Identifying and Fixing Performance and Logic Issues

Chris B. Behrens
SOFTWARE ARCHITECT
@chrisbbehrens

Logging and Instrumentation with SSIS

Logging Data in SSIS

We're going to attach a log provider to the event hooks that SSIS fires

The events always fire, but without a delegate, they go nowhere and do nothing

Logging Providers in SSIS

- Sql Server
- Text File
- XML File
- Event Log
- Sql Profiler
- I don't recommend Event Log or Sql Server...
- Because those are very hard to log ship

The Text Providers

Text file (CSV)

Xml file

Sql Profiler

When in doubt, use text file

The Sql Trace Provider

The ability to *replay the log*

This allows you to re-execute the output of the package against a target db

Demo

Configure logging for our package

- With the text file provider
- Select some events

Execute our package

- Review the logs
- Talk about what it means

An Anti-pattern for Integration Services Performance

A Massive Data File

The file had dozens of fields

The size of the file was around a gigabyte (and growing)

We load the entire file and table into memory

Then join them together to perform INSERT, UPDATE and soft-DELETE operations

How Our Approach Changed

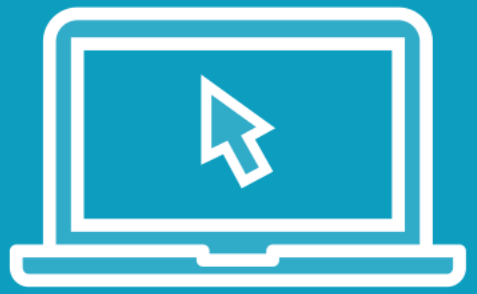
We began with a small set of sample data

Our approach didn't hold up under the full size load

So, we loaded the entire file into a temporary table, and resolved it inside

Big Table

Demo



- Modify our ETL package to import the data to a temporary table
- Create a view that reconciles the imported data
- Created a stored proc that imports the new data
- Execute our new package
- Talk about a problem with this approach

Half-splitting

A Kill to A View

Our view transforms the data types

But if the source data doesn't conform to the target data type...

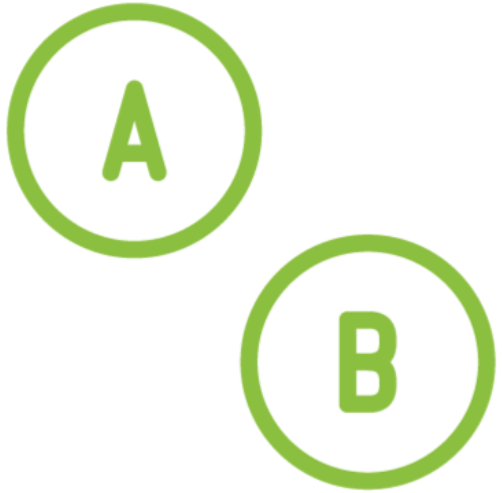
The view will fail with an exception

And you have to find the needle in the record haystack

A Tool for Your Toolbox

We've got a tool for this kind of problem

And it can be applied widely to problems in technology and elsewhere



Half-splitting allows us to isolate the problem

Half-splitting a View

A view is just a query

We can limit the scope of the query with different syntaxes

And figure out which half of the records contains the problem

Half-splitting Queries

```
SELECT TOP 3 Row = ROW_NUMBER()  OVER (ORDER BY Id), ID, Title, Description, Points FROM vw_TempActivities

SELECT TOP 3 Row = ROW_NUMBER()  OVER (ORDER BY Id), ID, Title, Description, Points FROM vw_TempActivities
ORDER BY ROW desc


WITH expr AS
(
    select   Id,
             Row = row_number() OVER ( order by ID)
    from     vw_tempActivities
)
SELECT  A.ID, A.Title, A.Description, A.Points
FROM    expr E JOIN vw_tempActivities A on E.Id = A.ID
WHERE   Row > 3 AND Row < 6
```

Fixing the Problem

Once you have found the record...

For example, an "N" in what should be a numeric field...

You can decide how to fix it

```
SELECT * FROM TEMP_Activities WHERE Points = ‘N’
```

Multiple Bad Records

If there’s one problem…

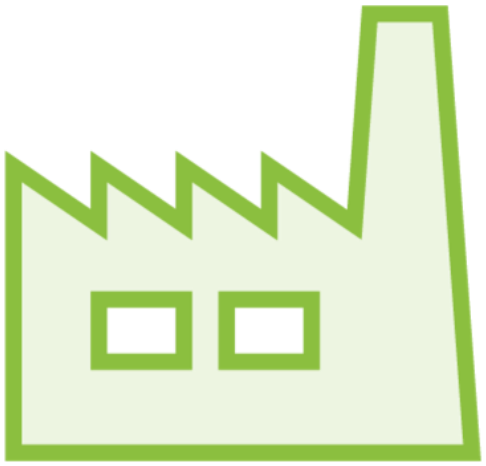
There are probably more

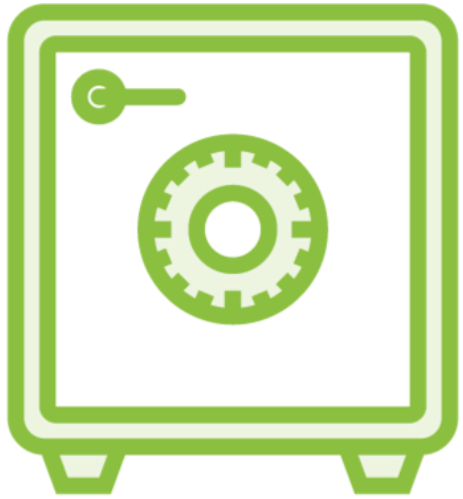
This is not a problem

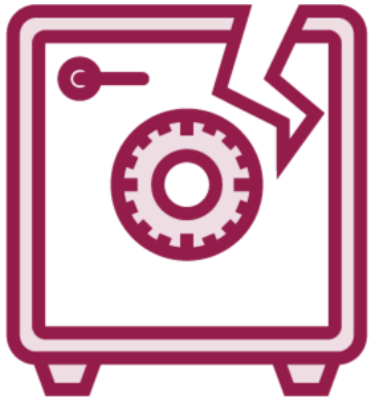
If it’s in both halves (or eighths, or 1/2048), just pick one, either one will work

The Worst Thing You Can Do

Why This Is so Bad







Exceptions Don’t Belong in SSIS

The Nature of Exceptions

Exceptions should be exceptional

Of course, errors are going to happen (often)

But practically, the only solution is to redesign your process to eliminate it

Summary

Getting good information out of our system

With logging from SSIS events

- OnError
- OnInformation

Log Providers

- Only three providers as useful
  - Text Provider
  - Xml Provider
  - Sql Trace

Common anti-pattern

- Big Table
- How to get around it
  - Temp tables
  - Views
  - Stored procs