



## From FDWs to Sharding

- Robert Haas | PGCon 2015

# Pushdown

- `SELECT * FROM ft1 WHERE x = 1` → Yes!
- `SELECT * FROM ft1 ORDER BY y` → No.
- `SELECT * FROM ft1, ft2 WHERE ft1.x = ft2.x` → No.
- `SELECT count(*) FROM ft1` → No!!
- `UPDATE ft1 SET x = 1 WHERE y = 2` → No!!

# Partitioning: No!

Join

→ Append

→ Scan server s1, table x1

→ Scan server s2, table x2

→ Scan server s3, table x3

→ Append

→ Scan server s1, table y1

→ Scan server s2, table y2

→ Scan server s3, table y3

# Partitioning: Yes!

Append

→ Join

→ Scan server s1, table x1

→ Scan server s1, table y1

→ Join

→ Scan server s2, table x2

→ Scan server s2, table y2

→ Join

→ Scan server s3, table x3

→ Scan server s3, table y3

# Partitioning: Aggregates

AggregateCombiner

→ PartialAggregate

→ Scan server s1, table x1

→ PartialAggregate

→ Scan server s2, table x2

→ PartialAggregate

→ Scan server s3, table x3

# Partitioning: Replication

```
SELECT * FROM sharded_table JOIN codes_table ON  
sharded_table.codes_table_id = codes_table.id
```

- Maybe there's a copy of codes\_table on every node!

# MVCC

## Atomic Commit

- Single update touches multiple shards, either all updates commit or all updates roll back.

- Atomic Visibility

- Single update touches multiple shards, no one can see an intermediate state where some but not all updates have committed.