FSA Integration Partner

United States Department of Education

Federal Student Aid

Data Strategy Enterprise-Wide
XML Framework Team
123.1.19 XML ISIR Performance Test and SAIG Capacity Analysis

Task Order #123

Version 1.0

August 21, 2003

Executive Summary

The Student Aid Internet Gateway (SAIG) Portal provides telecommunications support for the delivery and administration of Title IV programs via the Internet. The U.S. Department of Education and its Integration Partners are sponsoring the SAIG Portal to promote the electronic exchange of Title IV information between educational and other entities over the Internet.

ISIRs (Institutional Student Information Record) are electronic records that are produced by CPS (Central Processing System) and provide schools with processed application information. ISIRs are transmitted electronically to destination points (schools, servicers, and state agencies) daily through SAIG. Currently these ISIR files are produced in a fixed-length file format. The purpose of the XML ISIR performance test plan is to assess the impact of moving the ISIR from a fixed length file format to XML.

The ISIR Performance Test

In order to get an accurate representation of the impact to SAIG, a number of performance test scenarios were executed. The purpose of the performance test is to understand the resultant effect of change from the fixed-length ISIR to an XML formatted ISIR. The goals of this performance test were to determine the end-to-end impact of the XML ISIR on SAIG as well as the capacity required to support the new file format.

The Data Strategy team leveraged the ITA performance testing lab to simulate a production-like environment for this test. The test scenarios employed ISIRs of varying sizes while a representative load of other SAIG traffic was running concurrently in the performance test environment. The testing team monitored a number of data points to generate their results, such as: SAIG system resources (CPU, disk space, etc), transmission time between a simulated CPS system and SAIG, compression ratios, and distribution time for SAIG to place the individual XML ISIR batches into the appropriate destination mailboxes.

Key Findings

This document contains the results of the performance testing cycles, as well as the calculations performed during the capacity analysis. The following points highlight the key findings from the work done in these two areas.

- *As the number of records within an XML file increases, the compression ratio also increases.* This figure is based on a comparison of the XML files with varying number of records. The compression algorithms take advantage of repeated text. Due to the repeated XML tags, files with more records undergo a better compression ratio. Small XML batches achieve compression (on average) of 90%, while larger XML batches achieve compression of 95% or better.

- *For 55% of the user community receiving ISIR batches, the compressed XML ISIR will be approximately five times larger than the compressed fixed-length ISIR.* This figure is based on comparing the compression ratio of fixed-length ISIR batches to XML ISIR batches with

less than 20 individual ISIRs per batch. Compressed size was a primary consideration since the files are stored and transmitted in compressed format. The amount of time an ISIR batch takes to download is directly proportional to the compressed size of the XML ISIR; therefore, the XML ISIR for these low volume recipients will increase by a factor of five. For users with the current FSA minimum standard connection speed of 56kbs will receive XML batch of 19 records in 8 seconds, versus the 2 seconds it would take for a fixed length file.

- *For the remaining 45% of the batches contained in the ISIR transmission file, the compressed XML ISIR will be about 2.5 times larger than the compressed fixed length ISIR.* This figure is based on comparing the ratio of compressed batches of each type. This 45% represents destinations receiving batches of 20 or more individual ISIRs. The smaller ratio is due to the 95% compression on larger files. Compressed size was the main consideration since the files are stored and transmitted in a compressed format. The amount of time an ISIR batch takes to download is directly proportional to the compressed size of the XML ISIR; therefore, the XML ISIR for these higher volume recipients will increase by a factor of 2.5.

- *The XML ISIR will increase the transmission time between CPS and SAIG, by a factor of two.* The transmission time was measured as total transmission network time to send the files between CPS and SAIG. This time did not include compression on CPS. While the percentage increase in transmission time was large, the absolute "wall clock" time increase is still very small – for example, in Cycle 1, a fixed-length file took 54 seconds versus 110 seconds for an XML file.

- *The XML ISIR will increase disk space requirements on SAIG, by about 220 gigabytes.* Due to the larger batch size of the XML ISIR, additional disk space is required to store the file. On average, a compressed XML ISIR file is five times larger than a compressed fixed length ISIR. By analyzing current disk space requirements, an increase of 220 gigabytes is required.

- *Mailboxing of the XML ISIR file vs. the fixed-length ISIR file will take approximately the same amount of time.* Once the XML ISIR transmission file is received by SAIG, SAIG divides the transmission files into batches and distributes the batches to the appropriate SAIG mailbox. The time to perform this process is mainly a function of the number of batches in the file. Since the two files contained the same number of batches, the file splitting processes took about the same amount of time.

- *SAIG will need additional CPU to support longer user download times.* Due to longer download times, a user will connect to SAIG for longer times, potentially increasing the number of concurrent FTP sessions. Insufficient data exists to determine the magnitude of the effect; however, SAIG performance testing indicates additional hardware may be required. Recommendation that the SAIG release scheduled for December include additional logging to gather statistics needed to determine the impact.

- *The VDC network infrastructure can support transmitting XML ISIRs with no upgrades.* The primary connection of the VDC to the Internet is a 155 Mbps OC3 link. Analysis shows that this connection provides ample networking bandwidth for the needs of FSA traffic through SAIG.

- *The majority of the institutions will be able to download a XML ISIR batch within a reasonable time frame by using FSA's current minimum standard connection speed of 56kbs.* Approximately 90% of the ISIR batch recipients receive ISIR batches with 123 or less individual ISIRs. Batches with 123 individual ISIRs or less can be downloaded within one minute. The remaining 10% high-volume users will need connection speeds greater than the current minimum standard in order to download ISIR batches within a reasonable timeframe.

Summary

The purpose of the XML ISIR Performance test and Capacity plan was to investigate the impact of instituting a new file format - XML ISIR. The performance test measured the attendant effects on the existing SAIG infrastructure, the data center network, and to the end users. The XML ISIR SAIG Performance test and Capacity Analysis demonstrates that, in most categories, the VDC infrastructure is currently capable of supporting the new XML ISIR file upon its implementation. The major area in need of upgrading is the storage capacity on the SAIG server. SAIG will also need added CPU capacity during the peak ISIR processing period to support the longer user download times. Furthermore, the analysis also revealed that most destination points receiving the XML ISIRs will not require a connection speed greater than the current FSA minimum standard connection rate, but longer download times are likely.

Amendment History

DATE	SECTION/ PAGE	DESCRIPTION	REQUESTED BY	MADE BY
07/31/03	All	Document submitted for FSA-wide review.	N/A	A. Lefur
8/20/03	SAIG Storage Analysis / 39	Incorporated FSA Comments.	Yolanda Brooks	J. Goldhirsch

Table of Contents

1 INTRODUCTION 9

1.1 BACKGROUND 9
1.2 BUSINESS OBJECTIVE 9
1.3 EXPECTED RESULTS 9
1.4 SCOPE 9
1.5 ORGANIZATION OF THE DOCUMENT 10
1.6 ASSUMPTIONS 10
1.7 ISSUES 10

2 PROCESS 11

2.1 PERFORMANCE TEST PROCESS 11
2.1.1 Performance Test Process 11
2.1.2 Performance Test Areas 11
2.1.3 Data and Parameters 11
2.1.4 Fixes and Changes 11
2.1.5 Specific Goals for Subsequent Test Runs 11
2.2 ROLES, RESPONSIBILITIES, AND DELIVERABLES 12
2.2.1 Performance Test Roles and Responsibilities 12
2.2.2 Performance Monitoring 14
2.2.3 Deliverables 15
2.3 XML ISIR SAIG BUSINESS PROCESS 15
2.3.1 Business Process 15
2.3.2 Business Process Simulation 16
2.4 PERFORMANCE TEST CYCLES 18
2.4.1 Test Cycle Overview 18
2.4.2 File Volume Figures 18
2.4.3 SAIG Mailboxes 19
2.4.4 Performance Test Cycles 19

3 RESULTS 22

3.1 OVERVIEW 22
3.2 ISIR FILE SIZING ANALYSIS 22
3.2.1 Summary 22
3.2.2 Data Analysis 23
3.2.3 Risks 28
3.3 XML ISIR PERFORMANCE TEST RESULTS 29
3.3.1 Summary – Cycle 1 29
3.3.2 Data Analysis 30
3.3.3 Summary – Cycle 2 32
3.3.4 Data Analysis 34
3.3.5 Summary –Cycle 3 35
3.3.6 Data Analysis 37
3.3.7 Risks 38
3.4 SAIG STORAGE ANALYSIS 38
3.4.1 Summary 38
3.4.2 Data Analysis 39
3.4.3 Risks 41
3.5 XML ISIR DOWNLOAD ANALYSIS – VDC PERSPECTIVE 41
3.5.1 Summary 41

3.5.2 Data Analysis *41*
3.5.3 Risks *43*
3.6 XML ISIR DOWNLOAD ANALYSIS – END USER PERSPECTIVE 43
3.6.1 Summary *43*
3.6.2 Data Analysis *44*
3.6.3 Risks *45*

APPENDIX A: SAIG TECHNICAL INFRASTRUCTURE 46

APPENDIX B: XML ISIR CALCULATIONS 47

Tables and Figures

Tables

Table 2.1 – Performance Test Roles and Responsibilities 14
Table 2.2 – Performance Test Monitoring 14
Table 2.3 – Performance Test Dates and Conference Call Numbers 18
Table 2.4 – EAI SAIG Varied File Size 18
Table 2.5 – ISIR Varied File Size 19
Table 2.6 – Non-ISIR Varied File Size 19
Table 3.1 – ISIR Batches per File 24
Table 3.2 – ISIR Batch Size Distribution 24
Table 3.3 – Number and Size of Batches in ISIR File 25
Table 3.4 – ISIR Volume Comparison – Historical Data vs. Capacity Analysis Calculation 25
Table 3.5 – COD XML Compression – Historical Data 26
Table 3.6 – XML ISIR Batch Compression – Test Data 27
Table 3.7 – Uncompressed Batch Size Comparison in bytes – Calculated Values 27
Table 3.8 – Compressed Batch Size Comparison in bytes – Calculated Values (Based on 90% XML Compression) 28
Table 3.9 – Compressed Batch Size Comparison in bytes – Calculated Values (Based on 95% XML Compression) 28
Table 3.10 – Cycle 1 - Global Metrics on HPL15 (SAIG server) 31
Table 3.11 – Cycle 1 - Application CPU breakdown on HPL15 (SAIG server) 32
Table 3.12 – Cycle 2 - Global Metrics on HPL15 (SAIG server) 34
Table 3.13 – Cycle 2 - Application CPU breakdown on HPL15 (SAIG server) 35
Table 3.14 – Cycle 3 - Global Metrics on HPL15 (SAIG server) 37
Table 3.15 – Cycle 3 - Application CPU breakdown on HPL15 (SAIG server) 38
Table 3.16 – SAIG Storage Requirements (Non-Archived Data) 40
Table 3.17 – SAIG Storage Requirements (Non-Archived Data) 40
Table 3.18 – Sample ISIR Bandwidth Requirements Calculations and Results 42
Table 3.19 – Actual Network Bandwidth Utilization during ISIR Performance Test Run 43
Table 3.20 – ISIR Download Times – Fixed Length Record (85% Compression) 44
Table 3.21 – ISIR Download Times – XML Record (90% Compression) 45
Table 3.22 – ISIR Download Times - XML Record (95% Compression) 45

Figures

Figure 2.1 – SAIG Production Business Process 16
Figure 2.2 – SAIG Performance Test Simulation 17
Figure 3.1 – ISIR Process 23

1 Introduction

1.1 Background

The Student Aid Internet Gateway (SAIG) Portal provides telecommunications support for the delivery and administration of Title IV programs via the Internet. The U.S. Department of Education and its Integration Partners are sponsoring the SAIG Portal to promote the electronic exchange of Title IV information between educational and other entities over the Internet.

Substantial cost savings to FSA is achieved by moving FSA data transmissions to the Internet. The SAIG Portal provides FSA with a non-intrusive solution for customers, reduces the overall cost of delivering student aid, and reduces operating costs by consolidating operations and systems. FSA to the Internet went live on September 20, 2001.

SAIG is critical to FSA and the Title IV programs because it provides the infrastructure for transferring data and messages. If the system is down or experiencing performance decreases for extended periods of time the exchange of critical data would be impacted and detrimental to the grant and loan processes on which schools and students depend.

ISIRs (Institutional Student Information Record) are electronic records that are produced by CPS (Central Processing System) and provide schools with processed application information. ISIRs are transmitted electronically to destination points (schools, servicers, and state agencies) daily through SAIG. ISIR files are produced by award year, once a day, Monday through Friday. Each file will contain batches for a specific award year. For example, during this spring's peak, CPS produced one file with all of the 2002-2003 award year batches, and one file with all 2003-2004 award year batches.

1.2 Business Objective

The purpose of this analysis is to study the impacts to SAIG and end-users of converting the ISIR from a fixed length record to XML. The performance test will study the impacts to SAIG of converting the ISIR from a fixed length record to XML. The results from this test will be evaluated in this document.

1.3 Expected Results

At the conclusion of each test cycle, a test report will be prepared with the results of the test cycle and confirmation that the specific objectives of the cycle were met. At the conclusion of all test cycles, the SAIG Capacity Analysis will evaluate how the results documented here will impact SAIG.

1.4 Scope

The XML ISIR Performance Test and SAIG Capacity Analysis (Del 123.1.19) lays the framework for the XML ISIR performance test. This document describes how the performance test will be

executed, what scenarios will be executed, what individuals are responsible for completing tasks during the performance test, and dates and times of the performance test cycles. This document also completes the analysis of what moving the ISIR from a fixed length record to XML will mean for SAIG. This document will evaluate the results of the performance test, as well as complete additional data analysis and methodology on file sizing, storage space, compression, and internet bandwidth.

1.5 Organization of the Document

The XML ISIR Performance Test and SAIG Capacity Analysis consists of the following sections:

- Section 1: Introduction – Provides the high-level overview, scope, business objectives, and assumptions for the XML ISIR Performance Test and SAIG Capacity Analysis.
- Section 2: Process – Provides the framework for the execution of the XML ISIR performance test. This framework includes background information about the SAIG Business process, how this business process will be simulated in the performance test environment, detailed goals for each of the cycles, and roles and responsibilities.
- Section 3: Results – Provides an analysis of what moving the ISIR from fixed length to XML will mean to SAIG and the end user. It includes an analysis of the results from the XML ISIR SAIG performance test, as well as additional data analysis about ISIR file sizing, storage capacity, and calculation of XML ISIR download times.

1.6 Assumptions

During the planning of the XML ISIR SAIG performance test the following assumptions were made:

- The XML ISIR Performance Test will use the same mix of data files as the COD (Common Origination and Disbursement) SAIG performance test used to simulate non-ISIR traffic through SAIG.
- In order to determine the size of the of the CPS ISIR test data, the 2003-2004 award year batches for the spring 2003 SAIG peak period will be analyzed.
- This performance test will only be using the network at the VDC. There will be additional data analysis in the Capacity Analysis document which will look at the impact of XML ISIR files being downloaded, both from the end-user's perspective, and from a VDC bandwidth perspective.
- For all sections of this analysis, unless otherwise stated, the SAIG file compression rate will be based on SAIG's production data's compression rate.
    - Fixed length batch compression is approximately 85%.
    - XML Batch compression is between 90% and 95% (both values will be used).

1.7 Issues

Any issues that occur during testing will be documented in this section of the document.

2 Process

2.1 Performance Test Process

2.1.1 Performance Test Process

A process is required to establish an XML ISIR baseline, set goals for the performance test, and achieve those goals. The steps below explain the process that will be undertaken through the performance test effort.

2.1.2 Performance Test Areas

Performance test areas include a wide range of items such as user response times, throughput, CPU and memory utilization, and number of concurrent users etc. These areas are established before the test so that monitoring points can be derived, and the teams have time to schedule the right people to monitor the tests. Specific monitoring points will be analyzed to identify issues/bottlenecks and will also provide the necessary data for analysis and response times. Table 2.2 includes all the performance test areas that need to be monitored and the responsible party for each.

2.1.3 Data and Parameters

Data will need to be collected for analysis during and after each test cycle. Table 2.2 outlines the type of data that should be monitored, collected, and the responsible party for each monitoring point. Section 3.3 outlines the data analysis deliverables that CSC, Pearson, and ITA are to submit at the conclusion of each test cycle.

2.1.4 Fixes and Changes

Based on the analysis and recommendations of the team, configuration changes may need to be made after each test run. These changes should provide higher performance results in the next test run.

2.1.5 Specific Goals for Subsequent Test Runs

Specific goals for each test cycle are outlined in Section 5 – Detailed Goals per Cycle**.** After data has been collected, and the required fixes made, there will be the opportunity to modify specific goals for the next test run.

2.2 *Roles, Responsibilities, and Deliverables*

2.2.1 Performance Test Roles and Responsibilities

This table lists the roles and responsibilities for the participants in XML ISIR SAIG performance test.

Role	Assigned	Phone	Responsibility
Performance Test Lead/ITA Resource	Roshani Bhatt	202-962-0740	• Coordinate with all teams and resources to ensure that the capacity planning, performance test planning, and performance test execution are completed on time. • Work with CSC to measure the performance of each component (CPU, Memory, and I/O). • Define the detailed technical targets for each business goals. • Define the expected concurrent users per server at optimized configuration. • Determine the length of each run and starting/stopping points. • Provide necessary VDC network utilization data to Data Strategy team for internet capacity analysis
SAIG Application Lead	Colleen M. Ward	202-962-0768	• Ensure that performance test requirements are met successfully. • Define and set the goals for each test cycle. • Verify data that came from capacity planning.
Data Strategy Resource	Andrew Smalera	202-962-0789	• Provide XML ISIR capacity planning input. • Provide XML ISIR files (sample test data). • Gather necessary information from ITA to analyze the impact of XML ISIR files being downloaded from the internet. • Provide XML ISIR Download Analysis – VDC Perspective.

Role	Assigned	Phone	Responsibility
Data Strategy Resource	Jeffrey C. Goldhirsch	202-962-0735	• Provide XML ISIR File Sizing. • Provide XML ISIR files (test data). • Provide SAIG Storage Analysis. • XML ISIR Download Analysis – End-User Perspective • Produce and package XML ISIR capacity plan final deliverable.
Performance Test Resources	Musab Alkateeb Chi-yen Yang	202-962-0729 202-962-0758	• Coordinate test dates with all testing resources (ITA, Pearson CSC, and Integration Partner). • Work with CSC to measure the performance of each component (CPU, Memory, and I/O). • Work with Pearson and Data Strategy team to execute performance test. • Work with all the teams to verify administrator contact information and attendance. • Set up pre-test and test day conference calls. • Participate in performance test calls, document results. • Provide XML ISIR Performance Test Results.
SAIG Application Resource	Jamie Steapp	319-665-7987	• Provide historical SAIG volume information to Accenture team for analysis. • Setup and configure SAIG application for testing. • Develop the test scripts (Perl). • Develop the test data. • Execute each performance test. • Monitor SAIG mailbox.
HP Hardware Administrator	Paul Izzo	203-317-2175	• Setup and configure HP hardware. • Provide support for HP machines during testing. • Provide results of stress test in graphical format at the end of each test.
Network Administrator	Georgery Lindstrom	203-317-5178	• Monitor Network to ensure optimal configuration. • Monitor capacity levels at different intervals throughout the test cycle. • Coordinate test dates with NCS and ITA.

Role	Assigned	Phone	Responsibility
CSC Solution Delivery Manager	Jerry Anderson	503-539-1039	• Ensure CSC activities are executed according to plan. • Act as primary liaison for VDC coordination.
CPS Lead	Ida Mondragon	202-377-3243	• Verify the size of XML ISIR and capacity plan.

Table 2.1 – Performance Test Roles and Responsibilities

2.2.2 Performance Monitoring

The table lists each performance area, a general description of the performance area, and the group responsible to monitor the area. This document will be supplied to those monitoring the test. It will ensure that all required areas would be monitored.

Performance Areas	Description	Monitored By:
FTP Connections/ Sessions	Indicates the number of "users" of the SAIG application. Data collected in increments of 1 and 5 minutes.	Pearson
SAIG Mailbox	Indicates the time to process messages going through bTrade.	Pearson
Memory Utilization	Indicates the percentage of physical memory in use. Data collected in increments of 5 minutes.	CSC
Disk Utilization	Indicates the percentage of time that the busiest disk was performing I/O transfers. Data collected in increments of 5 minutes.	CSC
CPU Utilization	Indicates the percentage of time the CPU was not idle. Data collected in increments of 5 minutes.	CSC
Priority Queue	Indicates the average number of processes blocked on priority (waiting for their priority to become high enough to be processed by the CPU). Data collected in increments of 5 minutes.	CSC
Swap Space	Indicates the percentage of available swap space that was being using by running processes. Data collected in increments of 5 minutes.	CSC
Network	Monitor capacity levels throughout the test cycle. Data collected in increments of 5 minutes.	CSC
EAI Adaptors	Indicates the time to process messages and CPU consumption.	EAI

Table 2.2 – Performance Test Monitoring

2.2.3 Deliverables

The following documents will be created by the conclusion of the XML ISIR performance test. Table 2.2 provides a guide as to the data that needs to be captured and presented to the team at the conclusion of each test cycle.

CSC Performance Test Data
The Performance Monitoring Areas, Table 2.2, provides a guide as to the type of data that will need to be collected. CSC should decide on the best format for providing this data and summary of test results. The data should be collected at the interval required in Table 2.2, unless otherwise suggested during the test. CSC is expected to deliver this data and summary of test results the same day after each test cycle is run.

Pearson Performance Test Data
The Performance Monitoring Areas, Table 2.2, provides a guide as to the type of data that will need to be collected. Pearson should decide on the best format for providing this data and summary of test results. The data should be collected at the interval required in Table 2.2, unless otherwise suggested during the test. Pearson is expected to deliver this data and summary of test results the same day after each test cycle is run.

Formal Deliverables
There is one deliverable that will be completed at the conclusion of the SAIG XML ISIR performance test. ITA will provide a document which will contain the performance test dates, plans, data, and results from the test runs. The XML Framework team will complete a document which will analyze the impact of moving the ISIR from a fixed length record to XML and its impact to SAIG.

2.3 XML ISIR SAIG Business Process

This section defines the business process that will be simulated during the XML ISIR performance test, and how this simulation will be executed.

2.3.1 Business Process

The XML ISIR performance test targets the following three processes:

- EAI SAIG Traffic – Data going between SAIG and COD
- ISIR SAIG Traffic – Data coming from CPS to SAIG
- Non-ISIR SAIG Traffic – Data going between SAIG and all other Entities

In order to accurately assess the impact of moving the ISIR from fixed length to XML, each of these business processes must be simulated in all test cycles.

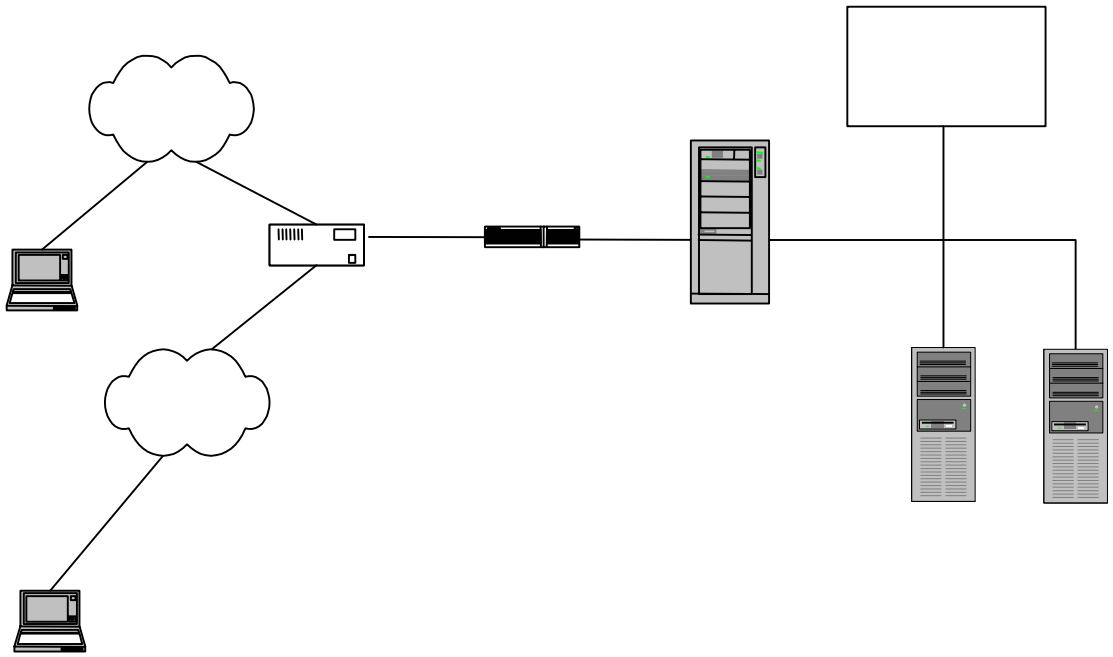

Figure 2.1 – SAIG Production Business Process

2.3.2 Business Process Simulation

Each business process will have its own simulation process. Figure 2.2 is a diagram of the SAIG performance test environment. Each of the servers mentioned in these simulations are included in that diagram.

The Internet

B/C

Internet Client

The Intranet

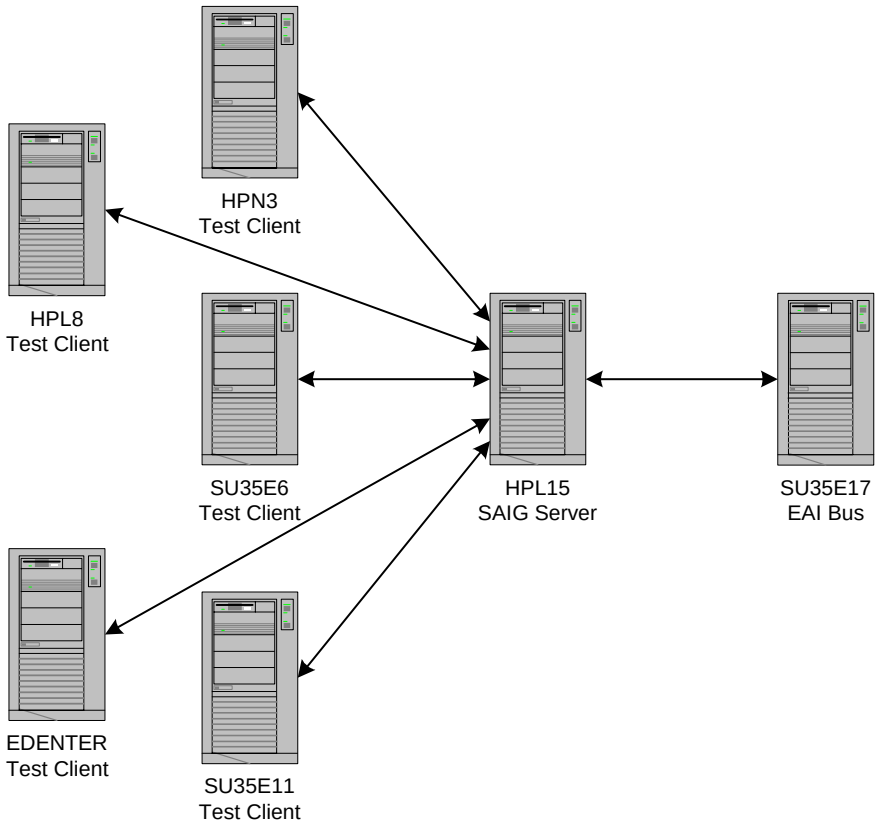

Figure 2.2 – SAIG Performance Test Simulation

EAI SAIG Traffic

This traffic is represented by the letter "A" in Figure 2.2. The EAI team will be sending and receiving messages using a test stub. The test stub will use the EAI SAIG Put and EAI SAIG Get to send and receive messages between the EAI Bus test server (E17) to the SAIG test server (HPL15). This will serve to simulate traffic going between SAIG and COD.

ISIR SAIG Traffic

This traffic is represented by the letter "B" in Figure 2.2. ISIR traffic will be simulated by building test files, both fixed length and XML. These test files will be built to match how CPS builds the ISIR files. Currently, Monday through Friday, CPS builds one large ISIR file per award year, per day. This file contains multiple batches that will be split up and mailboxed at SAIG. The ISIR file will be sent from one of the Test Client servers to the SAIG test server (HPL15).

Non-ISIR SAIG Traffic

This traffic is represented by the letter "C" in Figure 2.2. Non-ISIR SAIG Traffic will be simulated by starting multiple client sessions to send and receive files to SAIG. Approximately 55% of the client sessions will be dedicated file senders, and 45% of the client sessions will be dedicated file receivers. For example, when there are 100 concurrent users, there will be

approximately 55 dedicated senders, and 45 dedicated receivers. This traffic will be generated from the Test Client servers, which will be communicating with the SAIG test server (HPL15).

2.4 Performance Test Cycles

2.4.1 Test Cycle Overview

This section will include detailed goals for each test cycle. The goals for each cycle are not static. Depending on how the test cycle goes, the goals for the current cycle and future cycles may expand or decrease accordingly. If issues are encountered on the scheduled performance test cycles, the performance test cycles will need to be rescheduled.

The test cycles dates, times and conference call numbers were as follows:

Cycle	Date	Time (Eastern)	Conference Call Number
Cycle #1	7/17/2003	9:30 AM – 12:30 PM	1-866-277-6700 Meeting ID: 2903
Cycle #2	7/22/2003	9:30 AM – 12:30 PM	1-866-277-6702 Meeting ID: 2903
Cycle #3	7/23/2003	9:30 AM – 12:30 PM	1-866-277-6702 Meeting ID: 2903
Cycle #4 (Optional)	7/24/2003	9:30 AM – 12:30 PM	1-866-277-6702 Meeting ID: 2903

Table 2.3 – Performance Test Dates and Conference Call Numbers

2.4.2 File Volume Figures

EAI SAIG Traffic Volume
For all test cycles requiring files sent between the EAI Bus and SAIG, simulating the traffic that goes between SAIG and COD via the EAI Bus, the following batch size breakdown will be used. This breakdown is based on the percentages used in the COD SAIG performance test for the August 2003 peak.

File Size	% of Total
5 Kb	70%
10 Kb	17%
50 Kb	12%
100 Kb	1%

Table 2.4 – EAI SAIG Varied File Size

ISIR SAIG Traffic Volume
For all test cycles requiring ISIR files, the following batch size breakdown will be used. This breakdown is based on the 2003-2004 award year data from the 2003 spring peak.

Number of Detail Records in Batch	% of Total	Number of Batches
19	54.54%	3,396
123	36.31%	2,262
254	4.84%	301
1,295	3.49%	217
2,596	0.45%	28
13,009	0.34%	21
30,000	0.03%	2
	Total	6,227

Table 2.5 – ISIR Varied File Size

Non-ISIR SAIG Traffic
For all test cycles requiring varied file sizes to simulate the non-ISIR SAIG traffic that is not destined for COD, the distribution used in the most recent COD SAIG Performance test will be used. These values are documented here.

File Size	% of Total
1 Kb	35.05%
5 Kb	33.51%
10 Kb	19.30%
50 Kb	10.80%
100 Kb	0.67%
500 Kb	0.47%
1000 Kb	0.07%
5000 Kb	0.05%

Table 2.6 – Non-ISIR Varied File Size

2.4.3 SAIG Mailboxes

During the SAIG performance test, 3001 different mailboxes will be used to send and receive files. EAI traffic will send and receive from one mailbox. This will simulate how EAI puts message to SAIG from COD and gets messages from SAIG for COD. The remaining 3000 mailboxes will be used to send and receive messages associated with both the ISIR SAIG traffic and the non-ISIR SAIG traffic.

2.4.4 Performance Test Cycles

Performance Test Cycle 1 - Scheduled - 07/17/2003 9:30 AM – 12:30 PM

Call in #: 1-866-277-6700 Meeting ID: 2903

Detailed goal for test cycle one:

> *To compare SAIG processing of fixed length ISIR batches to XML ISIR batches with 100 concurrent users.*

The following steps will be executed:

1. Ramp up to steady state of 100 concurrent users sending varied file sizes, and have EAI send and receive 13 files a minute for twenty minutes.
2. Send fixed length ISIR to SAIG for processing. Run until all ISIR batches have been received by simulated users.
3. Return to Steady State for twenty minutes.
4. Send XML ISIR to SAIG for processing. Run until all ISIR batches have been received by simulated users.

Expected Results:
SAIG should be able to process both types of ISIR files. These results will be used in the SAIG Capacity Analysis.

Performance Test Cycle 2 - Scheduled - 07/22/2003 9:30 AM – 12:30 PM
Call in #: 1-866-277-6702 Meeting ID: 2903

Detailed goal for test cycle two:

> *To compare SAIG processing of fixed length ISIR batches to XML ISIR batches with 200 concurrent users.*

The following steps will be executed:

1. Ramp up to steady state of 200 concurrent users sending varied file sizes, and have EAI send and receive 13 files a minute for twenty minutes.
2. Send fixed length ISIR to SAIG for processing. Run until all ISIR batches have been received by simulated users.
3. Return to Steady State for twenty minutes.
4. Send XML ISIR to SAIG for processing. Run until all ISIR batches have been received by simulated users.

Expected Results:
SAIG should be able to process both types of ISIR files. These results will be used in the SAIG Capacity Analysis.

Performance Test Cycle 3 - Scheduled - 07/23/2003 9:30 AM – 12:30 PM
Call in #: 1-866-277-6702 Meeting ID: 2903

This cycle will be used to test if SAIG can process ISIR files 15% larger than those used in the previous test. In the case that this cycle will be running ISIR files that are 15% larger, the number of concurrent users may be scaled down based on the results from the previous two cycles.

The following steps will be executed:

1. Ramp up to steady state of 200 concurrent users sending varied file sizes, and have EAI send and receive 13 files a minute for twenty minutes.
2. Send fixed length ISIR to SAIG for processing. Run until all ISIR batches have been received by simulated users.
3. Return to Steady State for twenty minutes.
4. Send XML ISIR to SAIG for processing. Run until all ISIR batches have been received by simulated users.

Expected Results:
SAIG should be able to process both types of ISIR files. These results will be used in the SAIG Capacity Analysis.

Performance Test Cycle 4 - Optional - 07/24/2002 9:30 AM – 12:30 PM
Call in #: 1-866-277-6702 Meeting ID: 2903

Detailed goal for test cycle four:

This cycle will be used to rerun any previous test that had technical issues.

The following steps will be executed:
If it is necessary to re-run any of the previous cycles, then the execution steps will be based on the failed cycle's execution steps.

Expected Results:
If it is necessary to re-run any of the previous cycles, then the expected results will be based on that cycle's expected results.

3 Results

3.1 Overview

The performance test and data analysis are summarized in the following points:

- ISIR File Sizing Analysis
  - Due to the use of the largest batch size from the SAIG batch size distribution data, the ISIR batches will be larger than those seen in production.
  - Test Data compresses at a higher percentage than production data.
- SAIG Performance Test Results
  - Cycle 1
    - Achieved objectives of running 100 concurrent FTP Sessions, 13 files per minute on EAI Bus, compressed, transmitted, and split fixed length and XML ISIRs.
  - Cycle 2
    - Validated the system's ability to support 200 concurrent FTP Sessions for all three business processes. The test simulated a worst-case-scenario by accounting for the August and March peaks. The system was slowed by the heavy load, but successfully processed all the files.
  - Cycle 3
    - The fixed-length ISIR and XML ISIR files prepared for this cycle were 15% larger. The system successfully processed the larger files, in addition to the 100 concurrent FTP Sessions and the 13 files per minute EAI business process. As expected, CPU utilization increased marginally as an outcome of employing the larger file sizes.
- SAIG Storage Capacity
  - Assuming no changes to the archiving process or compression, SAIG will require additional storage space. Assuming 90% compression, SAIG will need an additional 220 GBs of storage.
- XML ISIR Download Analysis – VDC Perspective
  - The VDC network, both behind the firewall and between the firewall and the internet, will be able to handle the XML ISIR with out any issues.
- XML ISIR Download Analysis – End User Perspective
  - The larger compressed files associated with moving the ISIR from a fixed length record to XML will result in longer download times. Assuming a 90% compression rate for the XML ISIR, download times will be 4.87 times longer than the fixed length ISIR.

3.2 ISIR File Sizing Analysis

3.2.1 Summary

ISIRs are electronic records that are produced by the CPS (Central Processing System) and provide schools with processed application information. ISIRs are transmitted electronically to destination points (schools, servicers, and state agencies) daily through SAIG. ISIR files are

produced by award year, once a day, Monday through Friday. Each file will only contain batches for a specific award year. For example, during this spring's peak, CPS produced one file with all of the 2002-2003 award year batches, and one file with all 2003-2004 award year batches. This large file is then sent to SAIG, where it is split into batches and stored in the destination mailbox. SAIG refers to this process of splitting up and loading messages as mailboxing. Figure 3.1 shows this process.

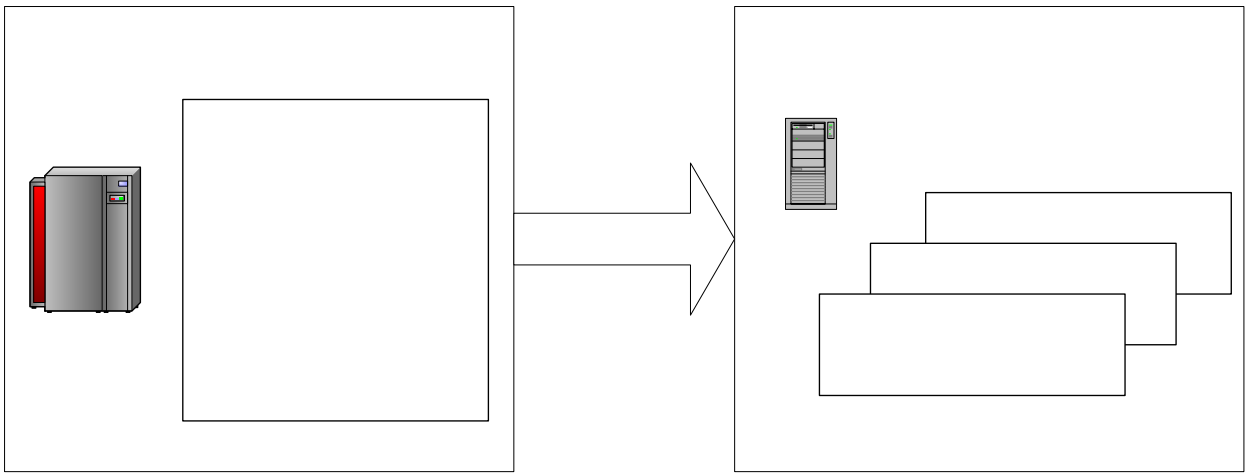

Figure 3.1 – ISIR Process

In order to properly estimate the size of the XML ISIR the following areas were targeted:

- Number of batches in each ISIR file
- Batch size distribution
- Number of ISIRs sent form CPS per week

These areas were targeted for the 2003-2004 award year from March of 2003. The month of March was chosen, because it has the largest volume of ISIR traffic. The number of batches in each in each file and the batch size distribution were combined to get an accurate picture of the current ISIR file sizing. These results were then validated against the number of ISIRs sent from CPS per week.

3.2.2 Data Analysis

Number of Batches in each File

SAIG collects daily data on the number of batches sent by message class. Since CPS only sends one ISIR file per day per award year, all of the ISIR batches for the 2003-2004 award year would have been included in one ISIR file. For this analysis, and the performance test, the number of batches in the ISIR file will equal the average number of batches plus one standard deviation. Table 3.1 shows these values; the data and calculations used to generate these numbers, can be found in Appendix B: XML ISIR Calculations.

CPS

ISIR BATCH #1
ISIR BATCH #1
...

Statistic Category	Number of Batches in File
Average	5,820
Standard Deviation	408
Average plus Standard Deviation	6,228

Table 3.1 – ISIR Batches per File

Batch size Distribution

SAIG collects monthly data on batch size distributions by message class. This data on ISIR batch size distributions was used to determine what percentage of batches will fall into each batch size window. Since this data is collected based on the compressed batch size, this data must go through a conversion process to take it from compressed batch size to number of ISIR detail records per batch. The conversion processes first estimates decompressed file ranges. The next step in the conversion process is to divide this decompressed volume by the size of each ISIR detail record size, currently 2,561 (2560 plus End of Line Character). This process yields a range of batch sizes in terms of the number of student detail records in the batch. To ensure no underestimation, the larger size of each batch size range will be used. Table 3.2 shows these percentages; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Number of Detail Records in Batch	Percentages of Batches in each range
19	54.54%
123	36.31%
254	4.84%
1,295	3.49%
2,596	0.45%
13,009	0.34%
30,000	0.03%

Table 3.2 – ISIR Batch Size Distribution

ISIR File Size Estimation

The size of the ISIR file is calculated by combining information about the number of batches with the batch size distribution. The previously calculated number of batches is multiplied by the percentage of batches in each range, to give a discrete number of batches for each range. Table 3.3 shows these percentages; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Number of Detail Records in Batch	Percentages of Batches in each range	Number of Batches
19	54.54%	3,396
123	36.31%	2,262
254	4.84%	301
1,295	3.49%	217
2,596	0.45%	28
13,009	0.34%	21
30,000	0.03%	2

Table 3.3 – Number and Size of Batches in ISIR File

Number of ISIRs sent from CPS per week

CPS collects weekly data on the number of ISIR files sent from CPS to SAIG. This data is used to validate the capacity analysis methodology. Due to the use of the larger size of each batch size range, the methodology for estimating the number of ISIR detail records should yield a higher number of detail records than actual number of detail records. Table 3.4 shows that the capacity analysis methodology yields a 15% higher number of detail ISIR records than were actually sent from CPS to SAIG.

Date Range	Number of ISIRs - MIS Report (CPS Data)	Number of ISIRs – Capacity Analysis Methodology	Difference	Percentage Difference
3/3/2003 – 3/9/2003	2,196,470	2,524,261	327,791	15%
3/10/2003 – 3/16/2003	2,364,452	2,677,543	313,091	13%
3/17/2003 – 3/24/2003	2,002,024	2,282,919	280,895	14%
3/25/2003 – 3/30/2003	1,686,033	2,024,602	338,569	20%
3/3/2003 – 3/30/2003	8,248,979	9,509,325	1,260,346	15%

Table 3.4 – ISIR Volume Comparison – Historical Data vs. Capacity Analysis Calculation

ISIR Batch Size Comparison: Fixed Length vs. XML

XML batches are larger than fixed length batches, because of a fundamental change in how data is represented. Fixed length data fields are represented by positions; while, XML data fields are determined by tag name. These tag names require additional bytes of data. For example, in the fixed length ISIR layout a student's Selective Service Registration status is determined by the one byte located at position 186 in the record. In the XML ISIR, this data is represented by 50 bytes in the XML record, "<SelectiveServiceMale>true</SelectiveServiceMale>."

However, this size difference is at least partially mitigated, because of XML's higher compression rate. This increase in compression rate is due to the fact that a large percentage of an XML document's volume is XML tags. These tags are repeated throughout an XML document and are easily compressed. Since XML batches have a higher compression rate than fixed length batches, the net ratio difference between compressed batch sizes and decompressed

batch sizes drops from 7.30 to 4.87 with an XML compression of 90%. This ratio is further decreased to 2.43 with an XML compression of 95%.

It should be noted that the smaller batches have a smaller ratio than the larger batches. The reason for the smaller ratios with the smaller batches is due to the difference in header lengths for the batch types. The fixed length files pad the header and trailer records to a length of 2,561 bytes. The XML files have a header and trailer length of 82 bytes. Both of the header and trailer character counts include the end of line character.

To analyze the compression rates for the XML ISIR, two different areas were reviewed. The first looked at the compression rates that the XML COD Common Record sees in production. The analysis from this showed an average compression rate of 93%. The data also shows a trend that the larger batches compress at a higher percentage than the smaller batches. Table 3.5 shows how the trend that larger batches have higher compression rates; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Uncompressed Batch Size (bytes)	Compressed Batch Size (bytes)	Compression Percentage
115,020,441	2,107,243	98.17%
113,502,276	3,790,737	96.66%
32,797,186	794,387	97.58%
11,731,906	711,645	93.93%
8,835,357	210,305	97.62%
6,610,074	144,561	97.81%
6,285,672	155,625	97.52%
2,236,329	126,403	94.35%
1,886,374	42,497	97.75%
1,086,405	20,408	98.12%
815,865	24,955	96.94%
760,732	40,032	94.74%
403,003	11,092	97.25%
309,399	16,500	94.67%
119,193	3,553	97.02%
91,455	3,060	96.65%
39,999	2,761	93.10%
34,372	4,900	85.74%
16,392	5,573	66.00%
16,080	1,821	88.68%
5,427	981	81.92%
	Average	93.44%

Table 3.5 – COD XML Compression – Historical Data

The second area for XML ISIR compression looked at how test XML ISIR batches would compress. These batches also showed the same trend that larger batches will compress at a higher percentage than smaller batches. It should be noted that these percentages are based on test data. Test data could have a higher compression rate than actual data, because there is less variability than production data. Table 3.6 shows the compression percentages of the XML ISIR test batches; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Number of Detail Records in Batch	XML ISIR Uncompressed Size (bytes)	XML ISIR Compressed Size (bytes)	Compression Rate
1	20,627	5,967	71.07%
2	40,472	6,351	84.31%
5	99,901	7,614	92.38%
10	199,075	9,181	95.39%
20	397,369	12,187	96.93%
30	595,605	15,257	97.44%

Table 3.6 – XML ISIR Batch Compression – Test Data

Table 3.7 compares the batch size of equivalent ISIR batches in fixed length and XML formats. The uncompressed file sizes are based on uncompressed test data; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Number of Detail Records in Batch	Percentage of Batches Range	Fixed Length Size (bytes)	XML Size (bytes)	Ratio
19	54.54%	53,781	377,506	7.02
123	36.31%	320,125	2,439,197	7.62
254	4.84%	655,616	5,036,382	7.68
1,295	3.49%	3,321,617	25,673,910	7.73
2,596	0.45%	6,653,478	51,465,824	7.74
13,009	0.34%	33,321,171	257,901,294	7.74
30,000	0.03%	76,835,122	594,742,222	7.74
			Net Ratio	7.30

Table 3.7 – Uncompressed Batch Size Comparison in bytes – Calculated Values

Table 3.8 compares the compressed batch size of equivalent ISIR batches in fixed length and XML formats. The compressed file sizes are based on an 85% compression rate for fixed width batches and 90% compression for XML batches; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Number of Detail Records in Batch	Percentage of Batches Range	Fixed Length Size (bytes)	XML Size (bytes)	Ratio
19	54.54%	8,607	37,751	4.68
123	36.31%	48,019	243,920	5.08
254	4.84%	98,342	503,638	5.12
1,295	3.49%	498,243	2,567,391	5.16
2,596	0.45%	998,022	5,146,582	5.16
13,009	0.34%	4,998,176	23,790,129	5.16
30,000	0.03%	11,525,268	59,474,222	5.16
			Net Ratio	4.87

Table 3.8 – Compressed Batch Size Comparison in bytes – Calculated Values (Based on 90% XML Compression)

Table 3.9 compares the compressed batch size of equivalent ISIR batches in fixed length and XML formats. The compressed file sizes are based on an 85% compression rate for fixed width batches and 95% compression for XML batches; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Number of Detail Records in Batch	Percentage of Batches Range	Fixed Length Size (bytes)	XML Size (bytes)	Ratio
19	54.54%	8,607	18,875	2.34
123	36.31%	48,019	121,960	2.54
254	4.84%	98,342	251,819	2.56
1,295	3.49%	498,243	1,283,696	2.58
2,596	0.45%	998,022	2,573,291	2.58
13,009	0.34%	4,998,176	12,895,065	2.58
30,000	0.03%	11,525,268	29,737,111	2.58
			Net Ratio	2.43

Table 3.9 – Compressed Batch Size Comparison in bytes – Calculated Values (Based on 95% XML Compression)

3.2.3 Risks

Areas for Overestimation

Since the batch size distribution is based on a batch size range, in order not to underestimate the batch size, the larger batch size from the range was used to calculate the batch sizes. This use of the larger batch size will yield a file that is larger than the current peak production ISIR file.

Table 3.7 shows the difference between the uncompressed batch sizes, based on the number of detail records in each batch. Table 3.8 and Table 3.9 shows the difference between the compressed batch sizes, based on the number of detail records in each batch. The XML compression percentages are based on the range of compression percentages for the COD Common Records. It should also be noted that the larger the COD Common Record batch, the higher the compression ratio. This suggests that the XML ISIR compression ratio will be closer to the 95% compression range, or even higher. Some of the largest COD Common Record batches have seen compression rates as high as 98%. That being said, COD and ISIR schema differences should be enough evidence to only use a conservative mapping from the COD compression percentage to the XML ISIR compression percentage for future calculations.

Areas for Underestimation
Test data tends to have a higher compression rate than actual production data. Due to this fact, when the individual test batches are compressed, a test batch's compressed size may actually decrease below the size of compressed production data. To mitigate this risk, compression rates in the capacity analysis, will be based on SAIG production compression rates.

3.3 XML ISIR Performance Test Results

3.3.1 Summary – Cycle 1

Test Outcome:
Cycle I of the SAIG XML ISIR performance test validated three business processes and established a baseline for further testing. The EAI-SAIG traffic was simulated along with ISIR SAIG and Non-ISIR SAIG traffic. The test validated the system's ability to support an ISIR with 100 concurrent users sending varied file sizes, while sending and receiving 13 files per minute on the EAI test server. The transmission of fixed-length ISIR file to SAIG was accomplished, as was the compression and transmission of XML ISIR file.

Cycle Background:
Cycle I of SAIG XML ISIR performance test established a baseline for the test cycles ahead.

Hardware and Configuration:
Five test clients were used during the test: SU35E6, SU35E11, HPL8, HPN3 and EDENTER. HPL15 was used as the SAIG mailboxing and FTP servers. The EAI Bus was hosted on SU35E17. The five test clients send fixed length ISIR files to the mailbox and about 13 files per minute are then sent to the EAI Bus.

Detailed Description:
Cycle I of SAIG XML ISIR performance test validated four preset criteria, which validated the business processes for this application.

The test commenced by establishing 100 concurrent FTP Sessions to simulate a state of existing SAIG traffic. The EAI process of sending and receiving 13 files per minutes followed several minutes later. The CPU climbed to a peak of 63% utilization during this segment of the test. In

the meantime the priority queue was well below the threshold at 0.3 for the entire test. The EAI file processing time ranged between 17-18 seconds.

After 30 minutes had elapsed and the system had stabilized, the fixed length ISIR was introduced. The fixed length ISIR was sent to HPL15 (SAIG server) from HPN3 (test client) to simulate the process employed by CPS. Before the file is transmitted by HPN3 (hence CPS) it is compressed, this process lasted 9 minutes, while actual transmission time was one minute. SAIG received a 215MB from the test client which compressed it from its original file size of 2.8GB -- a 92.5% file compression rate. The SAIG server divided the 215MB file into batch files, and distributed these files to mailboxes. The duration of this process was 3 minutes. The CPU utilization during these proceedings was 63% and memory utilization was 53.7%. The system exhibited stability throughout the test.

The XML ISIR file follows the same path in the system as the fixed length ISIR; the difference between the two files is one of size. The XML ISIR file in uncompressed form was 21.9GB, the compression process for this file lasted 53 minutes, and the resultant, compressed file measured 375MB, which equated to a file compression rate of 98.3%. The file was transmitted from the test client (HPN3) to the SAIG server (HPL15) in 2 minutes. The duration of the message splitting, and distribution to mailboxes, process for the XML ISIR was 13 minutes. The CPU utilization peaked at 78.3% at the message splitting process. Memory utilization remained steady at 53.2%.

For this test cycle and those that will follow, the principle goal is to compare the performance of the system while it processed the fixed length ISIR and the XML ISIR file. SAIG is expected to have the capacity to process both types of ISIR files, as the results will be used for the SAIG Capacity Analysis.

Cycle I accomplished the pre-test goals in running 100 concurrent FTP Sessions, while EAI sent and received 13 files per minute, the fixed length ISIR and XML ISIR file were compressed on HPN3 (test client simulating CPS), transmitted to HPL15 (SAIG server), split into batches and stored in destination mailboxes.

3.3.2 Data Analysis

The Global Metrics graph for HPL15 (SAIG server) demonstrates three key metrics: CPU utilization, memory utilization, and priority queue.

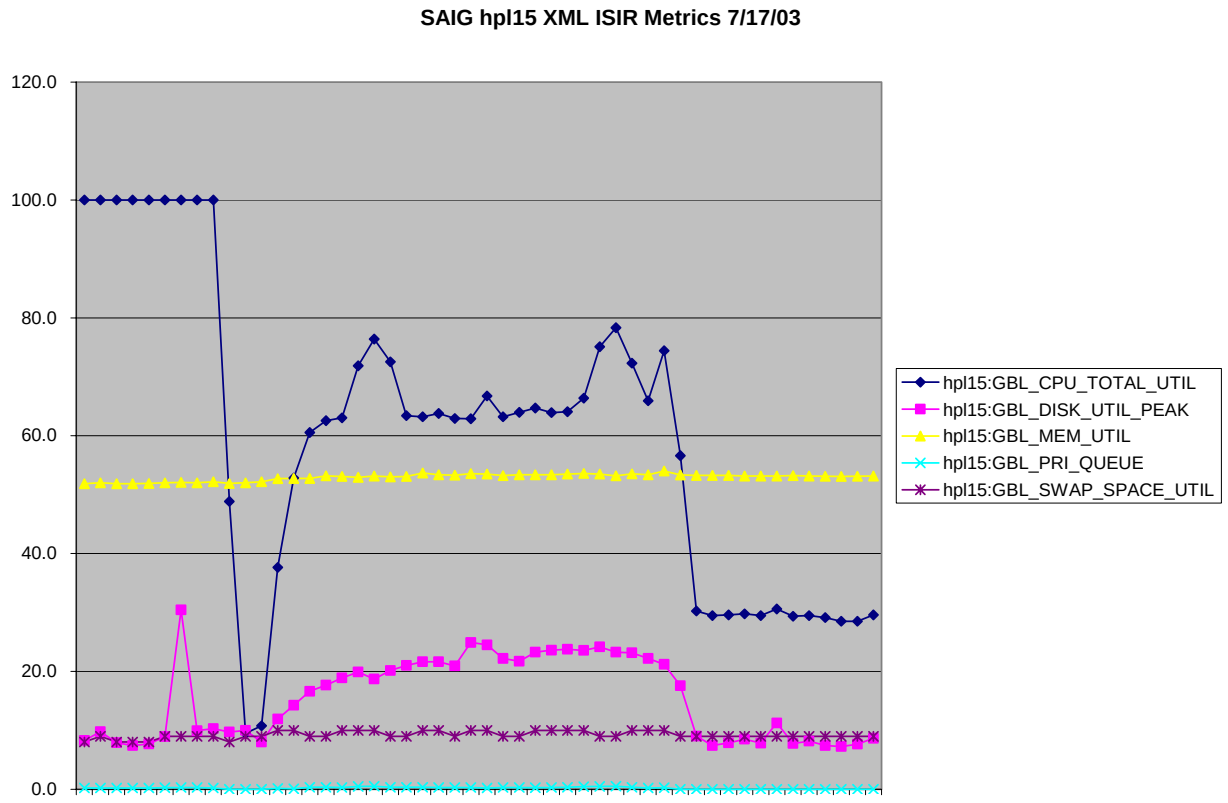

Table 3.10 – Cycle 1 - Global Metrics on HPL15 (SAIG server)

The Application CPU graph shows a break down of CPU utilization by process.

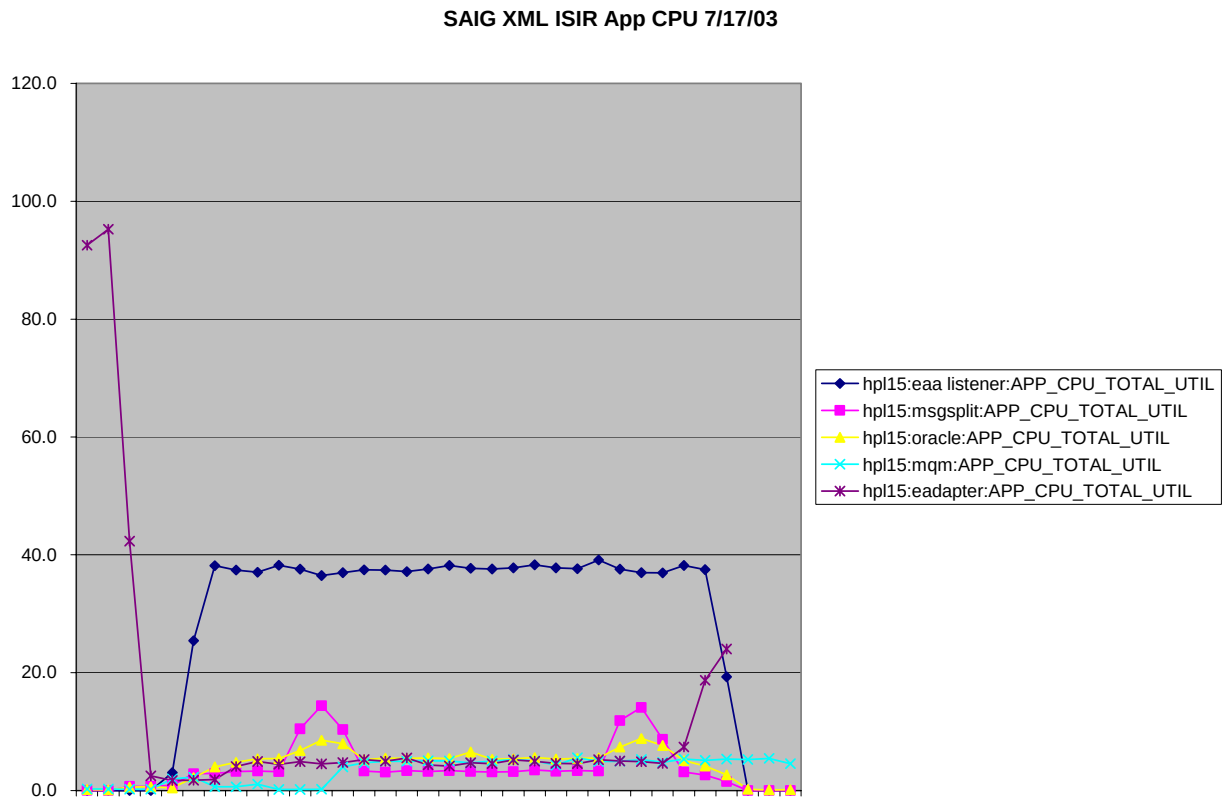

Table 3.11 – Cycle 1 - Application CPU breakdown on HPL15 (SAIG server)

3.3.3 Summary – Cycle 2

Test Outcome:
Cycle II of the SAIG XML ISIR performance test validated the application's three business processes. The application's processes are comprised of EAI-SAIG traffic, which was simulated for testing purposes, along with ISIR SAIG and Non-ISIR SAIG traffic in the form of FTP Sessions. The Cycle II test validated the system's ability to support 200 concurrent users sending varied file sizes, while sending and receiving 13 files per minute on the EAI test server. The transmission of fixed length ISIR file to SAIG was accomplished, as was the transmission of XML ISIR file.

Cycle Background:
Cycle II of the SAIG XML ISIR performance test established the system's capacity to run 200 concurrent users (FTP Sessions) while compressing and transmitting fixed length and XML ISIRs. EAI Bus traffic was simulated by sending and receiving 13 files per minute.

Hardware and Configuration:
Five test clients were used during the test: SU35E6, SU35E11, HPL8, HPN3 and EDENTER. HPL15 was used as the SAIG mailboxing and FTP servers. The EAI Bus was hosted on SU35E17. The five test clients send fixed length ISIR files to the mailbox and about 13 files per minute are then sent to the EAI Bus.

Detailed Description:
Cycle II of the SAIG XML ISIR performance test validated four test criteria in keeping with the Performance Test Plan.

The test set in motion 200 concurrent FTP Sessions to simulate a state of peak SAIG traffic. Forty-five minutes had elapsed before the FTP Session process was stabilized. As an outcome of the large number of FTP Sessions, the EAI process of sending and receiving 13 files per minute reached the breakeven point at approximately 175 FTP Sessions, prior to reaching the 200 concurrent FTP Session stabilization point. The EAI Bus performance, however, did not degrade further at the 200 concurrent FTP Session level, although the tendency for file queuing continued. Whereas, in Cycle I the EAI process of sending and receiving ranged between 17-18 seconds, the same process lasted one minute or longer for Cycle 2, which caused files to queue. The CPU peaked at 100% utilization, and memory was at 33% utilization for this part of the test. The priority queue hovered at the 4.0 mark for this segment of the test.

The fixed length ISIR is sent to HPL15 (SAIG server) from HPN3 (test client) to simulate the process employed by CPS. Hence, the test client compressed a 2.8GB file into 215MB, and then transmitted this file to the SAIG server. The compression rate was 92.5%. SAIG received the 215MB file from the test client, which was split into batch files, and distributed these files to mailboxes. The compression time was 11 minutes, while actual transmission time was 8 minutes; the file splitting process interval was 3 hours and 14 minutes. The CPU utilization for this part of the test was 100% on HPL15, while memory was approximately 33%. The priority queue was 4.0. The EAI file transfer rate remained high at one minute. Ample time was provided for the system to return to steady-state condition after each application of load.

The XML ISIR is sent to HPL15 (SAIG server) from HPN3 (test client) to simulate the process employed by CPS. The XML ISIR file in uncompressed form was 21.9GB, the compression process for this file spanned 52 minutes, while the compressed file measured 375MB. The compression rate was 98%. The XML ISIR file was transmitted from the test client (HPN3) to the SAIG server (HPL15) in 14 minutes. The duration of the message splitting, and distribution to mailboxes process for the XML ISIR was 3 hours and 7 minutes. The CPU utilization reached its maximum of 100%, while memory wavered at the 34% mark. The priority queue was steady at 4.0.

The goal of this test is to compare the performance of the system while it processed the fixed length ISIR and the XML ISIR file. SAIG has demonstrated the capacity to process both types of ISIR files.

Cycle II performance test achieved its objectives. 200 concurrent FTP Sessions were applied to the system, while EAI sent and received 13 files per minute, the fixed length ISIR and XML ISIR file were compressed on HPN3 (test client simulating CPS), transmitted to HPL15 (SAIG server), split into batches and stored in destination mailboxes. Running 200 concurrent users simulates the SAIG COD August peak, the XML ISIR file size represents March peak. The test simulated the August and March peak periods and represents a worst case scenario for the system to support. The system was slowed by the heavy load but succeeded in processing the files.

3.3.4 Data Analysis

The Global Metrics graph for HPL15 (SAIG server) demonstrates three key metrics: CPU utilization, memory utilization, and priority queue.

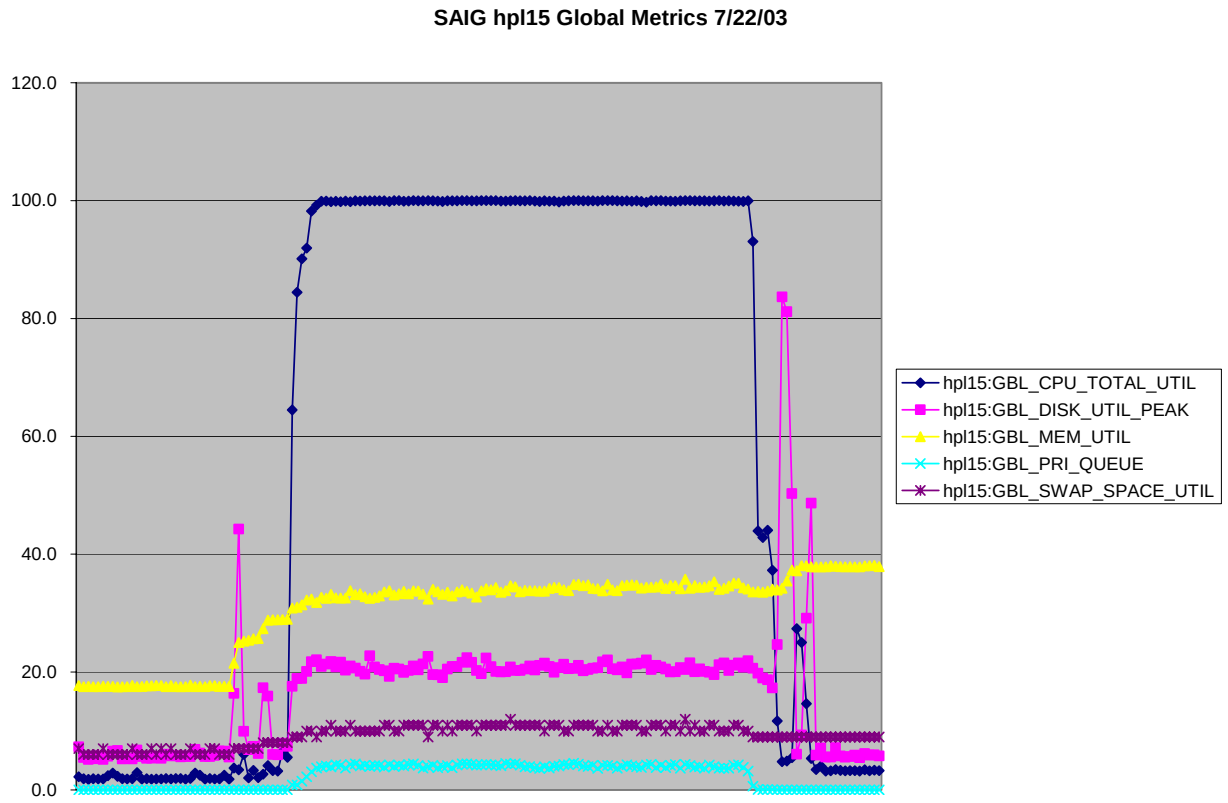

Table 3.12 – Cycle 2 - Global Metrics on HPL15 (SAIG server)

The Application CPU graph shows a break down of CPU utilization by process.

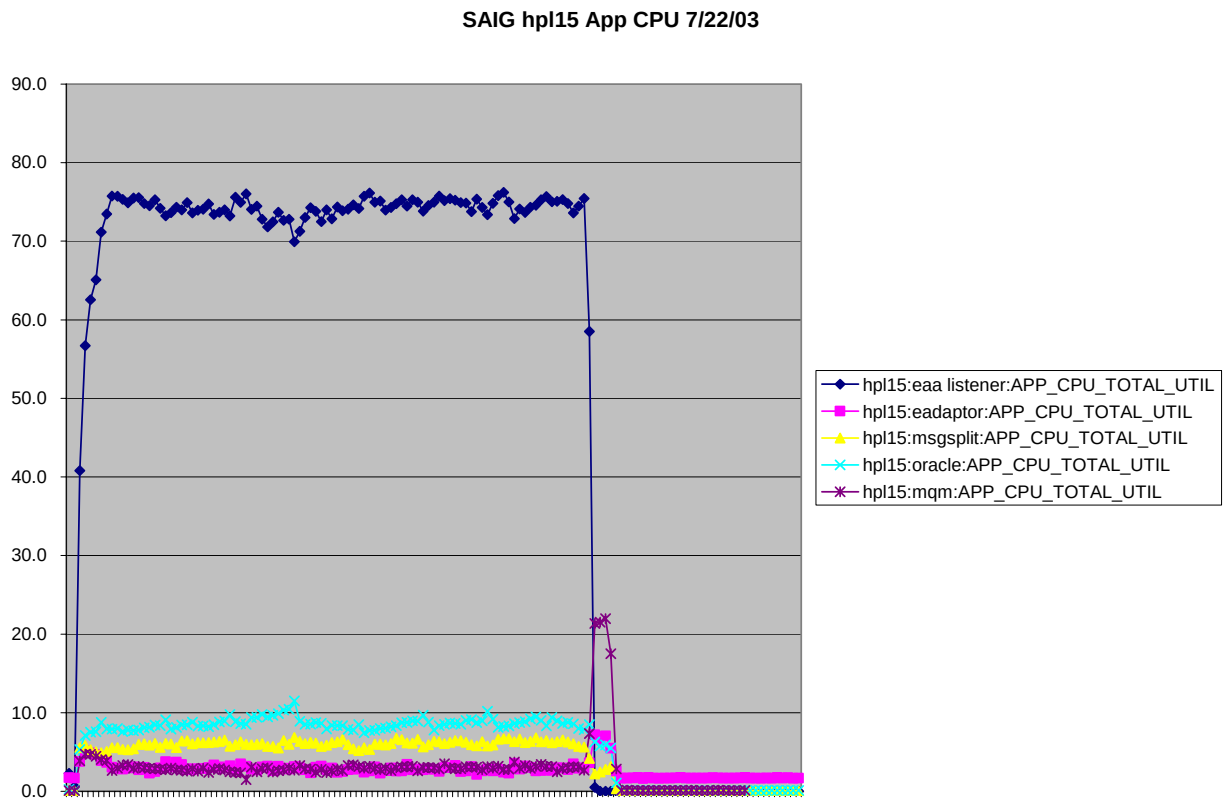

Table 3.13 – Cycle 2 - Application CPU breakdown on HPL15 (SAIG server)

3.3.5 Summary –Cycle 3

Test Outcome:

Cycle III of the SAIG XML ISIR performance test validated the application's three business processes. The application's processes are comprised of EAI-SAIG traffic, which was simulated for testing purposes, along with ISIR SAIG and Non-ISIR SAIG traffic in the form of FTP Sessions. The Cycle III test validated the system's ability to support 100 concurrent users sending varied file sizes, while sending and receiving 13 files per minute on the EAI test server. The fixed length and XML ISIR files were 15% larger than those employed by the previous cycles. The transmission of fixed length ISIR file to SAIG was accomplished, as was the transmission of XML ISIR file.

Cycle Background:

Cycle III of the SAIG XML ISIR performance test established the system's capacity to run 100 concurrent users (FTP Sessions) while compressing and transmitting fixed length and XML ISIRs that are 15% larger than files used in previous cycles. EAI Bus traffic was simulated by sending and receiving 13 files per minute.

Hardware and Configuration:

5 test clients were used during the test: SU35E6, SU35E11, HPL8, HPN3 and EDENTER. HPL15 was used as the SAIG mailboxing and FTP servers. The EAI Bus was hosted on SU35E17. The five test clients send fixed length files to the mailbox and about 13 files per minute are then sent to the EAI Bus.

Detailed Description:
Cycle III of the SAIG XML ISIR performance test validated four test criteria in keeping with the Performance Test Plan.

The test spawned 100 concurrent FTP Sessions to simulate a state of peak SAIG traffic. The EAI process of sending and receiving13 files per minute was initiated simultaneously. The CPU peaked at 100% utilization at initial ramp up and stabilized at the 75% mark. Memory utilization remained at 40%. The priority queue was well below the 3.0 and remained flat at zero for the entirety of the test. The EAI file processing time ranged between 20 – 26 seconds throughout the entire test.

The fixed length ISIR is sent to HPL15 (SAIG server) from HPN3 (test client) to simulate the process employed by CPS. The test client was supplied a 15% larger file to compress, hence, the 3.22GB fixed length ISIR was compressed to 247MB. The compression rate was 92.5%. Subsequently, the test client transmitted the fixed length ISIR to the SAIG server, which received the 247MB file from the test client, split the ISIR into batch files, and distributed these files to SAIG mailboxes. The compression time was 10 minutes, while the transmission time was 54 seconds. The duration of the file splitting process was 18 minutes. The CPU utilization for this portion of the test, the fixed length ISIR processing, peaked at 82% on HPL15, while memory was approximately 40%.

After each business process was introduced, the system was provided with sufficient time to return to steady-state condition.

The XML ISIR is sent to HPL15 (SAIG server) from HPN3 (test client) to simulate the process employed by CPS. The XML ISIR file in uncompressed form was 24.9GB, but following the compression by the test client it was reduced to 428MB. The compression process for this file spanned one hour and one minute. The compression rate was 98%. The XML ISIR file was transmitted from the test client (HPN3) to the SAIG server (HPL15) in 1 minute and 50 seconds. The duration of the message splitting, and distribution to mailboxes process for the XML ISIR was 17 minutes. The CPU utilization reached its maximum of 72%, while memory wavered at the 40% mark.

Contrary to Cycle I, Cycle III performance test employed file sizes 15% larger for both the fixed-length ISIR and the XML ISIR file types. The increase in file size yielded different performance results. In general, the larger file size consumed slightly more system resources in terms of CPU, although the number of concurrent FTP Sessions was the same at 100. The same was true for the EAI send/receive process, in that the duration of the file transfer rate increased marginally.

Specifically, CPU utilization differential between Cycle I fixed-length ISIR and that of Cycle III was an 8% increase in the latter test. Similarly, the XML ISIR file consumed 5% more CPU utilization for Cycle III than Cycle I.

Cycle III performance test achieved its objectives. 100 concurrent FTP Sessions were applied to the system, while EAI sent and received 13 files per minute, the 15% larger fixed length ISIR and XML ISIR files were compressed on HPN3 (test client simulating CPS), transmitted to HPL15 (SAIG server), split into batches and stored in destination mailboxes.

3.3.6 Data Analysis

The Global Metrics graph for HPL15 (SAIG server) demonstrates three key metrics: CPU utilization, memory utilization, and priority queue.

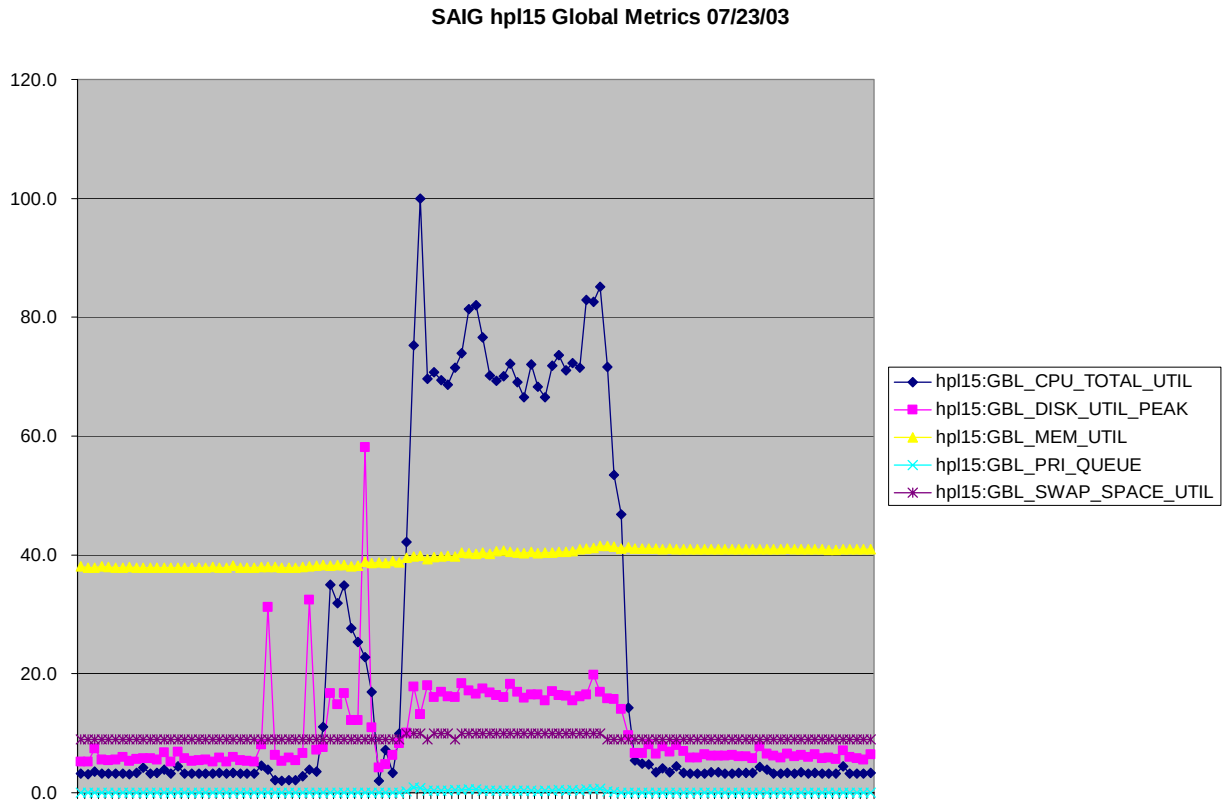

Table 3.14 – Cycle 3 - Global Metrics on HPL15 (SAIG server)

The Application CPU graph shows a break down of CPU utilization by process.

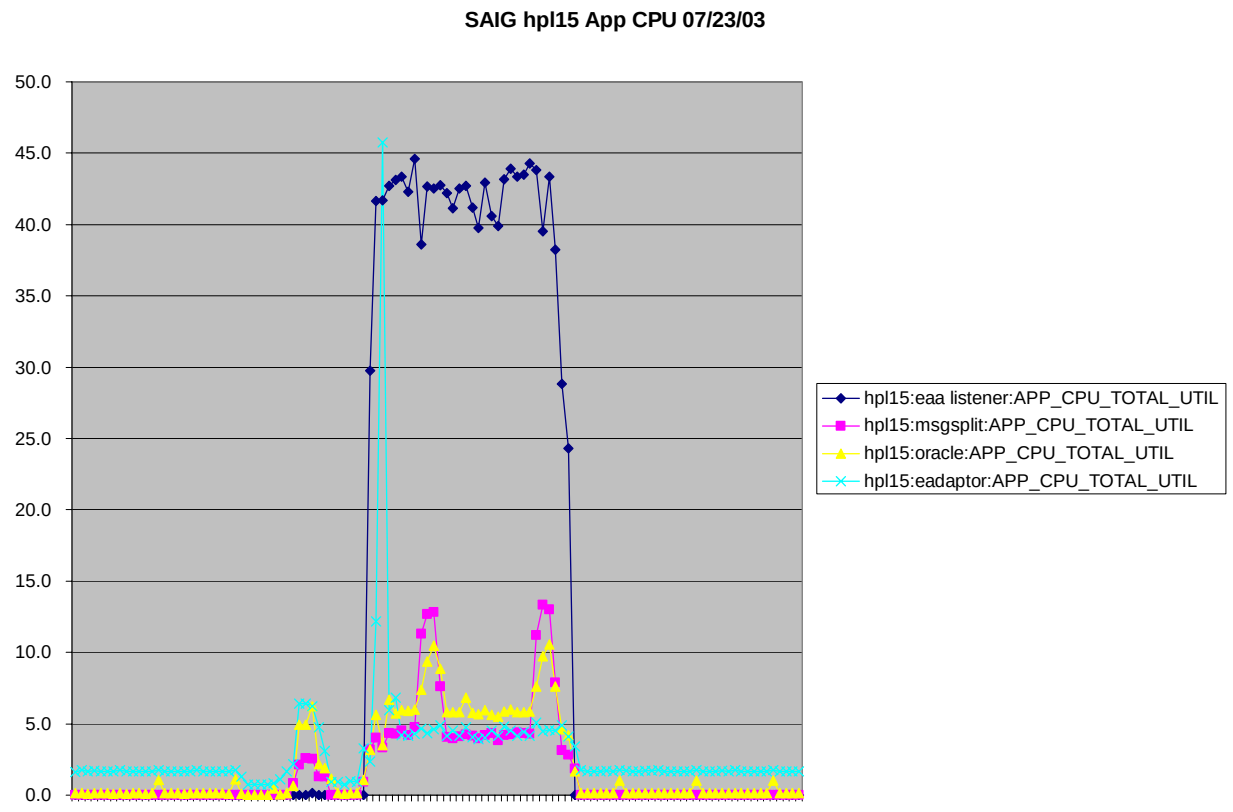

Table 3.15 – Cycle 3 - Application CPU breakdown on HPL15 (SAIG server)

3.3.7 Risks

The test data that was used for the ISIR files were not production data. Thus, it possible that repetitive data may have been included, which may have contributed to a higher compression rate for the performance test.

3.4 SAIG Storage Analysis

3.4.1 Summary

Moving the ISIR to XML will take more storage space on the SAIG server. There are two ways to mitigate the size increase: better compression and shorter archiving durations. If XML compression were to reach 98%, there would be no size difference. While this is possible in some cases, currently compression is typically in the 90% to 95% range. Reducing the duration batches, particularly ISIR batches are archived on SAIG, would reduce the amount of storage space required by SAIG. In the event that compression rates remain constant with today's production data and no changes are made to the archiving produces on SAIG, additional storage space will be required.

Compression
Improving compression can have an impact on the amount of storage space required at SAIG. There are compression techniques which are specifically designed for XML data. The XML compression algorithms provide economies of scale, meaning the larger the original batch, the higher the compression rate. This is because the larger batches have a larger percentage of repeating bytes. However, it is important to note that these higher compression rates do come at a cost. They require more CPU, either time or larger processors, to complete the compression process and the decompression process.

Archiving
Changing the rules for archiving is another approach that could alleviate the need for increases in required disk space. Currently ISIR files are archived for 180 days. Scaling this number back to fewer days, will lead to reductions in how much storage capacity is required. The addition of the ISIR Data Mart, may allow for changes in how SAIG archives the ISIR message classes. Currently, SAIG does have the capability to reduce the amount of time specific message classes are stored within SAIG.

3.4.2 Data Analysis

Non-Archived Data and Archived Data Evaluation
In order to evaluate how much additional space is required by SAIG, both the archived and non-archived volumes must be evaluated. Currently, SAIG keeps data in a non-archived state for approximately ten days. SAIG keeps data in an archived state for approximately 180 days. The non-archived data and archived data are stored on different volumes. For operation, SAIG would like to maintain 50% disk utilization for non-archived data, and 70% disk utilization for archived data. The following is an example scenario. Assume SAIG makes no changes to the archiving procedures for XML ISIRs, and the XML ISIR Compresses at a rate of 90%, in order to maintain the optimum file system utilization percentages, SAIG would approximately need an additional 220 GB (Non-Archived 90 GB + Archived 283 GB – Current Allocation 153 GB) of space. The total required space to store non-archived data and archived data would be 373 GB (Non-Archived 90 GB + Archived 283 GB).

Non-Archived Data
Table 3.16 compares the storage requirements for non-archived XML ISIR batches based on an XML compression rate of 90% and 95%; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Archive Scenario	Compression Percentage	Total Storage Space (GB)	Used Storage Space (GB)	Utilization Percentage	Total Storage Space Used with XML ISIR (GB)	XML ISIR Utilization Percentage	Total Storage Space Required to return to Reasonable Utilization Percentage (GB) – 50%
Non-Archived	90%	55.66	15.59	28%	45.14	81%	90.28
Non-Archived	95%	55.66	15.59	28%	26.51	48%	53.01

Table 3.16 – SAIG Storage Requirements (Non-Archived Data)

Archived Data

Table 3.17 compares the storage requirements for archived XML ISIR batches based on an XML compression rate of 90% and 95%. This table shows how the storage requirements fluctuate based on changes to the length of time XML ISIR batches are stored in the SAIG archives; the data and calculations used to generate these percentages can be found in Appendix B: XML ISIR Calculations.

Archive Scenario	Compression Percentage	Total Storage Space (GB)	Used Storage Space (GB)	Utilization Percentage	Total Storage Space Used with XML ISIR (GB)	XML ISIR Utilization Percentage	Total Storage Space Required to return to Reasonable Utilization Percentage (GB) – 70%	Net Increase Increase (Archived + non-Archived – Current Capacity) (GB)
180 Day Archive	90%	97.66	68.36	70%	197.99	203%	282.84	219.80
180 Day Archive	95%	97.66	68.36	70%	116.26	119%	166.08	65.78
90 Day Archive	90%	97.66	68.36	70%	116.43	119%	166.32	103.29
90 Day Archive	95%	97.66	68.36	70%	75.56	77%	107.94	7.64
30 Day Archive	90%	97.66	68.36	70%	62.05	64%	88.64	25.61
30 Day Archive	95%	97.66	68.36	70%	48.43	50%	69.18	N/A

Table 3.17 – SAIG Storage Requirements (Non-Archived Data)

3.4.3 Risks

Assuming that compression XML Compression rates remain between 90% and 96%, and no changes are made to the archiving process for XML ISIRs, SAIG will require additional storage space. How much additional storage space is a combination of what the actual XML ISIR compression rate will be, combined with how much of a storage cushion SAIG server needs to have.

3.5 XML ISIR Download Analysis – VDC Perspective

3.5.1 Summary

There are many network segments that exist between an SAIG end-user and the SAIG system located at the VDC. These network segments include the phone, DSL, or T1 lines between the institution and its ISP; the networking infrastructure of the ISP itself; and intermediate network segments across the Internet before it reaches the VDC. The "VDC Perspective" of the ISIR download analysis refers to the VDC networking pipe between SAIG and the Internet. It is the only segment of the network between the end-user and SAIG that FSA has any control over. Therefore, it is very important that FSA ensures the networking capacity of the VDC to the Internet is more than adequate to handle spikes in network traffic.

The primary connection of the VDC to the Internet is 155 Mbps OC3 networking. The analysis in this section shows that this connection provides ample networking bandwidth for the needs of FSA traffic through SAIG. This is shown through some basic calculations, as well as the results of monitoring network traffic during the performance test runs.

3.5.2 Data Analysis

Calculated Results

A straightforward set of calculations shows that the 155 Mbps pipe provides more than an adequate amount of networking bandwidth to support the ISIR file transfer through SAIG. To start, assume an extremely large XML ISIR file at about 500 MB, compressed. This would represent a file with many more records than CPS currently produces. Then assume all of the potential recipients of those ISIR batches connect to SAIG rapidly, to download all of the data in a very short time of 1 hour, or an even shorter time of ½ hour.

The following table traces through calculating the networking bandwidth needed to move the data file to the Internet. First it converts the file size from MB to bits. Next, it calculates the average download rate the Internet "pipe" would need to support in order to allow everyone to download their ISIR batches in 1 hour, and in ½ hour. It shows that, even if everyone downloaded their ISIRs (for a total of 500 MB of data) over ½ of an hour, that extremely aggressive load would only use 1.5% of the total bandwidth of the 155 Mbps OC3 pipe.

File Size (MB)	File Size (Bytes) factor of : 1,048,576 Bytes / MB	File Size (Bits) factor of : 8 Bits / Byte	Download Rate (Mbps) for 1 hour of time to download factors of: 3600 sec/hr and 1,000,000 bits/Mbit	% of Bandwidth utilized factor of: 155Mbps available in OC3 Pipe	Download Rate (Mbps) for 1/2 hour of time to download factors of: 3600 sec/hr and 1,000,000 bits/Mbit	% of Bandwidth utilized factor of: 155Mbps available in OC3 Pipe
500	524,288,000	4,194,304,000	1.165	< 1%	2.330	1.5%

Table 3.18 – Sample ISIR Bandwidth Requirements Calculations and Results

Measured Results

During the ISIR SAIG Performance test runs, actual networking statistics were captured to verify the behavior as calculated above. The chart below shows the average total network traffic into and out of SAIG during the test. It shows two peaks, which represent the network traffic while the fixed length ISIR and the XML ISIR were run through the system. The bandwidth utilization for the 2 files were about 4.4 Mbps and 7.3 Mbps, respectively. This is consistent with the calculations above because the file loading processes took about 10-15 minutes each.

To summarize, this file shows that when SAIG is moving data in and out as fast as it can, the networking utilization from the perspective of the VDC to the Internet is still only about 5% of the total OC3 capacity. Furthermore, all of the ISIR traffic for the day can be moved at this bandwidth, in about 15 minutes. All of these determinations indicate that the 155 Mbps OC3 connection from the VDC to the Internet provides more than enough networking capacity for ISIR traffic.

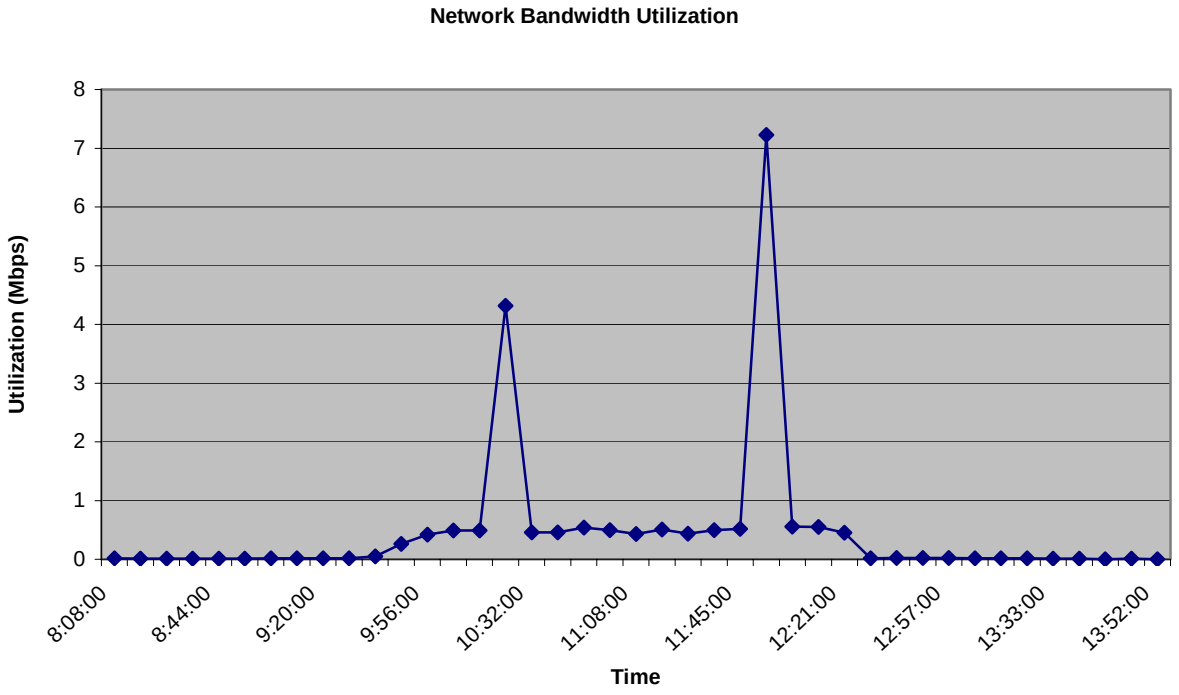

Table 3.19 – Actual Network Bandwidth Utilization during ISIR Performance Test Run

3.5.3 Risks

The risk with this analysis is that the actual networking hardware from the VDC to the Internet – the physical OC3 cable, the routers, the firewalls, etc., - were not tested during the performance test. Therefore, the results presented above do not represent traffic that actually ran through the VDC pipe to the Internet. The tests were not conducted through the VDC-Internet pipe, because it would have been cost-prohibitive to set up the necessary testing environment. Instead, networking statistics were collected between the test machines, all connected on the VDC internal network. This approach was determined to be a reasonable alternative, since the actual networking infrastructure was stress tested before, during the FAFSA performance testing cycles at the end of 2002. Also, the networking has been continually monitored, and shown to be operating efficiently and well below its capacity. For these reasons, the risk of underestimating the impact of the ISIR traffic on the VDC-Internet pipe is negligible.

3.6 XML ISIR Download Analysis – End User Perspective

3.6.1 Summary

Assuming that compression rates remain inline with today's data analysis, the larger size of XML ISIRs will cause download times to increase. The analysis in this section assumes that

there are no issues with SAIG Server handling the larger XML ISIR batch sizes and there are no issues with the VDC pipe to the internet handling the larger XML ISIR batch sizes.

3.6.2 Data Analysis

Since compression rate is directly linked to compressed batch size, and since compressed batch size is directly linked to download times, the analysis of download times will be based on different compression rates. The evaluated compression rates will be 85% for fixed length batches and for XML Batches 90% and 95% will be evaluated. It should be noted that as XML compression rates approach 98% the compressed batch sizes of XML batches and fixed length batches are roughly equal; hence, download times at the percentage would also be equivalent.

Table 3.20 evaluates the download times for the fixed length batch sizes evaluated in the earlier sections of the SAIG Capacity Analysis. This table assumes an 85% compression rate for fixed length ISIR batches and all times are measured in seconds; the data and calculations used to generate these download times can be found in Appendix B: XML ISIR Calculations.

Batch Distribution			Download Times - Seconds							
Number of Detail Records in Batch	**Percentage of Batches Range**	**Fixed Length Size (bytes)**	**14.4 kbps**	**28.8 kbps**	**56.6 Kbps**	**64 kbps ISDN**	**128 kbps ISDN**	**640 Kbps DSL**	**T1/ 1.5 Mbps DSL**	**T3/DS3**
19	54.54%	8,607	7	3	2	1	1	<1	<1	<1
123	36.31%	48,019	39	19	10	9	4	1	<1	<1
254	4.84%	98,342	80	38	20	17	9	2	1	<1
1,295	3.49%	498,243	405	195	101	88	45	9	4	<1
2,596	0.45%	998,022	812	390	203	177	89	18	7	<1
13,009	0.34%	4,998,176	4,068	1,952	1,017	887	448	90	37	1
30,000	0.03%	11,525,268	9,379	4,502	2,345	2,046	1,033	207	85	3

Table 3.20 – ISIR Download Times – Fixed Length Record (85% Compression)

Table 3.21 evaluates the download times for the XML batch sizes evaluated in the earlier sections of the SAIG Capacity Analysis. This table assumes a 90% compression rate for XML ISIR batches and all times are measured in seconds; the data and calculations used to generate these download times can be found in Appendix B: XML ISIR Calculations.

Batch Distribution			Download Times – Seconds							
Number of Detail Records in Batch	**Percentage of Batches Range**	**XML Size (bytes)**	**14.4 kbps**	**28.8 kbps**	**56.6 kbps**	**64 kbps ISDN**	**128 kbps ISDN**	**640 Kbps DSL**	**T1/ 1.5 Mbps DSL**	**T3/DS3**
19	54.54%	37,751	31	15	8	7	3	1	<1	<1
123	36.31%	243,920	199	95	50	43	22	4	2	<1
254	4.84%	503,638	410	197	102	89	45	9	4	<1
1,295	3.49%	2,567,391	2,089	1,003	522	456	230	46	19	1
2,596	0.45%	5,146,582	4,188	2,010	1,047	914	461	92	38	1

Batch Distribution			Download Times – Seconds							
Number of Detail Records in Batch	Percentage of Batches Range	XML Size (bytes)	14.4 kbps	28.8 kbps	56.6 kbps	64 kbps ISDN	128 kbps ISDN	640 Kbps DSL	T1/ 1.5 Mbps DSL	T3/DS3
13,009	0.34%	23,790,129	20,988	10,074	5,247	4,579	2,311	462	191	7
30,000	0.03%	59,474,222	48,400	23,232	12,100	10,560	5,328	1,066	440	15

Table 3.21 – ISIR Download Times – XML Record (90% Compression)

Table 3.22 evaluates the download times for the XML batch sizes evaluated in the earlier sections of the SAIG Capacity Analysis. This table assumes a 95% compression rate for XML ISIR batches and all times are measured in seconds; the data and calculations used to generate these download times can be found in Appendix B: XML ISIR Calculations.

Batch Distribution			Download Times – Seconds							
Number of Detail Records in Batch	Percentage of Batches Range	XML Size (bytes)	14.4 kbps	28.8 kbps	56.6 kbps	64 kbps ISDN	128 kbps ISDN	640 Kbps DSL	T1/ 1.5 Mbps DSL	T3/DS3
19	54.54%	18,875	15	7	4	3	2	<1	<1	<1
123	36.31%	121,960	99	48	25	22	11	2	1	<1
254	4.84%	251,819	205	98	51	45	23	5	2	<1
1,295	3.49%	1,283,696	1,045	501	261	228	115	23	10	<1
2,596	0.45%	2,573,291	2,094	1,005	524	457	231	46	19	1
13,009	0.34%	12,895,065	10,494	5,037	2,624	2,290	1,155	231	95	3
30,000	0.03%	29,737,111	24,200	11,616	6,050	5,280	2,664	533	220	8

Table 3.22 – ISIR Download Times - XML Record (95% Compression)

3.6.3 Risks

The download times here are based on calculated compressed batch sizes, not actually compressed batches. The download times are not based on actual downloads, but are calculated based on estimated download rates based on ISP type.

Appendix A: SAIG Technical Infrastructure

Appendix B: XML ISIR Calculations