



Lessons Learnt Datastage 11.3



January 21, 2016



Title

U06

Domain/ Service/ Technology /Geo: Datastage 11.3

Name of the Project/Support Fn: Chrysler

Name of the Author: Dinesh Aila

Date Created: 28/12/2015

Description

Project/Context

- The objective of this project is to design Datastage jobs to replace the current COIN system which is holding the Customer Owner information used in the Campaign Management Services.

- This will aid in the reduction of Mainframe MIPS Utilization which will benefit in the cost reduction

How it happened

Process Adopted

1. **Understating the existing process at the mainframe end**
2. **Designing the Jobs in Datastage 11.3 accordingly.**
3. **Extracting the data form the Mainframe server.**
4. **Applying the existing the business transformation on the data.**
5. **Loading the data into Target tables(Bulk Load).**

Issues & Analysis

Issues Faced

- Reading the EBCDIC mainframe format file.
- DB2 Bulk load of data.
- Records were dropped in the join stage due to improper datatype for reading integer data as char.
- Inconsistent in record length for the input files leading to rework.
- The varchar fields were populated with special characters because of issues with the APT_STRING_PADCHAR.
- Records were dropped in the pivot stage when it is incorporated in joining with the other data.
- Reading the Binary data from EBCDIC format file

Causal

- Char Fields were not trimmed properly
- Null handling not present in some transformations.
- Timestamp conversion for null values.
- Data extraction from multiple files and using merge stage in a single job takes long execution time.
- Improper format of the EBCDIC File.

Results

Justification

- The varchar fields were not populated correctly and were populated with special characters.
- The EBCDIC format files were able to read in datastage by using CFF file
- Exported the metadata by using the COBOL copybook export
- Able to read the binary data.
- Changes in requirements lead to rework and modification of jobs in some cases
- Performed the Bulk load by loading the Huge volume of data

Benefits

- Reduction in MIPS at mainframe end
- error free work.
- Less rework.
- Better Code Quality.
- Reusable Component.
- Eliminated the staging space area of about 150GB.

Solutions & lessons

Solutions

- The varchar fields in all the tables were populated with special characters. The environment variable APT_STRING_PADCHAR was set to 0x20 which was previously set to 0x0

Lessons Learnt

- The environment variable APT_STRING_PADCHAR was set to 0x0, so the pad character (ASCII null value 0x0) was not overridden. The environment variable should be set to 0xn (where n is the maximum possible value for that field, in our case 20), so that the varchar column are populated correctly.

Solutions & lessons

Solutions

- The Sequential file stage is not able to populate all the columns for the EBCDIC format file as the input is having Group Columns, Redines, Binary data from the COBOL Copybook layout.
It is flattening the group columns.
Also not able to read the COMP files perfectly.

Lessons Learnt

- Used the CFF(Complex Flat File) stage and it kept all the group and redefined columns.
It was able to read the zoned and packed decimal values.
Binary data were read correctly by using the text option as BigIndian

Solutions & lessons

Solutions

The join stage was silently dropping records (without warnings) due the char filed type when the data in Integer format.
This is due to reading from the fixed length record.

Lessons Learnt

Implicitly converted required char fields to integer by using the copy stage.

Solutions & lessons

Solutions



The tmp path option should have write access permission in Bulk load mode.


Lessons Learnt



Valid path should be provided for the tmp path option.


Solutions & lessons

Solutions



Pivot of data were removed and created new job. The output dataset were used in actual job for the joining the data.

Lessons Learnt



Pivot of data should not be incorporated in the job when it is doing the join operation from multiple files or datasets.

How this may be prevented elsewhere

Prevention

- The environment variables should be properly set while creating the datastage project.
- The string fields should always be trimmed.
- Some of the functions like sorting, trimming can be done in the database stage itself, so that less number of stages are required in the job and its execution time is less.
- Proper null handling should be present wherever required.
- The warnings should be taken care of during development.

Contact Info

Name: Dinesh Aila

Employee Id: 6656101

Contact email id:
dinesh.aila@tcs.com



Thank You

IT Services
Business Solutions
Consulting