WHITE PAPER

INTELLIGENT BUSINESS STRATEGIES

Big Data - Why Transaction Data is Mission Critical To Success

By Mike Ferguson
Intelligent Business Strategies
July 2014

Prepared for: IBM

Table of Contents

THE TWO SIDES TO BIG DATA

The business interest in big data is driven by the demand for greater insight to remain competitive

In the last two years, the frenzy around big data has reached fever pitch with investment and technology advances moving at a staggering pace. Much of the focus in this fast moving world is on the use of analytics to analyse structured, semi-structured and unstructured data in big data platforms like Hadoop in order to produce new high value insights to add to what is already known about customers and business operations. For many companies, one of the top priorities in the use of big data is to deepen customer insight to improve customer experience. In addition, there is also a drive to improve operational *effectiveness* through better operational decision making. Also there is a rise in demand for real-time analytics. Some examples here include real-time analysis of markets data for more timely and accurate trading decisions, tackling financial crimes such as credit card fraud or anti-money laundering, monitoring of asset and utility grid usage for field service optimisation and preventative maintenance, and improving click through in on-line advertising.

However if the promise of big data is to be realised, operational transactional and non-transactional systems need to scale to capture the data needed for analysis

While there is no doubt that all of this is extremely important in maintaining competitive edge, none of it would be possible if we don't pay attention to making sure that the systems needed to capture, stream and store all of the required transactional and non-transactional data, are in place and capable of scaling to support peak transaction volumes, peak data arrival rates and peak ingestion rates. In addition, in order to maximise the value of new insights produced in big data analytical environments, these analytical systems need to be integrated back into core operational transaction processing systems so that prescriptive insights are available to all that need them to continuously optimise operations and maximise operational effectiveness.

Transaction processing systems provide the data needed by big data analytics systems

Therefore, although much of the focus in the world of big data is on analytics, it is clear that the role of transaction processing systems is mission critical to big data success. Transaction systems have always been mission-critical and so speed, availability, and scalability are central to their operation. Transaction processing systems, and the non-transactional data systems, such as click stream web logs that accompany them, provide data needed by traditional and big data analytical systems. They also make insights produced by analytical systems, available to the right people at the right time in the context of every day operational activities. What this implies is that, even with the new focus on big data analytics, the basic business requirement is still the same. Companies more than ever need to close the loop between analytical systems and their core operational transaction processing systems to maximise success.

Yet, despite this, the focus is often on defining requirements for big data analytics with not enough attention being given to defining requirements associated with capturing transactional and non-transactional data needed or in delivering right-time, in-context insights to front-line decision makers via integration with transaction processing systems.

In this context, you can break a big data strategy into two major parts:

- Big data analytical processing
- Big data operational transaction and non-transactional event processing

The focus of this paper is on the latter.

Big Data Analytics – The Dependency On Transaction Data

Structured transaction and master data is needed in big data analytical environments to provide context

In order to analyse big data unstructured or semi-structured data on platforms like Hadoop, many organisations also load structured transaction and master data into this environment to integrate with multi-structured data in order to provide a context for analysis.

A good example here would be analysis of clickstream data from web logs which often requires the need for both customer and product master data as well as customer transaction data (e.g., on-line purchases) to put clickstream into context. For example, all this would be needed to answer the question "Which navigational paths were taken through our website by our most profitable customers that resulted in a purchase of a particular product?" The same would be needed to answer the question "What products on different pages on our website do people consistently view that would indicate they should appear together on the same web page?"

Non-transaction data such as click stream is in high demand to understand customer on-line behaviour prior to initiating a transaction or abandoning one

Click stream data is now a strategic data source for big data analysis in order to understand customer on-line behaviour. The reason this is important is because the impact of the web on business has been profound. In some cases, on-line competition has even resulted in entire industries being "re-wired" e.g. travel, car insurance. Given that this is the case, we are now at the point where transaction data is not enough to provide comprehensive customer insight. We also need to understand on-line customer behaviour and not just what they bought. On the web the customer is king. They can surf around comparing products, services and prices any time anywhere all from a mobile device. In addition, with new 'lighter weight' web-based companies springing up everywhere, the customer has much more choice and can easily switch loyalty with a few clicks of a mouse.

Both inbound and outbound customer interactions are needed to understand and improve customer experince

In order to analyse customer on-line behaviour, companies often want to combine customer data, product data, non-transactional click stream data and transaction data. For example, many organisations often want to analyse:

- The paths taken on a website that led to purchasing of specific products
- The paths taken on a web site that led to abandoned shopping carts
- The products placed into and taken out of shopping carts en route to purchasing
- Customer profitability on-line versus other channels

Another example is the need to understand customer sentiment on in-bound email and on social networks in relation to products purchased by customers and also in relation to customer service. In this case both customer service transaction data and sales transaction data are needed to be able to time-sequence transactions and sentiment interactions. In fact for true customer experience management across all channels, all interactions and all transactions associated with a customer are needed. In a survey conducted in late 2013[1], customer experience management came out as the top priority use case for big data analytics.

Customer interactions and customer transactions are the most common data sources in big data analytics projects

Also in a 2014 Big Data Survey[2] of over 750 respondents, emails (52 percent) and customer transaction databases (49 percent) were the current most common big data sources.

It is clear from both of these examples and surveys that big data analytics is very much dependent on transaction data. In addition, given that many organizations using Hadoop are now offloading data warehouse ETL processing to Hadoop, it is also the case that transaction data may end up on Hadoop systems first for data cleansing and data integration en-route to traditional data warehouses.

---

[1] Survey Analysis: Big Data Adoption in 2013, Gartner September 2013

[2] Big Data Survey, 2014 IDG Enterprise

THE IMPACT OF BIG DATA ON TRANSACTION SYSTEMS

On-line transaction processing systems have grown to record every aspect of business activity

Since the late 1950s and the birth of mainframe technology, software applications have been built to process and record business transactions in database management systems (DBMSs). Initially this was done with application programs processing groups of transactions in batch with input and output files being used in process execution. Then on-line transaction processing (OLTP) systems emerged and user interfaces appeared on the scene to capture transaction data. Today these user interfaces have progressed to being web and mobile based with OLTP systems being used to record transaction activity across almost every aspect of business operations including sales force automation, order-to-cash, fulfilment (materials-to-finished goods), shipment-to-delivery, procure-to-pay, customer service, billing, finance and HR e.g. hire-to-fire/retire.

TRADITIONAL TRANSACTIONS VS BIG DATA (NON) TRANSACTIONS

On-line transaction processing systems run on relational and non-relational systems

Historically transaction data was recorded in non-relational database systems such as IBM IMS, CA DatacomDB, Burroughs DMSII and Cincom Total until the arrival of relational DBMSs such as IBM DB2 in the early 1980's. Today although non-relational products like IBM IMS are still very much alive and well, relational DBMSs underpin the vast majority of the world's transaction systems supporting on-line transaction processing.

From a business perspective transactions are everything from orders, shipments, payments, returns, refunds, purchases, inventory level changes, financial adjustments, etc. These and many, many more transactions are being recorded in OLTP systems. Technically speaking, a traditional online transaction is defined as:

"An input message to a computer system dealt with as a single unit of work performed by a database management system".[3]

On-line transaction processing systems support full transactional integrity properties known as ACID properties

Traditional on-line transaction processing systems usually support ACID properties[4], which stipulate that a transaction must be:

- Atomic – the transaction executes completely or not at all
- Consistent – the transaction must obey legal protocols and preserve the internal consistency of the database
- Isolation – the transaction executes as if it was running alone with no other transactions
- Durable – the transaction's results will not be lost in a failure

---

[3] Source: Wikipedia

[4] As defined by Dr Jim Gray

Traditional ACID transactions are typically associated with structured data. All relational DBMSs support ACID properties to guarantee transaction integrity.

Since the emergence of the web in the 1990's there has been explosive growth in the rate (or velocity) that data is being generated. Data velocity has increased both in terms of traditional on-line transaction rates and also in terms of so-called 'non-transaction' data.

Non-transactional operational systems have also emerged to capture clickstream and machine generated data at scale

Operational non-transactional data includes things like sensor readings, user clicks, user profiles, machine generated log data, gaming actions, blogs, news feeds, shopping cart contents, reviews, etc. This data is being created at scale and is often stored in self-describing format using BSON, JSON or XML.

Non-transactional data includes, shopping cart data used in scalable, high availability ecommerce systems

A good example is shopping cart data, which is regularly held in JSON format by e-commerce applications. These applications have to deal with large numbers of concurrent users and be very highly available to cater for the 24x365 nature of business on the web. They also have to deal with the fact that the contents of shopping carts change a lot prior to a purchase transaction occurring. Traditional transaction processing is not needed until a purchase is made when updates to accounts and inventory have to be guaranteed. Prior to that, it is about performance, scalability and availability. If you consider that there could be thousands, tens of thousands or even millions of users on-line, there needs to be a way to rapidly capture changes to shopping carts in an application and to serve up up-to-date shopping cart data to all users many, many times *before* a transaction occurs. This data does not constitute a transaction until the user decides to go to checkout and buy what is in the cart. These kinds of requirements often result in NoSQL databases being used to capture data quickly. Some of the reasons of often cited by companies, for the adoption of NoSQL document databases, are that they support schemaless data, automatic sharding (partitioning), high availability and the ability to easily add nodes to a cluster to support scalability.

Non-transactional operatonal systems often only need eventual consistency and not full transactional integrity

Typically, distributed NoSQL databases use automatic sharding and replication techniques in order to scale, guarantee consistent performance and support high availability. However because of this, they often do not support ACID properties. Instead they support what is known as BASE (Basically Available, Soft state, Eventual consistency). This is when the system guarantees the availability of the data but the state of the systems is always 'soft' in that even when there is no data being input, changes may be still occurring due to eventual consistency. Eventual consistency means that the system will eventually become consistent when input stops and all replicas of updated shards have been updated to keep them synchronised.

Big Data Transactions - The Impact of Devices, Sensors and The Internet of Things

The emergence of the web and mobile devices has caused the number of on-line transactions to soar

Since the birth of the Internet there has been a relentless rise in online shoppers and on-line transaction processing as more and more websites offer e-commerce capabilities. However, it is the arrival of mobile devices, together with the convenience of mobile commerce that is taking the growth in transaction volumes to new levels. It is therefore no surprise that transaction data is rapidly on the increase but even that is small compared to what is happening to click stream data in web logs created from online browsing that happens prior to transactions occurring. Transaction levels and accompanying click stream are reaching unprecedented volumes. In addition, edge devices in a network such as network switches, routers, multiplexers or devices at the edge of a utility grid such as meters, re-closers, capacitor banks, solar inverters, street lights, and even buildings or homes, now support the embedding of traditional DBMSs to catch data on the edge for further analysis.

Similarly, the amount of non-transactional data being recorded has also skyrocketed

In a similar way that the web and mobile/edge devices have impacted the growth of traditional transaction rates, so the arrival of sensor networks and sensors embedded in 'smart' products has had a similar impact on the non-transactional big data generation rates.

The emergence of sensor data and the Internet of Things will drive non-transactional data volumes even hgher and fuel demand for even more analysis

Sensor data can be created at very high rates. 50000 sensors emitting sensor readings at 5 times per second creates 250000 pieces of data per second. That is 15,000,000 pieces of data a minute. Now consider the emerging world of smart products such as smart phones, smart meters for use in energy, smart cars, smart buildings, and smart production likes. Smart products have sensors built-in that can send back data for collection and analysis. Just imagine the GPS sensor data from all the mobile smart phones in the world or from all the vehicles on the move or all the smart meters in the world. The need to analyse this type of data means that the systems that capture it have to scale to handle both the volumes of data being generated and the velocity at which it is being created.

As more and more smart products are manufactured, the Internet of Things (IoT) comes alive resulting in a tsunami of data coming over the horizon. The "Internet of Things" refers to the growing number of devices and sensors that communicate and interact via the Internet, offering businesses new data and insights that can lead to new customers and revenue opportunities. Harnessing data from billions of connected devices lies in the ability to capture, store, access and query multiple data types seamlessly and use that data in meaningful ways. The impact of IoT is still to be fully understood but clearly the need to increase the rate of capture and ingestion is already obvious and so yet again, we realise that the capture and use of this data in every day business operations is paramount to the success of big data analytics.

Richer Transaction Data For Competitive Advantage

Modern transaction systems are adding new types of real-time data to their user interfaces

So far we have seen that transaction volumes are increasing rapidly and that the digital exhaust leading up to transaction processing taking place is also being recorded.

Analytics and recommendations are also being added to help optimise business operations for competitive advantage

But it is much more than that. The user interfaces of modern transaction processing systems today are increasingly demanding access to other real-time data. For example financial trading platforms are providing traders with real-time analytics, real-time news and media feeds (including video), social media sentiment and access to communities to provide traders with enriched data they need to help them make better real time decisions. On-line media firms are analyzing and optimizing advertising pages, images, video, and links on their web sites to keep you on a site and to get you to click more ads. On-line gaming firms track real-time usage on thousands of on-line users in-real time to determine when to make 'in-play' offers to drive transactions. On-line retailers want more and more data about customers to make personalised offers in real-time to continue to grow their business. In addition we now have in-app purchases in mobile applications. The days of just the on-line form and transaction data being used in the users interfaces of OLTP systems are fading rapidly and being replaced by OLTP systems with "data enriched' user interfaces providing everything needed to optimise transaction processing. This includes integration with traditional, streaming and big data analytical systems together with prescriptive analytics to guide internal operations personnel and to tempt customers and prospects.

THE NEED TO MODERNISE AND SCALE TRANSACTION SYSTEMS FOR BIG DATA CAPTURE AND EXTREME TRANSACTION PROCESSING

Given what we have discussed so far, there is no question that for many organisations, an important factor in contributing to business success with big data is in modernising the data platform underpinning transaction processing systems. This needs to happen to:

Modern OLTP systems need to scale to deal with higher transaction volumes

- Scale to capture and serve up non-transactional data prior to and during transaction processing
- Scale to handle ever-increasing transaction rates from desktop and mobile devices with full transactional integrity

Extensions are also needed to capture non-transactional data at scale

- Scale to capture, ingest and record non-transactional structured, semi-structured and unstructured data for big data analysis e.g. click stream data and sensor data, to produce the new insights needed for competitive advantage
- Capture, ingest, record and serve up structured, semi-structured and unstructured non-transactional data at scale to enrich OLTP system user interfaces

Transactional applications need to be enriched by incorporating new information services and insights produced by analytical systems

- Leverage enriched data in enhanced OLTP user interfaces to provide everything a user needs for informed transaction processing. This includes
    - Integrating transaction processing systems with non-transactional information services
    - Integrating transaction processing systems with traditional and big data analytical systems to leverage real-time analytics and actionable prescriptive analytics to guide operational decision making (including recommendations to tempt customers)

Real-time analytics may also be needed

It may also be the case that real-time analytics on streaming data in-motion may become an increasingly important requirement.

EXPLOITING RICHER TRANSACTION DATA

With the ability to capture non-transactional and transactional data together in real-time and enrich transaction data, there are new opportunities to gain considerable business value whether that be in the form of improved customer satisfaction and loyalty, increased revenue, reduced cost, optimised operations, reduced risk or better compliance. This is especially the case if you also add real-time analytics into the mix.

COMBINING RELATIONAL AND NOSQL TRANSACTION DATA FOR REAL-TIME ANALYTICS AND NEW BUSINESS VALUE

New data is needed to drive up business value

Consider business operations. Many companies lack insight in the area of operations mainly because:

- Data is simply not being collected to monitor operational activity
- Analytical queries on live transaction data is not possible or is often discouraged (e.g. to avoid analytical queries impacting on the performance of transaction processing)
- Insights and recommendations produced are not integrated into operational business processes

Today, business requirements dictate that all of these are needed to improve operational effectiveness. Lack of data about what is happening in operations is perhaps the most urgent requirement.

Many companies are now instrumenting business operations to understand and monitor business activity and deliver new services

To address this, organisations are starting to:

- Instrument business operations and grids using sensor networks
- Embed sensors in so called 'smart' products (e.g. GPS sensors in smart phones, smart drilling on oil wells, consumption sensors in smart meters)
- Monitor live transaction activity
- Monitor live market activity (e.g. financial and energy markets)
- Monitor events during business process execution

Being able to detect events and patterns means companies can improve responsiveness, reduce risk and optimise business operations

This extends data requirements to include the capturing, transforming, integrating and analysing:

- Transaction data (which is increasingly high volume)
- Real-time streaming data from sensors, smart devices and markets
- Events occurring during operational business process execution
- Web feeds e.g. news, weather, etc.
- Rich media streams e.g. video streams

By capturing and analysing this kind of data, companies can exploit richer transaction data for competitive advantage and to keep business operations optimised. For example, monitoring data from sensors in equipment or vehicles makes it possible to predict performance problems and perform proactive maintenance. Running operational analytical queries in OLTP systems allows staff working in operations to prioritise preventative maintenance as field service personnel become available to keep equipment running rather than having to respond to unplanned outages. It is also possible to optimize logistics and supply chain operations the same way.

TYING UN/SEMI-STRUCTURED DATA TO TRANSACTIONS AND CUSTOMER DATA FOR COMPETITIVE ADVANTAGE

Enriching customer data in customer facing OLTP systems is needed to turn OLTP systems into systems of engagement that truly understand customers

In a similar manner, customer facing OLTP systems can leverage transaction data enriched using customer on-line behaviour and customer social activity insights to better engage with customers for competitive advantage. Understanding the “interaction fingerprint” of each and every customer and prospect provides the deeper insight needed to make more personalised offers and provide better customer service. To that end, real-time analysis of on-line clickstream of a customer on your web site together with the customer’s transaction activity can result in more accurate next best offer recommendations to drive up customer value. Also being able to monitor what is put into and taken out of an on-line shopping cart in real-time would also allow OLTP systems to leverage real-time basket analysis to help serve up more personalised offers that tempt customers into buying multiple products and to reduce the number of abandoned shopping carts. Both of these mechanisms can increase transaction value.

In addition, unstructured data such as images, video and text can be used to enrich transaction systems to help customers with buying decisions, all of which can relate to better informed purchases, better customer satisfaction and repeat business.

Leveraging Operational Analytics In Transactional Environments

Operational analytics in transactional systems can create significant competitive advantage

As mentioned previously, analytical queries on live transaction data is often discouraged to avoid the impact that analytical queries may have on the performance of transaction processing. In this scenario, data is copied from the transaction system to an analytics system (e.g. data warehouse, data mart or Hadoop system), introducing a time lag or latency between a transaction and when the data is available to be analyzed. For many of the real-time analytical applications we have discussed, eliminating this time lag can create significant competitive advantage:

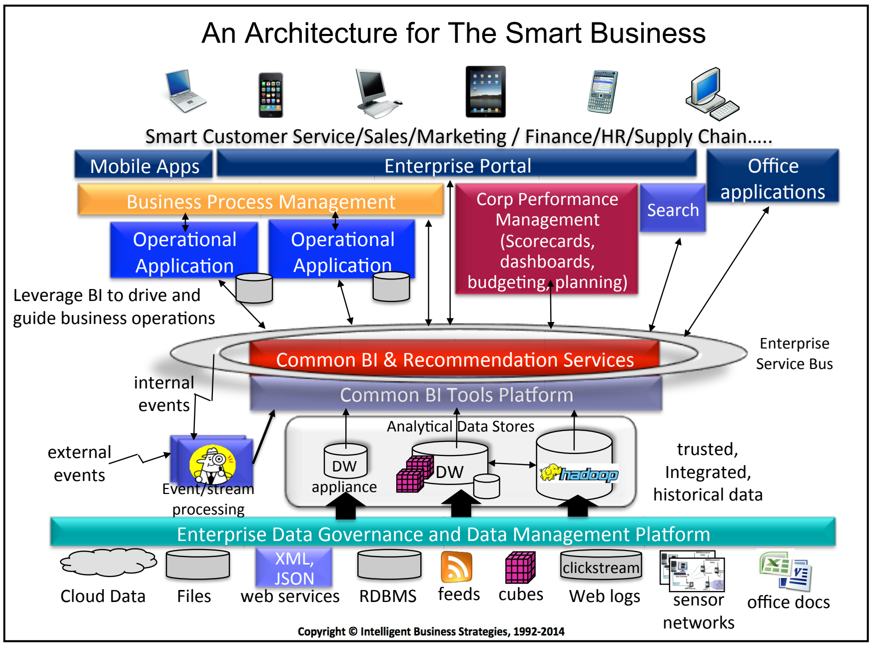
- Faster analytics on new transaction data in financial trading platforms could mean the difference between trading gains or losses
- Advertising and promotions can be served up more rapidly to help on-line media organizations improve ad response rates and increase revenue
- Gaming firms can better optimize their user experience and gain additional 'in-play' offers accepts

Smart business will require faster analytics on transactional data

- On-line and mobile retailers can personalise offers more quickly to grow their business

It is the emergence of these *smart* transaction systems that can capture data at scale and leverage insight in real-time that will drive business value in top performing companies. Smart OLTP systems will be able to leverage transactions and big data analytics on-demand, on an event-driven basis and in real-time for competitive advantage. The figure below shows how operational OLTP applications can integrate with traditional and Big Data analytical platforms to facilitate smart business.

Smart business requires closed loop integration between OLTP and analytical systems



About Intelligent Business Strategies

Intelligent Business Strategies is a research and consulting company whose goal is to help companies understand and exploit new developments in business intelligence, analytical processing, data management and enterprise business integration. Together, these technologies help an organisation become an *intelligent business*.

Author

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an analyst and consultant he specialises in business intelligence and enterprise business integration. With over 33 years of IT experience, Mike has consulted for dozens of companies on business intelligence strategy, big data, data governance, master data management and enterprise architecture. He has spoken at events all over the world and written numerous articles. He has written many articles, and blogs providing insights on the industry. Formerly he was a principal and co-founder of Codd and Date Europe Limited – the inventors of the Relational Model, a Chief Architect at Teradata on the Teradata DBMS and European Managing Director of Database Associates, an independent analyst organisation. He teaches popular master classes in Big Data Analytics, New Technologies for Business Intelligence and Data Warehousing, Enterprise Data Governance, Master Data Management, and Enterprise Business Integration.

INTELLIGENT BUSINESS STRATEGIES

Water Lane, Wilmslow
Cheshire, SK9 5BG
England
Telephone: (+44)1625 520700
Internet URL: www.intelligentbusiness.biz
E-mail: info@intelligentbusiness.biz