FUJ00176513

From: Ashford, Edward[/O=EXCHANGELABS/OU=EXCHANGE ADMINISTRATIVE GROUP (FYDIBOHF23SPDLT)/CN=RECIPIENTS/CN=D359CE8E14A94C9BA3EB1066E6C67C41-ASHFORD, ED]

Sent: Fri 03/12/2021 1:52:15 PM (UTC)

To: Barnes, Gerald GRO ; Browell, Steven GRO ; Gauntlett, Paul GRO ; Boardman, Phil GRO

Cc: Mistry, Manisha GRO ; Gibson, Andrew R GRO ; Wilson, Simon GRO

Subject: RE: Historical Issue with Audit Data

Can the network link to the Eternus be made faster

- essentially no, although we might be able to group up a couple of interfaces now POLSAP is no longer there, but it's only getting us twice the speed

From: Barnes, Gerald GRO
Sent: Friday, December 3, 2021 1:01 PM
To: Ashford, Edward GRO ; Browell, Steven GRO ; Gauntlett, Paul GRO ; Boardman, Phil GRO
Cc: Mistry, Manisha GRO ; Gibson, Andrew R GRO ; Wilson, Simon GRO
Subject: RE: Historical Issue with Audit Data

Hi Ed,

Thank you.

Answered in line.

Regards,
Gerald Barnes

From: Ashford, Edward GRO
Sent: Friday, December 3, 2021 8:20 AM
To: Barnes, Gerald GRO ; Browell, Steven GRO ; Gauntlett, Paul GRO ; Boardman, Phil GRO
Cc: Mistry, Manisha GRO ; Gibson, Andrew R GRO ; Wilson, Simon GRO
Subject: RE: Historical Issue with Audit Data

Hi Gerald,

It's a rather resilient cluster of linux servers presenting storage, but that is after all what it is, a large NAS server. We wouldn't install anything on it, what's inside the cabinet is considered as an appliance supplied by TPG.

We could present the shares to a backup server, and BSW could run SQL-Server, but really for the sake of appearances I think we should continue to only permit access to the NAS shares by the ARC servers and nothing else, even though you and I know that the seals are all the tamper evidence needed. We can readily increase the spec or ARC, but that's not so easy with BSW which is rather long in the tooth.

That would be useful. Can the network link to the Eternus be made faster?

FUJ00176513

If we do find a gap at one side is it possible to arrange for that ARC to ship the track to the other ARC for sealing and storage there so we can get to the point where we are confident that both sites have a copy of all the data?

Yes that would make sense. I know of one GAP. I would just store the files on the MiscArcs shares specially developed for this purpose.

It makes some kind of sense given we are approaching a migration to validate the data for HNG-X gaps. The last thing we need is a protracted process after migration if gaps are found but actually they are genuine gaps.

I guess the concern is that ultimately we are looking to deploy a single instance in Post Office Cloud and it needs some kind of tick to say "it's all there".

Because the two data stores are different we will probably import each separately and end up with two servers in AWS for tactical. For strategic we are just going to have one.

If we install a Snowball on site in Belfast does it need to copy 11PROD, 19PROD or both?

We would copy both Eternuses separately. Whether or not we have one or two snowballs I do not know. One snowball can have two buckets.

Me I would have a complete copy in each of the major cloud providers, but luckily it's not my problem.

Regards

Ed

From: Barnes, Gerald <GRO>
Sent: Thursday, December 2, 2021 11:04 PM
To: Browell, Steven GRO; Gauntlett, Paul GRO; Boardman, Phil GRO; Ashford, Edward GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Hi Steve,

I think I need to explain the architecture of audit.

All files are stored on an Eternus. This is a specialized computer to store large numbers of files. They are all stored on it compressed. It is not designed to have applications run on it. I doubt you can even uncompress files there (though I will copy Ed on this in case it is more sophisticated than I describe). The way things are done is that when there is an ARQ request the first thing that happens is that the relevant files are copied across the network to the audit server where they can be manipulated. We do not have a “data lake”. Therefore doing detailed analysis of every file stored is not something the system is designed for.

The program for making sure that the original files retrieved are the same as when stored, transactions are not corrupt, have no gaps and duplicates in the transaction run already exists. It is fully automatic. Every time an ARQ is run these things are done. No spreadsheets are submitted in evidence unless these checks pass (apart from duplicates where a list of duplicates is supplied and there is a clear warning on the spreadsheet).

The trouble with what you ask is you would have to run an ARQ for each FAD code and for all months.

FUJ00176513

Normally an ARQ is 1 month. There are 23,674 FADs. The earliest file is 10/12/2007. I am not sure whether you are worried about GAPs in the old Riposte software or the new HNGx software but if I were you I would be more worried about the latter since they only have one harvester whereas for Riposte there were two totally independent systems and therefore you are most unlikely to have any GAPs (in the totality of things since you can check both systems). So you will need to run about 4,000,000 ARQs.

And what would we do if we found a GAP? For Horizon there would not really be a GAP because you will always find the other server works fine. For HNGx you would just be completely stuck since only one copy of the file is made and that copy is duplicated.

Now I know about these legal cases but there is no fault at all in the audit software. If there is a fault in the accounting data it has been accurately preserved by audit. It is not the fault of audit that it correctly gathered a file with invalid transaction data in it due to some programming bug outside of the domain of audit.

There have been some PEAKs raised where gaps (very very rare) and duplicates (very rare) mainly if not exclusively with the Horizon data. They have all been looked at and none show a fault with the audit system. They all indicated a harvester bug. A very detailed analysis was done of the duplicates and they were found to be exactly that – complete and utter duplicates. The harvester had saved the transactions twice in some cases for reasons best know by the people who used to look after the harvester.

Hi Ed,

Have I maligned the Eternus in saying it is a computer designed for storing stuff and not much else? Would it be possible to unzip files directly on the Eternus and analyse them there and hence avoid the need to copy them all across the network to the audit server to analyse them?

The trouble though even if possible the meta data of the files is in a SQL table on the audit server. You would need to import that table to the Eternus to do sensible things like check that each files checksum was as when it was originally stored. This table can only sensibly manipulated by SQL because of its indices. Does SQL Server run on the Eternus?

Regards,
Gerald Barnes

From: Browell, Steven GRO
Sent: Thursday, December 2, 2021 4:20 PM
To: Barnes, Gerald GRO; Gauntlett, Paul GRO; Boardman, Phil GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Thanks again.

The 'cost' is man time to create the scripting and run it – then compare the output. The POL charging model won't apply.

Is it technically possible? How would we do it? How much actual hands on effort do we think it will need?

I believe we have got to do this. So we need to decide how – and do it quickly.

FUJ00176513

Steve Browell
Mob: GRO

From: Barnes, Gerald GRO
Sent: Thursday, December 2, 2021 4:17 PM
To: Browell, Steven GRO; Gauntlett, Paul GRO; Boardman, Phil GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Hi Steve,

Answered in line as there were too many questions.

Regards,
Gerald Barnes

From: Browell, Steven GRO
Sent: Thursday, December 2, 2021 3:29 PM
To: Barnes, Gerald GRO Gauntlett, Paul GRO; Boardman, Phil GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Thanks Gerald,

I will remove the bullet point re FADs using specific DCs.

As for the p.s. I am afraid I am not sure what that means and how it affects what we are describing. If you are simply enlightening me as to how it was configured – noting there were some differences to how other things were done – then thank you. However, if you are saying that the way it was configured was wrong then that warrants more checking.

I am saying that there is an option "Transfer Data" which automatically copies files gathered to the other audit server each evening. It was a deliberate design policy. When they know that there were two copies of things anyway by other mechanisms it was switched off.

Also, how can we pro-actively confirm no gaps in the combined audit archives for pre-HNG-X - instead of just waiting for ARQ anomalies to pick up on something? Can we write a script/program to hunt for any possible issues?

The audit system is not designed for this. The actual files are stored on the Eternus zipped. The normal way of getting them out is an ARQ which brings them back to the audit server, unzips them and checks for GAPS. 103 ARQs (a month each for one FAD) costs **£17,913.76 . By the normal process you would be talking tens of millions of pounds.**

Steve Browell
Mob: GRO

From: Barnes, Gerald GRO

FUJ00176513

Sent: Thursday, December 2, 2021 3:19 PM
To: Browell, Steven GRO Gauntlett, Paul GRO Boardman, Phil GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Hi Steve,

Apologies for the delay in replying. Too much going on including one of our cars in the garage!

- By design, some FADs used Bootle as their primary data centre, and others used Wigan. Some used both

This is not right,

My understanding is that Wigan and Bootle are equal.

There were separate harvesters on each.

Any transaction is saved both in a Wigan file and a Bootle file. Both audit servers were active. The Wigan audit server saved the Wigan file and the Bootle audit server saved the Bootle file.

Regards,
Gerald Barnes

p.s. Both audit servers by a configuration change could in principal have copied each file one to other but my guess is that the designers thought that you would end up with 4 copies of each transaction which was overkill. From the notes in DEV/INF/ION/0001 it looks like most other files other than these transaction files were copied one to other. I attach an earlier copy of DEV/INF/ION/0001. You will see most of the time in the table at the end "Transfer Data" is "YES" apart from the below

55	Cluster1B	8	TMS	8	TMSA	TMS transactions gather from Correspondence Server Cluster 1	\\MBOCOR01\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE
56	Cluster2B	8	TMS	7	TMSB	TMS transactions gather from Correspondence Server Cluster 2	\\MBOCOR02\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE
57	Cluster3B	8	TMS	8	TMSA	TMS transactions gather from Correspondence Server Cluster 3	\\MBOCOR03\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE
58	Cluster4B	8	TMS	7	TMSB	TMS transactions gather from Correspondence Server Cluster 4	\\MBOCOR04\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE
59	Cluster1W	8	TMS	8	TMSA	TMS transactions gather from Correspondence Server Cluster 1	\\MWICOR01\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE
60	Cluster2W	8	TMS	7	TMSB	TMS transactions gather from Correspondence Server Cluster 2	\\MWICOR02\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE

FUJ00176513

61	Cluster3W	8	TMS	8	TMSA	TMS transactions gather from Correspondence Server Cluster 3	\\MWICOR03\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE
62	Cluster4W	8	TMS	7	TMSB	TMS transactions gather from Correspondence Server Cluster 4	\\MWICOR04\D$\Audit\Done	HOARD+4320	84	0	1	tmp	NO	NONE
371	Cluster5B	8	TMS	14	DUMMY	Unused	\\MBOCOR05\D$\Audit\Done	HOARD+1440	84	0	1	tmp	NO	NONE
372	Cluster6B	8	TMS	14	DUMMY	Unused	\\MBOCOR06\D$\Audit\Done	HOARD+1440	84	0	1	tmp	NO	NONE
373	Cluster7B	8	TMS	14	DUMMY	Unused	\\MBOCOR07\D$\Audit\Done	HOARD+1440	84	0	1	tmp	NO	NONE
374	Cluster8B	8	TMS	14	DUMMY	Unused	\\MBOCOR08\D$\Audit\Done	HOARD+1440	84	0	1	tmp	NO	NONE

From: Browell, Steven GRO
Sent: Thursday, December 2, 2021 1:34 PM
To: Gauntlett, Paul GRO; Barnes, Gerald GRO; Boardman, Phil GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

I attach my amended notes for your critique please before I pass to various parties for review and permission to release.

Steve Browell
Mob: GRO

From: Gauntlett, Paul GRO
Sent: Wednesday, December 1, 2021 5:01 PM
To: Barnes, Gerald GRO; Boardman, Phil GRO; Browell, Steven GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Hi All,

Please see updated statement below.

- Accepted the changes that Phil added directly
- Orange updates based on comments below from Phil/Gerald.

Steve/Gerald/Phil – please confirm if happy with the below. I'm guessing there may be some further tweaks needed so will setup a short call for tomorrow am so we can finalise - I can cancel if not needed!

Problem Summary:
Post Office counter audit transaction files gathered prior to the HNGx software rewrite in 2010 are not consistent across both IRE11 & IRE19 servers. Holistically no data is missing but in rare instances transaction data exists only on one server or the other. The problem was caused by the occasional malfunction of the Harvester process of the previous Horizon system - Riposte.

FUJ00176513

Background:

Audit Gathering Method in Horizon up to approx 2010

- Originally a system called Riposte was used, as part of Horizon, to gather audit transactions from all the Post Office counters.
- Riposte was a big distributed database.
- At that time Riposte was deployed into Bootle & Wigan data centres.
- Each evening all Post Office counter transactions were harvested and stored in files.
- By design transactions were either harvested to Wigan or Bootle or both data centres.
- Different files were produced in each data centre. The files were different because although the same transactions were harvested they were processed in a different order and for different FADs. Filenames were also different.
- The "CopyData" option (robocopy) was not set for transaction files which were gathered to both data centres.
- It is now known that due to the occasional malfunction of the Harvester process, for those transactions harvested to both datacentres, data may be missing from the files gathered in one or other data centre.

  - There is no recorded instance of both harvesters failing at the same time so although data may be missing from one server it will be present in the other.
  - NOTE: The data gathered using this method, in accordance with the contract, should have been deleted by now. It is only being retained currently, beyond Fujitsu Services' contracted obligations, at Post Office's request (below text from CWO0395b)
    Post Office have requested under RTQSR0003106 (previously RTQSR0002349 and RTQSR0002456) that Fujitsu Services continue to preserve data (including Post Office Personal Data) generated on the HNG-X System as per the previous work orders (CT2616a & CW0251a) until 30 April 2022.

Audit Gathering Method for HNG-X from 2010 onwards

- Around 2010 the contract came up for renewal.
- Fujitsu's proposal was a rewrite called HNGx which eliminated Riposte (for which there was a big annual licence fee) and to migrate the datacentres from Wigan and Bootle to IRE11 and IRE19.
- As a part of this rewrite each Post Office counter transaction was only processed into a one file on a single server.
- This drove the decision to change the Audit gathering approach.
- From this point forward all transaction and non-transaction files were gathered by the IRE11 Audit Server only and robocopied to the IRE19 Audit Server.
- Therefore from 2010 all audit files (transaction and non-transaction) are consistent across both audit servers.

Impact on ARQ requests

- When ARQ requests for a FAD code are made all relevant files are retrieved from the target audit server .
- Then the files are processed by the Query Manager service on the audit server.
- Each Horizon or HNGx transaction has a unique number associated with it. The Query Manager checks that the transactions
  - have not been tampered with
  - that there are no duplicates or gaps in the sequence of transactions.
- Due to the Riposte harvester issue, for ARQ queries from 2010 or earlier, there may be gaps in the results from one or other of the servers.
- If this occurs then a Peak is raised and the ARQ request is rerun on the other server. This resolves any issues with data gaps.
- There have been no reported instances of unsuccessful ARQ request once a query has been executed on the second server.

FUJ00176513

From: Barnes, Gerald GRO
Sent: Wednesday, December 1, 2021 1:29 PM
To: Boardman, Phil GRO; Gauntlett, Paul GRO; Browell, Steven GRO
Cc: Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Hi Phil,

Answered inline.

Regards,
Gerald Barnes

From: Boardman, Phil GRO
Sent: Wednesday, December 1, 2021 12:45 PM
To: Gauntlett, Paul GRO; Browell, Steven GRO
Cc: Barnes, Gerald GRO; Mistry, Manisha GRO
Subject: RE: Historical Issue with Audit Data

Hi Paul

I think these two butllet-points (in the pre-2010 section) are liable to cause confusion/consternation …

- The Bootle audit server gathered Bootle transaction files and the Wigan audit server gathered Wigan transactions files.
- Although there was an option to robocopy files from one server it was switched off because there were already two copies of each transaction (one in each data centre).

… as I understand it (from Gerald's explanation (and I may have mis-understood)), by design transactions were either harvested to Wigan or Bootle or BOTH … and the robocopy was ONLY turned off for those transactions configured to be copied to both datacentres … I think it's important that we show that this was not designed to leave single copies of data anywhere. Please correct me if I'm wrong there Gerald.
My understanding (from reading historic copies of DEV/INF/ION/0001 and from practical experience) is that the main case of the "CopyData" option (the robocopy) not being set was these transaction files. The logic behind that I guess (I was not the designer) is that two copies of each transaction were made anyway – one on Bootle and one Wigan. If we need more detail on this I will need to get out from Dimensions historic copies of the configuration file to examine.

If I'm not wrong, then this statement "due to the occasional malfunction of the Harvester process in one or other of the data centres transactions may be missing from the files gathered in that data centre" needs to be clarified that that's only true for those transactions configured to be harvested to BOTH datacentres.

That is right. Pretty certain that was all Horizon (pre 2010) transaction data.

I've also proposed some changes to the text inline below (in this colour), trying to make it more clear that this was a change instigated by the Horizon to HNG-X change.

I think we also need to consider how we should include the details that the data in question is only being retained currently, beyond Fujitsu Services' contracted obligations, at Post Office's request (below text from CWO0395b) …

> Post Office have requested under RTQSR0003106 (previously RTQSR0002349 and RTQSR0002456) that Fujitsu Services continue to preserve data (including Post Office Personal Data) generated on the HNG-X System as per the previous work orders (CT2616a & CW0251a) until 30 April 2022.

FUJ00176513

… and (in accordance to our contract) should have been deleted, by now.

Yes that is right – normally the files would have been deleted long ago!

Regards, PhilB

From: Gauntlett, Paul [GRO]
Sent: Wednesday, December 1, 2021 11:51 AM
To: Browell, Steven [GRO]
Cc: Barnes, Gerald [GRO]; Mistry, Manisha [GRO]; Boardman, Phil [GRO]
Subject: Historical Issue with Audit Data

Hi Steve

I am Migration Lead for the Audit Migration to AWS.
Myself and Gerald Barnes (Audit SME) are currently working with POL to define requirements for the migration of historical Audit data
An historical issue has been identified and is detailed below.
This issue was discussed yesterday with John Nelis the POL PM & also Dean Bessell who I understand is your POL opposite.
POL requested that we write up what was communicated in the meeting so they can take it to their legal team ahead of making a decision regarding the scope of the data to be migrated.
On that basis I don't wish to send anything over without it being reviewed and agreed by relevant parties. Happy to have a call to discuss further.

Problem Summary:
Post Office counter audit transaction files gathered prior to the HNGx software rewrite in 2010 are not consistent across both IRE11 & IRE19 servers. Holistically no data is missing but in rare instances transaction data exists only on one server or the other. The problem was caused by the occasional malfunction of the Harvester process of the previous Horizon system - Riposte.

Background:

Audit Gathering Method in Horizon up to approx 2010

- Originally a system called Riposte was used, as part of Horizon, to gather audit transactions from all the Post Office counters.
- Riposte was a big distributed database.
- At that time Riposte was deployed into Bootle & Wigan data centres.
- Each evening all Post Office counter transactions were harvested and stored in files.
- The Bootle audit server gathered Bootle transaction files and the Wigan audit server gathered Wigan transactions files.
- Different files were produced in each data centre. The files were different because although the same transactions were harvested they were processed in a different order and for different FADs. Filenames were also different.
- Although there was an option to robocopy files from one server it was switched off because there were already two copies of each transaction (one in each data centre).
- It is now known that due to the occasional malfunction of the Harvester process in one or other of the data centres transactions may be missing from the files gathered in that data centre.
- There is no recorded instance of both harvesters failing at the same time so although data may be missing from one server it will be present in the other.

FUJ00176513

- TBC - In a DR situation data could have been lost was this captured as an operational risk and communicated to POL?

Audit Gathering Method for HNG-X from 2010 onwards

- Around 2010 the contract came up for renewal.
- Fujitsu's proposal was a rewrite called HNGx which eliminated Riposte (for which there was a big annual licence fee) and to migrate the datacentres from Wigan and Bootle to IRE11 and IRE19.
- As a part of this rewrite each Post Office counter transaction was only processed into a one file on a single server.
- This drove the decision to change the Audit gathering approach.
- From this point forward all transaction and non-transaction files were gathered by the IRE11 Audit Server only and robocopied to the IRE19 Audit Server.
- Therefore from 2010 all audit files (transaction and non-transaction) are consistent across both audit servers.

Impact on ARQ requests

- When ARQ requests for a FAD code are made all relevant files are retrieved from the target audit server .
- Then the files are processed by the Query Manager service on the audit server.
- Each Horizon or HNGx transaction has a unique number associated with it. The Query Manager checks that the transactions
    - have not been tampered with
    - that there are no duplicates or gaps in the sequence of transactions.

- Due to the Riposte harvester issue, for ARQ queries from 2010 or earlier, there may be gaps in the results from one or other of the servers.
- If this occurs then a Peak is raised and the ARQ request is rerun on the other server. This resolves any issues with data gaps.
- There have been no reported instances of unsuccessful ARQ request once a query has been executed on the second server.

Regards,

Paul Gauntlett
Customer Solution Architect
Cloud Transformation & Development - AMCS

Fujitsu
Central Park, Northampton Road, Manchester, M40 5BP
United Kingdom
GRO
www.fujitsu.com/uk