



Welcome to the Virtual IMS user group newsletter. The Virtual IMS user group at www.fundi.com/virtualims is an independently-operated vendor-neutral site run by and for the IMS user community.

Virtual IMS user group presentation

The latest webinar from the Virtual IMS user group was entitled, "IMS to Big Data - Common Traits of Successful Implementations", and was presented by SQData's Scott Quillicy.

Scott is the founder of SQData, a US-based software company that specializes in high-performance data movement and changed data capture (CDC) for IMS, VSAM, and relational databases, such as DB2 and Oracle. With over 30 years of experience, Scott specializes in the real-time replication of IMS databases to a variety of downstream databases and big data platforms. Scott and his team at SQData have been involved with many IMS related projects over the

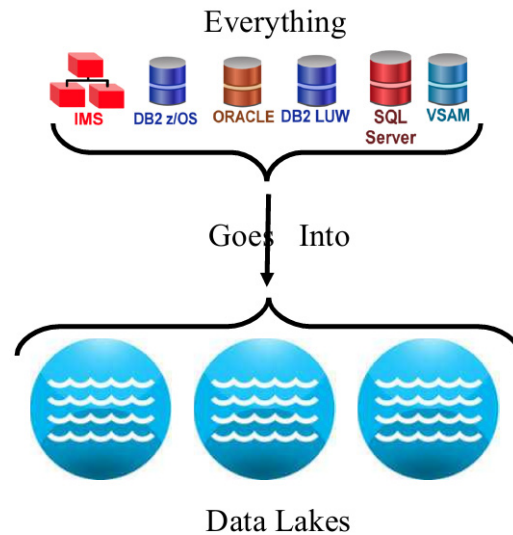


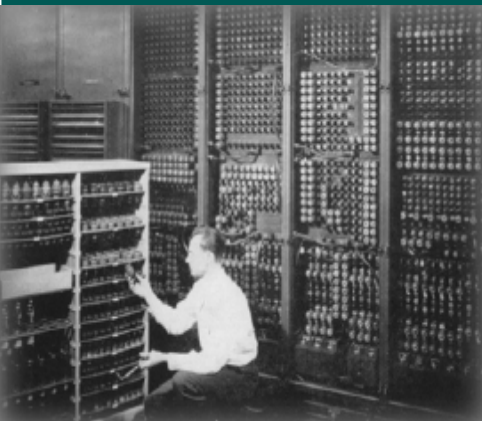
Figure 1: The wrong way!

years, compiling a set of best practices for success along the way.

Following his introduction, Scott asked, "Why stream IMS to Big Data?" His suggested answers were: real time analytics; decisions based on current information

Contents:

Virtual IMS user group presentation	1
Meeting dates	4
Arcati Mainframe Yearbook	4
Sponsorship opportunity	4
About the Virtual IMS user group	4



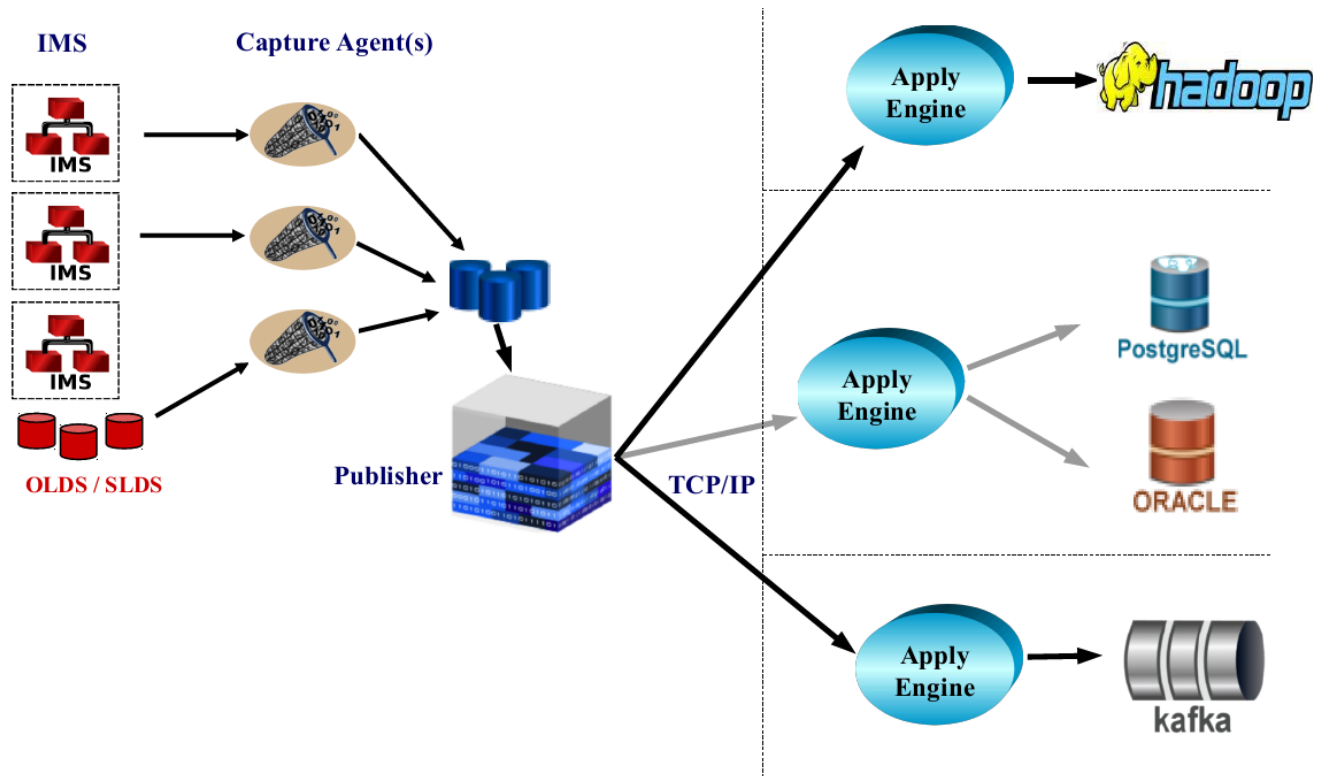


Figure 2: Streaming IMS to Big Data

as opposed to 24+ hour old data; to quickly detect key events / trends; to maintain a competitive advantage; to provide better customer service; and to increase revenue / profitability.

Figure 2 illustrates streaming IMS to Big Data. The best way to do this, according to Scott, is to: have sub-second latency from capture to apply; be able to handle high-transaction volumes; be multi-purpose (a major plus); not require any extra parts (ie no staging tables and no queues) when publishing; and be resilient / fault tolerant.

Scott then looked at the

common pitfalls of converting IMS data to Big Data. Firstly, he suggested was the lack of a holistic strategy. People adopted a “We can do it ourselves” approach. Often organizations would have multiple departments going into Big Data with small projects. They’d launch projects with the minimal structure in terms of methods, tools, and support. And it ended up being significantly more expensive in terms of time and money.

A second pitfall was not focusing on business needs. People took a “Build it and they will come” approach, only they didn’t. There were no clear use cases, and

this was often caused by pressure to deploy a Big Data solution.

A third pitfall was often data collection overkill. Sites seemed to think that “Everything needs to be in data lakes” (see Figure 1); there was minimal understanding of how to relate the data to business problems; and people spent a LOT of time moving data that was of little value to the business

A fourth pitfall was the result of not setting proper expectations. There was a ‘We can have something for you in no time’ approach. There was no guaranteed

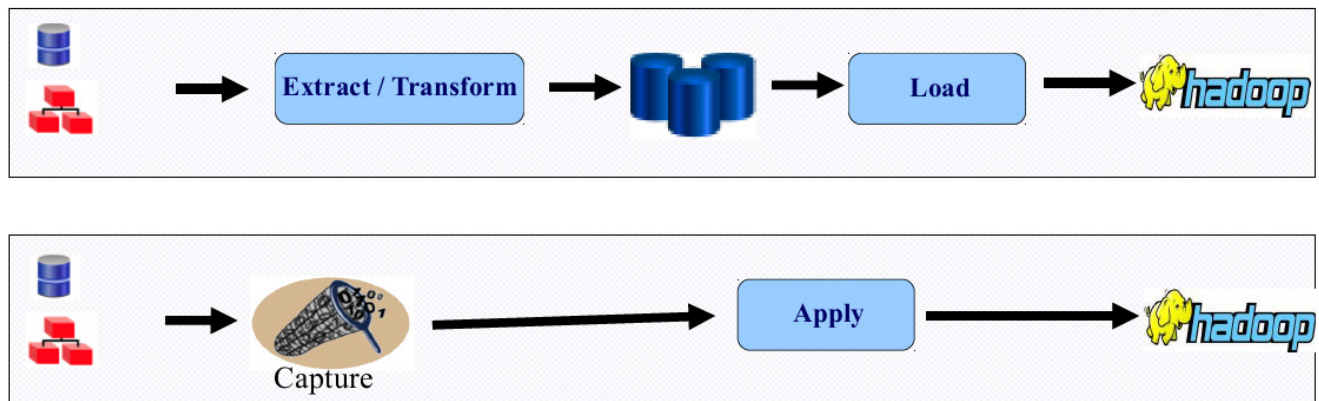


Figure 3: ETL and CDC

project timeline and there were always cost overruns.

The fifth pitfall was that people didn't understand mainframe data, particularly IMS. They would say things like, "Just take the data and copy it into Hadoop". They didn't understand the nuances of non-relational data, and there are many of them.

People with no IMS background will often not understand:

- IMS structures in general
- Repeating groups (Occurs)
- Redefines
- Dates
- Invalid data
- 'Special' fields (bits, Y2K, etc.)

They also have problems with:

- Code page translation
- Transaction consistency
- Streaming versus ETL
- Target apply concepts / streaming
- Normalization versus denormalization.

Another culture shock comes with the difference between ACID and BASE. On a mainframe we're used to consistency in the data. ACID guarantee that database transactions are processed reliably. ACID stands for:

- Atomicity – all or nothing, either the transaction commits or it doesn't
- Consistency – a transaction brings the database from one valid state to another
- Isolation – concurrency
- Durability – once a transaction commits, it

remains committed.

BASE works on the principle of eventual consistency

- Basically Available – data is there, but there are no guarantees of consistency
- Soft state – data may change over time. It may not reflect the commit scope
- Eventual consistency – the data will eventually become consistent.

As Figure 3 shows, there is a difference between ETL (Extract, Transform, Load) and CDC (Changed Data Capture). Typically, ETL needs to be done only the once to move the data from IMS to Big Data. Afterwards, CDC will keep the data stored in a Big Data repository up to date.

Scott Quillicy concluded with his five factors for success. Firstly, approach

with a holistic strategy. Use common infrastructure / tools / support. Use established methods (eg DevOps / Agile). And beware the “fiefdoms”.

His second factor for success was to involve the business from the start. He said that they understand the source data, they know the order of importance, and they can assist in design validation, QA, etc.

Thirdly, avoid the data collection overkill, which is a time and money killer. Organizations need to focus on the most important data first. They can then iterate through the remaining data, prioritizing by importance.

The fourth factor is to set proper expectations. In reality, 2 to 3 years is the minimum time expected for the completion of an entire project. Scott recommended delivering in increments, starting with the most important data.

Scott’s fifth factor for success was to understand that the IMS data is ‘special’. He suggested that patience is key, and organizations should not hesitate to ask for help.

A copy of Scott’s presentation is available for download from the Virtual IMS user group Web site at www.fundi.com/virtualims/presentations/

[IMStoBigDataOct16.pdf](#).

You can see and hear the whole user group meeting by downloading the WMV file from www.fundi.com/virtualims/presentations/2016-10-04meeting.wmv.

Meeting dates

- On 6 December 2016, IBM’s Suzette Wendler will be discussing “IMS Infrastructure and the CSL”.
- The following meeting will be on 7 February 2017.

Arcati Mainframe Yearbook

How do you know what’s really going on at other mainframe sites? One answer is to complete the Arcati mainframe user survey at www.arcati.com/usersurvey17. Tell us what it’s really like where you work. Your identity and company information will be treated in confidence and will not be divulged to third parties.

Sponsorship opportunity

Are you missing a great opportunity to advertise your IMS software?

The Virtual IMS user group

is now offering software vendors the opportunity to advertise their products in a number of ways. You could have an advert on the home page of the Web site (at www.fundi.com/virtualims), you could advertise in the newsletter, and/or you could advertise in the monthly e-mails sent to members of the user group.

E-mail trevor@itech-ed.com for full information about marketing opportunities with the Virtual IMS user group.

About the Virtual IMS user group

The Virtual IMS user group was established as a way for individuals using IBM’s IMS hierarchical database and transaction processing systems to exchange information, learn new techniques, and advance their skills with the product

The Web site at www.fundi.com/virtualims provides a central point for coordinating periodic meetings (which contain technically-oriented topics presented in a webinar format), and provides articles, discussions, links, and other resources of interest to IBM IMS practitioners. Anyone with an interest in IMS is welcome to join the Virtual IMS user group and share in the knowledge exchange.

To share ideas, and for further information, contact trevor@itech-ed.com.

The Virtual IMS user group is free to its members.