Small Overview of Skype Database Tools

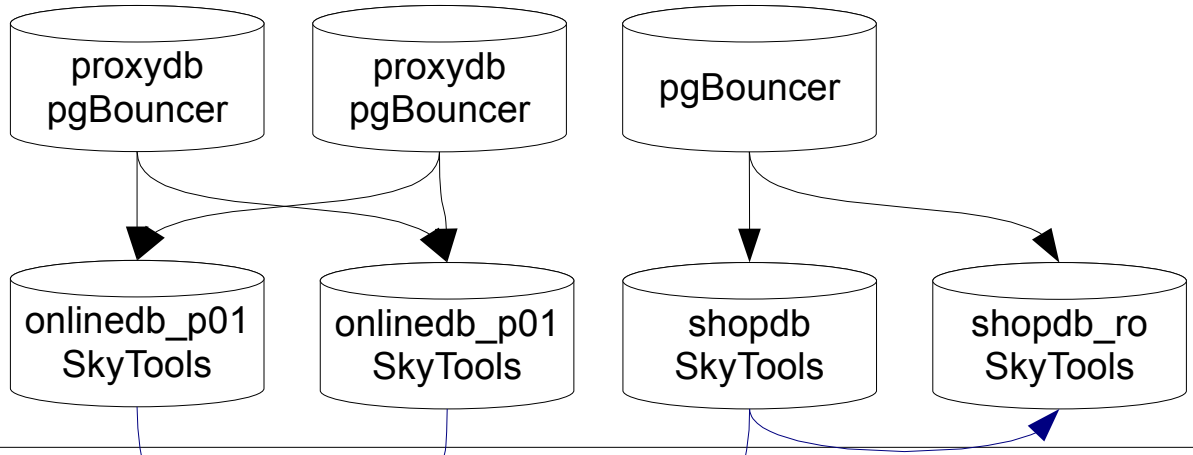
Asko Oja

Skype Databases

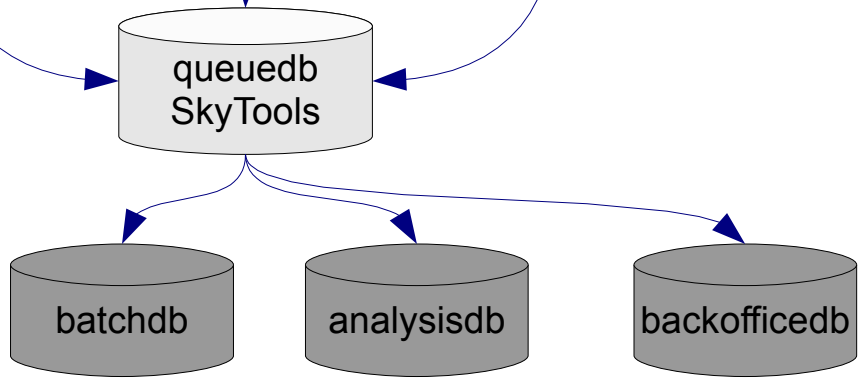
- Skype is communications company that provides calling and chatting services over internet. Thanks to the fact that a lot of activity goes on in p2p network not everything that user does hits databases.
- We have used PostgreSQL from the beginning for all our OLTP databases.
- Over 100 database servers and 200 databases
- Largest OLTP table has several billions of records.
- Over 10,000 transactions per second.
- Databases are used mainly for web store, billing and other centrally provided services.
- All access through stored procedures.

Layers of Databases

Online databases
- Proxy db's
- pgBouncers
- OLTP db's
- read only db's

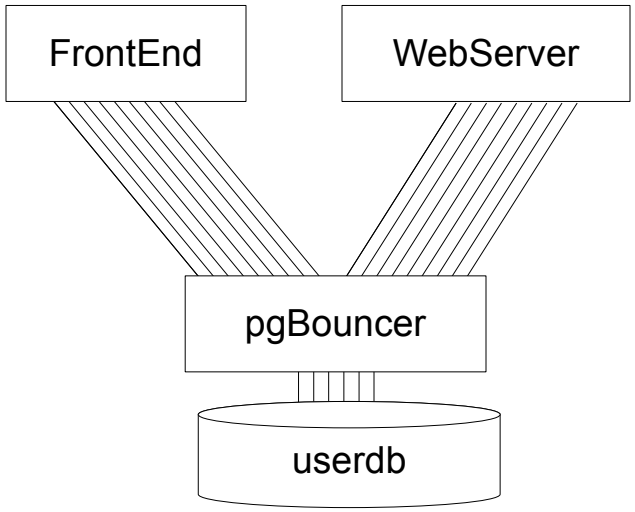


Support databases
- Queue db's
- Datamining
- Batchjobs
- Backoffice
- Greenplum



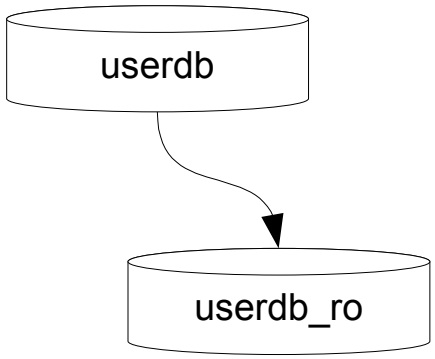
pgBouncer – PostgreSQL Connection pooler

- pgBouncer is lightweight and robust connection pooler for PostgreSQL.
- Reduces thousands of incoming connections to only tens of connections in database.
- Low number of connections is important because each connection uses computer resources and each new connection is quite expensive as prepared plans have to be created each time from scratch.
- We are not using pgBouncer for load balancing.
- Can be used to redirect database calls (database aliases).



plProxy: Remote Call Language

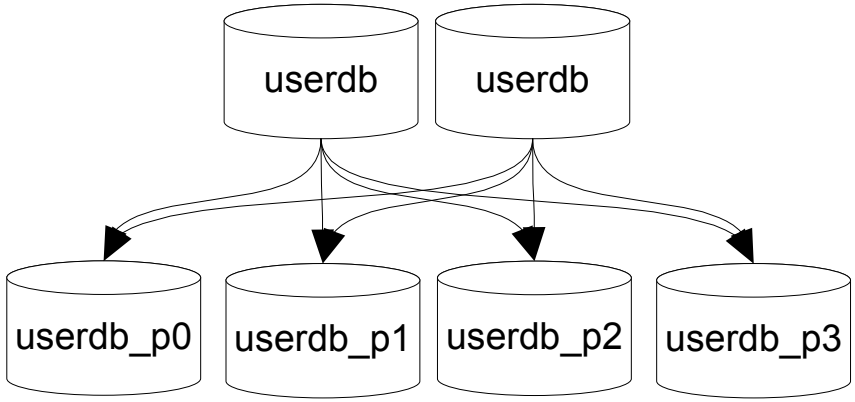
- PL/Proxy is compact language for remote calls between PostgreSQL databases.
- With PL/Proxy user can create proxy functions that have same signature as remote functions to be called. The function body describes how the remote connection should be acquired.
- plProxy adds very little overhead when used together with pgBouncer.
- On the other hand plProxy adds complexity to development and maintenance so it must be used with care but that is true for most everything.




```
CREATE FUNCTION get_user_email(username text) RETURNS text AS $$
    CONNECT 'dbname=shopdb_ro';
$$ LANGUAGE plproxy;
```

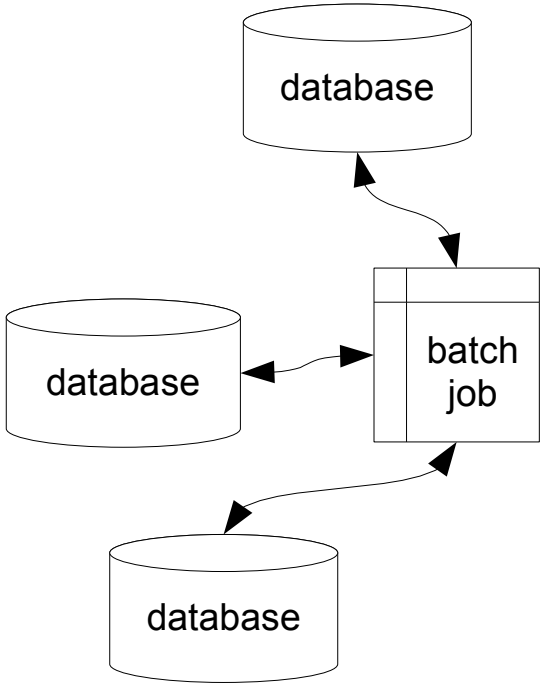
plProxy: Horizontal Partitioning

- We have partitioned most of our database by username using PostgreSQL hashtext function to get equal distribution between partitions.
- When splitting databases we usually prepare new partitions in other servers and then switch all traffic at once to keep our life simple.
- As proxy databases are stateless we can have several exact duplicates for load balancing and high availability.



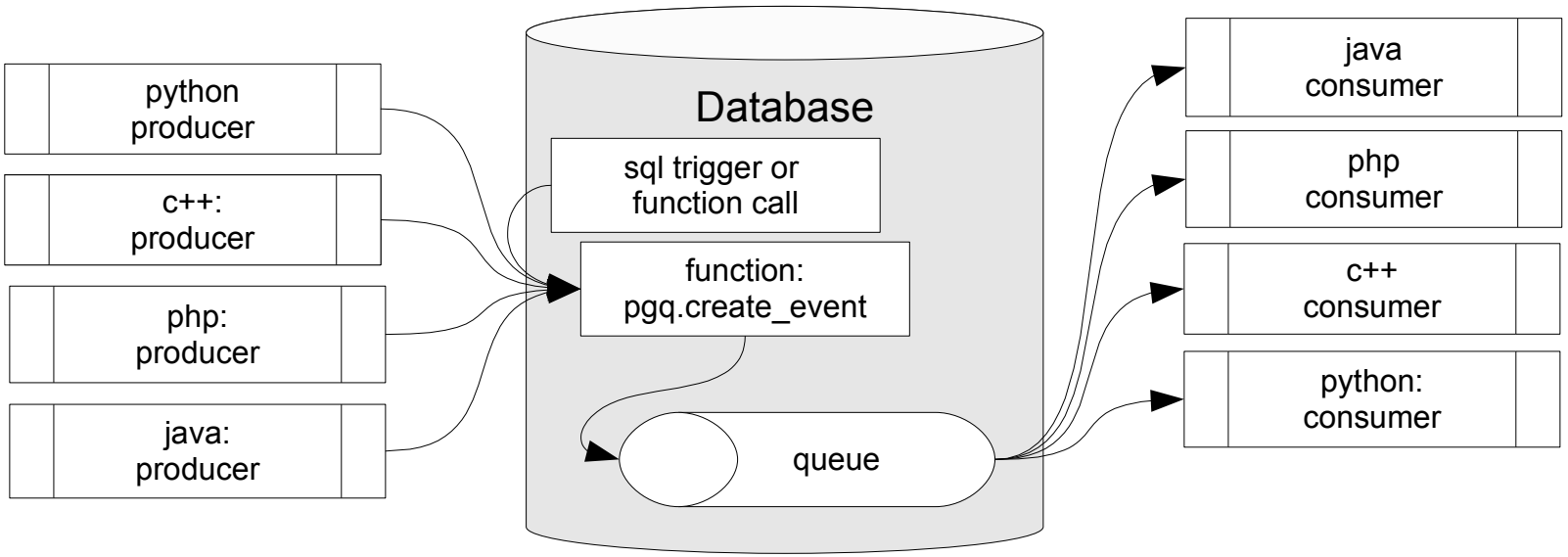
SkyTools: Package of Database Tools

- Contain most everything we have found useful in our everyday work with databases and PostgreSQL.
- PgQ that adds event queues to PostgreSQL.
- Londiste replication.
- Walmgr for wal based log shipping.
- DBScript framework which provide database connectivity, logging, stats management, encoding, decoding etc for batch jobs. Developers need only to take care of business logic in batch jobs all the rest is handled by batch jobs.
- SkyTools contains tens of reusable generic scripts for doing various data related tasks.



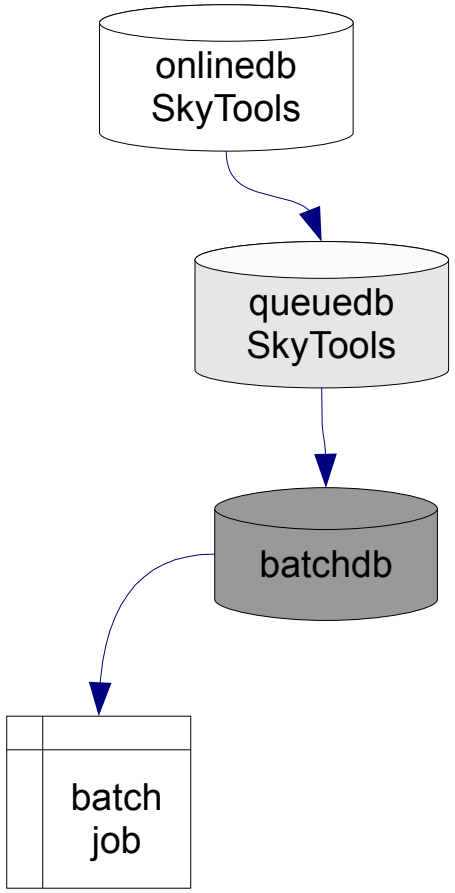
PgQ: PostgreSQL Queueing Implementation

- **PgQ** is PostgreSQL based event processing system. It is part of SkyTools package that contains several useful tools built on it.
- **Producers** - applications that write events into queue. Producer can be written in any language that is able to run stored procedures in PostgreSQL.
- **Consumers** - applications that read events from queue. Consumers can be written in any language that can interact with PostgreSQL.



PgQ: PostgreSQL Queuing

- **Transactional.** Event insertion is committed or rolled back together with the other things that transaction changes.
- **Efficient**. Usually database based queue implementations are not efficient but we managed to solve it because PostgreSQL makes transaction state visible to users.
- **Fast**. Events are processed in batches which gives low per event overhead.
- **Flexible**. Each queue can have several producers and consumers and any number of event types handled in it.
- **Robust**. PgQ guarantees that each consumers sees event at least once. There several methods how the processed events can be tracked depending on business needs.
- Ideally suited for all kinds of batch processing.

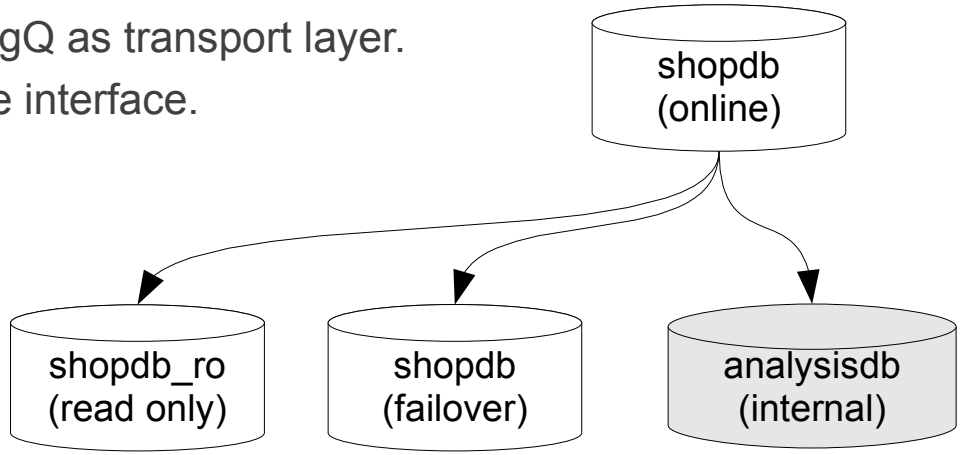


PgQ: Use cases

- Batch jobs. PgQ is very convenient media for providing events for background processing.
  - email senders
  - sms sender
  - external partners communication handling (payment handling)
- Moving data out of online databases into internal databases (logs, detail records).
- Replication. Copying data between databases.
- Partitioning data into daily batches for business intelligence.

Londiste: Replication

- We use replication
  - to transfer online data into internal databases
  - to create failover databases for online databases.
  - to switch between PostgreSQL versions
  - to distribute internally produced data into online servers
  - to create read only replicas for load balancing
- Londiste is implemented using PgQ as transport layer.
- It has DBA friendly command line interface.



Summary

- We have managed to create a beautiful mess of vertically and horizontally split databases that are connected by hundreds of replication processes and remote calls.
- But all this mess is still quite easy to manage and we see no architectural limits for Skype to grow some more magnitudes except need for better management tools.
- No license fees..