

PostgreSQL Extensibility

Ozgun Erdogan

PGCon 2016

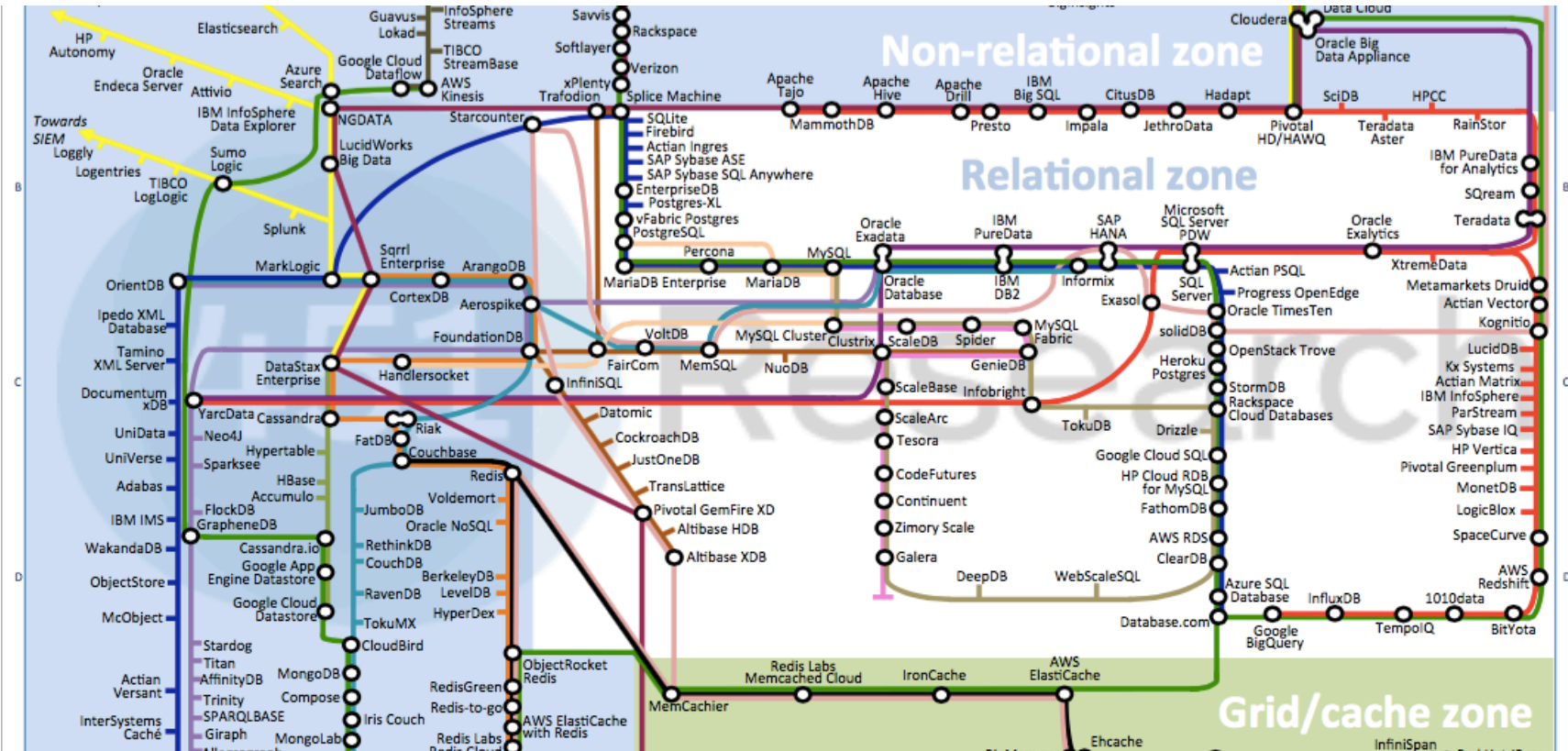
Punch Line

- PostgreSQL extensions are a game-changer for databases.
- The technology (extension APIs) is here. You can build your “own database” with Postgres.
- The challenges with extensibility aren’t about technology.

Why is Extensibility huge?

- Every decade brings new workloads for databases
- The past decade was about capturing more data, in more shapes and form. It was also about empowering developers
- This led to Hadoop, NoSQL, and in-memory databases. We would have dismissed these databases ten years ago

Evolution leads to specialized databases



How are these 300 databases built?

1. Fork an existing database
2. Build a new database from scratch
3. Extend PostgreSQL (new)

Technical Arguments Against Extensions

Technical due diligence feedback

1. Forking vs extensions: Is this even possible?
2. Building from scratch vs extensions:
PostgreSQL is an RDBMS from an old era.

Technical Arguments Against Extensions

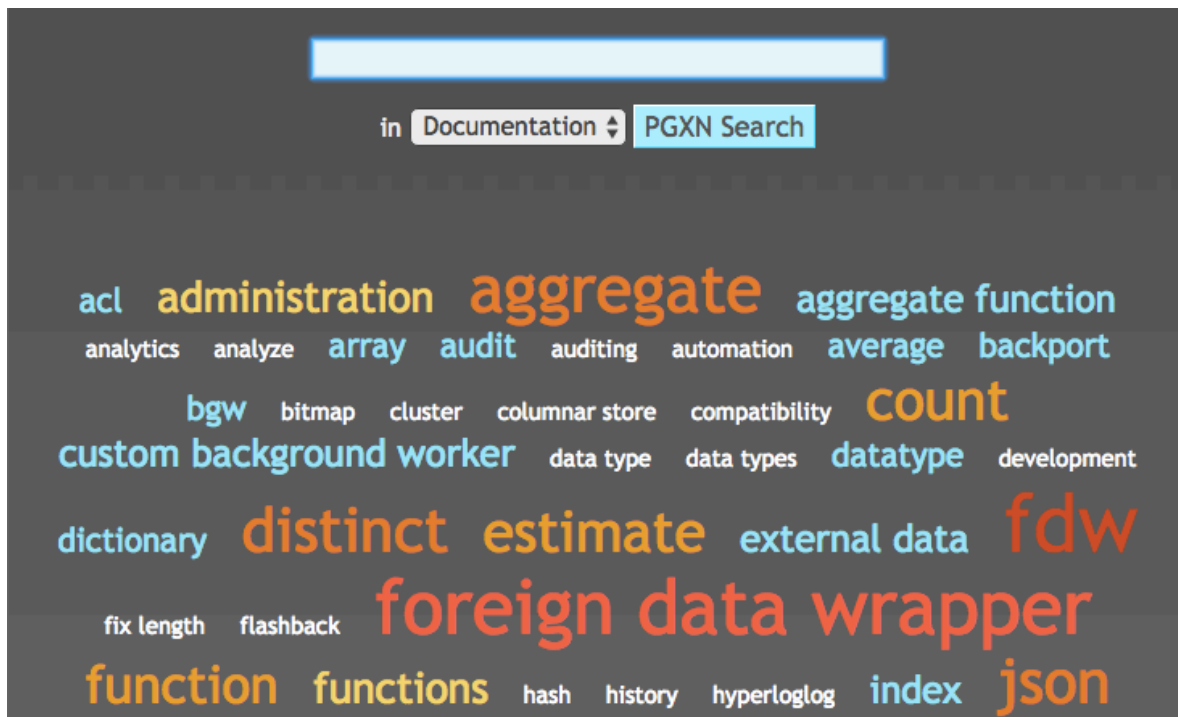
- Postgres can't be real-time: HyperLogLog and other sketch algorithms can provide approximate results
- Postgres' planner can't compete on analytical queries: Vitesse-X speeds up TPC-H queries by 3x
- Postgres can only store relational data: 99 foreign data wrappers and counting
- Postgres can't scale out: Citus shards and replicates your data; and parallelizes / distributes your queries

You can use all of this together, today

- HyperLogLog and other sketch algorithms for real-time approximate results
- Vitesse uses LLVM for faster analytical queries
- `pg_stat_statements` to capture stats
- `cstore_fdw` to compress columnar data
- Citus to shard and replicate your data; and parallelize your queries in a distributed cluster

If the technology is here,
why isn't there an Extensibility
revolution?

Users don't know what to search for



Developer to Community feedback

Data Source	Type	Licence	Code	Install	Doc	Notes
Database.com	Multicorn	BSD	GitHub			
Dun & Badstreet	Multicorn		GitHub			Access to the Data Universal Numbering System (DUNS)
DynamoDB	Multicorn	GPL	GitHub			
Facebook	Multicorn		GitHub			
Fixer.io	based on www_fdw		GitHub			
Google	Multicorn	PostgreSQL	GitHub	PGXN		
Heroku dataclips	Native		GitHub			
Mailchimp	Multicorn	PostgreSQL	GitHub			Beta
Parse	Multicorn	MIT	GitHub			
S3	Native		GitHub	PGXN		
S3CSV	Multicorn	GPL 3	GitHub			This is meant to replace s3_fdw that does is not supported on PostgreSQL version 9.2+
Twitter	Native		GitHub	PGXN		A wrapper fetching text messages from Twitter over the Internet and returning a table
Treasure Data	Multicorn	Apache	GitHub			
Google Spreadsheets	Multicorn	MIT	GitHub			

Big Data Wrappers

Data Source	Type	Licence	Code	Install	Doc	Notes
-------------	------	---------	------	---------	-----	-------

Competing Databases Communicate More

- NoSQL databases and Hadoop-based technologies raised \$1-2B in total
- They communicate that “the new era of databases mean capturing all data, in all shapes and form. RDBMS belongs to an old era.”
- The community needs to work together on making the Extensions better *for the users*.

Extensions Are Unique to PostgreSQL

- Extensions are unique to PostgreSQL. Other relational, Hadoop, and NoSQL databases don't have them.
- PGConf US Panel question: "Is PostgreSQL a better database than Oracle?"

PostgreSQL isn't yet the best
database in the world.
But, it could be.

PGConf SV (pgconfsv.com) Call for Papers



HOME **CALL FOR PAPERS** **PROGRAM** **SPONSORS** **VENUE** [REGISTER](#)

PGConf Silicon Valley 2016

Nov. 14-16 at the South San Francisco Conference Center

[Sign Up Here](#)