SAS Migration from Mainframe to UNIX

Western Load Research Association Conference
September 23, 2009
San Diego, CA

Sotan Im
Southern California Edison

Why the Migration?

- Company-wide Implementation of SAP-ERP system
  - SAP Analytics, BOBJ became available
  - SAS is no longer needed (false claim).

  We succeeded in convincing IT to keep SAS
- Mainframe will be decommissioned in 2010
  - Must move SAS from Mainframe to UNIX
- Mainframe and DB2 were upgraded instead

  SAS license more expensive with more powerful mainframe
  - Move SAS from Mainframe to UNIX
  - Customer info, billing, load data,… remain on Mainframe

SAS-SAP Integration and Migration Project

- **Purpose:**
  - Interface SAS with ERP BI data
    - ERP BI is the Enterprise Data Warehouse for SCE
    - Supplement Business Explorer Reporting Tool (BEx)
    - Produce analytics and ad hoc reports (GRC, Regulatory, Compliance, and Operations)
  - Move SAS system and applications from the Mainframe to Unix
    - Reduce O&M expenses by $500,000 per year (Jan, 2010)

- **SAS User Requirements for ERP BI (BEx is the first reporting option)**
  - Analytics, Modeling, and Statistical Analysis
  - Complicated reports (BEx not capable)
  - Large data retrievals (> 65,000 rows)
  - Retrievals utilizing multiple databases (BI, DB2, Oracle, SQL SVR, SAS , and flat files)

Running SAS on UNIX

- **SAS Enterprise Guide**
  - Point-and-click Interface to SAS on UNIX or PC
  - Generates code and submits it to SAS for execution
- **SSH Tectia**
  - Logon to UNIX
  - FTP capability between UNIX and PC
  - Run SAS on foreground
  - Run SAS on background (batch mode)
- **Hummingbird Exceed**
  - Exceed enables an X-Windows display on PC
  - Can run SAS interactively in windowing environment

Differences between z/OS and UNIX

- **Terminology**
  - z/OS: CLIST, JCL, JES, ISPF…
  - UNIX: Shell (sh, ksh, bash), Script, vi, …
- **Character set**

  The alphanumeric characters are not in the same order
  - z/OS: EBCDIC --- a-z, A-Z, 0-9
  - UNIX: ASCII --- 0-9, A-Z, a-z
- **File Structure**
  - z/OS: sequential, partitioned datasets
  - UNIX: file, directory (like DOS)
- **Access Permissions**
  - z/OS: RACF to set permissions for specific users
  - UNIX: CHMOD to set permissions for owner, group, and others

Problems encountered

- Passwords
  - Need to store passwords in SAS code to connect to DB2
  - This is a problem when sharing codes with others
- Formats
  - 'TOU-PA-A' - 'TOU-PA-4' = 3.4 OK on z/OS not on UNIX
  - Start is greater than end (since A > 4) on UNIX
- Numeric variable length
  - It took us some time to understand this problem. More details later.
- Packed decimal
  - Load data stored in packed decimal in DB2 table
  - Extracts on UNIX had wrong kW values
- Printing from UNIX to Mainframe printers or network printers
  - IT is still working on this problem

Password Problem

To connect to DB2 one can use libname statement.

- z/OS: LIBNAME CXSQ DB2 SSID=DB04 AUTHID=CXSQ;

- UNIX: libname cxsq db2 datasrc=db04
  schema=cxsq user="&mf_uid"
  password="&mf_pw";

- The macro variables mf_uid and mf_pw (encrypted) are stored in a file to be included in SAS programs.

  Use proc PWENCODE to encrypt password.

Problem with numeric variables length

Numeric Variable of Length 4				Largest Integer Represented Exactly		
Mainframe Value		Downloaded to UNIX		Length in Bytes	Mainframe	UNIX
1111111		1111111		2	256	N/A
1234567		1234567		3	65,536	8,192
2001125		2001125		4	16,777,216	2,097,152
2097151		2097151		5	4,294,967,296	536,870,912
2097152		2097152		6	1,099,511,627,776	137,438,953,472
2097153		2097152		7	281,474,946,710,656	35,184,372,088,832
2639246		2639246		8	72,057,594,037,927,900	9,007,199,254,740,990
4597018		4597016				
6700156		6700156				
9990012		9990008				

Packed Decimal Problem

DB2 packed decimal data are read incorrectly by UNIX/SAS

- Interval load data are stored in packed decimal in DB2
  - Two decimal digits stored in one byte
  - Sign in 1st half byte (ASCII), or 2nd half (EBCDIC)
  - Up to 40 days of data stored in a VARCHAR(30720)
- The DB2, being smart, performs an EBCDIC-to-ASCII conversion before it gives the data to SAS on UNIX. This conversion corrupts the data.
- Need to use a Pass-Through SQL with a CAST clause to prevent the DB2 from performing the conversion.

  SAS can now read the data correctly.

Some Preliminary Performance Results

Use Case	Ad-Hoc/Urgent or Heavy Data	Mainframe	Unix
Client needs listing of customers who are receiving low income rates.	Ad-Hoc/Urgent	2 min 58 sec	1 min 53 sec
Distribution of accounts by service plan code and revenue class code	Heavy Data	1 min 13 sec	1 min 57 sec
Compute monthly usage by tiers for residential accounts in baseline 15	Heavy Data	10 min 84 sec	9 min 56 sec
List of service accounts that have 15-min load data in 2009.Distribution by service plan code and status	Heavy Data	17 min 36 sec	12 min 07 sec
Select a sample of 50,000 residential single/multiple accounts. Calculate some basic statistics	Heavy Data Stat procedures	4 min 36 sec	3 min 15 sec
Testing an advanced statistical method: Mixed models using proc mixed on simulated data	Ad Hoc/Urgent	4 sec	1 sec

Some good news

- SAS/GRAPH, point-and-click TSF System (Time Series Forecasting) seem to be working on the X-windowing environment almost as well as on PC/SAS.
- Able to create report in EXCEL, GIF, PDF, … file on UNIX and download it to PC.
- Able to work with one year of monthly billing data for all residential customers on UNIX (51M rows, 3.9 GB). Extracted billing data and calculated the monthly usage by tier in 43 minutes (real time).

On the Mainframe we had to do it by quarter.

Migration Plan

SAS license on the Mainframe will expire by May 2010. Without SAS, MF SAS datasets will be useless thereafter.

- Convert all SAS datasets to transport files.
- FTP the transport file from mainframe to UNIX as needed.
- Read the transport file into a SAS data set on UNIX.

There is no problem with SAS codes. They are just text files and can be downloaded, modified, and run on UNIX whenever we want.