GroupBy

Free Online Training for Data Professionals.
By the Community, For the Community.

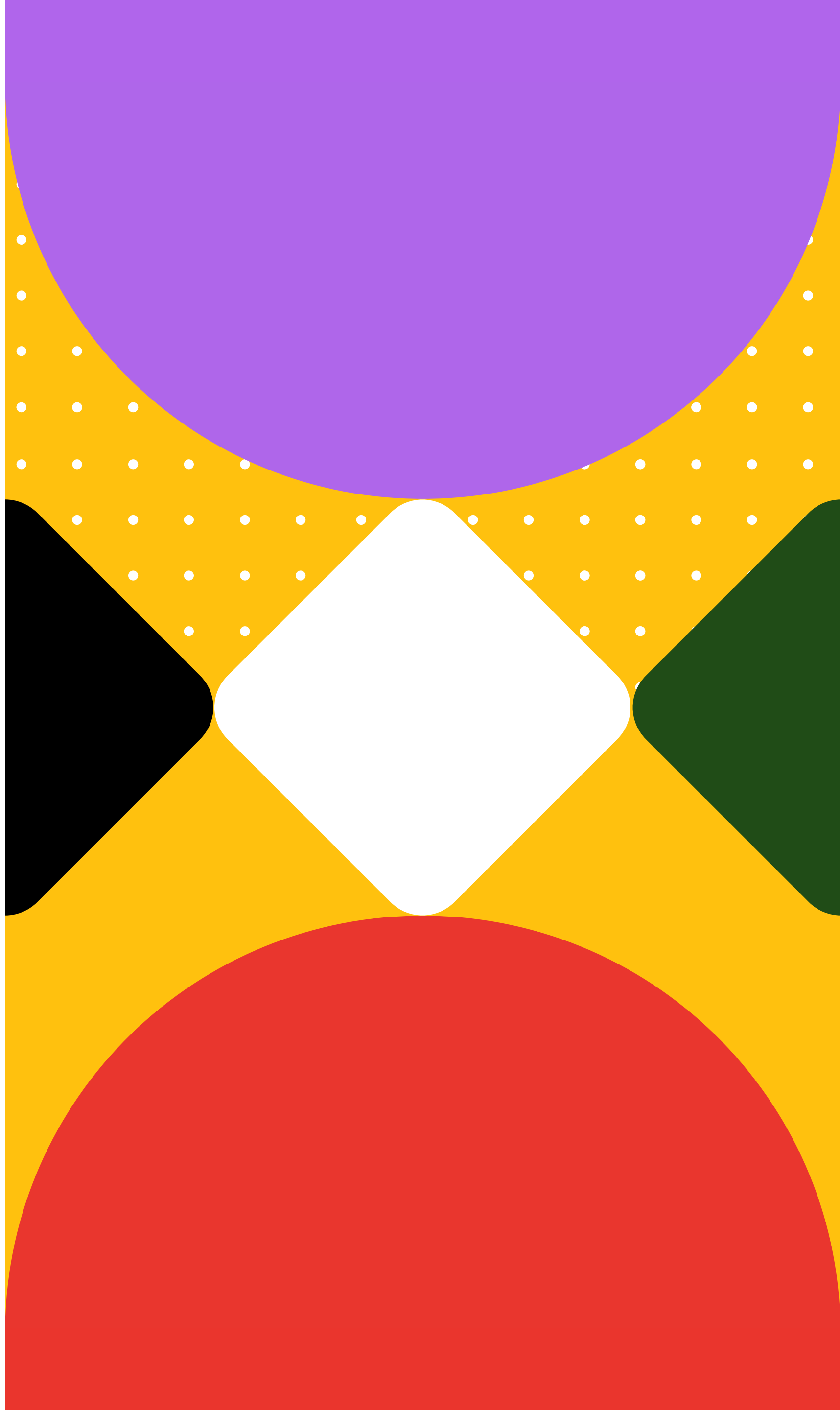
GroupBy Code Of Conduct

The Quick Version

We are dedicated to a harassment-free experience for everyone, regardless of who you are and what makes you *you*. We recognize the right of any individual to attend and participate. Anyone. This is included but not limited to gender identity and expression, sexual orientation, disability, physical appearance, body size, race, religion, or any other classification, affiliation, or label.

- **We do not tolerate harassment in any form.** For the duration of your engagement with GroupBy and its programs, you are expected to act appropriately and to adhere to this Code of Conduct. This includes conduct in-person and online, at the conference itself, as well as any non-conference programs that may include participants: including talks, workshops, parties, on social media, and other online forums. GroupBy participants violating these rules may be sanctioned or expelled without a refund (if that applies) at the discretion of the conference organizers.

- You can review the full policy at: **GroupBy.org/Code-of-Conduct**



GroupBy

Media Sponsor:

IDERA

IDERA.com

GroupBy

ETL Without Permission

Bulk-loading data quickly and easily, without elevated permissions or expensive tools

Courtney White

CursorsAreEvil.com

Courtney White

Lead Business Initiatives Consultant @ Wells Fargo

SQL – VBA – C# – process engineering

Linked in linkedin.com/in/courtney-white-mba-pmp/

CursorsAreEvil.com

Accustomed to almost never having the *right* tool for the job, Courtney has come to enjoy building workarounds and solving problems with the best *available* tool for the job.

That tool is never a cursor.

Courtney lives outside Baltimore, Maryland, USA, with his wife, two elementary-school-aged children, cat, and unfinished garden projects.

(that I actually use)

Resources

- Sessions from past conferences
- SQL Data Partners Podcast: https://sqldatapartners.com/podcast/
- Brent Ozar's blog: https://www.brentozar.com/blog/
- Local user group events
- SQL Saturday
- Code from presentation is available at http://code.CursorsAreEvil.com

GroupBy

Approach and Principles

- Minimize network operations
- Prefer sets over rows
- Minimize resource contention
- Perform expensive operations exactly once
- Prefer large batches over small
- Use your knowledge of data & structure to optimize queries
- Load only what you need
- If you need an index, add it

Steps

Code

- Perform local transformations

SQL
Code

- Create temp tables

SQL

- Add constraints if necessary

Code

- Insert data to temp tables

SQL

- Index as necessary

SQL

- Generate final data set

SQL

- Load to target

- Match primary / unique keys on target table to indexed fields on temp tables where possible

Insert Data – Naïve Approach

- One row at a time
- Cursors or loops
- Common in non-SQL languages
- Dynamic or parameterized SQL
- Implicitly part of many applications (e.g. Access via linked tables)

- Repeated network calls
  - Reduced with batching
- Repeated IO calls
  - Batching doesn’t help
- Usually synchronous & sequential
- Query plans can usually be reused

Insert Data – Batch Approach

- Multiple rows at a time
  - Multiple single-statements
  - Single multi-row statement
  - Multiple multi-row statements
- VALUES limited to 1000 rows
  - … UNION ALL SELECT * FROM VALUES(…) AS X(Field1, Field2…)
- Usually largely dynamic SQL
  - Requires care to avoid injection

- One network call / batch
- One IO bundle / statement
- Creating batches relatively slow
- Individual server operations are relatively slow
- Individual query plans probably required

Weird Approaches

- Tokenizing duplicative inputs
  - Employee tables
- JSON/XML inputs
  - Send as a single parameter
- COMPRESS / DECOMPRESS
  - Compress your JSON/XML

- Proprietary formats
  - No.
- Open formats
  - Only if it’s reusable
- Micro-optimizations
  - I try not to. You should, too.

Performance Considerations

- MERGE vs DELETE/UPDATE/INSERT
    - Compare performance
    - Consider risks from MERGE errors
- Watch your locks
    - Minimize eXclusive locks on production tables
    - Sometimes 2 locks are faster than one
    - Even Shared locks can cause problems (if the query is slow enough)

GroupBy

ETL Without Permission

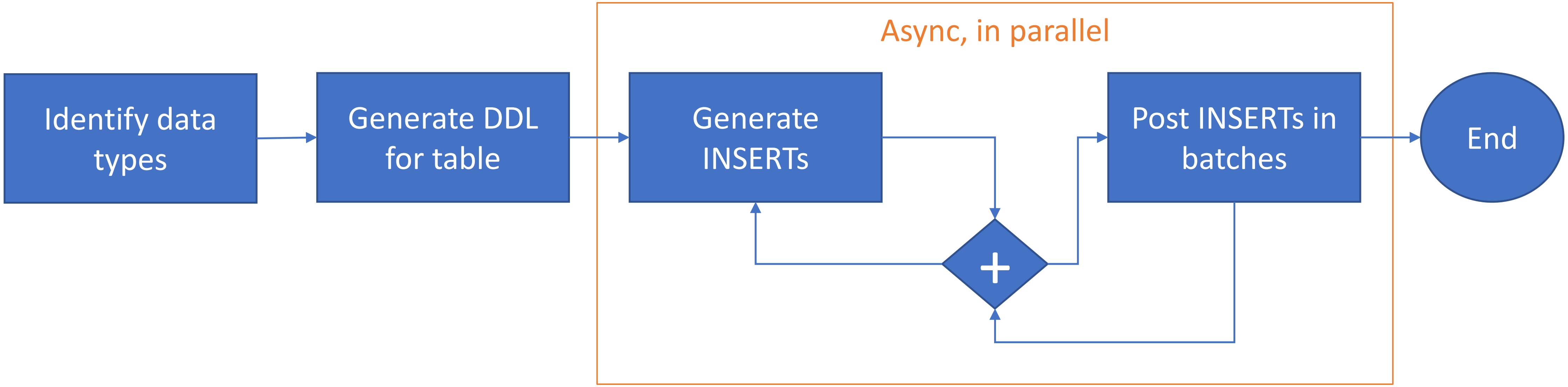
Bulk-loading data quickly and easily, without elevated permissions or expensive tools

SSMS Demo

Troubleshooting Performance

- Are your joins indexed?
  - Could you be using the clustered index or primary key?
- Are you performing unnecessary updates?
  - Use EXCEPT to eliminate meaningless UPDATEs
  - A rare time that it’s good to DUI: DELETE → UPDATE → INSERT

What is the code doing?



GroupBy

ETL Without Permission

Bulk-loading data quickly and easily, without elevated permissions or expensive tools

Excel Demos

Dynamic SQL Options

Sanitize your inputs (escape strings, etc.)

- Assuming you control the collations, this is entirely doable

Use parameters for everything

- Upper limit on parameters is ~2100
- The SQL injection risk from a numeric type is zero

Use parameters for all string values

- Practical if you have duplicate values and are tokenizing everything anyway

Cast all string fields to VARBINARY

Bonus – Handling Slow Linked Server Queries

Query a replica

- Not always available ☹

Use snapshot isolation

- Not always available ☹ ☹

Cache your OPENQUERY results to temp tables

Index your temp tables appropriately

Cache and index local tables if necessary

GroupBy

Free Online Training for Data Professionals.

By the Community, For the Community.

GroupBy

We're taking a short break.

Sessions will resume soon!