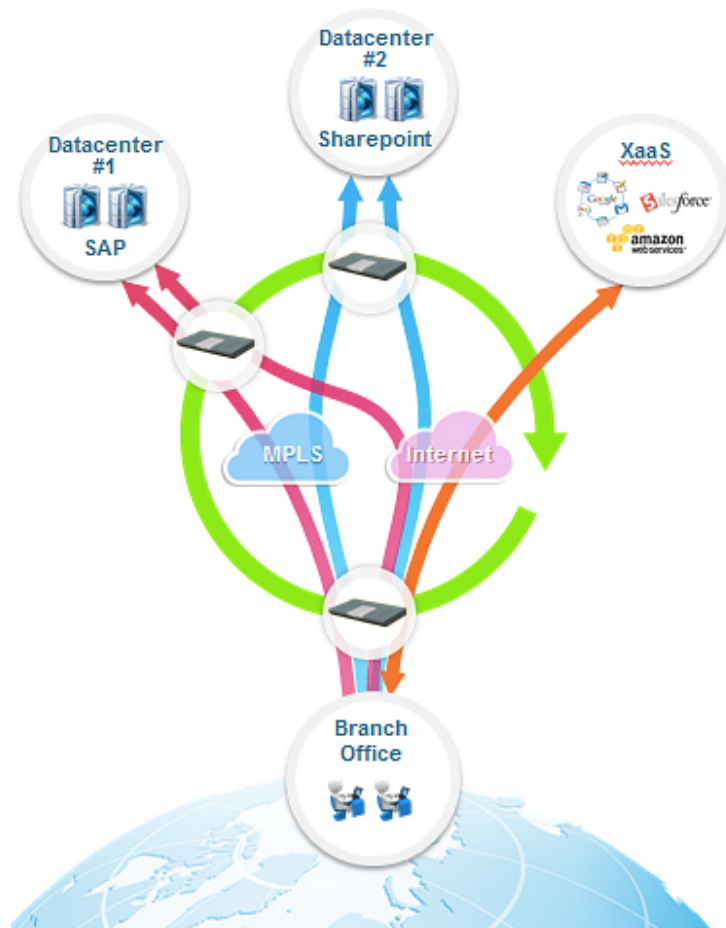




PostgreSQL Meetup

Agrégation temporelle sous
contrainte d'iops

Meetup, Paris 09/10/2014



Ipanema Technologies

The only fully integrated & automated system to easily guarantee application performance over your WAN

Sonde réseau



Type de réseau

- Orange Business Services
 - 250 clients
 - 2500 sondes
- PME
 - 1 client
 - 2-3 sondes
- 500 mesures / minute / sonde

Objectifs

- MySQL -> PostgreSQL
- 1 million de mesure / minute
- Historique
 - 1 min -> 72 heures
 - 8 To
- Agrégation temporelle (heure, jour, mois)
- Hardware entrée de gamme

Disque

- 8 To -> 1 To / 72 h = 14 Go/h
- 2 unsigned int4 -> real
- Timestamp -> int
- Pas de data inutilisée
- Field_A IS NULL XOR Field_B IS NULL

Bande passante

- 1 To / 72 h = 14 Go/h = 4 Mo/s
- Xlog
- Base
- Background: vacuum, hints bits, etc.
- Journal EXT3/4: data=writeback

Hardware

- ◉ RAM 16 Go
- ◉ Core 8
- ◉ RAID 1 (no BBU)
- ◉ Disque SATA 7200
- ◉ IOPS 120

- ◉ Random iops: $120 * 8k = 1 \text{ Mo/s}$

1 millions de mesures

- INSERT
 - 8 cores
 - 60s
- COPY
 - 1 core
 - 12 s (x40)

Partitionnement

- Partition de la taille de la période
- Librairie SQL de gestion des partitions
- Pas de trigger, de rules, ni DELETE
- Processus externe asynchrone
- Gestion de la place libre (FREE)

Agrégation / UPSERT

- PL/SQL: Trop d'erreurs, aléatoire

```
INSERT  
IF ERROR THEN UPDATE
```

```
UPDATE  
IF ERROR THEN INSERT  
IF ERROR THEN UPDATE
```

Agrégation / UPSERT

- CTE, aléatoire

WITH upsert AS ()
INSERT SELECT NOT EXIST

Agrégation / UNION ALL

- UNION ALL, nombreux batchs, async

```
CREATE TABLE tmp AS  
SELECT * FROM (  
    SELECT * FROM agg_data  
    UNION ALL  
    SELECT * FROM new_data  
) x GROUP BY
```

Agrégation / UNION ALL

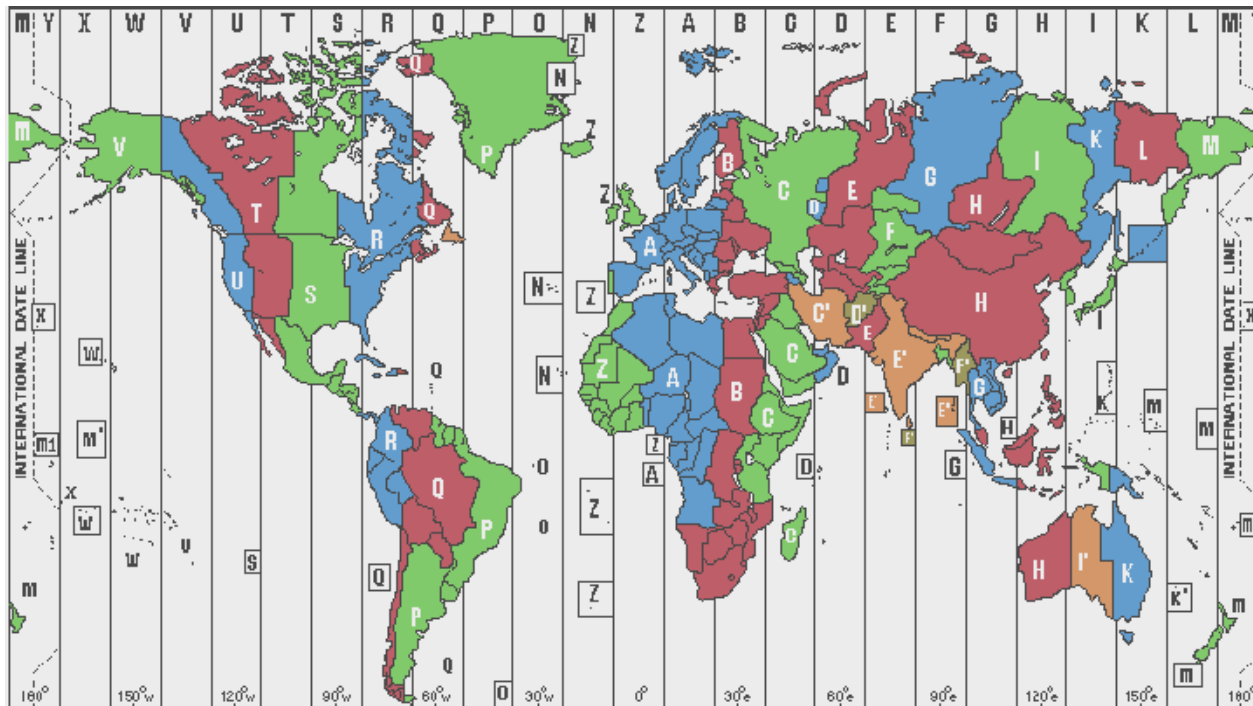
- UNION ALL, minimum batchs, async

SET work_mem TO 1GB;

- Moins de 60 secondes

Agrégation: timezone

- Timestamp -> int
- Time zone: 15 minutes



Performance

- OBS
 - 250 clients
 - 2500 sondes
 - 500 mesures / minute / sonde
- AVG: 200 IOPS
- MAX: 250 IOPS

postgresql.conf

- shared_buffers = 1/8 RAM
- wal_level = minimal
- synchronous_commit = off
- checkpoint_segments = 300
- checkpoint_completion_target = 0.9
- IOPS: 100-150

Partition & agrégation

- Optimisation de la période en cours
- UNLOGGED

Synthèse

- ◉ COPY
- ◉ Partition (No DELETE)
- ◉ UNION ALL + GROUP BY
- ◉ work_mem = 1 GB
- ◉ Time zone: 15 minutes
- ◉ UNLOGGED hot data
- ◉ postgresql.conf

Conclusion

- Données: 1 To
- Renouvellement: 72 heures
- Agrégation temporelle
- IOPS: AVG=100, MAX=150
- PostgreSQL 9.2

Questions

- Pourquoi est-ce 25% plus lent sous Windows ?