

PostgreSQL in Mission-Critical Financial Systems

May 20th 2010



Flavio Henrique Araque Gurgel



<http://creativecommons.org/licenses/by-sa/3.0/deed.pt>

www.4linux.com.br



4Linux in Caixa

- 2nd level support, 24/7
- Configuration, tuning
- Tests and documentation
- Knowledge transfer (training)
- Commitment: employees should attend communities
- Send code to the upstream in the same license

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Caixa Econômica Federal

- All cities of Brazil
- Social benefits
- “Every Brazilian has used or will use Caixa services”
- Caixa is a bank too!

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment

What is it?

Multicanal is the Brazilian word for “Multichannel”

- 1) Banking and Social transactions
 - Today: ATM self-service
- 2) Monitoring
- 3) SMS

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment in Numbers

- More than 22.000 ATM
- Peak: more than 7 million transactions in a day
- Peak: more than 18 million database transactions in a day
- More than R\$ 1 billion per month
(US\$ 1 ~ R\$ 1.8 in May 14th, 2010)

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment Characteristics

- High availability
- High performance
- High reliability

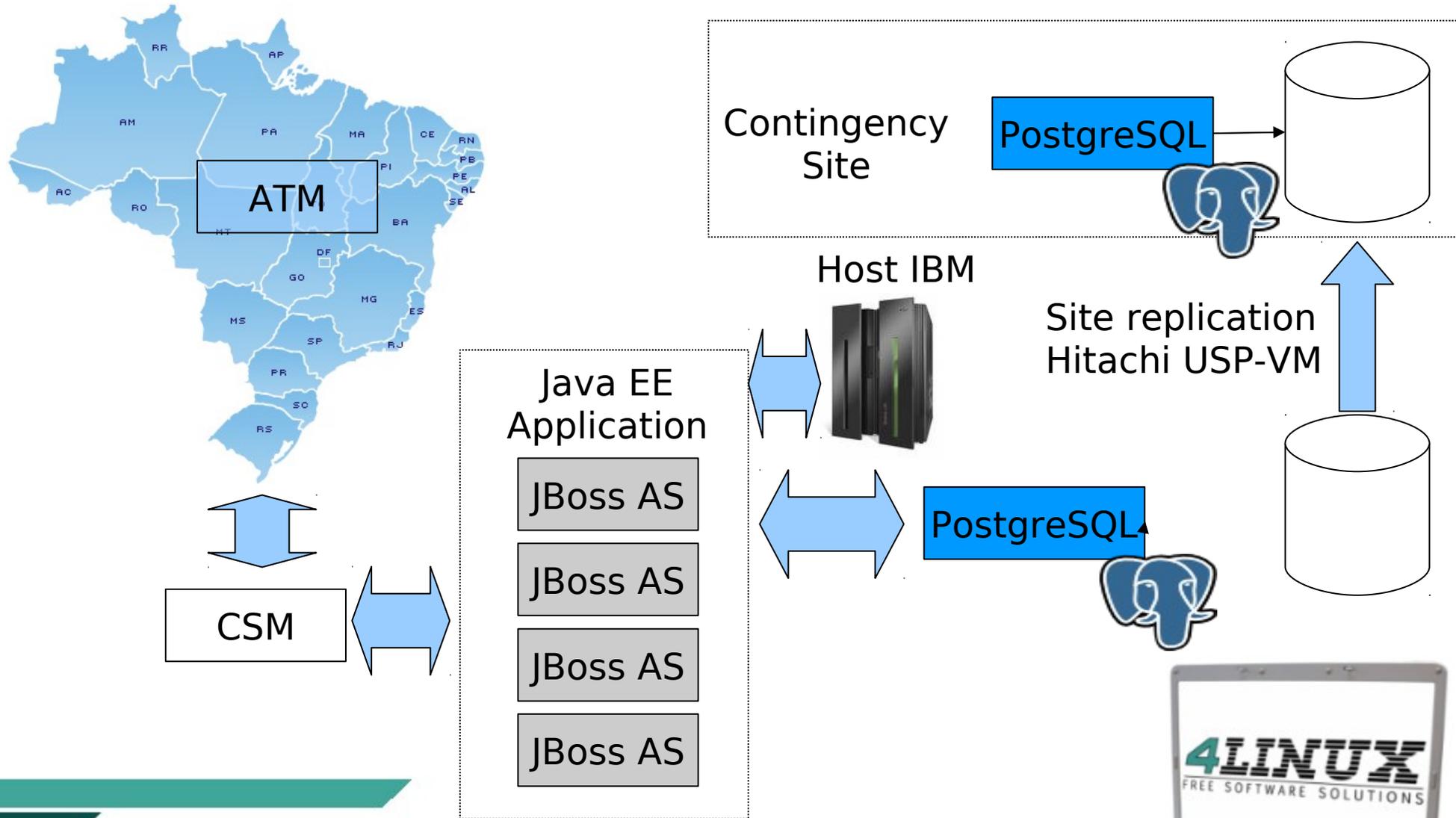
MISSION-CRITICAL!!!

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment Transactions (self-service)



www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment

Tests made with the system

750 TPS Minimum
Financial Transactions per Second

zOS + WebSphere + DB2

Debian + JBoss + PostgreSQL

Solaris + Sun Java System + Oracle

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment

The choice

- zOS + WebSphere +DB2
Best TPS.
- Debian + JBoss + PostgreSQL
TPS was ok.
1% of costs compared to zOS
Certified
- Solaris + JBoss + PostgreSQL
Least total cost
Solaris is widely used in Caixa
Definitive Platform

www.4linux.com.br

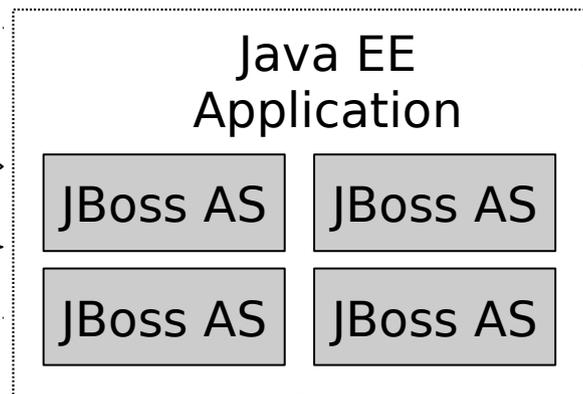
FREE SOFTWARE SOLUTIONS



Multicanal Environment Monitoring

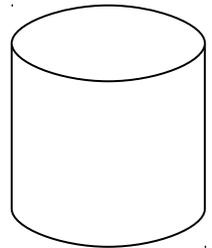


Transactional Application



Site replication
Hitachi USP-VM

PostgreSQL



Managers and Agencies

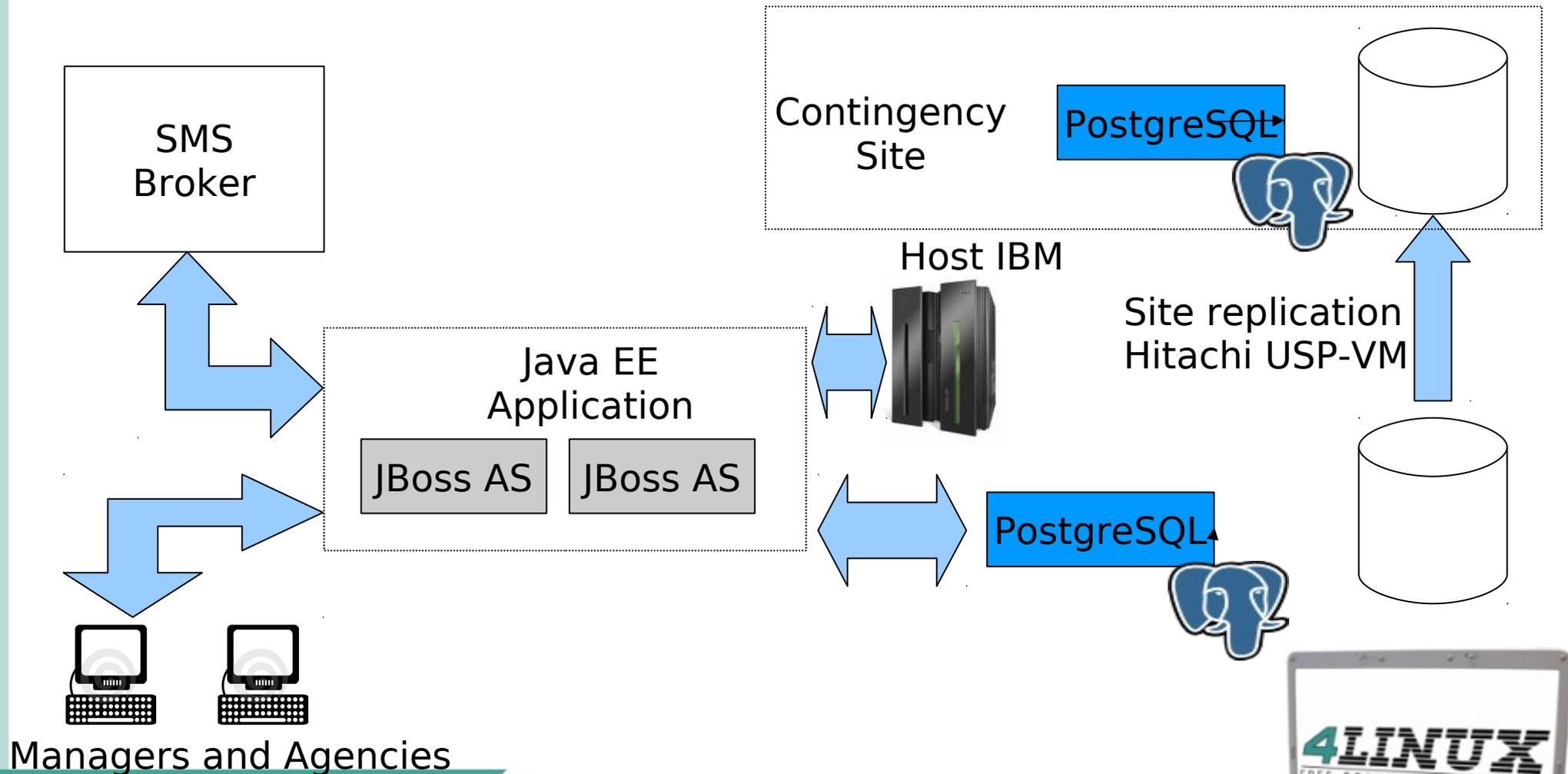


www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment SMS



www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment Database Infrastructure

- Infrastructure until final scale
 - Debian Etch Linux kernel 2.6.24/64bit
 - PostgreSQL 8.3.6 - community code compiled in-house
- Final infrastructure
 - Migrated to Solaris 10u6/Sparc 64bit
 - PostgreSQL 8.3.7 - Sun Package - 64bit
 - Transactions System - OLTP - 250 GB database
 - Monitoring System - DW - 1 TB database
 - SMS System - OLTP - 20 GB database

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment Hardware

- Dell Servers:
 - 2x Quad Xeon HT (2x4x2HT cores) @1.8GHz
 - 16 GB RAM
- HP Servers:
 - 2x Quad Xeon (2x4 cores) @2GHz
 - 32 GB RAM
- Sun Fire E25K (per system board):
 - 2x Ultra Sparc IV+ (4x2 cores) @1.5GHz
 - 32 GB RAM

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment Storage Sub-system

- Characteristics:
 - Storage Area Network – Fiber Channel
 - Hitachi Virtualizer – 16 GB cache
 - McData Switches
- Production:
 - Hitachi USP-VM – 8 GB cache
 - FC disks 15000 rpm – RAID5 – 6 and 7 disks
- Backup:
 - Online: EMC CLARiiON CX Series Storage
 - Tape: IBM Tivoli

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multicanal Environment Replication

- Hitachi Virtualizer Synchronous Replication
- Fiber Optics
- Site-to-site distance:
approximately 4 km (~ 2.5 miles)

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges

- Availability
- OLTP Stability
- DW Performance
- Maintenance
- Other DBMS knowledge

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges

Availability

- Online Backup and PITR strategy
- One initial copy each day
- Two days in disk
- Initial copy and archived WAL are replicated

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges

Stability

- Keep the Throughput!!!
- Tuning PostgreSQL
 - shared buffers
 - work_mem
 - Connection count
 - checkpoints
 - bgwriter
 - sync method
 - **autovacuum**

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges

Stability

- Keep the Throughput!!!
- Debian Tuning
 - I/O scheduler
 - kswapd kernel patch
- Solaris Tuning
 - max_queue_msg
 - forcedirectio (UFS)
 - ARC cache (ZFS)

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges

Stability

- Keep the Throughput!!!
- Disks Tuning
 - tablespaces for mostly accessed tables
 - tablespaces for mostly accessed indexes
 - Exclusive disk array for pg_xlog
- Synchronous replication tuning
 - High concurrency in remote storage
 - Fiber parameters and cache

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges

DW performance

- DW query tuning
- When more than 10 million live tuples:
 - Table partitioning
 - Automatic partitioning function
 - Development collaboration
- XML data in tables
 - Development collaboration
 - Indexing tables

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges

Database maintenance

- **All systems are 24h!!!**
- **autovacuum**
- How to schedule REINDEX
- How to schedule dump/restore
 - *Or VACUUM FULL + REINDEX???*
- “state machine” → partitioned
 - Development collaboration
- Purge (database and application)

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Daily Challenges Paradigms

- Reality in data-centers **is not** PostgreSQL (yet)
- Knowledge transfer
- Relative limitations
- Relative advantages
- No DBMS is the same as another
- How to “think PostgreSQL”

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



After challenges, praises

- Performance, high TPS
- Data was **never** lost
- Extensive documentation
- Open source software model
- **Issues are solved quickly**
- Cost relatively low

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Needs

Performance and availability

- Scale out
 - Multi-master cluster (Postgres-XC)?
 - Transparent memcached
 - Use more resources in maintenance routines
- SQL administration (configuration)
- System catalogs for performance monitoring
- Corporate audit flexibility

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Multi-master for availability

Tests 4Linux and Caixa

- pgpool-II and Sequoia
- OLTP – performance was not enough
- PostgreSQL stand-alone – 750 TPS*
- pgpool-II + 2 PostgreSQL – 170 TPS*
- Sequoia + 2 PostgreSQL – 50 TPS*

* Financial Transactions per Second, Caixa minimum is 700 TPS

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Final needs (and more important) Tools

- Integrated and graphical monitoring and administration tool
- All the database infrastructure
- Incident prevention alerts, automatic reports

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Conclusion

- PostgreSQL is ready for mission-critical
- but we need to evolve
- we know how to evolve
- **we can evolve**

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



Special Thanks!

- **Euler Taveira de Oliveira**
- **Caixa Econômica Federal**
- **PostgreSQL Community**
- **4Linux**

www.4linux.com.br

FREE SOFTWARE SOLUTIONS



- Flavio Henrique Araque Gurgel
flavio@4linux.com.br

www.4linux.com.br

FREE SOFTWARE SOLUTIONS

