From SAS® Data Management to
Big Data Appliances
How SAS/ACCESS® Makes Life Easier
From SAS® Data Management to Big Data Appliances: How SAS/ACCESS® Makes Life Easier

Magali Thésias
Senior Specialist - Deloitte

What is the issue?
Many (large) companies are struggling in dealing with these technologies and more importantly on integrating them in their existing data management processes. Moreover, they also want to rely on the knowledge built by their teams in existing products and implementing change to learn new technologies can therefore be a costly procedure.

How SAS is helping?
SAS® is perfectly fitting in this situation by offering a suite of software that can be set up to work with any third-party database through the usage of the corresponding SAS/ACCESS®. Indeed, for every new database technology SAS® is releasing a specific SAS/ACCESS® allowing users to develop and migrate SAS® solutions almost transparently. Only few techniques have to be known by your users to combine the power of SAS® with a third-party (big) database.

90% of all the data in the has been generated over the last 2 years

New technologies are developed to help companies handle this huge volume of data. Amongst them, the Massively Parallel Processing (MPP), enabling splitting data and queries across a large number of nodes in order to perform simultaneous computation.

Context

In-Database Processing

New technologies are developed to help companies handle this huge volume of data. Amongst them, the Massively Parallel Processing (MPP), enabling splitting data and queries across a large number of nodes in order to perform simultaneous computation.

What is the issue?
Many (large) companies are struggling in dealing with these technologies and more importantly on integrating them in their existing data management processes. Moreover, they also want to rely on the knowledge built by their teams in existing products and implementing change to learn new technologies can therefore be a costly procedure.

How SAS is helping?
SAS® is perfectly fitting in this situation by offering a suite of software that can be set up to work with any third-party database through the usage of the corresponding SAS/ACCESS®. Indeed, for every new database technology SAS® is releasing a specific SAS/ACCESS® allowing users to develop and migrate SAS® solutions almost transparently. Only few techniques have to be known by your users to combine the power of SAS® with a third-party (big) database.

90% of all the data in the has been generated over the last 2 years

New technologies are developed to help companies handle this huge volume of data. Amongst them, the Massively Parallel Processing (MPP), enabling splitting data and queries across a large number of nodes in order to perform simultaneous computation.

What is the issue?
Many (large) companies are struggling in dealing with these technologies and more importantly on integrating them in their existing data management processes. Moreover, they also want to rely on the knowledge built by their teams in existing products and implementing change to learn new technologies can therefore be a costly procedure.

How SAS is helping?
SAS® is perfectly fitting in this situation by offering a suite of software that can be set up to work with any third-party database through the usage of the corresponding SAS/ACCESS®. Indeed, for every new database technology SAS® is releasing a specific SAS/ACCESS® allowing users to develop and migrate SAS® solutions almost transparently. Only few techniques have to be known by your users to combine the power of SAS® with a third-party (big) database.

90% of all the data in the has been generated over the last 2 years

New technologies are developed to help companies handle this huge volume of data. Amongst them, the Massively Parallel Processing (MPP), enabling splitting data and queries across a large number of nodes in order to perform simultaneous computation.

What is the issue?
Many (large) companies are struggling in dealing with these technologies and more importantly on integrating them in their existing data management processes. Moreover, they also want to rely on the knowledge built by their teams in existing products and implementing change to learn new technologies can therefore be a costly procedure.

How SAS is helping?
SAS® is perfectly fitting in this situation by offering a suite of software that can be set up to work with any third-party database through the usage of the corresponding SAS/ACCESS®. Indeed, for every new database technology SAS® is releasing a specific SAS/ACCESS® allowing users to develop and migrate SAS® solutions almost transparently. Only few techniques have to be known by your users to combine the power of SAS® with a third-party (big) database.

90% of all the data in the has been generated over the last 2 years

New technologies are developed to help companies handle this huge volume of data. Amongst them, the Massively Parallel Processing (MPP), enabling splitting data and queries across a large number of nodes in order to perform simultaneous computation.

What is the issue?
Many (large) companies are struggling in dealing with these technologies and more importantly on integrating them in their existing data management processes. Moreover, they also want to rely on the knowledge built by their teams in existing products and implementing change to learn new technologies can therefore be a costly procedure.

How SAS is helping?
SAS® is perfectly fitting in this situation by offering a suite of software that can be set up to work with any third-party database through the usage of the corresponding SAS/ACCESS®. Indeed, for every new database technology SAS® is releasing a specific SAS/ACCESS® allowing users to develop and migrate SAS® solutions almost transparently. Only few techniques have to be known by your users to combine the power of SAS® with a third-party (big) database.

90% of all the data in the has been generated over the last 2 years

New technologies are developed to help companies handle this huge volume of data. Amongst them, the Massively Parallel Processing (MPP), enabling splitting data and queries across a large number of nodes in order to perform simultaneous computation.

What is the issue?
Many (large) companies are struggling in dealing with these technologies and more importantly on integrating them in their existing data management processes. Moreover, they also want to rely on the knowledge built by their teams in existing products and implementing change to learn new technologies can therefore be a costly procedure.

How SAS is helping?
SAS® is perfectly fitting in this situation by offering a suite of software that can be set up to work with any third-party database through the usage of the corresponding SAS/ACCESS®. Indeed, for every new database technology SAS® is releasing a specific SAS/ACCESS® allowing users to develop and migrate SAS® solutions almost transparently. Only few techniques have to be known by your users to combine the power of SAS® with a third-party (big) database.
From SAS® Data Management to Big Data Appliances:
How SAS/ACCESS® Makes Life Easier

Magali Thésias
Senior Specialist - Deloitte

What is the issue?
Many (large) companies are struggling in dealing with these technologies and more importantly on integrating them in their existing data management processes. Moreover, they also want to rely on the knowledge built by their teams in existing products and implementing change to learn new technologies can therefore be a costly procedure.

How SAS is helping?
SAS® is perfectly fitting in this situation by offering a suite of software that can be set up to work with any third-party database through the usage of the corresponding SAS/ACCESS®. Indeed, for every new database technology SAS® is releasing a specific SAS/ACCESS® allowing users to develop and migrate SAS® solutions almost transparently. Only few techniques have to be known by your users to combine the power of SAS® with a third-party (big) database.

90% of all the data in the has been generated over the last 2 years

New technologies are developed to help companies handle this huge volume of data. Amongst them, the Massively Parallel Processing (MPP), enabling splitting data and queries across a large number of nodes in order to perform simultaneous computation.

ELT with SAS/ACCESS

In Database processing

SAS® Server

In Database processing

Third-party database

Push queries

Push queries

Extract & (Bulk)Load

Data and Document Stores

Operational Systems

Landing

Transform

DWH
How to push code for in-database processing?

Explicit pass-through

- Use the PROC SQL CONNECT – EXECUTE statements.
- The SQL query has to be able to work AS-IS in the database.
- Database will generates errors if a query contains anything SAS specific:
  - SAS formats.
  - SAS functions.
  - DATE or DATETIME actual numeric values.
  - INTO: macro variable.
  - SAS options.
  - Using multiple librefs.

```sql
PROC SQL;
   CONNECT to NETEZZA (
      DATABASE=SANDBOX
      SERVER=SNZ_SERVER
      USER=SNZ_USER
      PASSWORD=SNZ_PSW);
   EXECUTE (CALL DROP_OBJECT_IF_EXISTS('D_DATE', 'TABLE')
       BY NETEZZA);
   DISCONNECT FROM NETEZZA;
QUIT;
```
How to push code for in-database processing?

**Implicit pass-through**

- Use the **LIBNAME** engine.
- Use the various DBMS data types and translate them into SAS formats.
- SAS attempts to generate a database specific SQL query that will be executed in-database.
- Not all SAS functions and procedures can be converted to DBMS-specific syntax. As example, Netezza engine supports:
  - 48 SAS functions
  - 7 SAS procedures

```
LIBNAME nz NETEZZA DATABASE='SANDBOX'
SERVER='XX.XX.XX.XXX'
SCHEMA='admin'
USER=admin
PASSWORD="XXXXX";
```
From SAS® Data Management to Big Data Appliances: How SAS/ACCESS® Makes Life Easier

Magali Thésias
Senior Specialist - Deloitte

SAS® DATA INTEGRATION STUDIO

Implicit pass-through code generation

Data Transforms  Apply Business Rules  Analytic Transforms

Code generation reside: Automatic

```
91 * Target Table: Extract (PDA.WI00B5CO) ADEL22PR.MUN00002 +
92 ************************************************************
93 let (tablesStr = values('ADEL22PR.MUN00002'));
94 let trans_rc = 0;
95 let etlis_pSpecStartTime = gettime('@datetime());
96 let etlitsminer = values('PDA.Extract_dataid');
97 let etlitsminer = 'PDA.Extract_dataid';
98 let ETLISMINER = values('PDA.Extract_dataid');
99 apl(pda) etlis_psql_pdb('sqlSnippet');
100 let etlsql_psql_pdb waive -1;
101 option SILENTDUMP;
102
103 */ Map the columns *****/
104 /* Ensure etlsql_psql_pdbExist; */
105 let etlsql_psql_pdbExist = "when(when(pda.extpName(SSI100B5CO, DATA)) or when(pda.extpName(SSI100B5C0, V200)))";
106 let etlsql_psql_pdbExist = "when(pda.extpName(SSI100B5CO, DATA)) or when(pda.extpName(SSI100B5C0, V200)))";
107 if [etlsql_psql_pdbExist] then
108 */
```

Diagram: [Image]
Best practices to create ELT data flows with SAS® DI Studio

- Redirection of the default SAS work
- Leverage bulk load capability
Best practices to create ELT data flows with SAS® DI Studio

Use database specific SQL functions to generate explicit SQL pass-through

Create explicit pass-through SQL join

From SAS® Data Management to Big Data Appliances:
How SAS/ACCESS® Makes Life Easier

Magali Thésias
Senior Specialist - Deloitte
From SAS® Data Management to Big Data Appliances:
How SAS/ACCESS® Makes Life Easier

Magali Thésias
Senior Specialist - Deloitte

Libname options

BULKUNLOAD=YES
DIRECT_SQL=
SQL_FUNCTION=ALL
Table Creation and Loading Options

BULKLOAD=YES

DBCREATE_TABLE_OPTS=

BL_USE_PIPE=NO

BL_DELIMITER=

BL_OPTIONS=
See what is happening in database: System Option

In a transformation

At job level

SASTRACE
From SAS® Data Management to Big Data Appliances: How SAS/ACCESS® Makes Life Easier
Magali Thésias
Senior Specialist - Deloitte

See what is happening in database: SASTRACE

```sql
CREATE TABLE PDA.DE_TABLE
(
    [DUPLICATE= YES DESCRIPTIVE_TABLE_OPTS='DISTRIBUTED ON RANDOM' DUPLICATE_FILENAME DUPLICATEIMITER=':', DUPLICATEOPTIONS='asterisk 100']
)
SELECT *
FROM SASDATA.VACATION as VACATION
QUIT;
```
How to use these options? A concrete example

Log result thanks to SASTRACE system option

• In-database processing cannot be executed, so an external table is created to download data on the SAS® server.

• The default delimiter (pipe) is used but this character is found in one of table columns.

Unload properly the data by using the bl_delimiter option.
Conclusions

• SAS/ACCESS® is automatically generating SQL code compliant with your third-party database.
  ➢ SAS® users can rely on their knowledge.
  ➢ SAS® development is back-end independent.

• The only extra knowledge needed is a set of options to fine tune the in-database processing.

• Nevertheless, explicit pass-through SQL code can be sent to the third-party database.