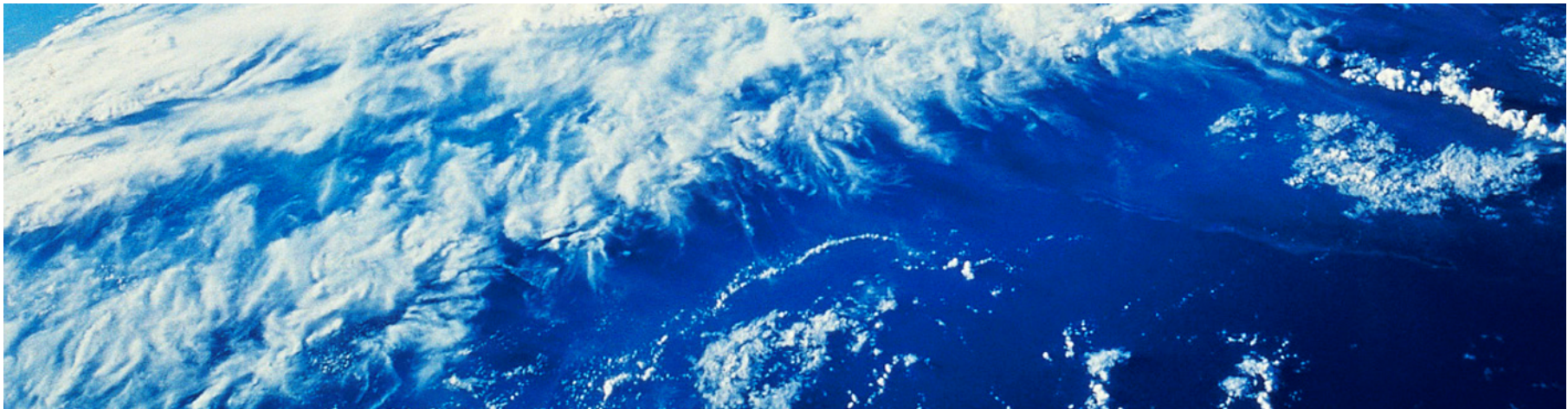


# IBM WebSphere Commerce V7 FEP8

Enhancement In Data Load Difference File Generation



Addy Leung

# Agenda

- Introduction
- Data Load Configuration
- Data Load File Difference Examples
- Performance
- Links

# Introduction

Data Load diff file generation was introduced in V7 FEP 6, it is a data reader preprocessor that runs at the beginning of the reader initialization when you run the Data Load utility. There are 2:

- CSVFileDiffPreprocessor
- XmlFileDiffPreprocessor

In V7 FEP 8, this tool is enhanced with the following:

- Ability to do columns compare, not just line by line compare
  - Ability to allow columns to be arranged differently in the new file
- Ability to ignore certain columns
  - Ability to include or exclude columns
- Ability to recognize indifference changes
  - Extra comma at the end of line is ignored
  - Optional double quote that quote the CSV token is ignored

# Data Load Configuration

To enable, in DataReader section in data load business object configuration file, e.g.

<WCS>\components\foundation\samples\DataLoad\Catalog\wc-loader-catalog-entry.xml

```
<_config:DataReader className="com.ibm.commerce.foundation.dataload.datareader.CSVReader"
firstLineIsHeader="true" useHeaderAsColumnName="true" >
  <_config:property name="keyColumns" value="PartNumber" />
  <_config:property name="dataReaderPreprocessOnly" value="true"/>
  <_config:property name="columnBasedCompare" value="true" />
  <_config:property name="includeCompareColumns" value="Name, ShortDescription,
LongDescription" />
  <_config:property name="excludeCompareColumns" value="Thumbnail, FullImage" />
  <_config:DataReaderPreprocessor
className="com.ibm.commerce.foundation.dataload.datareader.CSVFileDiffPreprocessor" />
</_config:DataReader>
```

If columnBasedCompare, includeComparecolumns, and excludeCompareColumns are not set, columnBasedCompare will default to false.

For best practice, only one of these 3 properties should be specified for activation.

# Data Load Configuration

includeCompareColumns	excludeCompareColumns	Behavior
<b>non empty value</b>	non empty value	<b>includeCompareColumns</b> takes priority and the other is ignored.
<b>non empty value</b>	empty value	<b>includeCompareColumns</b> takes priority and the other is ignored.
empty value	empty value	<b>includeCompareColumns</b> takes priority and the other is ignored.
empty value	<b>non empty value</b>	<b>excludeCompareColumns</b> takes priority and the other is ignored.

Key columns cannot be in the includeCompareColumns or the excludeCompareColumns

When includeCompareColumns has an empty value, none of the columns will be used for the compare. Lines with the same key columns are considered same and will not be in the difference file.

# Data Load File Difference Examples

New file			Old file		
KeyCol	Col1	Col2	KeyCol	Col1	Col2
A	A1new	A2new	A	A1	A2
B	B1	B2new	C	C1	C2
C	C1new	C2	B	B1	B2
E	E1	E2	D	D1	D2

`<_config:property name="includeCompareColumns" value="Col1" />`  
`<_config:property name="excludeCompareColumns" value="Col2" />`

is exactly same as

`<_config:property name="includeCompareColumns" value="Col1" />`

Result:

A, A1new, A2new

C, C1new, C2

E, E1, E2

# Data Load File Difference Examples

New file			Old file		
KeyCol	Col1	Col2	KeyCol	Col1	Col2
A	A1new	A2new	A	A1	A2
B	B1	B2new	C	C1	C2
C	C1new	C2	B	B1	B2
<code>&lt;_config:property name="includeCompareColumns" value="" /&gt;</code> E			D		
	E1	E2		D1	D2

Result:

E, E1, E2

Note: record with KeyCol=D will not be included

# Performance

For CSV file difference, expect column based compare to be slower than line by line compare, depending on the file size.

For the (CSV like) XML file difference, there are 2 cases:

If all columns are to be compared, there will be no performance impact because the XML lines need to be rebuild anyway

If only some columns are included or excluded, the performance of the column based compare maybe slightly better as less columns are needed in the rebuild resulting in smaller XML line to be compared.

Use column based compare when

- Input contains frequently updated but optional data, e.g., last update timestamp
- Input file is large
- Data columns orders could be inconsistent between last one and new one.



# Troubleshooting

## Trace

Same as always, update the dataload logging.properties,

-Dcom.ibm.commerce.foundation.level=ALL

-Dcom.ibm.commerce.catalog.level=ALL

## Links

[Data Load file difference preprocessing](#)

[Configuring the Data Load utility to run a file difference preprocess](#)

Thank You!

## Trademarks, disclaimer, and copyright information

IBM, the IBM logo, ibm.com, Coremetrics, DB2, PowerVM, Rational, WebSphere, and z/VM are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of other IBM trademarks is available on the web at "

Copyright and trademark information" at <http://www.ibm.com/legal/copytrade.shtml>

Other company, product, or service names may be trademarks or service marks of others.

THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS OR SOFTWARE.

© Copyright International Business Machines Corporation 2014. All rights reserved.