the *Availability Digest*

So You Think Your System is Robust?

August 2007

A lot of people are convinced that their systems are robust. Sometimes, however, their best laid business continuity plans can go astray. The following sad tales are all about those who learned the hard way that business continuity planning is harder than they thought. And all of these snafus occurred to major enterprises in the first half of this year, 2007.[1]

The Dow Drops 200 Points in Seconds

Talk about a trader's nightmare. Tuesday, February 27, 2007, was already a bad day. Shortly before 3 PM, the Dow Jones Industrial Average, the most closely watched market index, was already down by 280 points from its open of about 12,450 due to uncertain domestic economic news and an apparent meltdown in the Chinese economy. Then, within seconds, the index dropped another 178 points. Such a rapid drop had never occurred before.

The sudden drop sent traders into a selling frenzy. The market was sent into a further tailspin until it started to recover from a 546 point drop, its largest loss since 9/11.

However, bad as the market drop was, it was not nearly as precipitous as it appeared. It turned out that the Dow Jones system that is responsible for calculating the Industrial Average and 400 other indices had started to fall behind around 1:50 in the afternoon. About 2 o'clock, Dow Jones noticed that a backlog of trades was building up and, suspecting a computer problem, switched over to a backup system. It took an hour for the backup to come fully online, during which time the backlog of trades continued to grow.

The Dow Jones Industrial Average, at one-minute intervals yesterday:

Official

Recalculated to adjust for tabulation delay

Yesterday's close: 12216.24

Source: Dow Jones Indexes

Finally, the backup system became fully functional at about 3 PM and rapidly processed the last 70 minutes of trades. The falling market had actually dropped 178 points over the previous hour and ten minutes, but the sudden processing of the backlog made it appear that this drop had occurred in the last few seconds.

Though Dow Jones management had notified the New York Stock Exchange of the backlog, evidently this message did not get to the traders. Many of them frantically sold positions that they normally would not have just to see the market recover in the next hour. This sudden, massive

---

[1] The information for all of these stories was taken from news reports in the days following the incidents.

increase in volume overwhelmed the NYSE systems; and trades had to be posted manually. Exchange staff worked overtime to clear up the backlog.

The Exchange may face several lawsuits by traders who claimed that they lost their collective shirts due to the erroneous reporting.

US Airways meets Merger Mania

It had been eighteen months since US Airways and Air West had consecrated their merger. Finally, the new reservation system that would merge the reservation systems of both of the original airlines, an upgraded version of the Air West reservations system, was ready to be put online. No longer would Air West passengers have to go to an Air West kiosk at the airport to check in, nor would US Airways passengers have to go to a US Airways kiosk. All kiosks would serve both airlines. Furthermore, there would be a single Web site for online check-in.

Hundreds of new check-in kiosks that could handle the flights for both airlines had been installed. After much planning and testing, the switchover was scheduled for midnight, Saturday, March 3, 2007. Switchover would take about two hours, during which time the reservation system would not be available. However, there were no domestic flights at this time anyway.

Accordingly, the system was shut down at midnight; and seven million passenger records were transferred from the old systems to the new reservation system. The new system was brought online in less than the scheduled two hours to the relief of management.

Then chaos hit. It turned out that the new reservation system worked flawlessly – almost. Unfortunately, it would not work with the new kiosks. The old kiosks still were functional, but most of those that had been installed at the old US Airways hubs – Charlotte, Philadelphia, Boston, as well as Las Vegas – had been decommissioned. The computer glitch affected 685 kiosks out of 930 that were in operation.

As passengers arrived for their flights on Sunday morning at these airports, they found that they could not check in at the kiosks but rather had to wait in line for an agent. Few agents were available to meet the unexpected demand, as staffing was still normal. Lines got longer and longer, and tens of thousands of passengers stood in line for hours. Waits of up to two or three hours were reported. The airlines did not initially hold planes, and many passengers missed their flights. The result – flaring tempers and overworked agents.

US Airways sent out an urgent plea to its passengers to check in online before they arrived at the airport. Problem was, the online check-in facility did not come up until Sunday afternoon.

In its wisdom, US Airways did at least drop rebooking fees for these hapless passengers. It also scurried to get additional staff assigned to agent positions and ultimately started holding planes so that all could make it aboard.

By late Tuesday, more than one-half of the affected kiosks were still not working; but lines were down to an average of ten or fifteen minutes due to the increase in staff. US Airways ordered 600 new kiosks to improve performance. These new kiosks were delivered in June and were added to the 800 kiosks already in service.

There is one positive side to this story. Unlike the Blackberry failure described in an earlier article,[2] US Airways management worked hard to keep passengers informed. H. Travis Christ, VP Sales and Marketing, began publishing a merger update sent to all US Airways frequent flyers. In one issue, he said:

---

[2] Blackberry Gets Juiced, *Availability Digest*; May, 2007.

"You may be weary of us apologizing for the tough times we had this spring due to the migration of our reservations systems, but here it is: We sincerely apologize for the numerous failures. After a pretty nice start to the merger, we've learned some tough lessons through all this and received some very helpful advice from many of you, some of which we're applying and some of which is unprintable. So thank you for sticking with us as we climb back on the horse."

The Week that Canadians Couldn't Pay Taxes

The Canada Revenue Agency (CRA) is the Canadian government arm that is responsible for collecting taxes. Over the last several years, the Internet-savvy Canadians have moved to filing their taxes online. CRA maintains several online tax filing systems – NETFILE, TELEFILE, and EFILE – which are marketed as convenient, quick, and utterly flawless.

Well, maybe "utterly flawless" is a bit of an overstatement since the system went down for ten days in early March of 2007. Unfortunately, this was the peak filing time for Canadians, whose returns were due on the 30th of April. During this time, millions of Canadian citizens could not file their returns online; and additional millions of mailed-in returns stacked up. Refunds due to Canadian citizens were going to be seriously delayed.

How could this have happened? It seems that a vendor patch was applied Sunday, March 4, to the CRA's Advantage CA-IDMS database management system. The patch was intended to correct an obscure problem which, if encountered, would crash the database management system. CRA claimed that this patch had been thoroughly tested, but problems started to be reported shortly after the patch installation. It was noticed that irregularities were occurring during the transfer of information between the system databases that store each Canadian's personal tax information.

Unable to resolve the growing problem, the CRA was finally forced to shut down their systems late Monday night, March 5. As they investigated the problem, they determined that the erroneous patch had impacted 76 applications comprising 7.5 million lines of code that updated 91 different databases. The recovery effort was massive.

Over 700 employees worked around the clock to repair the damage. As the department reeled in crisis, the Revenue Minister was absent on a photo-op with the Prime Minister in Saskatchewan, despite the pleas for her return.

It was ten days before the system could be restored to service.

FAA Modernization Takes a Step Backwards

The traffic control systems employed by the U.S. Federal Aviation Administration (FAA) are extremely complex, and failures could be life-threatening. Therefore, once a system is put into service and has proven to work well, it stays there for a long time,

Such is the case with the computers in the nation's Air Route Traffic Control Centers (ARTCCs). These are the twenty centers that control flights that are generally longer distance flying at higher altitudes. The centers use a system called Host to consolidate the radar returns from several radar sites serving each ARTCC. The Host computers are ten years old, and critical parts are no longer available. In fact, the systems are not even Y2K compliant.

The FAA is replacing them under a broad modernization program with a new system called Host Computer Replacement System, or Hocsr (we applaud the FAA for their imaginative naming conventions). These systems, supplied by IBM, are not new; but parts are still available. In

addition, a new display subsystem is being installed and features consoles that can display more information with much higher resolution.

The first of the new Hocsr systems was installed at the Long Island ARTCC and began operation on February 24, 2007.

On Sunday, May 6, at 6:25 AM, as technicians were attempting to attach the new display subsystem, the Hoscr system crashed. However, the old Host system was still configured as a standby; and it immediately took over. Unfortunately, it, too, malfunctioned. The IT staff was able to bring a third system into service, but it was much less capable. Controllers had to double the spacing between planes from 10 miles to 20 miles. As a result, 274 flights at LaGuardia, Newark, and Philadelphia were delayed up to an hour or more, inconveniencing tens of thousands of passengers.

Henry Brown, an official with the union that represents FAA electronics technicians, claimed that the new system had not been thoroughly tested and that the FAA had rushed it into service to demonstrate the success of its modernization program.

FAA – Yet Again

Almost exactly one month later, the FAA got hit again. This time it was their flight planning computer in Atlanta.

Before a commercial flight can take off, the pilot must file a flight plan indicating his route, departure time, time enroute, altitude requested, and several other details. Flight plans are filed electronically with the Atlanta system, which then routes them to the appropriate air traffic controllers. Each day, approximately 50,000 domestic flight plans are filed in addition to those required for international flights.

About 9 AM on Friday, June 8, 2007, the Atlanta flight-plan system went down. No worries – the mirrored backup system in Salt Lake City immediately took over flight-plan processing as planned, and all flight plans were rerouted to it. However, this system was underconfigured and was quickly overwhelmed. Air traffic controllers had to input tens of thousands of flight plans manually, a very time consuming process. Since a plane could not take off until it had a valid flight plan filed, delays across the country affected tens of thousands of passengers.

The Atlanta system was recovered two hours later, at 10:52 AM; but it took another two hours for the New York systems to reconnect to the Atlanta system. The New York airports, which are in the most heavily congested airspace in the United States, experienced delays of up to four hours.

This already terrible situation was aggravated by a severe line of thunderstorms stretching from Canada to Texas.

Oh, well. If you have time to spare, go by air.

M&T Bank Snafu Delays Deposits by 24 Hours

M&T Bank is a highly regarded regional bank with 650 branches in New York, Pennsylvania, New Jersey, Delaware, Maryland, Virginia, West Virginia, and Washington, D.C.

But even the best aren't perfect. On Thursday, June 6, 2007, the bank's automated system that updates customer accounts at the close of the business day failed. Once the problem was corrected, it took until Friday afternoon at 5:30 PM to complete the postings.

The problem especially affected customers who counted on withdrawing money immediately after regularly scheduled direct deposit of paychecks into their accounts.

The bank did all they could to notify customers of the problem via their Web site, through their telephone banking services, and at their local branches. They even sent area managers to the local branches to help customers access their funds.

As a result, said M&T Bank spokesman Mike Zabel, "People are generally very understanding when you make a mistake or when you have a problem. Everybody knows what it means to have a computer glitch. It's happened to everybody."

All Nippon Airways Glitch Delays or Cancels Hundreds of Flights

All Nippon Airways (ANA) is Japan's largest domestic carrier. Its computer systems for reservations and departure control are provided by Unisys, whose equipment ANA has been using for the last 25 years.

Its system includes a host system (currently being upgraded by Unisys) that, among other tasks, connects to intermediate computers that serve check-in terminals at the airports served by ANA. In mid-May of 2007, three of the six intermediate computers were replaced with new systems. However, after this upgrade, it was suspected that there was a problem with the new computers; and the old systems were moved back into place.

It was only after this move that it was determined that the problem was not a computer problem after all but rather a problem with the communication channels between the host computer and the intermediate computers.

It looks like this problem caught up with ANA on Sunday, May 27, when data flowing between the host computer and the intermediate computers slowed to a crawl. This lead to a backlog of data that soon overwhelmed the system. The impact of the resultant delay in data delivery led to the cancellation of 130 flights and the delay by more than one hour of 306 flights. Almost 70,000 passengers were inconvenienced.

By the next day, ANA was able to return operations to normal.

And Then There Was BlackBerry

Perhaps the failure that was most obvious to the most people occurred when the BlackBerry system went down on Tuesday, April 17, 2007. We reported on this failure in detail in our May, 2007, *Never Again* article, Blackberry Gets Juiced. Millions of BlackBerry subscribers in North America were without service until Wednesday morning, when queued-up emails from the huge backlog of messages started to trickle through. It wasn't until Thursday that service was returned to normal.

BlackBerrys are handheld devices that are used to send and receive emails almost anywhere in the world. The BlackBerry service has been highly reliable and has attracted bankers, lawyers, journalists, law makers, and business people.

The BlackBerry devices and their associated services are provided by a Canadian company, Research in Motion Limited (RIM), of Waterloo, Ontario.
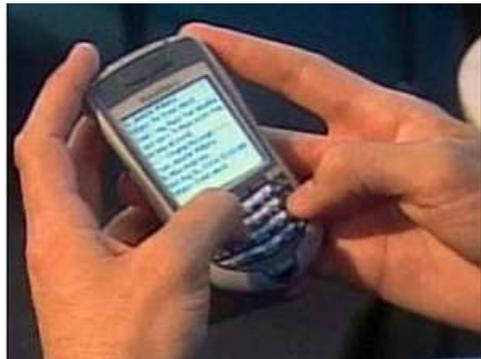
On Tuesday, April 17, at 5 pm PDT, BlackBerry messages stopped flowing to subscribers in North America. The problem was in BlackBerry's Canadian Network Operations Center (NOC), which serves North America.

However, RIM management refused to acknowledge that they had a problem (though I suspect that the computer room was in mayhem).

It wasn't until twelve hours later, on Wednesday morning, that RIM management acknowledged that they had, in fact, had an outage. But they gave no reason for it and no prognosis of what to expect.

By Wednesday morning, the system had been returned to service; but it had to deal with a huge email backlog that had accumulated during the outage. These emails started to trickle out Wednesday morning, but new emails that came in entered the end of the queue and were delayed for hours. It was not until Thursday that operations returned to normal.

Once operations normalized, RIM management began to release details of the outage. They reported that the outage was triggered by the "introduction of a new noncritical system routine" designed to optimize cache performance. They had not expected this change to affect the regular operations of BlackBerry.

However, despite previous testing, the new system routine set off an unexpected chain reaction. It triggered a series of interaction errors between the system's operational database and cache and resulted in the outage of the NOC.

After the RIM technicians isolated the database problem and tried unsuccessfully to fix it, they began the failover process to the backup system. But horror of horrors, the failover attempt failed, despite failover having been previously tested. The consequence was the extended outage that affected millions of subscribers.

The system failure was just one of the failures during this outage. The other was a near lack of communication from RIM management to its subscribers. RIM's tight-lipped response angered many. Grumbles could be heard from the White House to the Canadian Parliament.

Lessons Learned

Half of these failures (Dow Jones, BlackBerry, and the two FAA failures) involved backup systems that didn't perform properly, either because they were underconfigured or because the failover attempt failed. Failover is a complex and risky business. It is imperative that full failover be frequently tested, as expensive as it may be, to ensure that the backup system is configured properly, that procedures are up-to-date, and that personnel are properly trained. A failover fault can cost a company a whole lot more than periodic failover testing.

Five of these failures occurred just after a reconfiguration (US Airways, Canadian Revenue Agency, FAA, All Nippon Airways, and BlackBerry). Reconfigurations are an especially risky process. A new configuration must be thoroughly tested before putting it into service.

These two failure modes accounted for seven of our eight stories. The lesson to be learned may seem trite when stated, but it is simply that the first three fundamental rules for continuous availability are test, test, and test.