AUTHORED BY
KAZUMI SAITO, MAKOTO KOIZUMI

DECEMBER 2013–FINAL

FOR
XBRL INTERNATIONAL, INC.

# PERFORMANCE TESTING PROCESS FOR LARGE XBRL INSTANCE PROCESSING

A Guidance Document



## ABSTRACT

Large XBRL Instance Processing, dealing with XBRL instances exceeded GB per file, was a new topic of discussion for the first time at the 25th XBRL International Conference in Yokohama in December 2012. So far, the international XBRL community have studied data creation / validation / usage / taxonomy architecture, and software. Some major users such as banking and tax authorities, however, are considering XBRL applications which require Large XBRL Instance Processing, and solutions in this theme are increasingly urgently required.

This paper discusses the proof of concept methodology related to performance measurement of an XBRL processing engine, in the case of the Bank of Indonesia's XBRL application predicated on Large XBRL Instance Processing. The paper will also refer to Taxonomy Architecture predicated on Large XBRL Instance Processing.

# TABLE OF CONTENTS

## INTRODUCTION– LARGE XBRL INSTANCE

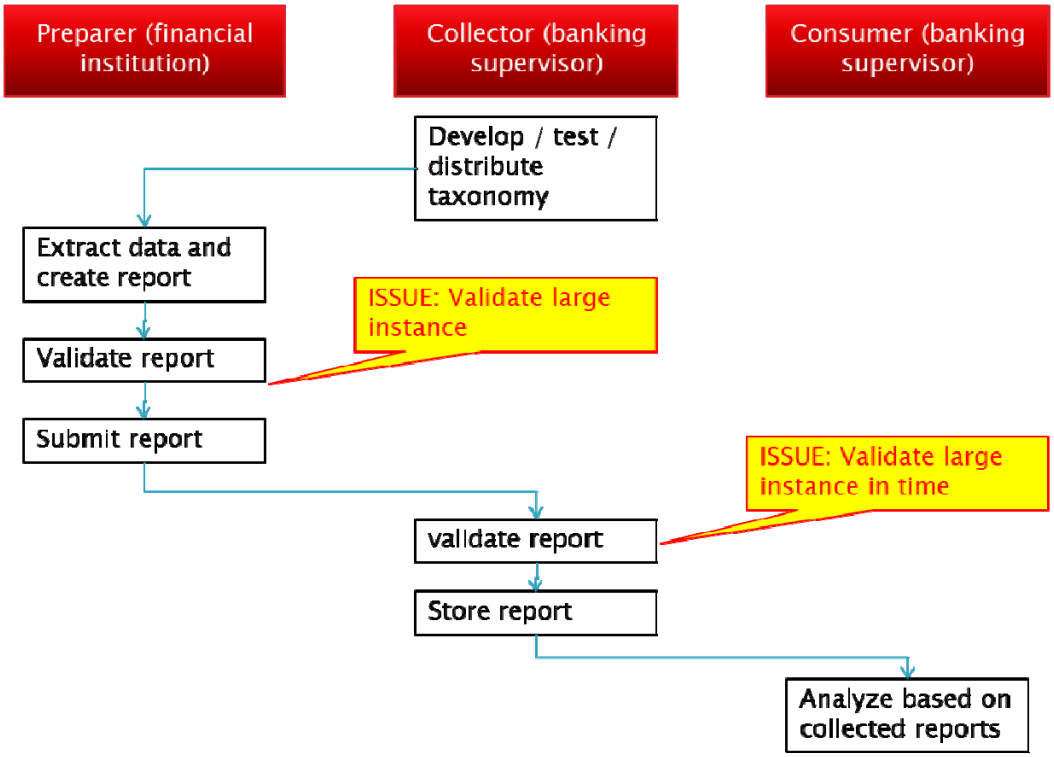
Large XBRL Instance is seen in reporting of statement information and detailed data such as in loan detail information, position reporting of derivatives products, rating information reporting of credit rating institutions, with data size of several GB per file or multi-dimensional instances. In these reports, reporting is carried out in the record format pivot X column n known as Open table, where notonly one pivot item has around 100 items, but also where records exceed 1 million columns. Further, if Dimension is used, the number of contexts exceeds several tens of thousands

Open table has 3 formats: Tuple, Open context, and Typed Dimension; the differences are shown in the table below. In bank supervisory reporting, most supervisory authorities provide XBRL Formula for data checking purposes

| | XBRL specification: tuple | XBRL specification: open context | Dimension specification: typed dimension |
|---|---|---|---|
| Taxonomy | Example: Taxonomy defines columns as primary items. Tuple (complex type) references (and binds) primary items as columns in one table. | Example: taxonomy may (but not necessarily) define a simple or complex XML item (or type) that is used in the instance document open context. | Example: taxonomy defines the typed dimension container as a simple or complex XML item (or type) that is used in the instance document as part of dimensional context definition. |
| Report | Pieces of data from cells in the open table are reported as values of facts. Facts are bound by their placement within the tuple tag as according to the tuple definition from the taxonomy.<br>The infinity factor is provided by infinite number of tuple (group of facts) instantiations.<br>Tables open from row and column perspective are addressed through tuple nesting. | Pieces of data from cells in the open table are most commonly reported as values of facts. Facts are bound by the context ID that references multiple facts to a single context with specific open context component which may be any XML-valid construct.<br>The infinity factor is provided by infinite number of contexts that are distinguished by open context components. | Pieces of data from cells in the open table can be reported either as values of facts or values of explicit or typed dimensions in contexts or in any combination. Facts are bound by the context ID that references multiple facts to a single context with specific typed dimension context component value.<br>The infinity factoris provided by infinite number of contexts that are distinguished by typed dimension context component. |

## EXPECTED BOTTLENECK IN LARGE XBRL INSTANCE PROCESSING

When processing a Large XBRL Instance, the main issue[1] is the memory volume and processing time needed to process the instance load and the Turn Around Time of the XBRL Formula Validation. For example, loading an XBRL file into DOM parser memory requires a memory of 10 to 60 times the file size.




In addition, Formula takes extra time, which may significantly lengthen overall Turn Around Time.

Currently it is desirable that performance testing is carried out at the evaluation stage when implementing XBRL based on Large XBRL Instance Processing, due to (1) constant improvement of hardware performance and the hardware and software components needed for Large XBRL Instance Processing such as XBRL processing engine performance, and (2) TAT requirements e.g. validation reporting parameters etc differing between each supervisory authority.

The next section deals with performance evaluation goals, criteria, and steps.

[1] It has been pointed out that the time needed to transfer files from the submitter to the supervisory authority, and data usage within the supervisory authority, may give rise to problems. File transfer time, however, can be cut to 1/100$^{th}$ or less as XBRL itself is a text file and therefore has high compression efficiency.

## GOALS AND STEPS FOR PERFORMANCE TESTING

The following points are examined at the XBRL application examination stage.

- The taxonomy structure for obtaining the optimal TAT for XBRL application object reporting
    - Tuple base taxonomy
    - Open context base taxonomy
    - Typed Dimension base taxonomy
- The XBRL Formula implementation method for obtaining the optimal TAT for XBRL application object reporting
- The processing method for obtaining the optimal TAT for XBRL application object reporting
    - Sequential processing on one large server
    - Multi-stage parallel processing in several small servers
- Confirming rules for the creation of files to be submitted which do not impair performance
- Evaluation and selection of the XBRL processing engine to realise the processing mentioned above.
- Evaluation of the support provided by the vendor the XBRL processing engine

Evaluation is considered in the following cycle:

1. Call for participation (to the Technology Provider)

   Invite performance appraisals from providers of XBRL processing engine and solution technology. When using production data in performance appraisal, all considerations such as NDA are advised

2. Tentative TAT targets

   Tentative setting of TAT targets covering current processing time, maximum acceptable TAT, etc

3. Taxonomy, test data preparation

   Prepare taxonomy, XBRL Formula, Instance documents to be used in the test. Instance considerations to cover maximum and average business size, and by a plurality of scenarios to obtain the most realistic measurements possible.

4. Performance appraisal

   Verification of taxonomy structure, implementation of XBRL formula, TAT for each processing method.

5. Evaluating results

   Here, the results obtained are used to review feasible TAT, and to make decisions on taxonomy architecture, XBRL formula, and system architecture.

## CASE STUDY – CASE OF BANK INDONESIA

Bank Indonesia is the central bank of Indonesia, and one of the first banks in the world to study the application of Large XBRL Instance Processing. Bank Indonesia collates data for offsite monitoring of 34 Shariah Bank (571 filers) as of October 2013, and in the case of loan statement data collection, collects data from multiple columns of CSV files where 1 column can be as much as several GB

The following points are confirmed when deciding on applying XBRL to the collated data format

1. The relative performance merits between different implementation methods on the same table

   (Typed dimension table or tuple table)

2. The relative performance merits between writing methods when applying different formula descriptions to the same check table.

3. The relative performance merits of using multiple XBRL processing engines in the same environment

4. Confirmation of feasible TAT and necessary hardware resources

Validation takes the following steps

### Step 1: Compare 2 different taxonomy architectures (April 2012, 2 weeks offsite testing)

Taxonomy, formula, and instance documents are created for tuple and typed dimension for the same open table and business rule with the aim of measuring processing time in differing taxonomy architectures. Load and processing time between the taxonomy architectures are sent to all of the technology providers who expressed an interest in participating in the testing, and processing time reports are received from each.

After the taxonomy architecture decision in Step 1, extra formula is created based on tuple based taxonomy aimed at measuring the processing time of each formula. Processing times between formula implementations are sent to all of the technology providers who expressed an interest in participating in the testing, and processing time reports are received from each.

### Step 2: Performance Test by short listed vendors (June 2012, 1 week offsite testing)

To conduct the performance test, BI asked vendors to test the performance using multiple instances.

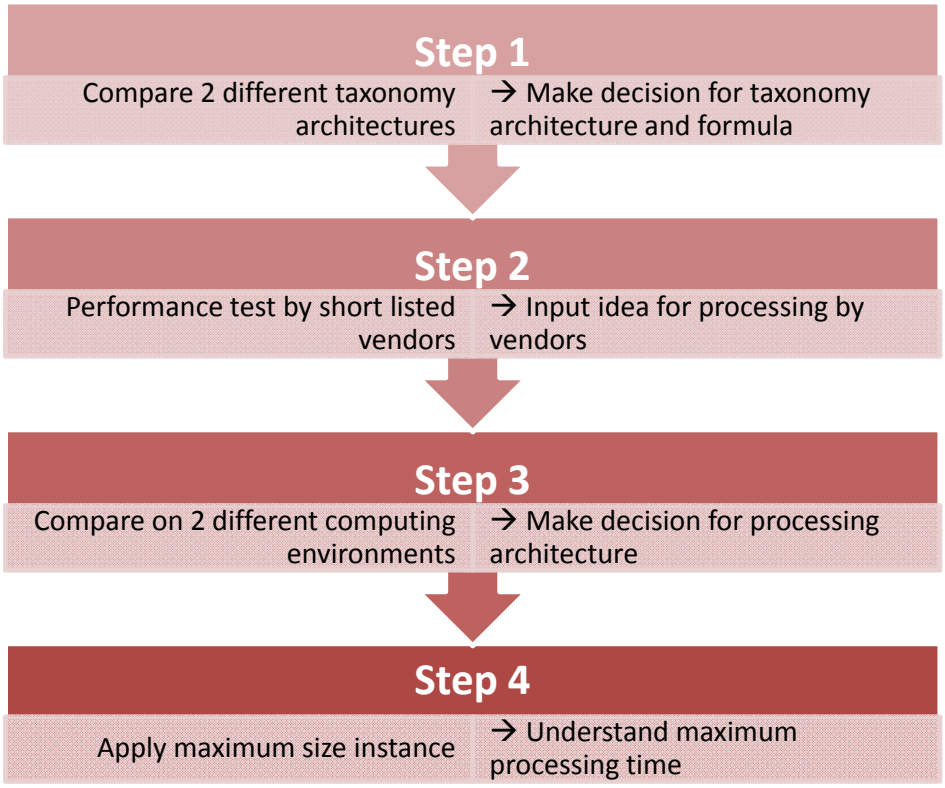
### Step 3: Compare on 2 different computing environments (October 2012, 2-3 days onsite testing)

The measurement results obtained in Step 1 and Step 2 are evaluated in the same computing platform, with a performance evaluation aimed at measuring the processing time difference between sequential processing on a single server and parallel processing on multiple servers, carried out onsite.

### Step 4: Apply maximum size single instance (October 2012, 2-3 days onsite testing)

This is measurement using a single instance with the maximum expected size of 14GB and 100 instance files averaging 100KB. Not only loading time and formula processing time, also the processing time it takes to load

to the database is measured. By using the same computing environment and data size as actually used, a final determination can be made on validity of scale and processing method of the production system.




## LARGE XBRL INSTANCE PROCESSING AWARE TAXONOMY ARCHITECTURE

A classic XBRL discussion– tuple or dimension in Large XBRL Instance processing

Large XBRL Instance processing contains multiple records in one instance. The structure of the record (thus, taxonomy architecture) has a large effect on the performance. What Bank Indonesia tested here was a method using tuple and a method using typed dimension. This document has not discussed actual performance figures, but when creating instances with both tuple and typed dimension, the instance using typed dimension is 1.5x larger than the instance using tuple. This is largely due to the context created in each record. Also, the performance of the applying of formula linkbase by load shows a marked preference for instance using tuple.

Bank Indonesia is expected to receive instances exceeding 10GB, so will therefore utilise applications using tuple.

# NOTE FOR LARGE XBRL INSTANCE FILE CREATION FOR PERFORMANCE TEST

## TAXONOMY AND INSTANCE CREATION PERIOD MUST BE CONSIDERED IN PROJECT TIMELINE

When creating a large XBRL Instance file for performance testing, efficiency is optimised by converting actual reporting data into XBRL. However, when planning the performance test timeline, consideration must be given to the time needed to create taxonomy, formula, and Large XBRL Instance files.

## SECRET DATA SHOULD BE CODIFIED

When the Large XBRL Instance file sent to the technology provider contains such commercially confidential data as loan balances of borrowers or financial institutions, such data should be converted to dummy code before sending. Creating and publishing dummy data advances the future development of high processing performance software, leading to receiving better proposals as a result.

## COLLABORATION WITH TECHNOLOGY PROVIDERS FROM EARLY PHASE

Due to the complexities of processing time and verification, Large XBRL Instance test data is hard to adequately verify before supplying to the vendor, and may turn out different from that intended. Therefore it is essential to take verification seriously, closely collaborating with the vendor on data accuracy from the first stage of verification.

# REFERENCE

**Symposium XBRL25:**

Large XBRL Instance Processing Symposium, Ashu Bhatnagar, XII, Michal Piechocki, BR-AG

http://archive.xbrl.org/25th/sites/25thconference.xbrl.org/files/TECH2Large%20instances%20session.pdf

**Notes on the Processing of Large XBRL Instances 1.0**

Working Group Note 31 October 2012

http://www.xbrl.org/WGN/Large XBRL Instance-processing/WGN-2012-10-31/Large XBRL Instance-processing-WGN-WGN-2012-10-31.html

# ACKNOWLEDGEMENT

Mr. Yudi Muliawirawan, Bank Indonesia

Ms. Laras Ayutirta, Bank Indonesia

## About Author

Kazumi Saito is Manager at Fujitsu Limited. He worked at Software division and in charge of XBRL related software product management. He has participated Large Instance Processing Proof of Concept held by Bank Indonesia. He is a member of XBRL Japan Financial and data consumption committee and Development Committee.

Makoto Koizumi is Managing Director at Mercury and Earth, KK. He worked for several XBRL taxonomy and disclosure platform development projects in and outside of Japan. Mr. Koizumi is Chair of the XBRL International Best Practices Board (2011 -2013), a member of Best Practice Committee in XBRL US and a member of XBRL Quality Review Team (XQRT) in IFRS Foundation.