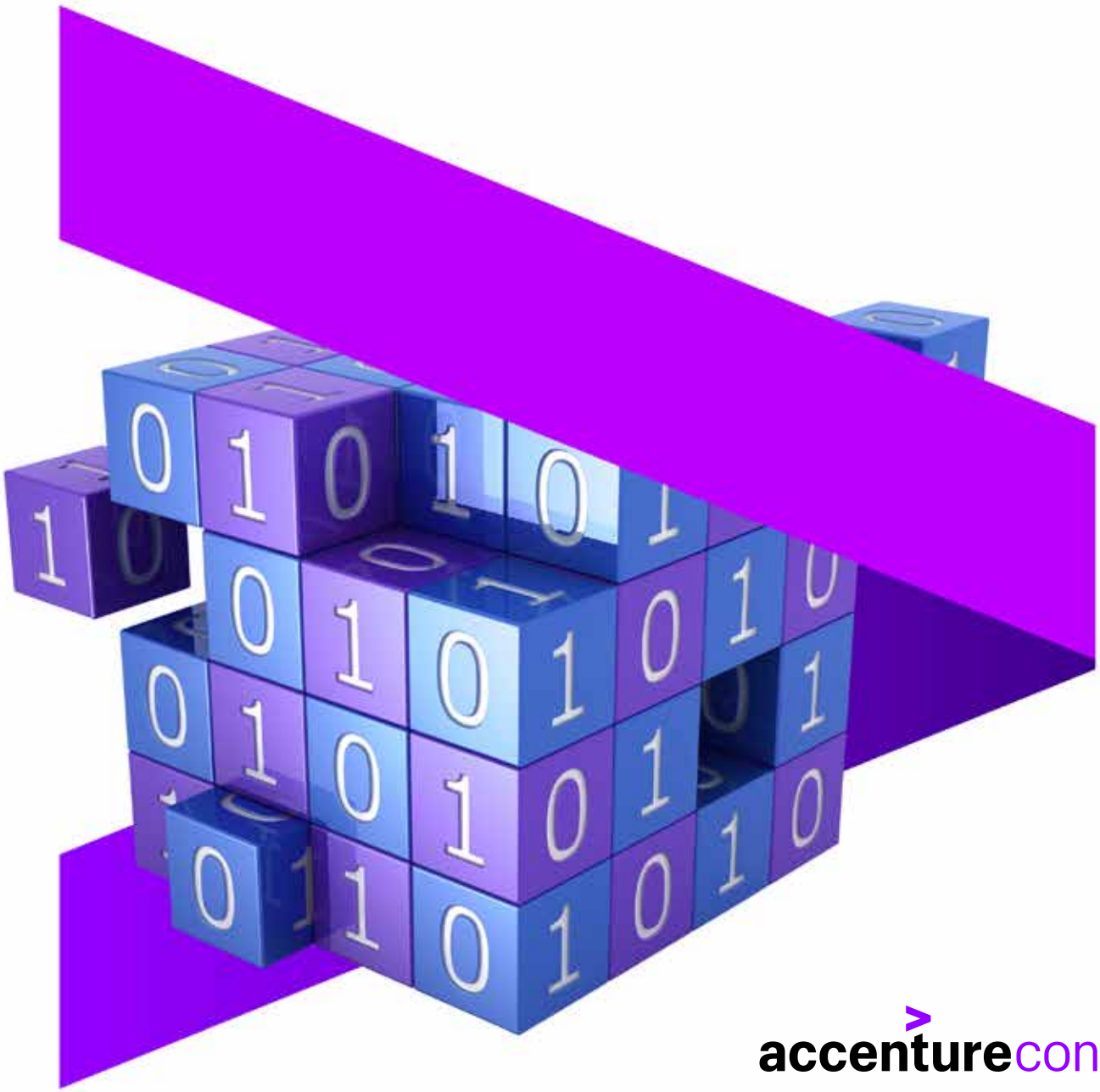
THE NEW DATA REALITY

FOR REVENUE AGENCIES



accentureconsulting

FOREWORD

These are fertile times for innovation driven by data analytics. New possibilities are emerging rapidly, creating significant opportunities for revenue agencies, while also increasing pressure to do more with data across a wide range of areas, from fraud detection to personalized experiences.

Over the next five-to-ten years, we estimate that revenue agencies will process over 100 times more data than paper/phone-based predecessors.[1] Much of this will, of course, flow rapidly over the internet, via online services and real time user engagement. In fact, government and business internet traffic will grow at 18 percent CAGR (compound annual growth rate) from 2015 to 2020.[2]

In addition, revenue agencies will increasingly tap into external data sources provided commercially or through the tax ecosystem partners. Blockchain based Distributed Ledger technology will likely serve to increase the breath of external data sources accessed by agencies.

To adapt to this data-driven world, and take the opportunities on offer, revenue agencies need to fundamentally rethink the way they manage and govern their data, in addition to assessing potential new sources of data that they do not directly control. This has to start by considering the implications of new systems architecture and data technologies, as well as the appropriate data governance model to support a data-driven, agile revenue agency. This paper explores these issues and sets out the fundamental systems architecture and data domain requirements for modern revenue agencies.

David Regan
Global Managing Director,
Accenture Revenue Industry Group

1

THE NEW REVENUE AGENCY SYSTEMS ARCHITECTURE

For decades, revenue agency systems architecture has centered on systems of record - securely and reliably holding taxpayer transactions, status and static information. But as taxpayer relationships have digitized, and as data volumes have grown, two other classes of systems have grown critically important.

The new revenue agency systems therefore has three major divisions:

SYSTEMS OF RECORD

Familiar to all revenue agencies, these systems hold taxpayer accounts, returns, historical data, forms and information resources.

SYSTEMS OF ENGAGEMENT

These are the systems that support data exchange and user-centric taxpayer interactions - including self-service systems, communications, continuous clarification, behavioral “nudges,” and virtual assistants.

SYSTEMS OF INSIGHT

This is where multiple data sources can coalesce to underpin descriptive, prescriptive and, increasingly, predictive analytics. These insights are increasingly driving decisions and innovations that are crucial to the future of tax agency strategy and operations.

THESE THREE CLASSES OF SYSTEMS WILL BE UNDERPINNED BY A MIX OF DATA STORES/ TECHNOLOGIES, EACH WITH TAILORED FUNCTIONAL AND TECHNICAL CAPABILITIES.

The **systems of record** are typically based on RDBMSs and/or file systems and will continue to support the core processes of revenue agencies. The real innovations though happen in the technologies that have evolved to support systems of engagement and systems of insight. The data stores for systems of engagement are supported by a suite of data technologies in many cases defined as operational data stores (ODSs). Data stores for systems of insight have evolved beyond data warehouses, to also include data lakes and enterprise data hubs.

2

WHY DOES DATA MANAGEMENT NEED TO CHANGE?

Harnessing the power of these evolving technologies and managing the interdependencies between the many internal and external data stores that interface with them creates new demands for the data domains serving the emerging systems architecture for revenue agencies.

But, stepping back for a moment, why do data systems need changing at all? Why is a new data architecture emerging?

CHANGING DATA DEMANDS

The answer spans resource constraints, expectations and opportunity, including:

- Fiscal pressure to do more with less
- Taxpayer expectations to build a better tax system to match their experience as consumers
- The opportunity to take advantage of a rapidly maturing set of technologies
- The large volumes of new types of data now available that can make agencies more efficient and more capable than ever before

Many revenue agencies have already committed to large-scale digital transformations that put data-driven systems and strategies at the forefront.

In 2015, for example, the Australian Tax Office published a "blueprint for change" which outlines plans to build a full set of integrated, data-driven digital solutions by 2020.[3] Many other countries have similar agendas, including in particular, the "Digital Five" – the United Kingdom, South Korea, Estonia, Israel and New Zealand – countries that have begun to collaborate on new approaches to digitalization of government data and services.

Data stores, methodologies and governance need to change as the revenue agency systems of engagement and systems of insight deal with ever increasing volumes of data and depend on an array of different data services to fulfil their function.

No single data store technology is optimal for supporting all four functions which leads to a phenomenon known as "polyglot persistence". Agencies therefore will achieve their goals with several different data storage and processing technologies working independently, and in concert, across all three data domains serving the emerging systems architecture.

Agencies need to develop ways to improve data management across four key areas:

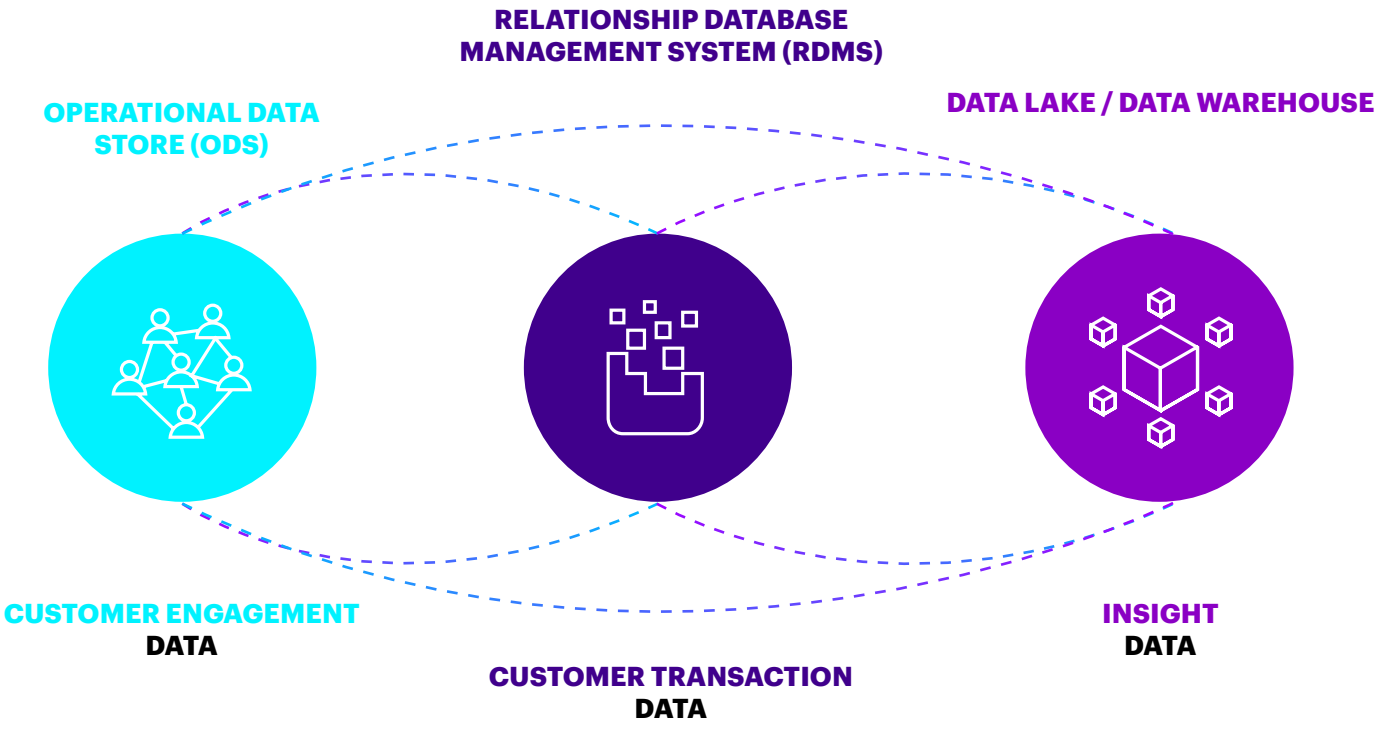
1. **Processing data at greater velocity, e.g. for real time fraud prevention, audits and payroll reporting**
2. **Using a wider variety of data sources, e.g. web logs and user click streams**
3. **Gathering insights from lower certainty or unstructured data, e.g. user generated content and social signals**
4. **Managing greater volumes of data, e.g. to support the capture of VAT or GST at transaction level**

3

THE TRADITIONAL AND THE TRANSFORMATIONAL: THE RIGHT MIX OF DATA STORES

Finding the right mix of data technologies to support systems of record, engagement and insight is a challenge for revenue agencies. The following describe the current state, the needs and technological directions for each one of those classes of systems.

THE DATA STORES UNDERPINNING THE NEW REVENUE ARCHITECTURE



SYSTEMS OF RECORD

Systems of record are well understood within the revenue agency context. As we touched on above, revenue agencies have largely relied on RDBMSs and traditional file systems to fulfil their duty to taxpayers and government stakeholders.[4] Systems of record are typically based on RDBMS and/or file-based implementations which remain at the core of the business.

Atomic transactions are stored in these systems which enforce high referential integrity and strict formatting requirements. Relationships, rules and constraints are applied to all new records as well as updates to existing records.

Although RDBMSs may be less flexible than newer data technologies, they are hard to beat when it comes to high volume transactional processing and accurate, reliable storage of high-certainty data. For these reasons, customer transaction data will, for the most part continue to be processed through traditional RDBMS.

But while RDBMSs are well-suited to maintaining robust records of tax processes, basic reporting and traditional customer service needs, they are not designed to manage the velocity, variety, volume and varied levels of certainty that are inherent in the data stores of the new revenue agency.

SYSTEMS OF ENGAGEMENT

Systems of engagement are increasingly important. Taxpayers expect their interactions with tax agencies to be similar in sophistication, effectiveness and speed to that which they enjoy from their banking, retail, media and entertainment services. Accenture research suggests revenue agencies currently fall short in this regard. For example, **while three-quarters of taxpayers visit their revenue agency's website, less than half of those find what they are looking for.** [5]

Revenue agencies are seeking to develop a digital user experience which engages and supports taxpayers while achieving compliance objectives. To do this, it is often necessary to combine large data sets, by replicating information from disparate systems, while providing API driven access.[6]

For example, current and past tax records need to come together with real time risk scoring (based on internal and third party data), plus social signals and online behavior analysis, to mention just a few. These need to integrate rapidly to confirm the taxpayer's position across relevant taxes and benefits, to assess fraud risk, provide personalized guidance and influence desired behavior. It is all about having the right information provided at the right time through the right channel.

Modern revenue agencies need to redesign user experiences and offer a new range of intelligent, contextual and adaptive interactions with each particular segment of the user base – class of taxpayers. Ideally, as soon as a taxpayer is identified, the system leverages his/her unique profile, interaction history, identified needs and implied or stated preferences in order to serve an optimized, personalized experience matching the particular timing, channel, request and session. The systems of engagement will allow conversational interactions, un-structured communication via text or even voice requests; they will support content search and access via natural user interfaces. They will provide consistent experiences across devices and channels to ensure continuity in the user (taxpayer) experience.
They will have to provide a simple user experience, focusing on the exact need and state of each particular taxpayer – and this requires complex technologies and powerful data models.

This new level of service now expected by taxpayers implies a dramatically increased volume of data, also including new data types. In offering advanced personalized experiences across channels, systems of engagement must have instant access to the exact state of each taxpayer, a suitably aggregated history of interactions and requests, user insights, rules and patterns which allow better serving each particular taxpayer. Suitably classified content, enriched with performance optimization metadata can be used to dynamically prepare content-driven, personalized experiences for each particular customer through a given touchpoint.

This new generation of experiences and the increased level of service depend on fast, streamlined processing of large volumes of complex, heterogeneous data and rapid insight generation and/or retrieval. At the same time, these new range of interactions will generate large volumes of data which need to be captured, handled and fed back to the intelligent insights layer for further processing.

The range, volume and heterogeneity of data cannot be effectively served by conventional RDBMS and DW technologies. This is where the new data technologies come into play: NoSQL operational data stores (ODSs) can play a significant role in handling vast amounts of structured/semi-structured/unstructured data and empower personalized services to individual taxpayers. Data lakes will help enable Agencies to accept and handle any kind of data, in any format, from binary large objects (for example voice requests captured in binary files, multilingual video content) to payroll data, scanned documents, forms, legal documentation etc. This allows the Agency to capture, store and prepare any type of document or data object for further streamlined analysis and modeling.

SYSTEMS OF INSIGHT

The 'Systems of insight' is where the magic happens (or, if we focus at the data layer, is stored). Systems of insight significantly benefit from new data technologies and capabilities as a wider range and increased volumes of data become available for modeling: Analytical models consume established or ad-hoc feeds of heterogeneous data, coming from different data sources, possibly including the ODS in addition to the data lake, to apply advanced modeling and analytical processing.

The output (patterns, insight, models, rules, metadata of several types, classification systems, ontologies etc.) is stored in special data structures and is shared across systems as corporate knowledge enabling optimization, automation and decision support.

The insights generated, depending on the case, could be stored back to special data models in the DW and/or in a NoSQL storage scenario possibly including an Operational Data Store. This way, knowledge and insights become actionable via standardized data integration processes, data access patterns and APIs.

Enterprise applications, system components, reporting systems and Business Intelligence tools to consume taxpayer analytics and metadata, scores, web traffic patterns and empower decision support functions, optimization components, personalization logic, web traffic analysis, content performance assessment, predictions and more.

4

THE DIFFERENTIATORS: NOSQL ODS AND DATA LAKES

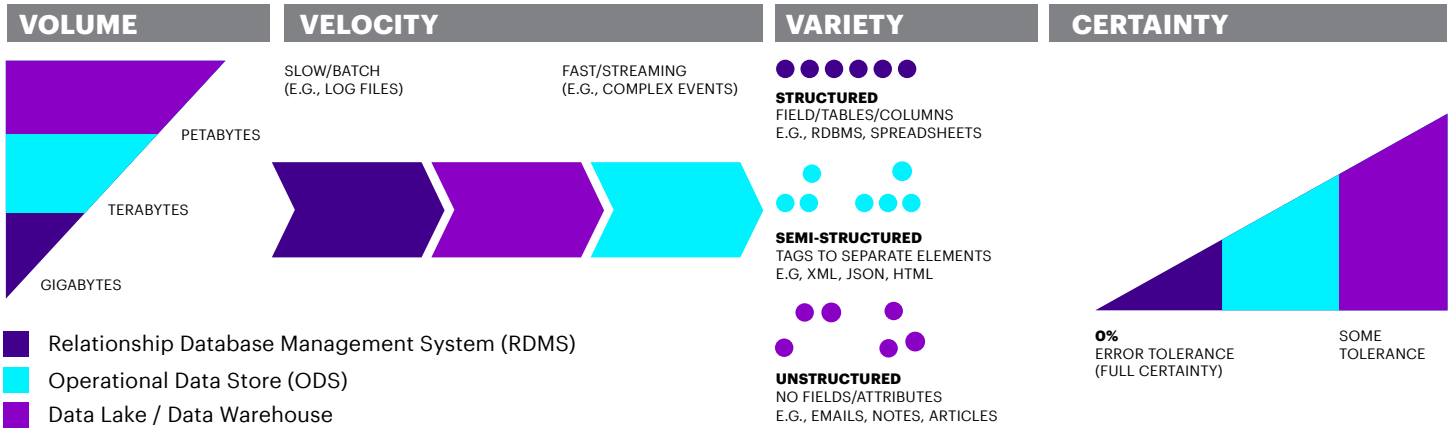
THE NOSQL ODS

Operational data stores underpin best practice

Drawing together multiple databases has already revolutionized our lives in multiple areas – beginning with services from Google, Amazon and Facebook. The use of ODSs has been critical to the success of these enterprises in rapidly providing the right information at the right time, for their systems of engagement. ODSs are now spreading to businesses as diverse as British Gas (an energy retailer), Activision (a video games publisher) and William Hill (a sports betting company). [7]

Systems of engagement today are often based on the open source ODS, Apache Cassandra, which has arguably become the most powerfulleading NoSQL data platform available. It is used by many of the world's leading companies, including financial services giants such as ING, UBS and Macquarie Bank.[8]

KEY DATA STORE CHARACTERICS - VOLUME, VELOCITY, VARIETY AND CERTAINTY



Availability and Scalability

An ODS works as efficient data consolidator, replicating data from disparate systems to create a single data store. While this is not new in itself, contemporary systems are now enabling highly personalized interactions, driven by hundreds of variables and real time analytics (e.g. recommendations, risk scoring, etc.). An ODS can deliver this while providing high availability and linear scalability as the number of users and volume of data grows over time.

An ODS forms a strong foundation for supporting many common revenue agency online service requirements, including the ability to capture taxable transactions closer to the source, as they happen. The immediate capture of tax data not only increases collections and reduces errors, fraud risk it also reduces the administrative burden on taxpayers. It helps enables increasing the frequency, volume and velocity of data exchanges with third parties – that is why an ODS is needed to underpin the efficient, real-time delivery of services.

While in a typical RDBMS scenario transaction data requires 100% veracity and is subject to strict formatting requirements a NoSQL customer engagement ODS receives data from a multitude of sources and can be designed in such a way as to decouple itself from the restrictions of the original system.

KEY BENEFITS OF OPERATIONAL DATA STORES FOR SYSTEMS OF ENGAGMENT

1. Scalable and flexible control over data through constantly changing business requirements
2. Management of large data sets, sourcing from one or multiple systems/databases
3. Fast turnaround with control from business applications
4. Data scrubbing and transformation capabilities, including modification of both data and master/metadata
5. Powerful query capabilities, i.e. against tables and views for quick answers from large, complex data sets

DATA LAKES

Data lakes and similar technologies can store a wide range of unstructured data, including information that has never before been available for use by revenue agencies. They are also cost effective basis for systems of insight, helping agencies manage the storage of the ever-expanding volume of data from all sources. As such, data lakes are another transformational data store technology – they will be key to the future success of analytics engines, machine learning and artificial intelligence systems that will ultimately power a new set of service and compliance capabilities in revenue agencies.

As data lakes and surrounding analytical methodologies mature, revenue agencies will be able to discover the answers to questions they had not yet thought to ask, and identify patterns that have, to date, been impossible to discover.

But data lakes will not typically replace traditional data warehouses, which still have their place for aggregating data at scale (for example, to support financial reporting) and supporting business intelligence (see Figure 2).

FIGURE 2: WAREHOUSES AND LAKES: WHAT IS THE DIFFERENCE?[9]

DATA WAREHOUSE	vs.	DATA LAKE
Structured, Processed	**DATA**	**Structured / seimi-structured / unstructured, raw**
Schema-on-write	**PROCESSING**	**Schema-on-read**
Expensive for large data volumes	**STORAGE**	**Designed for low-cost storage and scalability**
Less agile, fixed configuration	**AGILITY**	**Highly agile, configure and reconfigure as needed**
Mature	**SECURITY**	**Maturing**
Business Professionals	**USERS**	**Data Scientitist et. a.**

Data lakes, often supported by, or incorporating, data hub technologies, will draw on all internal data stores related to taxpayers and their services. This includes RDBMSs (i.e. all current and historical tax filings and activities), graph databases and ODSs (i.e. including clickstreams, social, web logs, geospatial data, etc).
At the same time, they will combine external/ third-party data sources, such as information from employers, other public services and other tax jurisdictions.

Data lakes are fundamentally low-cost data management systems, rather than processing systems, so archiving in the normal sense is typically not required. By contrast, for systems of record, policies and processes for auditing, long-term storage and deletion of data are fundamental.

5

IMPACT ON DATA GOVERNANCE

As keepers of highly sensitive information, revenue agencies already have advanced data governance procedures. While these procedures are a valuable foundation, they are designed to govern structured, high certainty data and relatively low velocity, e.g. tax accounting and after-the-fact compliance. Facilitating digital self-service and real time compliance will require a more nuanced approach.

CERTAINTY STANDARDS VARY ACROSS DATA STORES

Fundamentally, revenue agencies need to look at their data domains through different lenses. Data certainty requirements, for example, are different on each type of data store.

The RDBMS that holds customer transaction data requires 100% veracity and data is subject to strict formatting requirements.

By contrast, a customer engagement ODS receives data from a multitude of sources and is designed in such a way as to decouple itself from the restrictions of the solely transactional system. The ODS needs to set the right level of validation and certainty for data records according to standards defined by each particular process. For example, there would be lower level of validation required for tailored informational services compared to accepting transaction data supporting refund requests.

Flexibility is a primary strength of data lakes, which have minimal requirements for veracity and formatting in order to accommodate data in a vast range of formats. Lower quality data is accepted and its usability is determined in future investigations.

KEY DIFFERENCES IN SECURITY AND ACCESS CONTROL STANDARDS

Security and access control requirements are also clearly different across data stores. For example, systems of record would typically sit behind multiple layers of firewalls and access to the underlying data would be controlled via rigid procedures and detailed access logs. High levels of authentication with strictly defined authorizations apply internally and externally.

By contrast, data lakes are typically accessed by a small number of users (e.g. those working with analytics models, machine learning, data mining etc.). Data is typically not directly exposed to the outside world, so data store security, while still critically important, may be less complicated. However, as the data is sourced from different domains with potentially different audit controls there may be a need for data segregation if such controls must be inherited by the data lake. In addition, as taxpayer data is typically accessed in bulk (unlike one-at-a-time transactional systems) creating full audits of access is potentially complex and demanding.

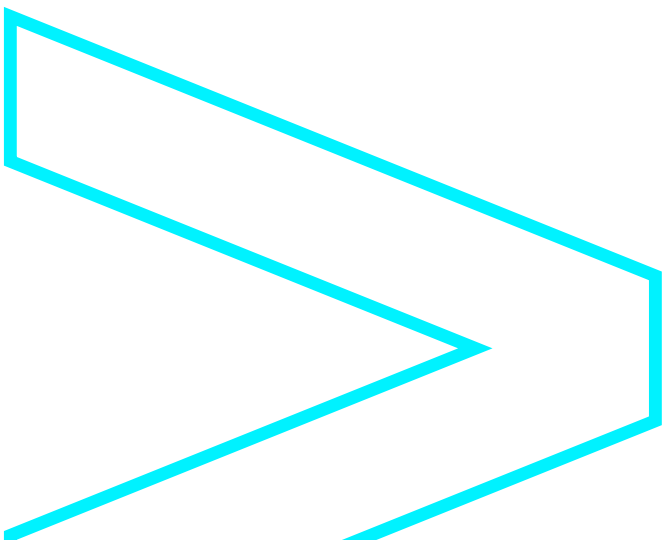
Data profiling, detecting fraud and data breach monitoring are all continuous undertakings across taxpayer data within a RDBMS, a customer engagement ODS and a data lake. However, the systems and technical skills required differ.

RETENTION, IMPACT ANALYSIS, METADATA MANAGEMENT – THE LIST GOES ON

Retention and archiving represents another example of difference. Data lakes are fundamentally low-cost, reporting systems, rather than processing systems, so archiving in the normal sense is typically not required. By contrast, for systems of record, policies and processes for auditing, long-term storage and deletion or archiving of data are fundamental. An ODS presents different needs again - having to retain data that may be relevant to future interactions, or other purposes, while also identifying data which can be purged.

The recruitment and retention of new technical and business skills, addressing the potential of all this data is already challenging agencies. Integrating these skills into the business and IT operations requires revised operating models and a faster change cadence based on constantly evolving data insights.

Revenue agencies need to consider the data governance complexities created by their evolving systems architecture. In this section we have just scratched the surface. In general, for most revenue agencies, data governance needs to be strengthened and expanded significantly to effectively manage the unique characteristics of different data domains without impeding data-driven innovation.



6

BANKING ON THE RIGHT MODEL

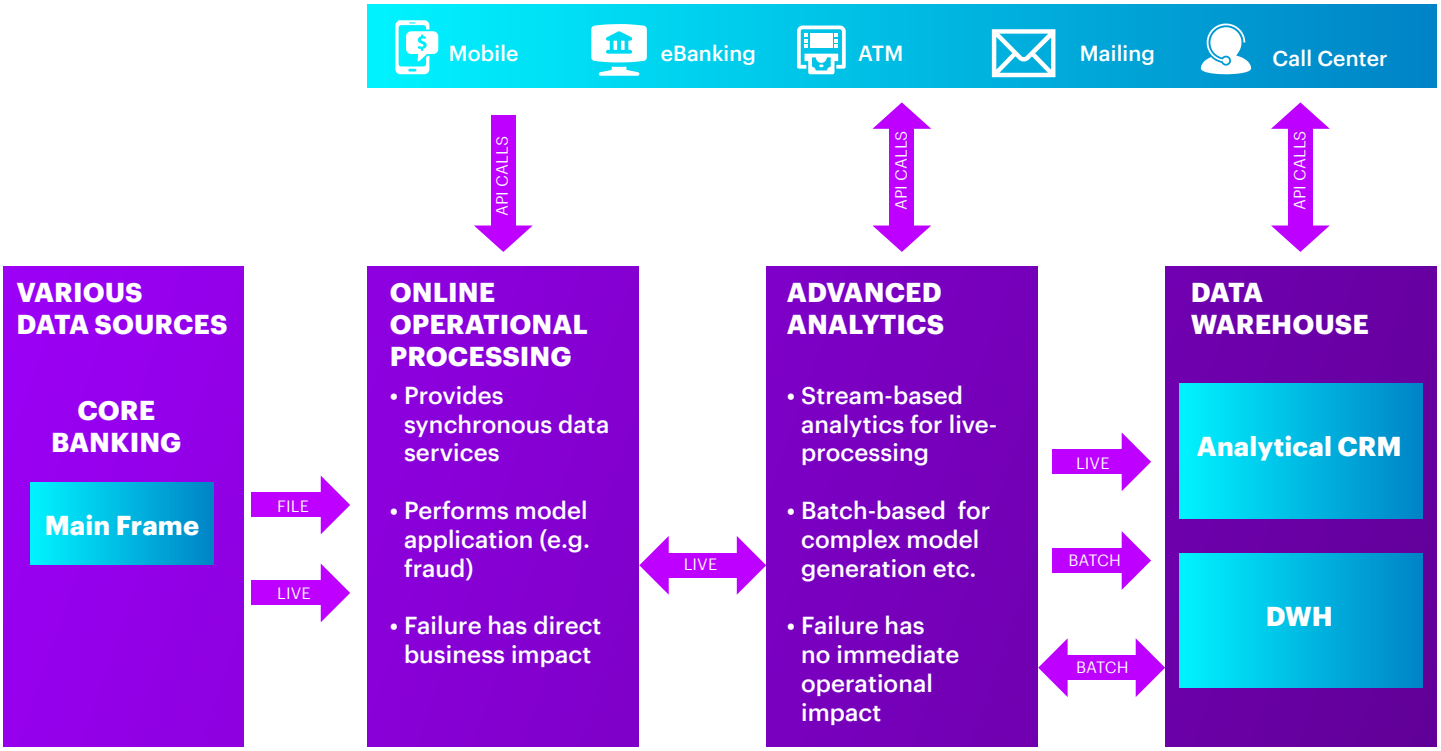
Accenture's experience with financial institutions illustrates similar data domain demands to revenue agencies. Indeed Accenture's engagements generally with the private sector serve to illustrate the acceptance and proliferation of a polyglot approach to data repositories with special data domains in place to support customer service, operational insights and core transactional processing.

The banking world offers perhaps the closest comparison to the revenue industry. While each bank has unique considerations influencing their data management approach, several common themes continually recur, including:

- Preserving the data in the core banking platform.
- Creating online operational analytics capabilities to support online customer services typically based on the replication of data from core platforms, integration of advanced analytical models and operational data captured from online interaction.
- Strengthening advanced analytics capabilities by taking advantage of file systems technologies such as Hadoop to create analytical models which service customer service and risk system demands.
- Enhancing business intelligence using technologies such as Hadoop to enrich traditional data warehouse capabilities.
- A hybrid infrastructure approach leveraging both on premise and cloud.

Supporting these architectures is combination of replication approaches both batch and real-time which reflect the requirements of each business function with regards to data velocity. A stylized example of such an architecture is provided below based on the solution adopted by an Austrian bank.

BANKING DATA ARCHITECTURE



For revenue agencies evolving data technologies create new possibilities and efficiencies, but also produce more complex data governance requirements. While data demands vary by agency as they do with banks, the characteristics of their systems architectures are shared and lead revenue agencies to a similar set of demands from a data management perspective.

CONCLUSION

Revenue agencies system and data architectures are evolving and agencies are starting to get to grips with the data governance approaches required to support the involved data domains.

Data management has never been more complex or more important. Systems of record, engagement and insight underpin everything the modern revenue agency does, including unprecedented advances in service delivery and analytical power. Revenue agencies will use multiple data storage technologies choosing different options for data domains supporting systems of record, engagement and insight. This polyglot persistance will require improved data management processes.

Cost will be a key driver as revenue agencies seek to put in place fit for purpose data solutions to meet their varying requirements. Technology options will not map neatly to the different data domains, overlapping and competing with differing strengths and weaknesses. However a focus on cost and fit for purpose will deliver a polyglot data architecture for revenue agencies suited to their needs.

It will allow agencies to be instantly connected to taxpayers and the economy, driven by ever-deeper insights, able to adjust to continuous change, and to orchestrate – not operate – a highly automated, secure, personalized and efficient tax system. In short, Revenue agencies need to pick the right data tools for the right use case.

AUTHORS

David Regan
Global Managing Director,
Accenture Revenue Industry Group

George Krasadakis
Senior Architect,
Software Engineering, The Dock
Accenture Ireland

John Baxter
Senior Technology Architect,
Accenture UK

Yaz Yazicioglu
Senior Technology Architect,
Accenture Federal Revenue Practice

ABOUT ACCENTURE

Accenture is a leading global professional services company, providing a broad range of services and solutions in strategy, consulting, digital, technology and operations. Combining unmatched experience and specialized skills across more than 40 industries and all business functions – underpinned by the world's largest delivery network – Accenture works at the intersection of business and technology to help clients improve their performance and create sustainable value for their stakeholders. With more than 394,000 people serving clients in more than 120 countries, Accenture drives innovation to improve the way the world works and lives.
Visit us at www.accenture.com.

REFERENCES

1. Accenture analysis.
2. Cisco. The Zettabyte Era: Trends and Analysis. July 2016.
3. Australian Tax Office. Program blueprint for the future tax. March 2015.
4. Complemented by satellite systems for business intelligence, analytics, content and capture.
5. Accenture Taxpayer Survey. 2015
6. This replication applies in many instances but could be minimized if all transactional data is stored on a single database (i.e. all registrations, tax types, benefits and cases data).
7. ComputerWeekly. Keeping up with Cassandra, the NoSQL database. October 2015.
8. www.datastax.com/customers
9. Tamara Dull, Director of Emerging Technologies, SAS Institute, blog August 28, 2015