Mainframes in "Big Data"

My part in it's downfall ? Or opening the floodgates ?

Colin Knight, Db2 Sys Prog Tech Lead
NatWest

Autumn 2025

Mainframes in Big Data - Agenda

- What is big data and what has a mainframe got to do with it ?

- Overview of Db2 z/os in NatWest

- Making Mainframe data available for analytics on big data in NatWest

- The future (it's only just begun)

- Summary, Puzzle, Musings

- Questions

What is big data and what has a mainframe got to do with it ?

How big is data ?
- Current estimate of the size of the global datasphere is around 181 zettabytes (ZB)
- A zettabyte is 1 trillion gigabytes – this is really BIG !
- Estimates of around 400 million terabytes per day (does that sound right ????)
- Could any of us get through a single day without creating any data ?

The biggie – what is Big data ?

What is big data ?

- Extremely large and complex datasets that are difficult to process using "traditional" data processing tools.
- Characterised by the 3 "V"'s
  - Volume…..It's massive, terabytes, petabytes or exabytes
  - Velocity…..Streamed at high speeds needs real-time processing
  - Variety….Data types like Structured (DBMS, Spreadsheets), semi-structured (XML, JSON) & unstructured (text, images, audio, video)
  - Veracity….quality and trustworthiness of the data

What is big data and what has a mainframe got to do with it ?

What is big data used for ?

- Finance: Detecting fraud, risk management
- Healthcare: Diagnosis, personalised medicine
- Retail: Customer behaviour, adverts
- Social media: More advertising
- Government: Health tracking, crime analysis
- Science: Research into climate change
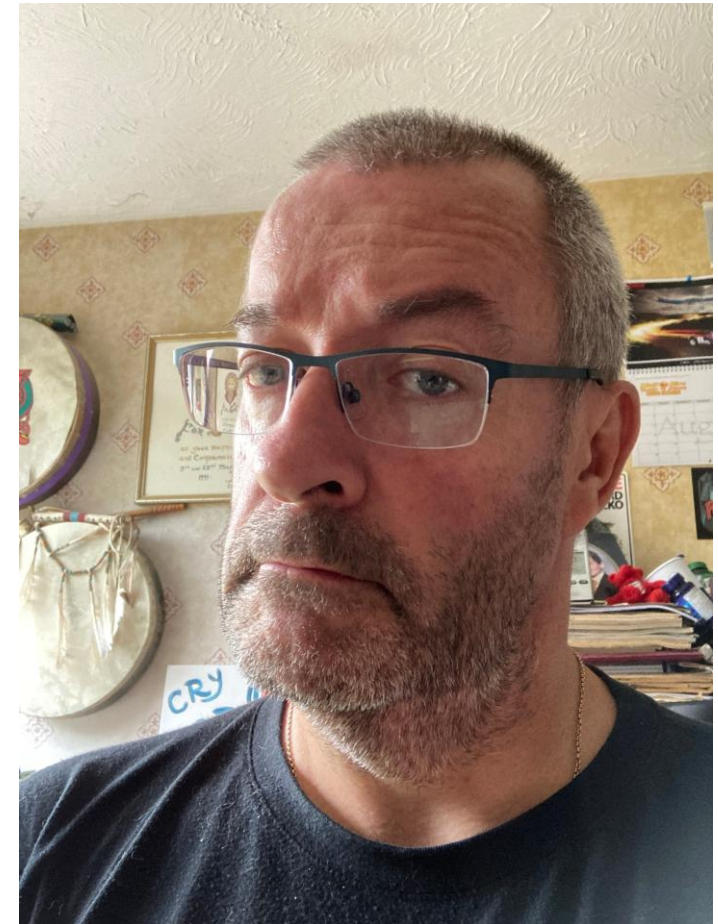- Much much more

Db2 z/os in NatWest

- Peak volume in 2025: 21,740 tps
- Biggest table in production: 65bn rows
- Db2 13 M500 is use across NWG
- First z17 in Europe installed in 2025
- 921k SQL statements per sec
- Over 20m customers
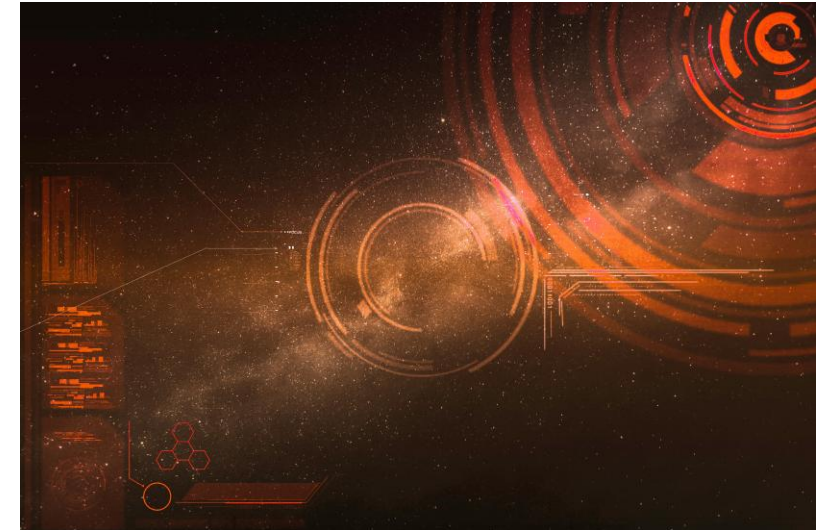- Operating across retail, commercial & private banking markets in UK.

Db2 z/os in NatWest – Who is Colin Knight ?

- Db2 Sys Prog Tech Lead: Colin Knight
- Worked at NatWest since 1994 (RBS)
- Worked on Db2 since V2.1 (just missing V1.3)
- Forty years in Mainframe
- 36 Years in Db2 from 1989
- Started as Cobol programmer on IMS DL/I
- colin.knight@natwest.com
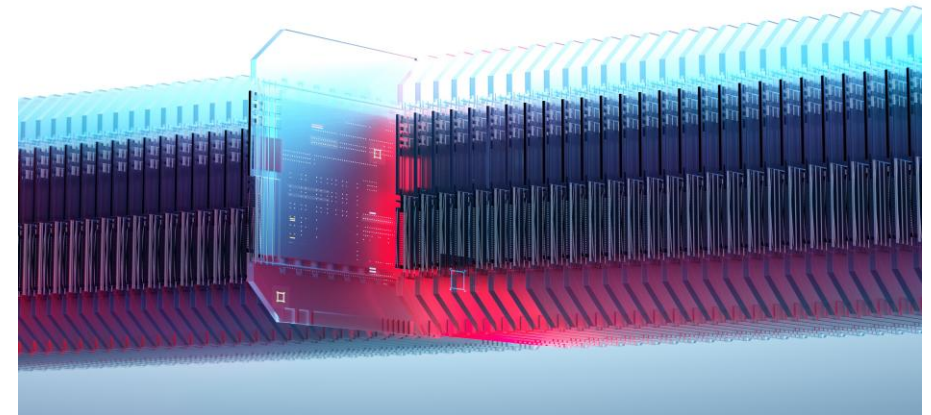- Current count – 61 orbits around the sun

Making MainFrame data available for analytics on big data in NatWest

- Extracts, unloads and HPU
- QREP
- Infosphere CDC
- Direct links eg Splunk
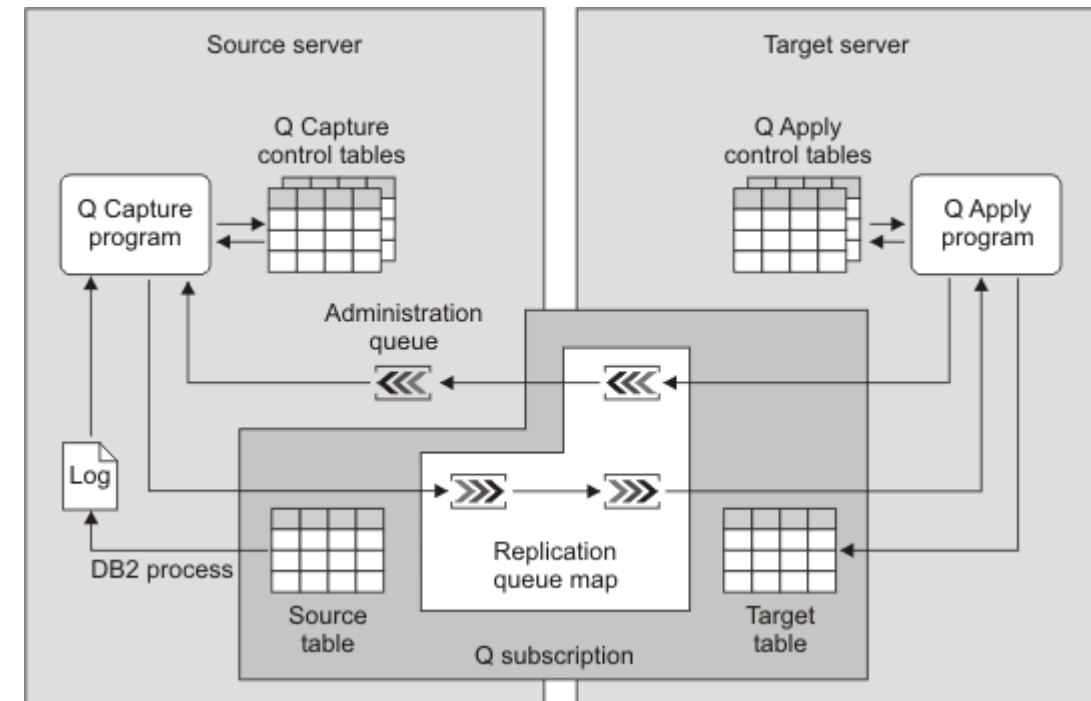- Teradata
- IDAA
- Oracle
- REST API's

Extracts, Unloads and HPU

- Multiple copies
- Transforms
- Moment in time – daily, weekly ?
- Cheap
- Once set up, many years pass, still used ??
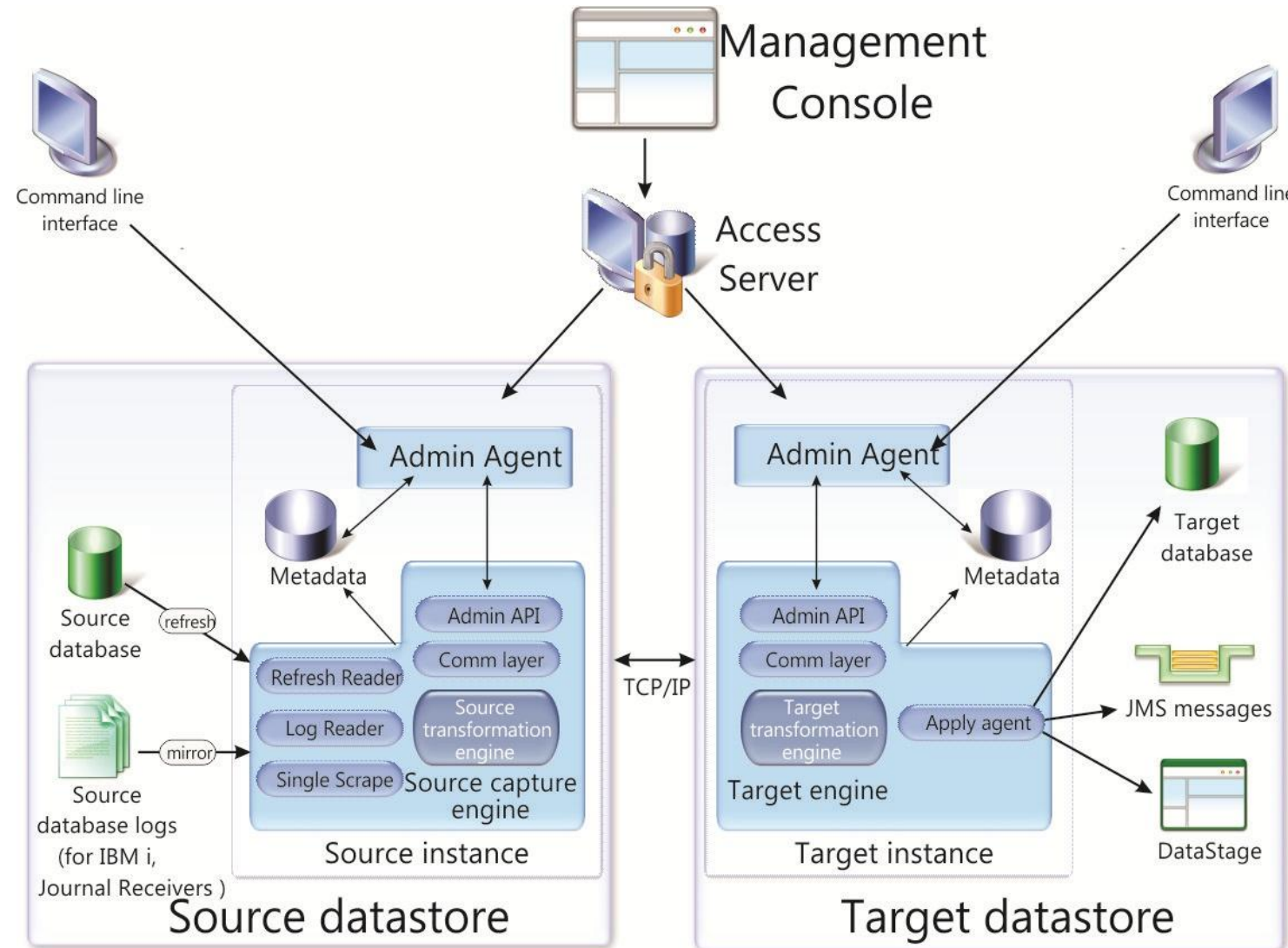- HPU and clever extracts
- Good for MIS reporting

Q Replication

- Db2 z/os to Db2 z/os
- Large volumes, low latency
- CPU costs
- Source changes especially in test systems
- Keeping source and target in sync
- Transforms, character sets, confusion
- Spill queues and consistent copy

Infosphere CDC

- High Volume, low latency
- Assynchronous
- CPU costs ?
- Refresh not refreshing
- Kafka targets
- Transforms efficiently
- Management
- Target vs Source – different versions of truth ?
- Problems – just a few

Infosphere CDC – There's more

- Clever – but how many moving parts ?
- Refresh is annoying
- Post filters vs Source Ignore
- Log scraping not as efficient as QREP
- Db2 data + VSAM + Oracle etc
- Target Kafka and then MongoDB etc
- Understanding the data at source
- Tail wagging the dog, changes at source
- Does the job and mostly low latency
- Meeting new people

Beads of sweat at 03:00 AM

IBM Db2 Analytics Accelerator

- IDAA SQL subset of Db2 z/os
- Fantastic InSync faster, cheaper
- Monitoring in V8.1
- IDAA Loader (Now 3.1) – SMF, non Db2
- Data studio – sad demise
- Archiving data
- SMF monitoring
- CDC SMF data a godsend in IDAA
- Orders of magnitude faster
- IBM Db2 Warehouse under cover

IBM Db2 Analytics Accelerator continues

- Copy data between IDAA's in 8.1
- Loading non Db2 data
- Complex queries on large datasets
- More expense to get more out of Db2 ?

Making MainFrame data available for analytics on big data in NatWest

Problems, issues, queries

- Security in transit and at target – cloud
- Synchronised data
- Latency, realtime
- Recovery at source & source changes
- CPU overheads, maintenance overheads
- Why not leave the data on source ?
- Moving data gives more units of fun

The future

- Data virtualisation
    - IBM Cloud Pak for data
    - IBM Db2 Data Virtualisation
- Structured and non-structured data view
- Simplification
- More REST API's
- AI & SQLDI – modelling data
- More IDAA less CDC ?
- Keeping the data at source and data gravity
- Quantum computing to revolutionise analytics on Big data ?

The future

- Grafana, Prometheus to view data from Broadcom Sysview for Db2 z/os APIs
  - Python for use with Sysview APIs
  - Dashboards for reporting, monitoring
  - Quite a way to go with help from Broadcom
- SQLDI easy to install....why ?
- If not SQLDI then what or how do we run AI on Mainframe Db2 data ?
- More AI with z17 mainframes from IBM
- What about golden source and data dictionaries ?
- Data visualisations

The future – more questions

- Do you need your own copy of data ?

- How close to realtime ?

- Is any replication cheap ?

- Is virtualisation and a holistic view of big data at source the final answer ?

- Do we know what data is actually available ?

- Do we understand or need to understand how data is used and changed at source ?

- How do I travel through the world of Big data ?

- Data lakes ?

- Hybrid cloud solutions ?

The future – a puzzle

- From a podcast on Big Data by Victor Mayer-Schonberger in 2016
- Walmart analysis of shopping behaviours before a hurricane
- What do people buy ?
- Why ?
- Sell sell sell
- Hung up on why ?

Summary – what happened so far and what did I miss when I nodded Off for 10 minutes ?

- Is there anything more peaceful than...
- Explanation of Big data and how mainframes fit

  into that ? Key system of record.
- Overview of Db2 z/os in NatWest
- Who the heck is Colin Knight anyway ?
- Making Mainframe data available for Big Data
- CDC & QREP + IDAA but many others
- Future, virtual, visual, hybrid cloud, AI
- Problems, issues, pain points along the way
- Can I stop copying data off mainframe ?
- Time for thought and a puzzle
- TIME UP>>>>

Questions, questions, questions please.

Colin Knight
Db2 Systems Programmer Technical Lead
NatWest Bank

Colin.knight@natwest.com

And this is my pal Thor >>>>>>>>>>>>>

This is the end – or is it just the beginning ??

Thanks for listening and have fun