redgate

Mastering PostgreSQL Partitioning: Supercharge Performance and Simplify Maintenance

Ryan Booz

redgate | About me

Ryan Booz

PostgreSQL & DevOps Advocate

@ryanbooz

/in/ryanbooz

www.softwareandbooz.com

youtube.com/@ryanbooz

Agenda

01 The What and Why

02 Types of Partitioning

03 Creating and Maintaining Partitions

04 Demo

05 SQL Tips and Best Practices

redgate

01/05

The What and Why

redgate

https://www.nj.com/bergen/2017/11/man_who_owes_17k_in_tolls_arrested_at_gwb_cops_say.html

Large tables don't scale well...

...and table partitioning helps to solve that scaling problem

What Are Partitions?

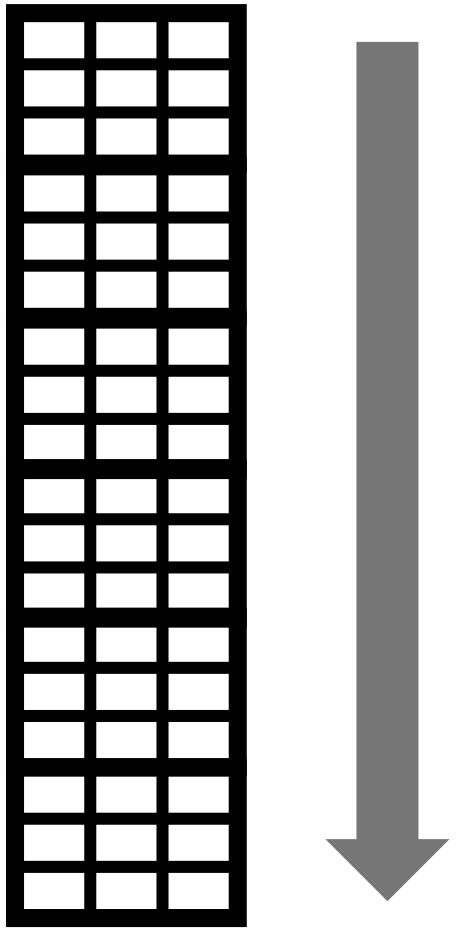
- Regular PostgreSQL tables
  - Schema
  - Tablespace
- Attached to a parent table (children)
- Self-contained indexes
- Can also be parent tables (sub-partitioning)

Declarative Partitioning

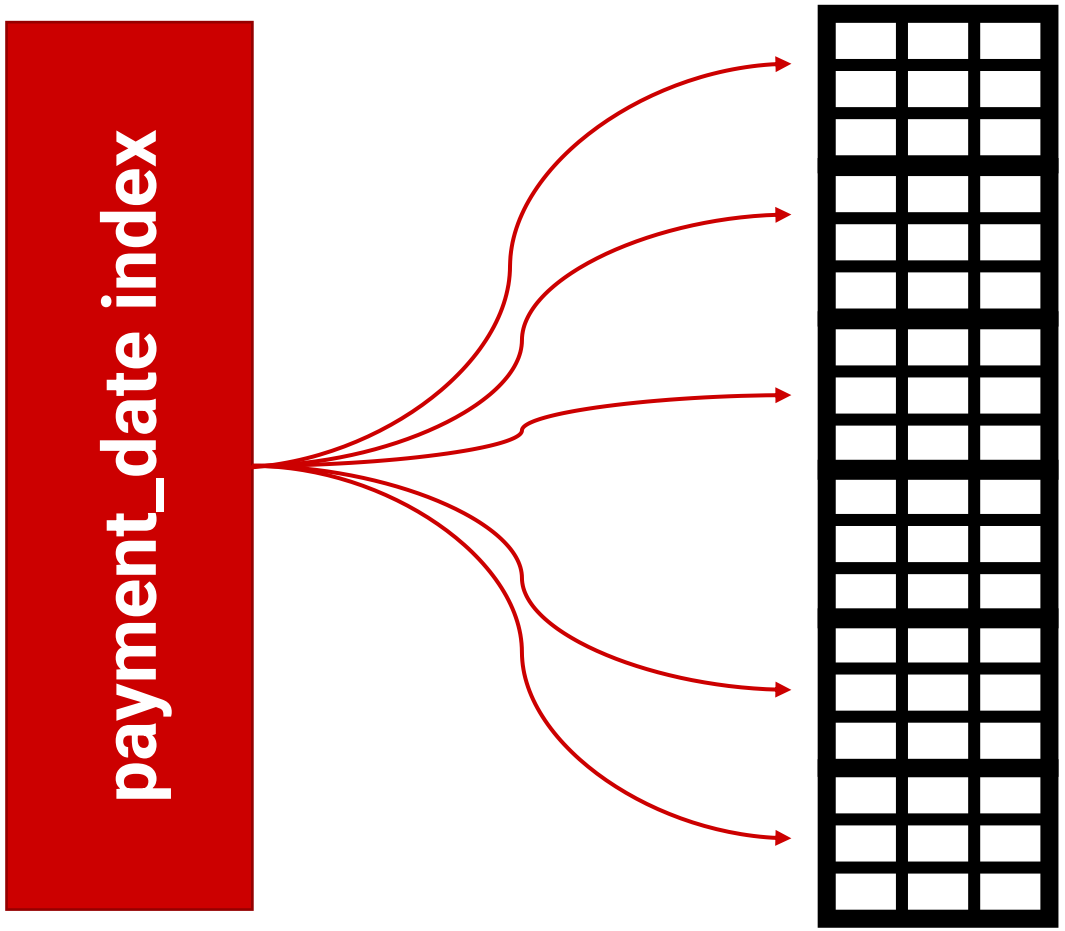
- PostgreSQL 10+
- Identify partitioning key column
- Specify method (RANGE, LIST, HASH)
- Partitions must be created before data arrives
- Default Partition = catch all
  - Can introduce challenging maintenance

```
CREATE TABLE public.payment (
   payment_id serial4 NOT NULL,
   customer_id int4 NOT NULL,
   rental_id int4 NOT NULL,
   amount numeric(5, 2) NOT NULL,
   payment_date timestamptz NOT NULL,
   CONSTRAINT payment_bak_pkey
      PRIMARY KEY (payment_date, payment_id)
);
```

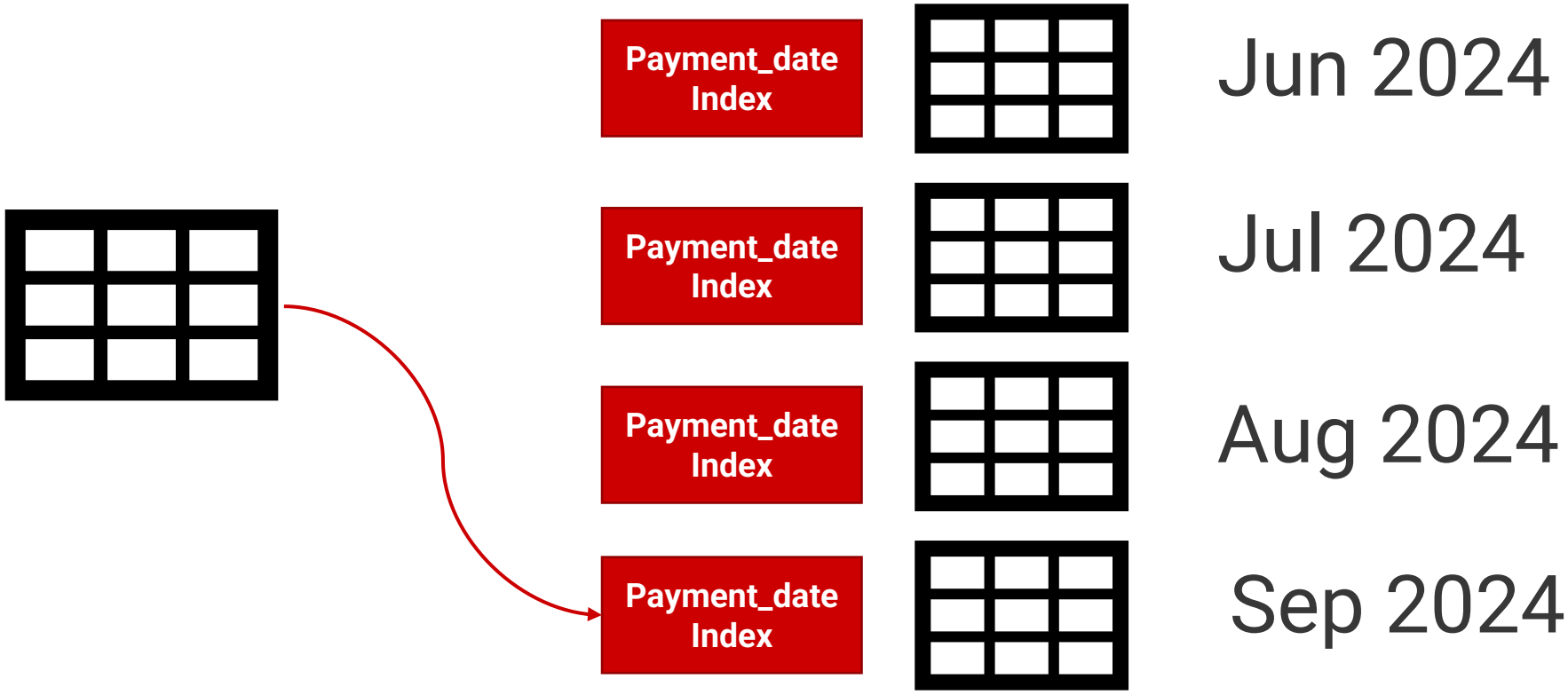
```
SELECT * FROM payment WHERE
    payment_date = '2024-09-11';
```

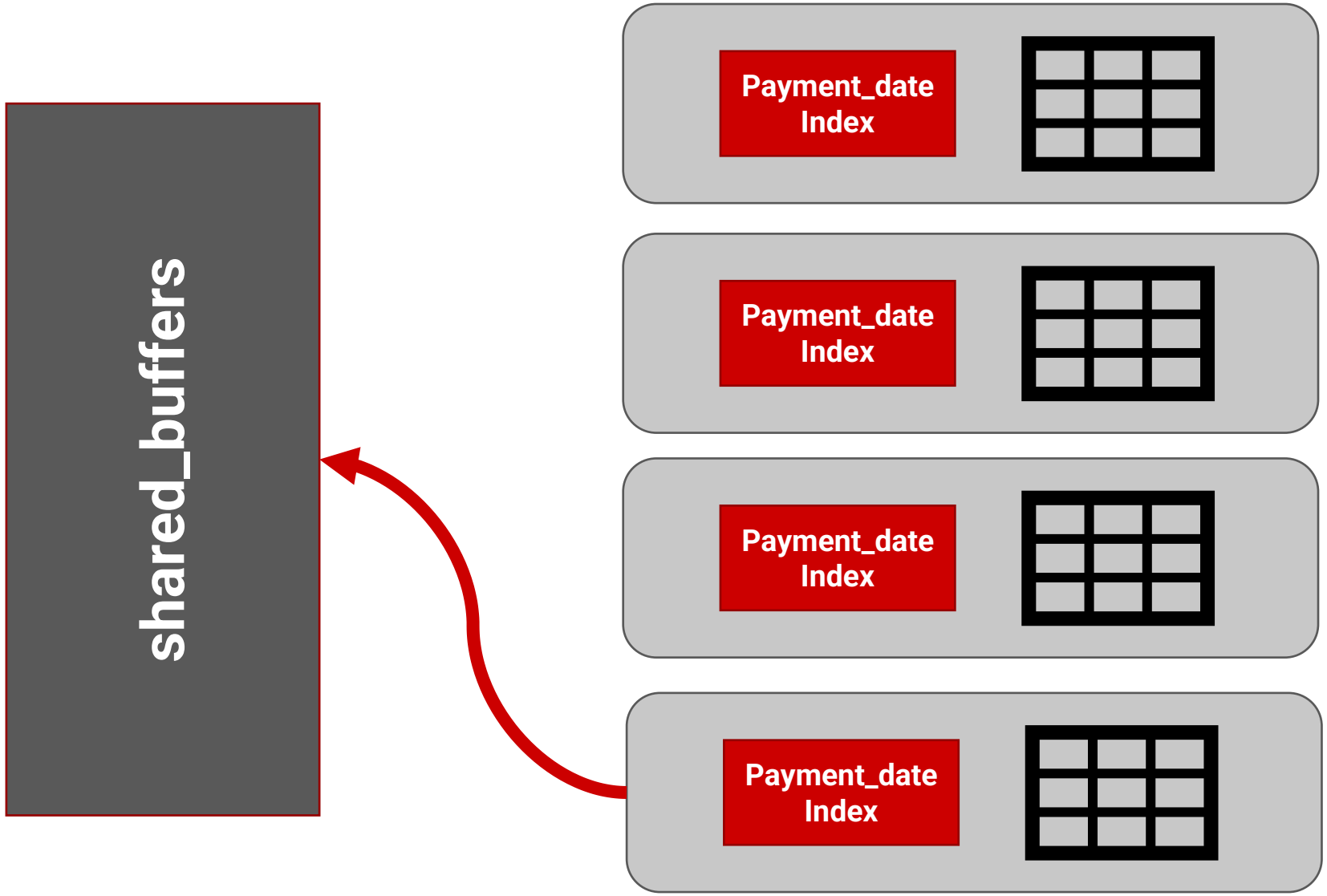


```sql
SELECT * FROM payment WHERE
    payment_date = '2024-09-11';
```



```
SELECT * FROM payment WHERE
    payment_date = '2024-09-11';
```




shared_buffers
Payment_date Index
Payment_date Index
Payment_date Index
Payment_date Index

Data Retention/Archiving

- Without partitions, DELETE adds significant overhead
  - MVCC bloat
  - Index maintenance
  - Potential blocking
- With partitions
  - DETACH PARTITION
  - DROP TABLE

02/05

Types of Partitioning

For all types:

The partitioned table (parent)
specifies column(s) to partition by...

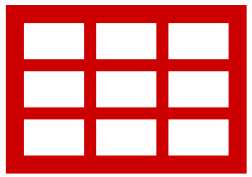
...the partitions (children) specify the values

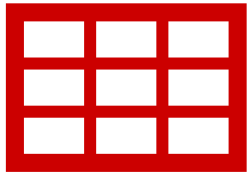
Range

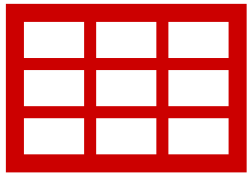
- Most common partitioning method
- Typically
  - Time-series = date/timestamp
  - Object identifiers = integers
- Inclusive of lower value, exclusive of upper value

Range

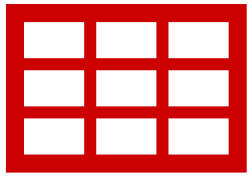
Time-series (logdate)

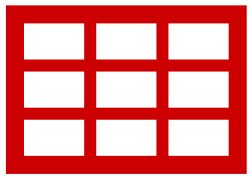

[Jan 1 – Feb 1)

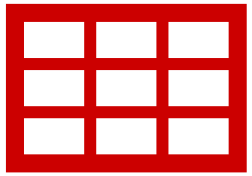

[Feb 1 – Mar 1)


[Mar 1 – Apr 1)

Numeric (category_id)


[1-10)


[10-20)


[20-30)

redgate

```
CREATE TABLE bluebox.payment2 (
   payment_id serial4 NOT NULL,
   ...
   payment_date timestamptz NOT NULL,
   CONSTRAINT payment_bak_pkey
      PRIMARY KEY (payment_date, payment_id)
)
PARTITION BY RANGE (payment_date);

CREATE TABLE bluebox.payment2_y2024m01
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-01-01') TO ('2024-02-01');

CREATE TABLE bluebox.payment2_y2024m02
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-02-01') TO ('2024-03-01');
```

List

- Known list of key values
- Product categories/regions
  - 'sports','board_game','rc_toys'
  - 'region_1','region_2','region_3'

```
CREATE TABLE bluebox.customer (
   customer_id int8 NOT NULL,
   store_id int4 NOT NULL,
   full_name text NOT NULL,
   email text NULL,
   ...,
   state text NOT NULL,
   ...
)
PARTITION BY LIST (state);

CREATE TABLE bluebox.customer_northeast
     PARTITION OF bluebox.customer
FOR VALUES IN ('NY','PA','VT','CT','RI','ME','NH','NJ');
```

Hash

- Specify a MODULUS and REMAINDER
- No clear partitioning key
- Generally even distribution of data
- Cannot re-partition in the future without rewrite

```
CREATE TABLE bluebox.payment2 (
    rental_id int NOT NULL,
   ...
   payment_date timestamptz NOT NULL,
   CONSTRAINT payment_bak_pkey
      PRIMARY KEY (rental_id)
)
PARTITION BY HASH (rental_id);

CREATE TABLE bluebox.payment2_p0
     PARTITION OF bluebox.payment2
FOR VALUES WITH (MODULUS 8, REMAINDER 0);

CREATE TABLE bluebox.payment2_p1
     PARTITION OF bluebox.payment2
FOR VALUES WITH (MODULUS 8, REMAINDER 1);
```

03/03
Creating and Maintaining Partitions

redgate

```
CREATE TABLE bluebox.payment2 (
   payment_id serial4 NOT NULL,
   customer_id int4 NOT NULL,
   rental_id int4 NOT NULL,
   amount numeric(5, 2) NOT NULL,
   payment_date timestamptz NOT NULL,
   CONSTRAINT payment_bak_pkey
      PRIMARY KEY (payment_date, payment_id)
)
PARTITION BY RANGE (payment_date);
```

```
CREATE TABLE bluebox.payment2_y2024m01 (
   payment_id serial4 NOT NULL,
   customer_id int4 NOT NULL,
   rental_id int4 NOT NULL,
   amount numeric(5, 2) NOT NULL,
   payment_date timestamptz NOT NULL,
   CONSTRAINT payment_bak_pkey
      PRIMARY KEY (payment_date, payment_id)
);
```

```
CREATE TABLE bluebox.payment2_y2024m01
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-01-01') TO ('2024-02-01');


CREATE TABLE bluebox.payment2_y2024m02
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-02-01') TO ('2024-03-01');


CREATE TABLE bluebox.payment2_y2024m03
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-03-01') TO ('2024-04-01');
```

Table DDL, and most admin tasks, should be applied through the parent table

Table/Partition Modifications

```
-- Altering the parent propagates to the children
ALTER TABLE payment2 ADD COLUMN status TEXT;
```

```
-- Adding an index propagates to the children
CREATE INDEX payment2_payment_date_customer_id_idx
  ON public.payment2
     USING btree (payment_date,customer_id);
```

NOT A GLOBAL INDEX!!

redgate

Data Retention

```
-- Partition is deleted and no longer queryable
DROP TABLE payment2_y2024m01;

-- Partition still exists but no longer queryable
ALTER TABLE payment2 DETACH PARTITION payment2_y2024m02;

-- Table must have the same schema and valid, checked
-- constraints that match the partition key
ALTER TABLE payment2 ATTACH PARTITION payment2_y2024m02
FOR VALUES FROM ('2024-02-01') TO ('2024-03-01');
```

Default Partition

- Special "catch all" partition
- Prevents data loss
- Maintenance headache as new partitions are created
- TL;DR; - often a necessary ~~evil~~ help

```
CREATE TABLE bluebox.payment2_default
     PARTITION OF bluebox.payment2 DEFAULT;

CREATE TABLE bluebox.payment2_y2024m01
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-01-01') TO ('2024-02-01');

CREATE TABLE bluebox.payment2_y2024m02
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-02-01') TO ('2024-03-01');

CREATE TABLE bluebox.payment2_y2024m03
     PARTITION OF bluebox.payment2
FOR VALUES FROM ('2024-03-01') TO ('2024-04-01');
```

Partition Automation

- Don't plan on manual creation or retention
- Common extensions:
  - pg_partman/pg_cron
  - timescaledb (not declarative partitioning)
  - citus
  - EDB Postgres Distributed Server

04/05
Demo

redgate

05/05

Tips and Best Practices

redgate

Merging or Splitting Partitions

- Not currently supported in Postgres core
- Nearly introduced in PG17, but reverted last minute
  - 🤞 for PG18!
- Can accomplish manually given previous discussions on attaching/detaching partitions
  - Remember to create appropriate constraints ahead of time

redgate

Partition size and number

- `shared_buffers` 25%+/- memory
- Hot partitions should fit in shared buffers
- Too many partitions can increase planning time
  - >~1,000 partitions may require constraint/range changes
  - <=PG13 may struggle with LW Locks
- For large partitions, ensure relevant indexes are available after partition exclusion

Partition size and number

- Don't try to target row count as your target
- Consider data retention

Always filter by the partition key

- Partition key should always be a predicate
- Query specific ranges if possible
- For dates, avoid interval math if possible
  - `payment_date > '2024-03-19'` vs.
  - `payment_date > now() - '1 month'::interval`

Use Schemas and Tablespace

- Consider creating partitions in a separate schema
- Separate tablespaces = data tiering
  - Hot data in fastest memory-based storage
  - Warm data in cheaper storage
  - Cold data to disk or object storage

Data Tiering

```
-- Partition to a new tablespace
ALTER TABLE payment2_y2024m02 SET TABLESPACE slow_ts;

-- Partition index to a new tablespace
ALTER INDEX rental2_y2023m01_rental_period_idx
      SET TABLESPACE slow_ts;
```

redgate

🎉 THANK YOU! 🎉

redgate

What Questions do you have?

redgate

🎉 THANK YOU! 🎉

github.com/ryanbooz/presentations

redgate