日経コンピュータ

判明、ANAシステム障害の真相

2016/04/12

井上 英明＝日経コンピュータ

大型のシステム障害の詳細が見えてきた。全日本空輸（ANA）が2016年3月22日に起こした国内線旅客システム「ANACore（エーエヌエーコア）」のシステム障害では全国49の空港で搭乗手続きができなくなり、ANAと提携航空会社5社の合計で719便、7万2100人以上に影響を及ぼした。インターネットや予約センターでの予約などもできなかった。

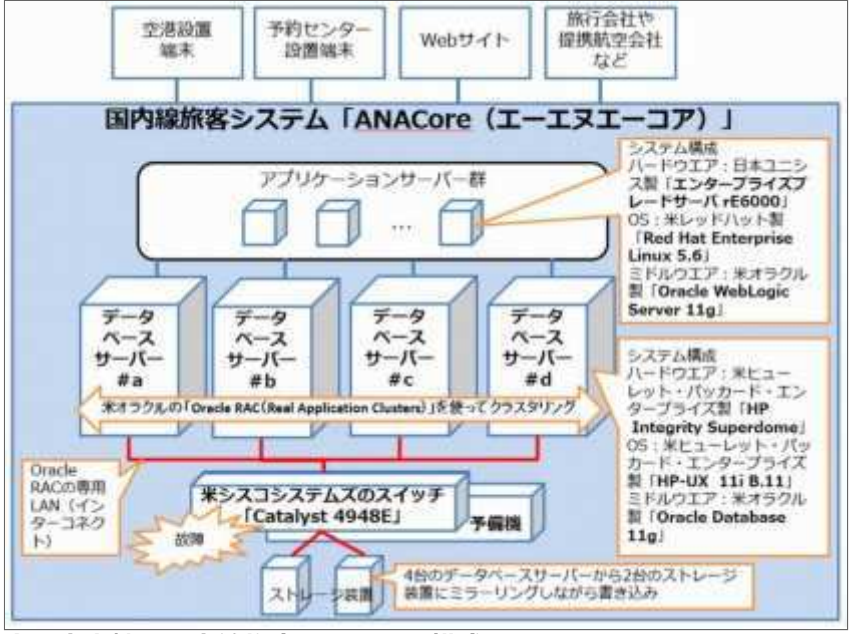
搭乗手続きなどでごった返す全日本空輸のカウンター（3月22日午前11時40分ころ、新千歳空港）
[画像のクリックで拡大表示]

ANAは障害発生から8日後の3月30日に経緯や原因を公表、さらに4月11日に弊誌のメール取材に応じ、一段詳しい真相が判明した。

4台のSuperdomeをRACでクラスタリング

今回のシステム障害の中身は3月20日のニュースで報じた通り、4台のデータベース（DB）サーバーが停止したというもの（関連記事：ANAシステム障害の原因判明、シスコ製スイッチの「世界初のバグ」でDBサーバーがダウン）。今回、弊誌の取材でシステム構成が明らかになった。

DBサーバーは米ヒューレット・パッカード・エンタープライズ（HPE）のUNIX「HP-UX 11i B.11」を搭載する「HP Integrity Superdome」を使い、データベース管理システム（DBMS）は米オラクルの「Oracle Database 11g」を使っていた。ANAが使うSuperdomeは1.66GHzのItanium2を12個と、64Gバイトのメモリーを搭載する。



全日本空輸の国内線旅客システムの構成図

全日本空輸や日本ユニシスの資料を基に編集部が作成

[画像のクリックで拡大表示]

4台のDBサーバーはオラクルの「Oracle RAC（Real Application Clusters）」を使ってクラスタリングして、可用性と性能を向上させていた。分散したDBサーバーが協調して処理を進める場合、ストレージ上のデータを共有する「シェアードエブリシング（共有ディスク、シェアードオールとも呼ぶことがある）」や、それぞれのDBサーバーにのみデータを持つ「シェアードナッシング」と呼ぶアーキテクチャーを採る。RACの場合は前者の「シェアードエブリシング」である。

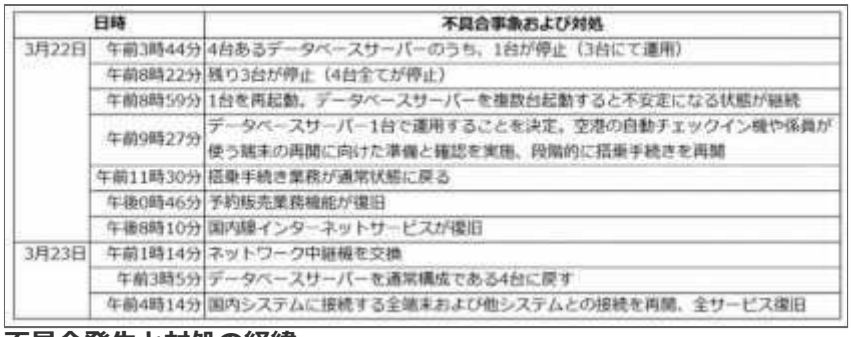
ANACoreではストレージは2台のミラー構成を使っている。4台のDBサーバーはそれぞれに同時に書き込む。この時、ストレージ上のデータが一貫性を保って参照・更新されるように、4台のDBサーバーは高速な専用ネットワーク（インターコネクト）を通して、メモリー上に展開したデータなどを転送し合う。今回、インターコネクトで使っていた米シスコのスイッチ「Catalyst 4948E」が故障し、最終的にDBサーバーの4台停止につながった。

1時間で縮退運転開始

ANAが3月20日に公表した資料と取材の回答結果、日本ユニシスがANACore稼働後に公表した技術論文集「ユニシス技法」の通巻118号「特集：エアラインリザベーション」を基に、改めてシステムダウンと復旧の経緯を時系列でみていく。なおユニシス技法の内容はANAも確認済みで、システム構成も基本的には変わっていないが一部で機器を増設しているという。

最初のDBサーバーが停止したのは3月22日の午前3時44分。ここから1台、また1台と停止し、約4時間40分後の午前8時22分には4台とも停止した。始発便はとうに出発している時間帯で、全国の空港で搭乗手続きに遅れが生じていた。最初に欠航したのは羽田空港を午前9時55分に出発する秋田空港行き403便だった。

羽田空港ではその後、欠航便が相次いだ。ANA広報は「欠航の判断については、（羽田空港など）代替交通機関を利用しやすい（空港にいる）お客様に対して早めに情報を提供し、お客様の時間ロスを最小限にするという点も考慮している」と話す。ただ欠航を判断する際の主目的は「最初は機材繰りによってダイヤの乱れが長引くのを防ぐためであり、その後は空港にお客様が滞留するのを防ぐためにやむを得ず決定する」と話す。



日時		不具合事象および対処
3月22日	午前3時44分	4台あるデータベースサーバーのうち、1台が停止（3台にて運用）
	午前8時22分	残り3台が停止（4台全てが停止）
	午前8時59分	1台を再起動。データベースサーバーを複数台起動すると不安定になる状態が継続
	午前9時27分	データベースサーバー1台で運用することを決定。空港の自動チェックイン機や係員が使う端末の再開に向けた準備と確認を実施、段階的に搭乗手続きを再開
	午前11時30分	搭乗手続き業務が通常状態に戻る
	午後0時46分	予約販売業務機能が復旧
	午後8時10分	国内線インターネットサービスが復旧
3月23日	午前1時14分	ネットワーク中継機を交換
	午前3時5分	データベースサーバーを通常構成である4台に戻す
	午前4時14分	国内システムに接続する全端末および他システムとの接続を再開、全サービス復旧

不具合発生と対処の経緯

全日本空輸の資料を基に編集部が作成

[画像のクリックで拡大表示]

DBサーバーの停止は「2パターンあって、両方とも仕様通り」とANAは取材で答えた。まず最初の3台が停止したのは「RACの管理通信がタイムアウトで異常終了した」（ANA）ためだ。データの同期処理が正常に進んでいないと判断してDBサーバーを自動停止する機能が働いた。最後の1台が停止したのは「Oracle DBを監視しており、タイムアウトが発生した」（同）ため。これもOracle DBが正常に動作していないとして自動停止機能が働いたという。

ANACoreは冗長化を徹底。さらにHPEのクラスタリングソフト「HP Serviceguard」でRACのクラスタリングを監視・構成し、日立製作所の運用管理ソフト「JP1/Integrated Management」でシステム全体の機器を監視していたようだ。今回の障害時、具体的にどのソフトでどういったアラートが出ていたかは明らかではない。

4台停止から約40分後の午前8時59分、ANAはDBサーバーを1台再起動した。だが複数台起動すると不安定になる状態が変わらなかった。そこでANAは4台停止から約1時間後の午前9時27分、DBサーバー1台での縮退運転を決めた。

ANACoreはもともと1台のDBサーバーでシステムの全機能を使える設計にしてあったという。ただし動かす機能を搭乗手続きに絞り、「ご迷惑をお掛けしているお客様への対応を最優先にした」（ANA広報）。予約や販売、Webサービス、他社連携といった各種機能は起動させなかった。

縮退運転後、自動チェックイン機や係員が使う端末が少ない小規模空港では搭乗手続き機能がすぐに復活したという。羽田空港など端末台数の多い空港でも端末の再起動を順次進めた。カウンターでの混乱は続いていたが、午前11時30分にシステム的には搭乗手続きが復旧した。

1日でシステム復旧、2日で再発防止

縮退運転後、ANAは原因の特定を急いだ。監視システムのログなどからDBサーバー、アプリケーションサーバーと順に障害を疑い、異常がないと判断した。残ったのがインターコネクトのスイッチ「Catalyst 4948E」だった。「本番環境と同等の作りにしてあるテスト環境にスイッチを持ち込んでテストしたところ、不具合が再現した」（ANA広報）。

スイッチも冗長構成を採っていた。本来は「スイッチが故障すると『故障シグナル』を発信し、予備機に自動的に切り替わる設計だった」（ANA）。だが、今回は故障しているにも関わらず、故障シグナルを発信しなかった。故障シグナルとはANAによれば「SNMP(Simple Network Management Protocol)によるメッセージ通知」という。これを運用監視システムで受け取っていた。

故障内容は厄介だった。「完全に停止したわけでなく、動作が不安定になった」（ANA広報）という“半死”の状態だったのだ。稼働開始から約3年、スイッチが故障により自動的に切り替わったことは一度もないという。

スイッチの故障が分かった時点でANAはすぐにシスコに連絡、代替機を取り寄せた。故障機と予備機、代替機は「同一型番、同一ファームウエア」（ANA）だったという。代替機を取り寄せた理由をANAは「念のためスイッチの健全性を確認するため」と説明する。予備機はオンライン状態で稼働しており、「事前（の健全性の）確認ができない状況だった」（ANA）。

午後0時46分には予約発券業務を、午後8時10分にはWeb予約やWebサービスを復旧させつつ、並行して代替機の健全性を確認し、翌3月23日午前1時14分に故障機と代替機を交換。午前3時5分にはDBサーバーを4台構成に戻し、午前4時14分には他社接続など全サービスを復旧した。

障害検知から全復旧まで24時間30分で済ませただけでなく、その翌日3月24日には再発防止策を打つ。「スイッチが故障シグナルを出さない場合でもDBサーバーからスイッチ故障を検知できるよう改善した」（ANA）。

1年に及ぶ製品のバグ出しテストをすり抜ける

ANACoreで使っていたCatalyst 4948Eはなぜ「故障シグナル」を発信しなかったのか。ANA広報によれば4月11日時点でもシスコで検証中という。「世界初の事象であり、機器固有の問題である可能性が高いという報告を受けている」と明かす。同スイッチは2010年6月の発売開始以降、世界で4万3000台、うち日本で8700台を販売しているという。

今回の障害は2013年2月にANACoreを稼働して以来、初めての大きなトラブル。ANACoreの開発ベンダーは日本ユニシスである。ANAは国内旅客システムを、1978年稼働の「RESANA」、1988年稼働の「able-D」と、米ユニシスのメインフレーム上でFortranで構築したシステムで稼働させ、日本ユニシスが構築を担当してきた。ANACoreの構築プロジェクトが始まったのは10年前、2006年4月のこと。「オープンシステムプラットフォームの環境でメインフレームと同等のサービスレベルを実現すること」（日本ユニシス）をゴールとした。

ANACoreのプロジェクトが始まった翌年の2007年と翌々年の2008年、大規模なシステム障害が起こる。2007年5月には約7万9300人に、2008年9月には約6万8000人に影響が及んだ。2007年5月に発生した大規模なシステム障害時もシスコのスイッチ不具合が原因だった（関連記事：【会見詳報】ANA障害の原因判明、「世界4例のスイッチ故障がきっかけ、対応も遅れた」）。

本来のゴールと発生した障害を踏まえ、ANAと日本ユニシスはANACore構築に当たり、製品に潜む不具合のたたき出しに注力していた。インフラ部分の製品テストを1年にわたって実施し、複数製品から30個以上の潜在的な不具合を発見したという。ANAによればこの製品テスト時には今回故障したCatalyst 4948Eを使っており、「スイッチは15項目にわたってテストした」という。さらにCatalyst 4948Eの保守サポートは2018年に終わることもあり、既に機器の更新計画も立てていた。

実はCatalyst 4948Eは当初想定の機器では無かった。設計時はCatalyst 4948Eと同じく1000Mbpsの処理性能を持つ下位機のCatalyst 2960を使う予定だった。日本ユニシスはベンチマークでインターコネクトのトラフィックが最大で数百Mbpsになると分かったため、これを最大100Mbpsに抑えるよう、便名や操作端末などによって処理するDBサーバーを事前に指定する工夫を施していた。だが、事前テストでDBサーバーの起動時に遅延する事象が見られたという。

そこでCatalyst 2960に加え、Catalyst 3750とCatalyst 4948EでDBサーバーの台数を増やしながら性能テストした結果、Catalyst 2960はDBサーバーが3台以上になるとインターコネクトで使うUDPパケットの処理能力が極端に低下することが分かった。これによりANACoreで使うスイッチをCatalyst 4948Eに決めた。「単位時間のパケット処理能力はメーカーが公表していない。機器選定の検証段階で確認する重要性が分かった」（日本ユニシス）。

ANAは「よくやった」のか

ANAホールディングスの片野坂真哉社長は2016年4月1日、ANAグループの入社式でこう話した。「全ての関係する役職員が全力で

対応と復旧にあたりましたが、多くのお客様にご迷惑をおかけし、厳しいお叱りをたくさん頂戴しました。原因を究明し、再発防止策をとりましたが、お客様の揺らいだ信頼を回復するため、引き続き全力を挙げていきます」――。片野氏は今回のシステムトラブルで1カ月20%の報酬を自主返上している。

今回のトラブルでANAは「3億6000万円の逸失収入が発生した」（ANA広報）。日本ユニシスに対し、損害賠償請求を検討している（関連記事：ANA、システム障害で日本ユニシスへの損害賠償検討）。ANACoreの瑕疵担保責任期間は「稼働後1年であり、既に期間は過ぎている」とした上で、ANA広報は4月11日時点で「損害賠償の根拠は日本ユニシスとの契約に基づくものであり、結論を出す時期も含めて現在検討中」と話す。

3月20日にANAが障害原因を公表したニュースには多くの反響があった。記者には「ANAの障害対応は称賛に値する」という識者からのメールが届き、ニュースに対するソーシャルメディアの反応を見ても障害の原因究明の早さや復旧までの早さに驚き、称賛する声が多かったように思えた。

スイッチの「世界初のバグ」を“踏み抜いた”ANAの不運に同情する声や、手作業で搭乗券を発行できる訓練を積んでいるというBCP（事業継続計画）の出来の良さを褒める声もあった。「年1回のeラーニングや着任時の座学などを通して、全空港の旅客係員全員がシステムを使わずに対応する訓練を最低1回は受講することを義務付けている」（ANA広報）。

記者も障害当日に取材しながら復旧の早さに驚き、原因公表が早かったことにも驚いた。「ANACoreのプロジェクトはコスト面で決して順風満帆ではなかった」。記者は過去に日本ユニシス幹部に聞いたことがあるものの、現場ではミッションクリティカルなシステムを運営する責任をステークホルダーが十分認識し、かつ過去の障害を踏まえて、障害対応手順を十分整備していたことがうかがえた。

一方で、「高信頼システムとしては仕組みが足りない」と指摘するアーキテクトもいた。日本有数のミッションクリティカルシステムをいくつも手掛けてきたこのアーキテクトは「ネットワーク機器の間欠故障は確かに厄介で頭が痛い」と認めつつ、「大規模システムであれば何度か経験する問題であり、高信頼性を追求するのであれば、複数手段での検知や切り替え手段、場合によっては手動での切り替え手順を持つべきだ」とした。

「ミッションクリティカルであれば製品の潜在バグを見つけるテストを当然実施すべきだし、いくら製品を“叩い”ても、『故障シグナル』の機能だけに死活監視を依存する限り、その機能自体がSPOF（Single Point of Failure：単一障害点）になる」。今回、DBサーバーからの監視を加えた再発防止策は、複数経路での監視に当たる

とこのアーキテクトは話す。間欠障害の検知には、業務部門の利用者と同じ経路、同じ操作でシステムの稼働状況を常時監視するような仕組みも有効と指摘している。

　障害対策・障害復旧でANAはよくやったのかそうでないのか。どの程度のコストを掛けて、どの程度の信頼性を、どういったアーキテクチャーで実現するのか。同じケースは一つとしてないが、自分の現場だったらどう振る舞えるのか。読者の皆さんはどう考えるだろうか。