CNAF PostgreSQL project
Philippe BEAUDOIN, Project leader
philippe.beaudoin@bull.net

2010, Dec 7
CNAF

- Caisse Nationale des Allocations Familiales
- Key organization of the French social security system
- Distributes benefits to help
  - Families
  - Poor people
- 123 CAF (local organizations) all over France
- 11 million families and 30 million people
- 69 billion € in benefits distributed (2008)
The project ... in a few clicks

Bull - NovaScale 9000 servers (mainframes)
DBSP

GCOS 8

CRISTAL (Cobol)

SDP (Cobol)

InfiniBand link

LINUX

S.C. (c+SQL)

S.C. (c+SQL)

PostgreSQL

S.C. = Surrogate Client
- CRISTAL + SDP: heart of the Information System
- Same application running on Bull (GCOS 8) and IBM (z/OS+DB2) mainframes
- Also J2E servers, under AIX, … and a lot of peripheral applications
- 6 to 8 CRISTAL versions per year!
Migration to PostgreSQL project plan

- A lot of teams all over France (developers, testers, production,...)
- 3 domains
  - CRISTAL application, SDP application
  - Infrastructure and production
- National project leading

Cristal | Development | Tests | 1 CAF C. | Deployment C. |
--- | --- | --- | --- | --- |
Production | Preparation | | | |
Sdp | Development | Tests | 1 | Depl S |

9/08 1/09 5/09 9/09 12/09 3/10
How BULL participated to the project

- Assistance to project leading
- Technical expertise (with Dalibo contribution)
- Development of the data migration tool
- Assistance to programs and tools changes
- Testing assistance
- PostgreSQL education
- Support
Focus on « Development » domains

- Data structures
- Program changes
- Tests
- Data migration
Data structures

- Same tables and indexes
- Identical or equivalent columns types
- 2 (minor) difficulties:
  - Encoding choice:
    • UTF8 is not suitable for our Cobol world
    • SQL_ASCII chosen for sort results compatibility
  - Complex Cobol data structures stored in single columns
    • => BYTEA needed instead of CHAR
    • => study to identify BYTEA columns
- 900 CRISTAL sources x 5 versions
- 250 SDP sources x 1 version
- Quite easy :
  - SQL concentrated inside a data access layer
    => business code unchanged
  - Existing CNAF tool to generate programs accessing RFM or DB2
- Very few SQL and Cobol changes
- More significant impacts on environments and on development and test tools
Testing

- Unit tests by development team
- Additional integration tests
- Heavy acceptance tests:
  - Batch chains tested on a PostgreSQL production mirror database during 1 month
  - Record and replay a full transactional day
  - A CAF was largely involved in the acceptance
Data Migration

- Direct migration by programs
- A tool to generate 7000 programs!
- Multi-row inserts and parallel execution
- RFM / PostgreSQL comparison after migration
- Data of the largest database moved in less than 24 hours
Focus on « Production » domain

- PostgreSQL instances and databases
- Batch chains
- Database administration
- High-Availability architecture
- Performance
PostgreSQL instances and databases

- 2 clusters and 2 databases per CAF
  - => 168 databases and clusters
  - on 10 partitions
- Volumes
  - All databases = 4 Tb
  - The largest database = 250 Gb
- Preliminary study to set clusters parameters:
  - Right values for common parameters to all clusters
  - Memory usage model taking into account relative CAF sizes
PostgreSQL Administration and monitoring

- PgAdmin III for developers
- phpPgAdmin for application testers and supports

- Recurring administrative tasks are fully automated:
  - Saves (at fs level, using Legato Networker)
  - Reorganizations (CLUSTER)
  - psql scripts for all operations to perform on all databases

- Monitoring:
  - Nagios already used
  - Add monitoring of Linux partition + DBSP link + PostgreSQL
    - check_progress.pl
Batch Chains

- No impact on the structure of the batch chains
- Few minor impacts on JCL
- New activities on Linux managed by the existing scheduler
- Same automation level
High-Availability architecture

For 1 CAF:

Production center

Disaster backup center

J2EE clients

Other clients

GCOS 8

Batch

TP

Linux

PG Cristal
PG SDP

Clone-1

Cristal
SDP

Clone-2

EMC-SRDF

Sync.inter-batch

Sync.18h

DBSP

TP

Batch

PG Cristal
PG SDP

STD

Cristal
SDP

Cristal
SDP

Cristal
SDP
Performance

- PostgreSQL much more efficient
- But inter-partition link adds a cost
- Net balance:
  - Transactional average response time a little lower
  - Batch elapse times often lower
  - But some large programs take longer time
  - Solved by parallel runs
- Simple SQL statements
- But about 1 billion statements per day!
Customer feeling

- All players appreciate PostgreSQL
- The project was on time
- All expectations are met
- Number of problems due to PostgreSQL engine = 0
- PostgreSQL participates to the payment of 3 Billion € per month
- Some short term evolutions
  - H.A. evolutions
    - WAL archiving
  - E-Maj contrib to suppress CLONE-1 mirror and group clusters?
  - PostgreSQL 9.0
Personal feeling

- What helped us:
  - All required PostgreSQL features available
  - Overall reliability and performance
  - SQL Cache option developed for ecpg (ecpg -r prepare)

- Difficulties:
  - Get significant references
  - Slow response on Cursor operations
    • Replace with multi-tuple-return SELECT INTO
  - Select true where substr('A '::char(3),2,1)='' returns no row!
What I would find nice...

- To decrease CLUSTERDB frequency
  - « maintain_cluster_order » patch
- To help optimization
  - GUC for a statement set as embedded comment
- For short term projects:
  - Include pg_lesslog in /contrib or in core
- For long term projects:
  - Per column encoding
  - SQL/MED (to integrate PostgreSQL databases inside a heterogeneous databases landscape)
  - Easy to use partitioning management
  - Parametrized buffer-pools (to get shared_buffers > 10 Gb)