Elastic Search Talk – Script

# Welcome

I’ve worked in the software industry for the past 20 years, coding in C# and using Microsoft’s SQL Server for storage

Recently we’ve been looking at other solutions such as Elastic Search across our group in order to solve some of our performance issues.

Who has used Elastic Search before?

Who has used a NoSql database?

Who has used a relational database?

# Tonight

Tonight we will be looking at Elastic Search from the point of view of a dot net developer. If this was a Java meetup then this would be a different talk!

What issues do we have with SQLServer which we like to solve?

What is elastic search?

How do we get started/install/add data/query it?

Does it solve our problems?

We only have an hour so it’s more of a getting started guide!

# The Story

So you are some coding, getting stuck as usual which requires you to use google and stackoverflow to find various answers

However, then disaster strikes, you lose your internet connection

You can no longer use stack overflow.

# The Application

To enable you to work offline I’ve download a SQLServer database containing 10GB of questions + answers

And written a windows form application in order to query it.

Open code and scope

Demo application, search for c#, sql and unicorn

Show number of results per page

Show wildcard – case insensitive search

# Architecture

Open Demo SQL.txt in SSMS

Look at the database in SSMS (StackOverflow2010)

Point out the size and main tables (Posts and Comments)

Our tables just have indexes on the ID

Do count(\*) to get the number of rows

select rowcnt from sysindexes where OBJECT\_NAME(id) = 'Posts'

We could create indexes, but on which columns, as we might want to search by any of them.

Even then we can’t create an index on the main content field.

We want to do a case-insensisitve search so need to convert ever thing to lower case. Also indexes can’t be used with do are doing a %wild card search

We could use full text search in SQLServer, but it’s limited in features and performance.

Finally, scale, the full stack overflow database is over 130GB, without any indexes. You can split sql server over multiple database but there are complications.

# The Problems

So to summarise, we have three problems

1. The ability to be able to search across multiple fields
2. The ability to be able to search within a block of text
3. The ability to scale out across multiple machines

# Introducing Elastic Search

The products are often referred to as either ELK Stack / Elastic Search

Elastic Search is built on top of lucence.

# E is for Elastic Search

Elastic search is the main query engine built on Lucence. It is typically accessed via REST calls.

L is Log Stash

This gives us an easy way to feed and transform log files into Elastic Search

K for Kibana

This is a visualisation tool build on top of elastic search

# Getting Started

docker pull sebp/elk

docker run –p 5601:5601 –p 9200:9200 –p 5044:4044 –it -–name elk sebp/elk

This starts running

Elastic Search (9200)

Log Stash (5044)

And Kibana (5601)

We will just look at Elastic Search for the moment.

Show http://localhost:9200/

We can see the version number and basic cluster info.

# REST

Open and scope to ElasticSearchRESTDemo tool

System Demos -> Overview

System Demos -> Cluster Health

Communication can be done via REST calls, for example to view the health of our 1 node cluster

<http://localhost:9200/_cluster/health>

Show code

Add document -> Describe URL format -> PUT method -> Schema less

Get Document -> Same URL -> GET method

Can re-add the first document -> version number increases

Add and Get a second document with a different URL/ID

Delete 2nd document – gets then returns a 404

If we wanted to store data about a different type of meetup, we might be tempted to use a different type within the index. However as you can see this fails. The advice about the use of indexes/types as changed. You shouldn’t think of types as tables, but rather a subset of rows in the same index

Delete entire index

Bulk Import action-data repeat – watch carriage returns – fails second time

Second bulk command – different actions. Note no data for delete, different format for update

LOAD STACKOVER FLOW DATA FROM COMMAND PROMPT

Searches

Show all data

Or single index

Can search on either Kevin or Kevin Smith or Kevin Betteridge(!) but we want to Kevin Smith to work but not David Betteridge. So the problem is that its analysing the text into separate words. Notice the score