



Reducing the Data Warehouse overnight build times (time it takes for automated processes to gather data from source systems)

SITE	PPS LEADER	TEAM MEMBERS	DATE	Signature	Approval
Leeds	Peter Watkins	Ben Arthur	21-Jul-21		

Step 1 PROBLEM CLARIFICATION

1.0 Background

Long pieces of code, collectively called a batch, run each night extracting data from source systems (SAP etc) and transforming and loading it into the Data Warehouse, ready to be used in reports used by the business.

This batch must complete outside of business hours to ensure reports are available for colleagues around the business with the most up to date information when they are needed.

Business hours are no longer 9-5 as people are working more flexibly and often starting early and/or finishing later - this reduces our time to update the data used in the reports by an estimated 3 hours per day.

As new reports are required, more data must be collected from more sources - therefore there must be space/time for the batch to grow.

The servers must be kept secure and require maintenance for software patches etc - there must be some 'spare time' for this to take place.

Optimising the code is necessary to reduce the batch run time and ensure that service expectations are met and planned outages are possible.

End users should never see this issue as reports always available when required.

1.1 Gap Analysis -

1.2 Problem Statement

Ultimate Goal: by 1st September 2022
Average nightly run times should not exceed 6 hours

Ideal Situation: by 31st December 2021
Average nightly batch run times should not exceed 8 hours

Current Situation:
Average nightly batch run times are around 10 hours

Problem Statement:
The GAP between the Current and Ideal situation is 2 hours per night

Step 2 CONTAINMENT

Containment Action (5W, 2H):

What: Create automated time stamp log which will populate when key components of the overnight batch complete

When: Start work immediately and have in place by end of March

Where: Add code for timestamps to each script

Who: BI Team (Ben Arthur, Chris Kehoe)

Why: Creates a picture of the overnight processes and to enable us to track various jobs and highlight opportunities for largest impact in reducing batch time

How: By tracking each individual part, we will see longest run times which could have the biggest impact on the GAP

How much: Longest running components will be investigated first to find root cause for time taken

Step 3 PROBLEM ANALYSIS & BREAKDOWN

1. Overnight DWH Batch Run Times

2. Daily run time breakdown

3. CRTBI Breakdown

4. CRTBI Data Population run times

5. Process Flow - Point of Cause of overnight process for Projects data in CRTBI

6. Direct Cause Investigation

Summary of Problem to Pursue: CRTBI Project Script takes 60 minutes to run at 1am - This accounts for 1hr or 50% of the GAP!

Step 4 TARGET SETTING

Target Statement

"Reduce the overnight run time of the CRTBI script component by 1hr by 30th September 2021."

Average run time of CRTBI Projects component = 60 minutes

The Projects component has been replaced so is duplicate work, but there are dependencies which need addressing before this can be switched off

Switching off the Project component should result in a reduction of the overall DWH population batch script time by 1hour or 50% of the 2hr Gap

Step 5 ROOT CAUSE ANALYSIS

Problem to Pursue: CRTBI Project Script takes 60 minutes to run at 1am

Direct Cause(s)

- Why? Data is being collected twice (in CRTBI component and DWH component)
- Why? Although replaced in DWH, the CRTBI component has not been switched off
- Why? Dependencies not known
- Why? No documentation or standards were in place for use of legacy BI databases

Check

- Therefore Data collected by Project script is still used (to feed the Project Master App)
- Therefore The app (used by the Programmes Team) is unable to talk to SAP
- Therefore There is no interface to SAP
- Therefore An interface would incur cost (CGI)
- Therefore BI database was used instead of creating a proper interface
- Therefore An alternative data feed for the Project Master App has not been built
- Therefore The Trust currently has no standard Data Integration practice
- Therefore No ownership for Data Integration

Step 6 DEVELOP & PLAN COUNTERMEASURES

Root Cause	Countermeasure Actions	Countermeasure Evaluation						Impact on Gap	Who	Schedule													
		Quality	Cost	Time	Risk	Effect	O/All			August			September			October			Nov				
1. No documentation or standards were in place for use of legacy BI databases	Map Project App dependencies on legacy databases	3	3	2	3	3	14	0	BA														
	Document dependencies	3	3	1	2	2	11	0	BA														
	Standardised process already developed and implemented	3	3	3	3	3	15	0	BA														
	Investigate/implement alternative solution to dependencies	2	2	1	1	3	9	0	BA														
2. No ownership for Data Integration	Turn off CRTBI Project script	3	3	3	3	3	15	1hr	BA														
	Define data integration standards	3	2	1	3	3	12	0	BA	Project roadmap being developed by September 2021													
	Define data integration owners	3	2	1	3	3	12	0	BA														

Countermeasure Evaluation Key: 3=Good, 2=So, so, 1=Poor. Highest total is best overall (O/All). Schedule Key: Plan, Actual, Delay

Step 7 CHECK RESULTS (MONITOR) - STATUS

1. Impact on overnight times when CRTBI Project script turned off

2. Overnight DWH Batch Run Times - Impact on GAP when CRTBI Project Script turned off

Benefit Summary

30-Sep-21 Date

1 hour per night Confirmed Benefit

3 hours per night Projected £ Benefit (Annualised)

3 hours per night Look Across Benefit

Step 8 STANDARDISE & SHARE

STANDARDISE	SHARE																														
<table border="1"> <thead> <tr> <th>Action</th> <th>Who and When</th> <th>Status</th> </tr> </thead> <tbody> <tr> <td>Document knowledge and best practice standard to use of BI databases</td> <td>BA by Aug</td> <td></td> </tr> <tr> <td>Highlight problems with mis-use of the BI databases to Data and Integration Project team</td> <td>BA by Sep</td> <td></td> </tr> <tr> <td>Restrict access to only allow BI team to make changes to BI databases</td> <td>BA by Sep</td> <td></td> </tr> <tr> <td>Identify, document and plan approach to dealing with all other 'hidden' dependencies falling outside of new standard</td> <td>BA by Jan</td> <td></td> </tr> <tr> <td>Data and Integration Project to define future data integration standards (confirm date once project roadmap defined Sep 2021)</td> <td>BA tbc</td> <td></td> </tr> </tbody> </table>	Action	Who and When	Status	Document knowledge and best practice standard to use of BI databases	BA by Aug		Highlight problems with mis-use of the BI databases to Data and Integration Project team	BA by Sep		Restrict access to only allow BI team to make changes to BI databases	BA by Sep		Identify, document and plan approach to dealing with all other 'hidden' dependencies falling outside of new standard	BA by Jan		Data and Integration Project to define future data integration standards (confirm date once project roadmap defined Sep 2021)	BA tbc		<table border="1"> <thead> <tr> <th>Action</th> <th>Who and When</th> <th>Status</th> </tr> </thead> <tbody> <tr> <td>Share knowledge and best practice with BI team members</td> <td>BA by Aug</td> <td></td> </tr> <tr> <td>Contact potentially impacted colleagues/business areas to share plan</td> <td>BA by Jan</td> <td></td> </tr> <tr> <td>Share output of Data and Integration project solution with impacted business areas</td> <td>BA tbc</td> <td></td> </tr> </tbody> </table>	Action	Who and When	Status	Share knowledge and best practice with BI team members	BA by Aug		Contact potentially impacted colleagues/business areas to share plan	BA by Jan		Share output of Data and Integration project solution with impacted business areas	BA tbc	
Action	Who and When	Status																													
Document knowledge and best practice standard to use of BI databases	BA by Aug																														
Highlight problems with mis-use of the BI databases to Data and Integration Project team	BA by Sep																														
Restrict access to only allow BI team to make changes to BI databases	BA by Sep																														
Identify, document and plan approach to dealing with all other 'hidden' dependencies falling outside of new standard	BA by Jan																														
Data and Integration Project to define future data integration standards (confirm date once project roadmap defined Sep 2021)	BA tbc																														
Action	Who and When	Status																													
Share knowledge and best practice with BI team members	BA by Aug																														
Contact potentially impacted colleagues/business areas to share plan	BA by Jan																														
Share output of Data and Integration project solution with impacted business areas	BA tbc																														