

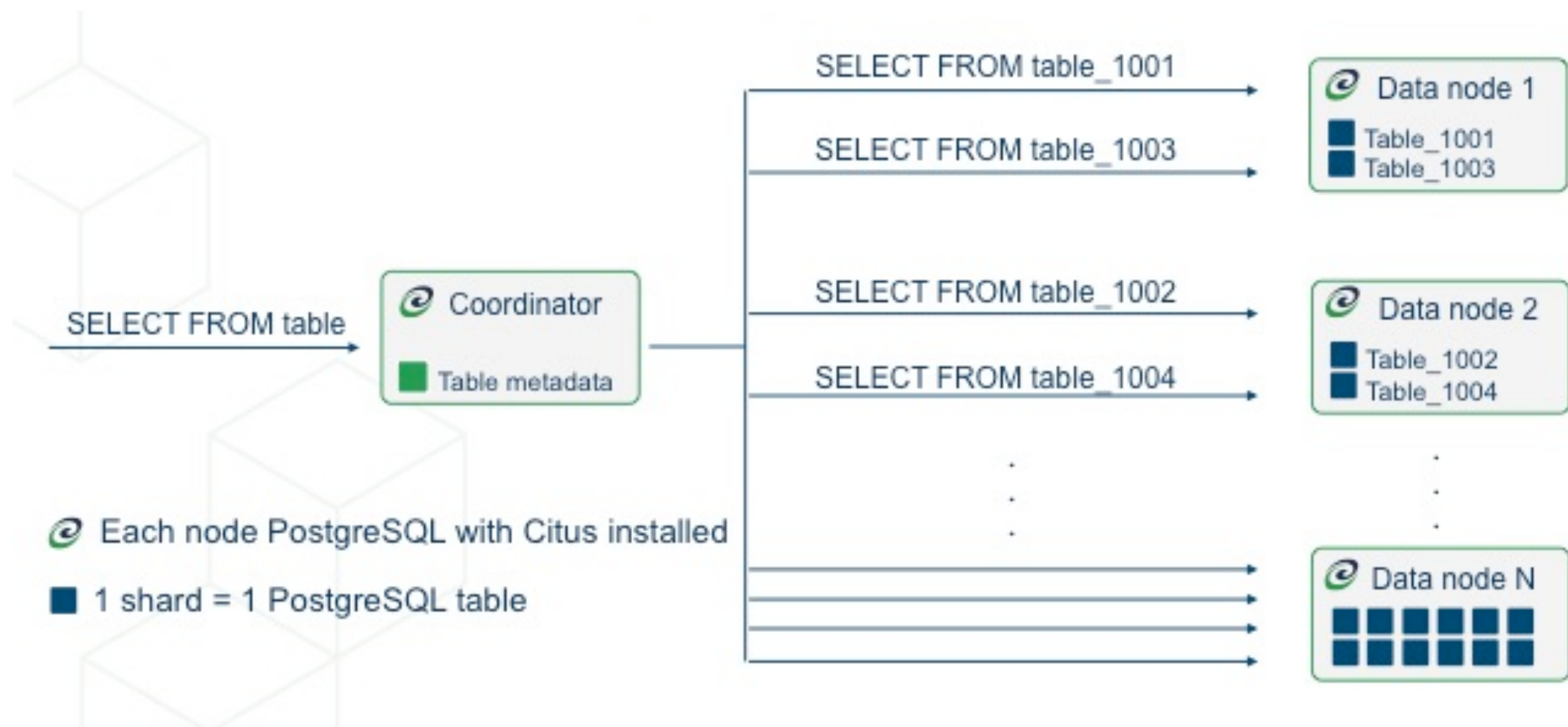


Citus

Artem Bakhtin

What is Citus?

- Open source **extension**
- Extends PostgreSQL with
 - Distributed tables
 - Reference Tables



Why use Citus?

- Multi-Tenant SaaS
- Real time analytics

Instalation

- Single-Node Citus

- Docker

- `docker run -d --name citus -p 5432:5432 -e POSTGRES_PASSWORD=mypass citusdata/citus:10.2`
 - `docker exec -it citus psql -U postgres`

- Ubuntu Or Debian

- `curl https://install.citusdata.com/community/deb.sh | sudo bash`
 - `sudo apt-get -y install postgresql-14-citus-10.2`

- Fedora, CentOS or Red Hat

- `curl https://install.citusdata.com/community/rpm.sh | sudo bash`
 - `sudo yum install -y citus102_14`

Citus SQL Language

- PostgreSQL

Example

Choosing column to distribute

```
CREATE TABLE actors (  
    id TEXT PRIMARY KEY,  
    data JSON,  
    movies TEXT[]  
);
```

```
CREATE TABLE movies (  
    id TEXT PRIMARY KEY,  
    data JSONB  
);
```

Example

Distributed Column

```
CREATE TABLE actors (  
  id TEXT PRIMARY KEY  
  data JSON,  
  movies TEXT[]  
);  
  
CREATE TABLE movies (  
  id TEXT,  
  actors_id TEXT,  
  data JSONB,  
  PRIMARY KEY (actors_id, id),  
  FOREIGN KEY (actors_id) REFERENCES actors (id)  
);
```



```
SELECT create_distributed_table('actors', 'id');  
SELECT create_distributed_table('movies', 'actors_id');
```

Example

Example

```
INSERT INTO movies (id,actors_id,data)
VALUES ('samotari', 'trojan', '{ "title": { "cs": "Samotari", "en": "Loners" }, "year": 2000, "rating": 84, "length": 103,
"actors": [ "trojan", "machacek", "schneiderova" ], "genres": [ "comedy", "drama" ], "country": [ "CZ", "SI" ] }');
INSERT INTO movies (id,actors_id,data)
VALUES ('medvidek', 'machacek', '{ "title" : "Medvidek", "year": 2007, "rating": 53, "length": 100, "director": { "first": "Jan", "last": "Hrebejk" },
"actors": [ "trojan", "machacek", "vilhelмова", "issova", "menzel" ], "genres": [ "comedy", "drama" ], "country": [ "CZ" ] }');
INSERT INTO movies (id,actors_id,data)
VALUES ('vratnelahve', 'schneiderova', '{ "title": { "cs": "Vratne lahve", "en": "Empties" }, "year": 2006, "rating":76, "length":99,
"director": { "first": "Jan", "last": "Sverak" }, "actors": [ "sverak", "machacek", "schneiderova" ], "genres": "comedy", "country": "CZ" }');
INSERT INTO movies (id,actors_id,data)
VALUES ('zelry', 'sverak', '{ "title": "Zelry", "year": 2003, "rating":81, "length":142, "director": { "last": "Trojan", "first": "Ondrej" },
"actors": [ ], "genres": [ "romance", "drama" ], "country": [ "CZ", "SK", "AT" ] }');
INSERT INTO movies (id,actors_id,data)
VALUES ('stesti', 'machacek', '{ "title": "Stesti", "year": 2005, "rating": 72, "length": 100, "director": { "last": "Slama", "first": "Bohdan" },
"awards": [ { "type": "Czech Lion", "year": 2005 } ] }');
INSERT INTO movies (id,actors_id,data)
VALUES ('kolja', 'sverak', '{ "title": "Kolja", "year": 1996, "rating":86, "length":105, "awards": [ { "type": "Czech Lion", "year": 1996 },
{ "type": "Academy Awards", "category": "A", "year": 1996 } ] }');
```

Advantages

- Fast queries for all tenants
- Hold more data than possible in single-node PostgreSQL
- Scale out without giving up SQL
- Isolate resource usage of large and small customers
- Parallelize SQL queries
- Use one database, not a patchwork-
- Rich PostgreSQL data types and extensions

Drawbacks

- When single-node Postgres can support your application and you do not expect to grow
- Offline analytics, without the need for real-time ingest nor real-time queries
- Analytics apps that do not need to support a large number of concurrent users
- Queries that return data-heavy ETL results rather than summaries