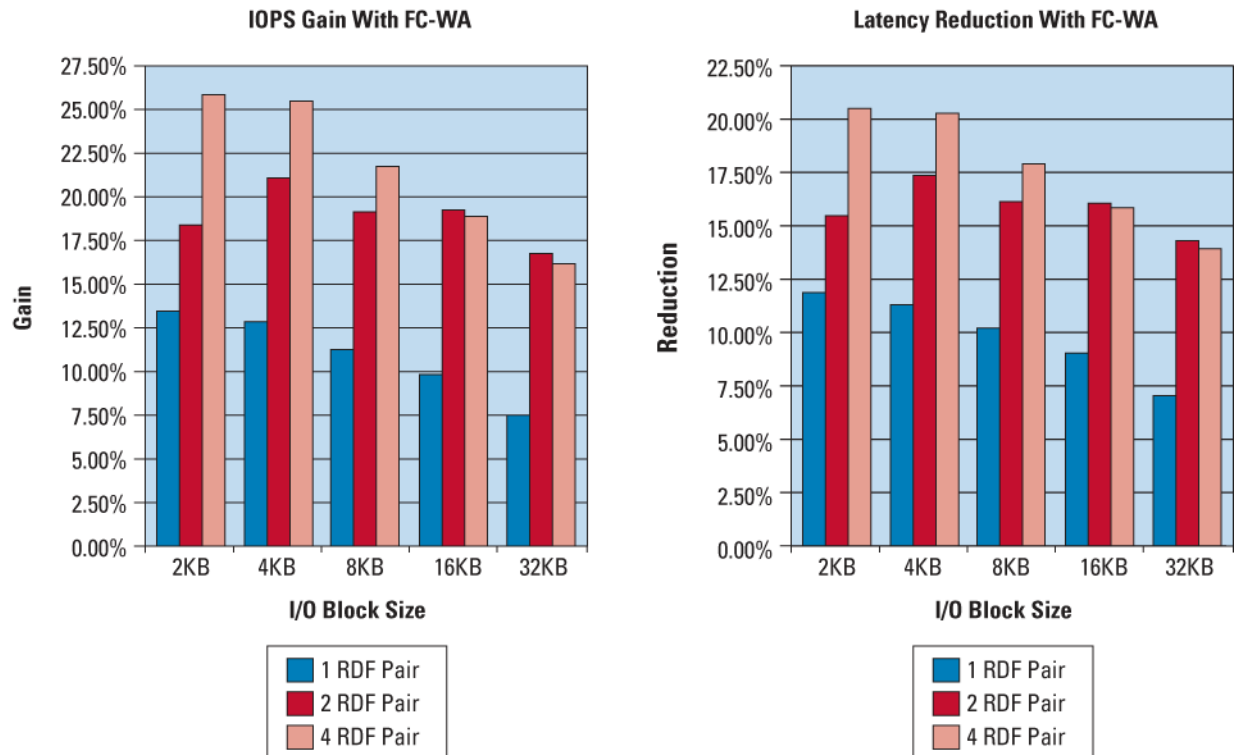
Figure 5. Oracle OLTP Application Performance with FC-WA

Oracle Application—TPS Gain
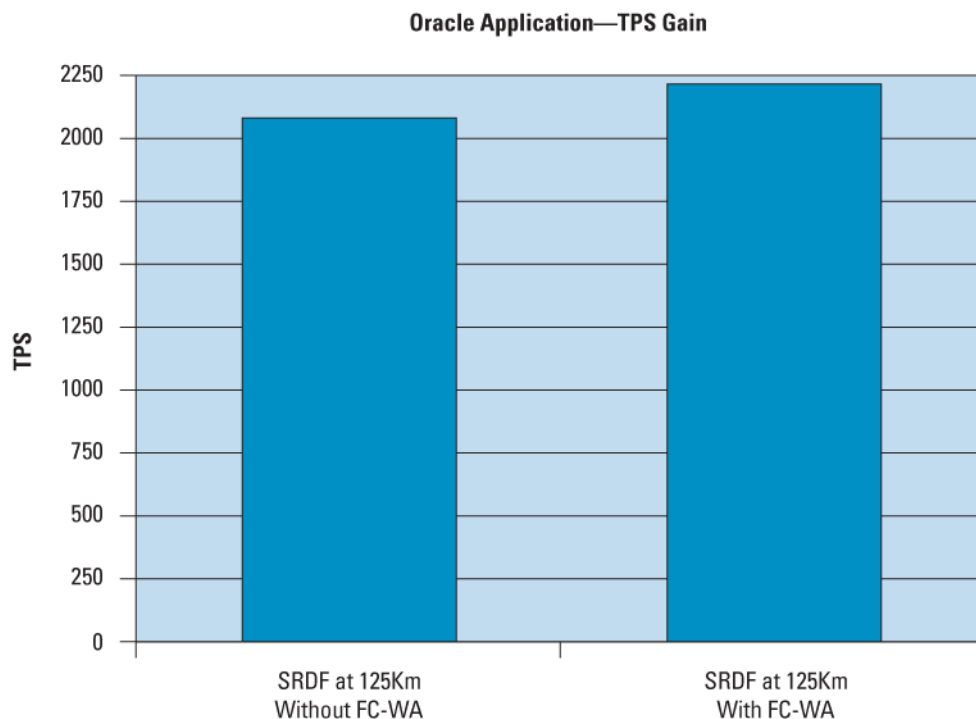
CONCLUSION

Using synchronous data replication as part of a business continuance plan can limit the distance between data centers. As demonstrated by the preceding test results, the Cisco MDS 9000 Family Storage Services Module can mitigate this disadvantage through the use of FC-WA, which significantly increases the distance between data centers without affecting application performance. The Cisco MDS 9000 Family Storage Services Module with FC-WA and its related intelligent features enhances the flexibility and resiliency of business-continuance solution.

CISCO SYSTEMS

Corporate Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-4000
    800 553-NETS (6387)
Fax: 408 526-4100

European Headquarters
Cisco Systems International BV
Haarlerbergpark
Haarlerbergweg 13-19
1101 CH Amsterdam
The Netherlands
www-europe.cisco.com
Tel: 31 0 20 357 1000
Fax: 31 0 20 357 1100

Americas Headquarters
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134-1706
USA
www.cisco.com
Tel: 408 526-7660
Fax: 408 527-0883

Asia Pacific Headquarters
Cisco Systems, Inc.
168 Robinson Road
#28-01 Capital Tower
Singapore 068912
www.cisco.com
Tel: +65 6317 7777
Fax: +65 6317 7799

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

Argentina • Australia • Austria • Belgium • Brazil • Bulgaria • Canada • Chile • China PRC • Colombia • Costa Rica • Croatia • Cyprus
Czech Republic • Denmark • Dubai, UAE • Finland • France • Germany • Greece • Hong Kong SAR • Hungary • India • Indonesia • Ireland • Israel
Italy • Japan • Korea • Luxembourg • Malaysia • Mexico • The Netherlands • New Zealand • Norway • Peru • Philippines • Poland • Portugal
Puerto Rico • Romania • Russia • Saudi Arabia • Scotland • Singapore • Slovakia • Slovenia • South Africa • Spain • Sweden • Switzerland • Taiwan
Thailand • Turkey • Ukraine • United Kingdom • United States • Venezuela • Vietnam • Zimbabwe