

## Does life happen in batches?

For over 25 years now, we’ve been conditioned to believe that wholesale replication and staging of ‘point solution’ data from operational systems is the right and proper way to build data warehouses and repositories to enable analytic and reporting capabilities. The technology and skills required to do this are easy to acquire, everyone else is happily doing it and frankly: ‘Nobody ever got fired for buying IBM’.

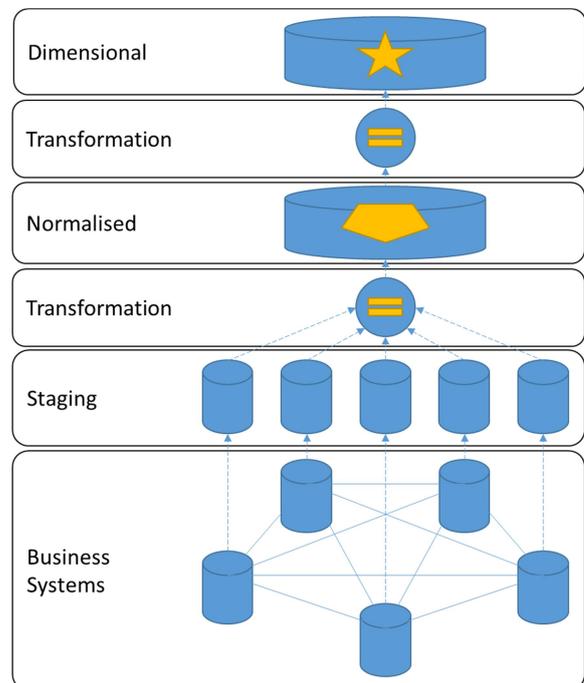
Ntegra Greenside looks beyond the fear, uncertainty and doubt and lazy conformity, to think about how and where new and emerging technology might be used to improve current architectures. In this short piece, we explore the use of Internet of Things (IoT) Event Data and Stream Processing to disrupt and improve conventional Business Intelligence (BI) and Management Information (MI) architectures.

This diagram depicts a design pattern and architecture approach popularised by Ralph Kimball in his 1996 publication ‘The Data Warehouse Toolkit’.

Business operational systems are ‘treasure troves’ of data that, when copied and staged, can be used to create a normalised data model reflecting the current (and historic) state of the business.

This foundational model is then further transformed, into a dimensional form enabling reports on subject areas such as Customer, Order, Product, etc. to be easily produced.

The pattern is simple enough, easy to implement, and provides a valuable resource for analysis and reporting purposes. By its nature, however, it’s **very** big (often more than doubling an organisation’s storage footprint), and it is difficult to support and extend.



Simply engaging with the operational system owners and implementing agreeable approaches to staging data can be challenging:

- *“We don’t have resource or capacity to build another interface”*
- *“You can’t have data from us because of its classification”*
- *“We’re only able to give you a feed between 2 and 4 in the morning”*
- *“We’ve just been rebadged as ‘business critical’ and have to implement a business continuity, disaster recovery capability, so can we use your staging area as a geo-redundant solution?”*
- *“We’re far too important keeping the show on the road to talk to you ‘BI’ guys – go away!”*
- *“Yes, no problem, you’ll have to take it as a disk array snapshot and reverse engineer the data and, by the way, we have 17 years of history that’ll need quite a bit of cleansing!”*

Having gathered, cleansed and modelled all the necessary data, change control becomes a critical issue. Any source system changing its internal data structures will have a knock-on impact to each of the layers in the architecture. In-life introductions of new systems and data definitions need to be tracked carefully and catered for because the smallest ‘coalface’ change have huge impacts on the data warehouse and report development / support teams.

And, how many successful data warehouses have become operationalised:

- *“No, you can’t run that ad-hoc query until we’ve closed-out the books at the end of the quarter!”*

## Life happens in streams

The ability to handle and process continuous streams of data has changed the way that organisations consider certain types of information flow. Being able to take advantage of streaming data is beginning to be an essential part of building a data-driven organisation. The main drivers are primarily coming from the Internet of Things (IoT) and the increasing use of, and dependence on, sensor data. Continuous streams of external data (such as social media) are also accelerating interest in the use of stream processing to enable organisations to better understand, for example, customer

experience and sentiment. However, new approaches to streaming designs can greatly improve the efficiency of an overall organisation and existing domains.

Many of the things we want to monitor and understand are continuous in nature and don't happen in batches: ocean currents; rainfall; machine metrics; Global Positioning System (GPS) positions; heartrate; flight dynamics; trajectories; oxygen uptake, oil viscosity... The list is essentially endless. It makes sense to collect and analyse information from these events as streams of data.

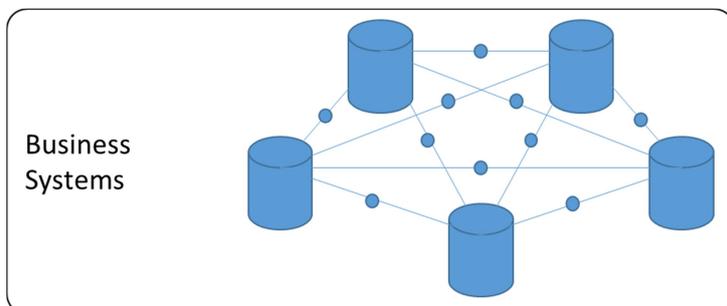


The benefits of handling streaming data are well understood. You can take a source of data, e.g. an anemometer, sampled at a high frequency, and stream both wind speed and direction data to real-time analytic functions to create live dashboards, showing the wind speed and direction in very near real-time. The analytic functions could also be used to, for example, combine the streams and watch for certain wind speed and direction combinations that might trigger, say, the closure of a bridge to high-sided vehicles.

**This is all well and good, but how can it be applied to the data warehouse architecture pattern above?**

Consider the business operational systems (the so called rich treasure troves of, what turns out to be the equivalent of data 'fat deposits').

The real value is not the systems themselves or the data they retain, but the integrations between them:



Integrations are:

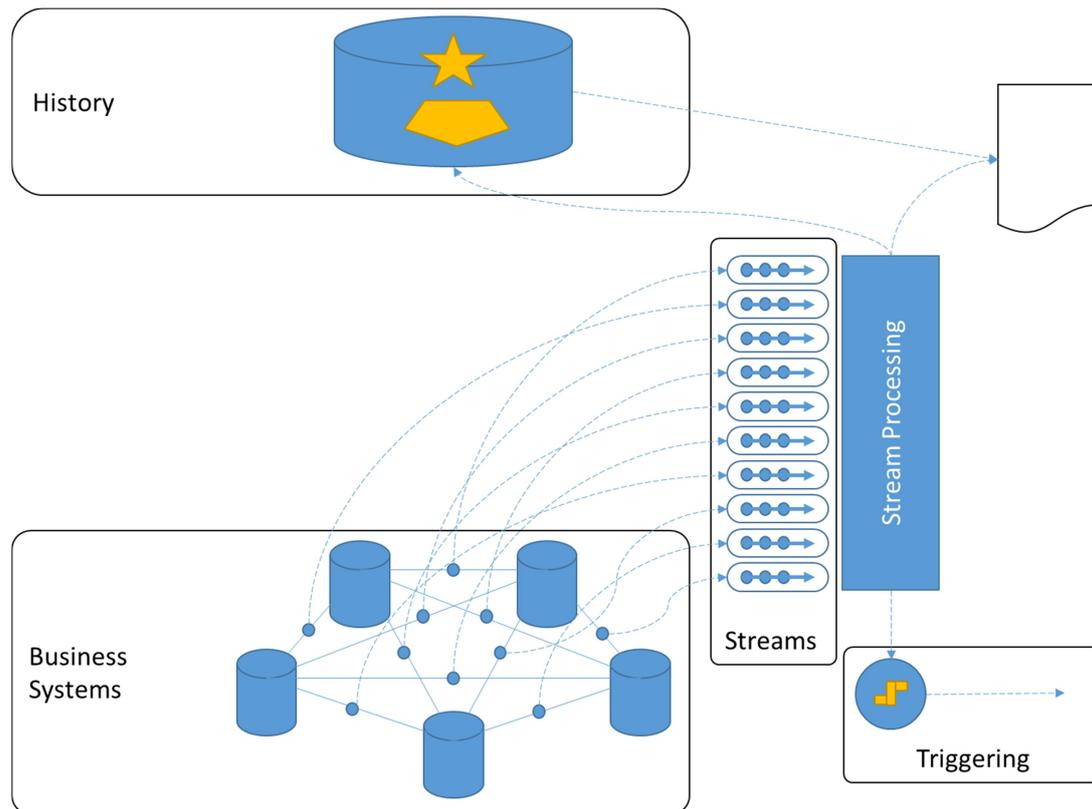
- Beautifully crafted
- Change managed
- Clean
- Robust
- Efficient

The existing integrations offer a perfect point to instrument the organisation's operations, using IoT and streaming methods. When looked at in this way, it becomes

obvious that integration events (the data exchanged between operational systems) *are*, in fact, *THE BUSINESS*...

So, what might a new architecture stack look like?

## Streaming BI Architecture



In the above schematic, there are several obvious opportunities to improve the nascent architecture and flows:

- The streams could be geographically distributed and events/transactions locked into a block-chain to establish integrity and availability
- If a 'History' repository/warehouse is required, it could be generated from a block-chain client view of the ledger created by the streams
- Real-time triggering to initiate business processes is extremely powerful

There are a host of other opportunities to think about, explore and make the subject of further research.

Please contact us if this interests you? [info@ntegra.co.uk](mailto:info@ntegra.co.uk)