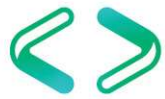
Tools for the SQL Server Accidental DBA

Ryan Booz

@RyanBooz

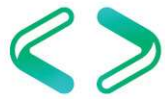
http://www.SoftwareAndBooz.com

ryan@SoftwareAndBooz.com



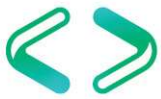
Who Am I?

- Husband and Father of 5
- Hobby farmer with 33 acres
- CTO at EnergyCAP, Inc. located in Central PA
- 20+ years of Relational DB experience
- Cut my teeth on a TRS-80 in the mid-80s
- Learned to program on a DEC 5000 in high school
- Data Science wannabe! (see bullet 1, 2 & 3…)



Agenda

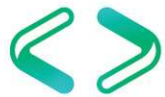
- EnergyCAP Overview
- Drink from the Firehose!
  - Monitoring
  - Server Setup
  - Constraints and FKs
  - Indexes
  - Query Tuning
  - Getting Help



EnergyCAP History

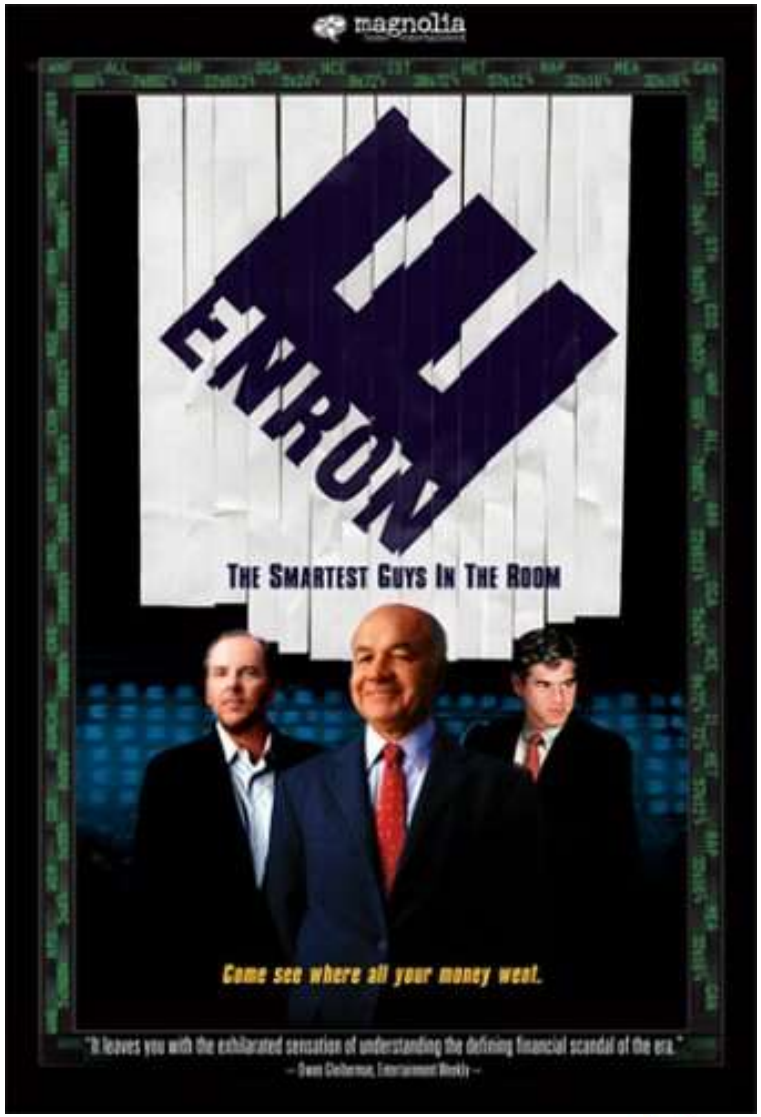
EnergyCAP®

- Utility Bill Management software started in 1980
- 1996 - Acquired by Enron
- 1998 – Relocated to Houston
- Started a brand new product but never released
- I.P. purchased back in 2002 to start EnergyCAP
- *Our DBA never came back to State College*



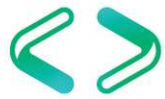
The smart guys at Enron…

- Sybase to SQL 6.5
- Erwin data modeler
- C++ installed application
- SQL XML Web Service
- Highly normalized database
  - ~200 tables
- Well done for the time



EnergyCAP Ecosystem

- Single DB accessed by three independent apps
  - Legacy desktop C++ client
  - (Almost) legacy Flash-based web app with .NET 4.5
  - Modern Angular app backed by .NET Core web services
- SQL 2008R2, 2012 & 2014 currently in use
- SaaS offering with ~300 production databases (representing ~2000 clients)

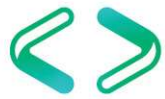


EnergyCAP Ecosystem

- Client-hosted DBs:
  - Often our largest databases
  - Almost always our limiting factor when it comes to new technologies

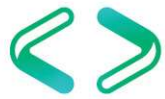
Main Challenges

- Database/SQL Server Knowledge Silo
- Didn’t upgrade tooling consistently
- Poorly done monitoring
- SaaS environment always worked better than large, self-hosted clients
- Simply came to accept “good enough” performance…



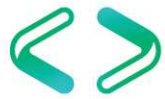
Two Caveats

- We're fortunate to host many of our client DBs
  - We get to know the data really well
  - We can observe client usage patterns and test accordingly
- We don't have a ton of Cloud DBA experience yet
  - Many of the improvements we have made are specifically preparing us for a cloud migration



One Number Changed My Journey

2371



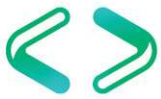
Monitoring

Monitoring


Overview
% Processor Time (7%)
100
80
8
6
4
2
0
10
Batch Requests/sec (2)
Processes
S..
U..
Login
Application
Wait Tim..
Wait...
B...
Resource Waits
Data File I/O
Recent Expensive Queries

Monitoring

- Baseline
- Essential to measure progress
- Free and Paid options we've tried
  - ELK Stack (OSS)
  - Splunk
  - SQL Sentry
  - RedGate
  - IDERA



My Top Picks

- Batch Request/sec
- Buffer Hit Cache Ratio
- Processor Queue Length
- Server CPU
- Page Life Expectancy
- Deadlocks
- Disk Read and Write
- Scans/splits
- (Re)-Compliations

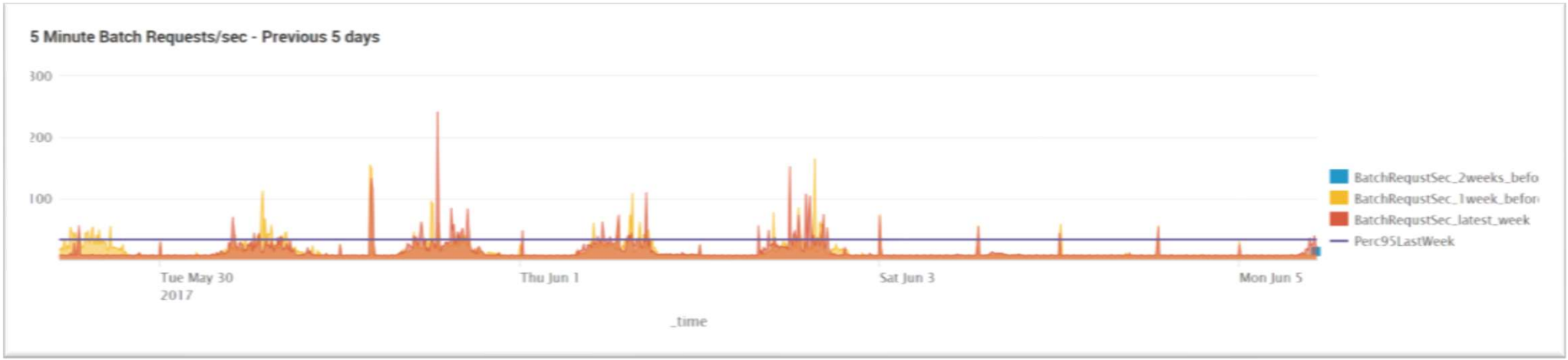
There are many great articles and suggestions. Start with most of these and grow from there.

Visualize these somewhere and check it regularly to get familiar with how your servers/DBs act

Be careful of the observer effect!

Bare Minimum

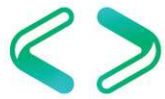
- Create some charts that compare current period to previous (this week vs. last week)
- Put key metrics in a dashboard of some sort that make it easy to surface and access



Bare Minimum

- Set a logging process with sp_WhoIsActive

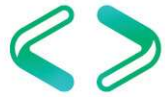
  https://www.brentozar.com/archive/2016/07/logging-activity-using-sp_whoisactive-take-2/
- Enable and configure Database Mail
- Long Running Queries alert
- Server Health alerts



Server Setup

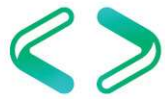
Don’t Reinvent The Wheel

- Find a setup guide that you can start with and build upon it
  - One included with Brent Ozar First Responder Kit
  - https://www.brentozar.com/first-aid/
- Learn about your current server setup
  - Glenn Berry Diagnostic Information Queries
  - https://www.sqlskills.com/blogs/glenn/sql-server-diagnostic-information-queries-for-september-2017/



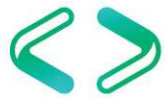
Trace Flags

- SQL Server's way of enabling new features
- Must be "turned on" as part of SQL Server startup
  - Added to startup command
  - Coded through startup PROC
  - DBCC TRACEON() for immediate need



Minimum Trace Flags To Consider

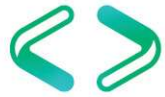
Trace Flag Number	What does it do?
1118	Forces allocation of Full Extents for object creation – usually helps TempDB the most and advised at least through SQL 2016
2371	Lowers the threshold for updating table statistics on larger tables
3226	Stops logging successful backup messages to SQL Server logs
1204 & 1222	Capture deadlock information in SQL Server logs - Read up carefully before enabling



Advanced Settings

- Most do not require a server restart
- Script for consistency

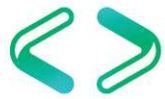
Setting	Recommended Value
Cost Threshold for Parallelism	Start at 50
Optimize for Ad Hoc Workloads	True (in many cases)
Max Degree of Parallelism	8 or less, very dependent on server hardware and setup. NUMA and Virtualized impact choices



TempDB

- Don’t ignore this, especially with OLTP data
- Work with IT to follow the guidance
  - Multiple data files
  - Separate drive
  - Consider expanding them to full size at creation time

- Any move requires a restart of SQL Server instance

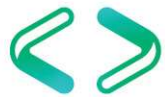


Right-Size Database Files

- Never grow by % (maybe OK in some situations)
- Know your data and take guidance from others
- Free space in files isn't bad if it will get used again

Database files:

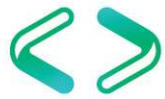
Logical Name	File Type	Filegroup	Initial Size (MB)	Autogrowth / Maxsize
modeldev	ROWS...	PRIMARY	512	By 256 MB, Unlimited
modellog	LOG	Not Applicable	128	By 128 MB, Unlimited



Maintenance Solution Scripts

- Check out Ola Hallegren's free scripts:
  - REORGANIZE or REINDEX databases
  - Database Integrity Checks
  - Database Backups

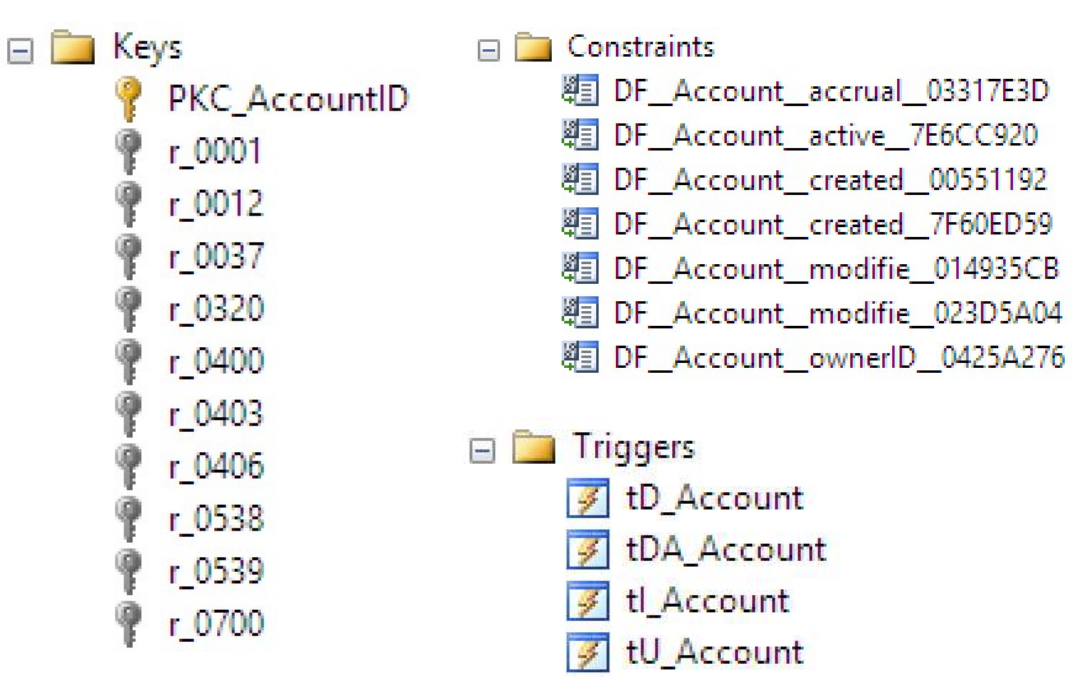
https://ola.hallengren.com/



Naming and Constraints

As important as code comments!

We used Erwin from the beginning, with (mostly) default settings



Foreign Keys:

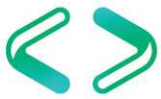
- FK_{TableName}_{FKColumnName}

Default Constraints:

- DF_{TableName}_{ColumnName}

Triggers (when necessary):

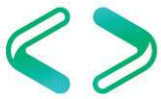
- tD_{TableName}_{Purpose}

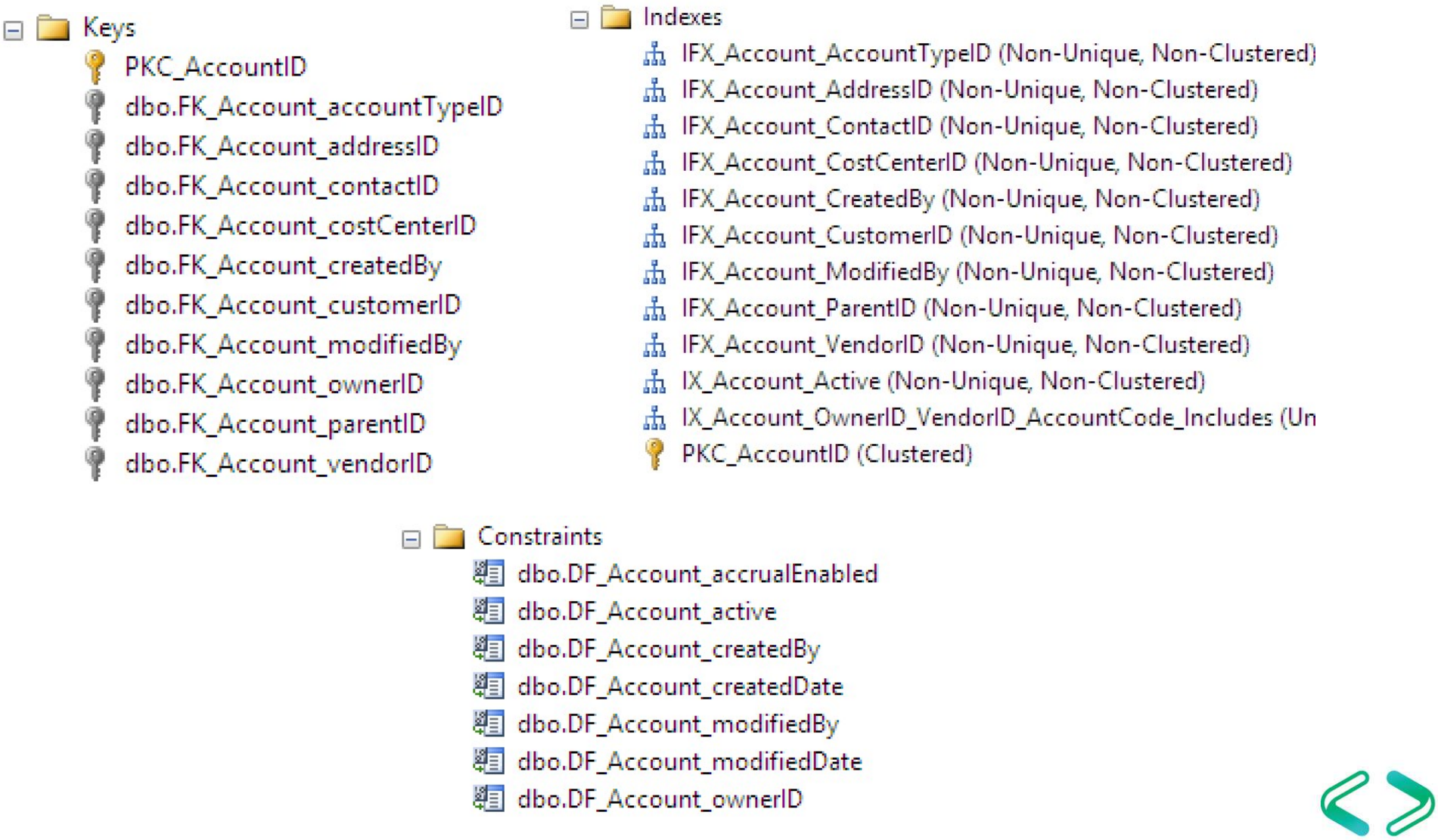


Indexes:

- {Prefix}_{TableName}_{Column(s)}_Includes/Filtered

Prefix	Type
PK	Non-clustered Primary Key
CX	Clustered Index
PKC	Clustered Primary Key
IFX	Foreign Key
IX	User-defined Index

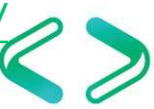



Keys
PKC_AccountID
dbo.FK_Account_accountTypeID
dbo.FK_Account_addressID
dbo.FK_Account_contactID
dbo.FK_Account_costCenterID
dbo.FK_Account_createdBy
dbo.FK_Account_customerID
dbo.FK_Account_modifiedBy
dbo.FK_Account_ownerID
dbo.FK_Account_parentID
dbo.FK_Account_vendorID
Indexes
IFX_Account_AccountTypeID (Non-Unique, Non-Clustered)
IFX_Account_AddressID (Non-Unique, Non-Clustered)
IFX_Account_ContactID (Non-Unique, Non-Clustered)
IFX_Account_CostCenterID (Non-Unique, Non-Clustered)
IFX_Account_CreatedBy (Non-Unique, Non-Clustered)
IFX_Account_CustomerID (Non-Unique, Non-Clustered)
IFX_Account_ModifiedBy (Non-Unique, Non-Clustered)
IFX_Account_ParentID (Non-Unique, Non-Clustered)
IFX_Account_VendorID (Non-Unique, Non-Clustered)
IX_Account_Active (Non-Unique, Non-Clustered)
IX_Account_OwnerID_VendorID_AccountCode_Includes (Un
PKC_AccountID (Clustered)
Constraints
dbo.DF_Account_accrualEnabled
dbo.DF_Account_active
dbo.DF_Account_createdBy
dbo.DF_Account_createdDate
dbo.DF_Account_modifiedBy
dbo.DF_Account_modifiedDate
dbo.DF_Account_ownerID

Foreign Keys

- Don’t use triggers for foreign key constraint logic
- Are your FKs trusted? Check!
- Prevents bad data, regardless of source
- Helps the query planner even when you write bad SQL!

Check out this article and the lively comments:
http://www.scarydba.com/2010/11/22/do-foreign-key-constraints-help-performance/



Other Constraints

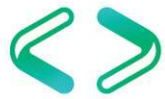
- Date constraints
  - StartDate < EndDate
- Min and Max-type columns

- Use discretion and check impact

Indexes

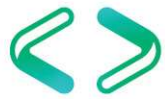
Never Stop Being an Index Parent

- Index tuning is not a once and done event
- Indexes ARE the data - unused or poorly used just cost you space and overhead
- SQL Server provides an abundance of data - use it!
- New query planner certainly changes the game and you probably need to react to it



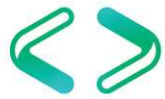
Dynamic Management Objects (DMO)

- 233 views and functions as of SQL 2016
- Provides information on:
  - Work/Executions
  - Transactions
  - Indexes
  - Database/IO
  - SQL Server Operating System



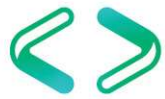
- sp_BlitzIndex script is invaluable!
  - Make sure you know when your statistics get reset so that you know your numbers mean something
  - Sample over time before making any major decisions and changes
  - Create a “back out” script if you can just in case

```
exec sp_BlitzIndex @databaseName='', @mode=4
```



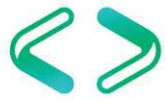
Missing Indexes

- These are only suggestions! With great power…
  - What patterns do you see?
  - Can current indexes be improved instead?
- Always measure the results and monitor impact
- Don’t apply old suggestions if you have modified indexes recently
- Search for ‘dm_db_missing_index_details’ to find example scripts



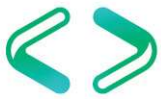
Clear Plan Cache after index changes

- Cached plans might not work well with modified indexes
- Clearing the cache of compiled plans will increase resource usage for some period of time. Be smart.



Our Results

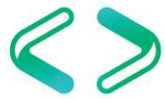
- ~320 indexes removed (25% reduction)
- Database size decreased 30% or more
- Work of key processes were often improved 40% or more



Query Optimization

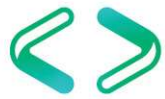
Life-long Learner

- Always new techniques and options to learn
- Find your most expensive queries and start there
  - Plan cache
  - Log sp_WhoIsActive output to find problem queries
- Learn how to read query plans
  - **Install SQL Sentry Plan Explorer!!** (Free)
  - https://www.sentryone.com/plan-explorer



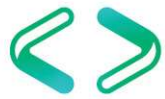
Life-long Learner

- Learn what each query plan operator does
  - Clash of the Row Goals – Adam Machanic (YouTube)
- Be thoughtful about your data and know what to expect
  - How many objects could there be?
  - How many pages of data is that?
  - Do the results and work fall in line?



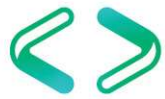
Query Statistics

- Query Statistics IO & Time are your friends!
  - SET STATISTICS IO ON
  - http://statisticsparser.com/
- Query Store
  - Introduced in SQL 2016
  - Check out Erin Stellato from SQLSkills.com



It’s (Almost) All About Reads & Time

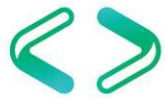
- Use statistics and plans
- Fewer READS means less work
- Sometimes more READS with more efficient plan will perform better



Getting Help

Scratching the Surface

- Twitter:
  - #sqlhelp – https://www.brentozar.com/archive/2009/12/i-need-sqlhelp/
  - #tsql2sday – Started by Adam Machanic
- Slack: sqlcommunity.slack.com
  - #sqlhelp
  - #general
- GroupBy.org – community chosen virtual conference



People on Twitter

- Adam Machanic
- Brent Ozar
- Kendra Little
- Paul Randal
- Kimberly Tripp
- Aaron Bertrand
- Jonathan Keyahias
- Glenn Berry
- Bob Ward
- Paul White

Blogs

- SQLBlog.com
- BrentOzar.com
- SQLSkills.com
- Blogs.sentryone.com
- Red-gate.com/simple-talk