Zetaris

Major Australian Institutional Bank

Zetaris Client Case Study

Introduction

A large Australian Institutional Bank, one of the top 4 in Australia, embarked on one of the largest Data Quality programs conducted. Zetaris supported the data strategy and initial deployment using its Virtual Data Warehouse (Agile Data Fabric) platform.

The Customer Data Quality program followed many years of project failures to detect the 'real' level and impact of poor data quality and the bank's ability to deliver a single view of its customers. The program was designed with the ultimate goal of fast-tracking new and improved Customer solutions to complex data sharing problems. The Institutional Bank required a platform to capture, analyse and identify Customer data with poor data quality and to report on the data quality dimensions to the regulator.

The Opportunity

The Institutional Bank has an established practice to capture and analyse Customer data sets; however, the traditional data quality processes were too cumbersome, short-term focused with unconnected data sets and services that were too costly to scale up to the levels needed.

A shift in how data is captured and used was proposed by building an information hub supported by a dynamic Big Data platform and delivered through the use of a solid Data Management approach to ensure sustainable operations for the long term.

Our Approach

Zetaris installed its Data Fabric platform and provided secure access services to the Bank's risk management team. A data quality strategy, framework and operating model were developed, and the following outcomes were delivered with solutions built and a framework to support them, so that the internal team could manage data quality issues going forward:

- A business-centric Customer data model was developed with a 'Fluid' data model for the customer domain, sub-typing customers into 'Legal Entity', 'Organisation' and 'Individual', with further sub-types developed depending on the business rules to be enforced by the Data Quality system. Refer to Appendices 1 and 2 for 'Business Rule Definition' and 'Sample Business Rules'.
- The Customer metadata from each of the 7 Customer master and reference systems, the operational Product systems, and the data warehouses were ingested in real-time by the active metadata manager (known as the Schema Store) and integrated with the business-centric Customer data model on the Data Fabric platform.
- Customer data were then validated against the business-centric Customer data model based on the business rules defined in the Customer data model. Refer to Appendix 3 for a 'Sample of Business Data Standards'.
- A Customer data 'Exception' set was defined and captured in the standard Exception Management schema, together with all the characteristics required to perform the root cause analysis and corrections needed to inform the Source systems of poor data quality. Refer to Appendix 4 on 'Adherence to Data Quality Metrics'.
- Exception Management dashboards and reports were automatically produced based on the Exception Management schema and Data Quality rules. Refer to Appendix 5 for definition of 'Data Quality Rules'.
- Predictive algorithms integrated with the Customer data set provided insights into customers' behaviour and likely reasons for the poor quality data.
- The same integrated Customer data set was used to perform Customer segmentation modelling to identify the likelihood of customer fraud and to manage their customer lifecycle.

Solution

Zetaris was engaged initially to perform a strategic review of the project. As part of this strategic review, Zetaris submitted several recommendations, including employing a sound data-centric approach with a solid Data Architecture for designing the solution, and selecting appropriate technology to ensure the future-proofing of the Data Quality solution, considering the very long-term horizon and complex Institutional Bank requirements.

The Zetaris Data Fabric platform (Virtual Data Warehouse) was chosen after a carefully considered process incorporating a review of Data Quality packages available, including those previously implemented by the Bank's Corporate division

> **The platform provided more consistent data in a more timely and trusted manner, enabling more informed decision-making.**
>
> *Data Governance Director*

At the end of the Proof-of-Concept (POC) phase the Institutional Bank completed a successful implementation focused on re-using previously collected data sets, which were not integrated and hindered users from sharing and re-using outcomes. Following the POC phase, the Zetaris Virtual Data Fabric platform was then confirmed as the Data Quality platform and the project proceeded to on-going phases and operational support.

The Results

Knowledge Transfer

A newly formed Data Management team was formed during the POC phase. Its members were trained on the project and easily adopted the fundamental practices and knowledge needed to support and enhance the Data Quality platform in on-going phases.

Process Optimisation

New processes supported by the Zetaris Data Fabric platform have been continuously discovered and implemented. They standardise and optimise data-related operations not possible before, such as managing Ontology classifications, managing Data Harmonisation and Data Derivation, ensuring Data Privacy by using Customer 'Keys', and integrating Customer data across business units.

Long-Term Vision

The vision for the Data Quality platform was established. It envisages, the use of the Zetaris Virtual Data Warehouse platform with opportunities to integrate 'virtually' the collected data sets, removing the need for multiple copies of the data and physical data movement; new use cases supporting structured and unstructured data sets (e.g. documents, images, audio and video from research), and improved self-service capabilities for customer functions to access and share data.

Appendix

1. Business Rule Definition:

- Business Rule is a constraint on the data "by the way we do business" and the way we wish to enforce integrity on that business data. Cf. Many of our existing data management procedures are driven by system constraints.
- There are 2 categories of Business Rules:

1. 'Business Data Constraints' – Business data constraints stem directly from the business and are defined in the data model. They pertain to the data, independently of application and organisation structure.
e.g. if an "Individual" entity exists then it cannot at once be an "Organisation" entity.

2. 'Data standards' – These standards must be consistently applied across systems; must support the business data constraints; and must remove data redundancy and maximise shareability of data across systems. There are generally 2 types of rules here:

(i) Taxonomy – Define 'business objects' using business glossary and business data model

(ii) Identify – Map data elements from applications and transform them to defined 'business objects'

• Data quality rules then enforce these rules across the ecosystem by way of Exception Management. Data quality rules are aligned to APRA's Data Quality metrics.
• Business rules also drive a systematic approach of implementing the right set of controls across the ecosystem. 'Business Data Constraints' are 'hard rules' which require preventative controls, while 'Data standards' include both hard and soft rules requiring a variety of rules (preventative and detective controls).

2. Business Rule Sample

1. Organisation, Sole-Trader & Business

I. An Organisation may operate as a Business under its own Organisation Registered Name. An Organisation may operate using additional Organisation Business Names, one for each alternative Organisation Business.

II. A Sole-Trader Individual may operate as a Business under its own full Individual name. A Sole-Trader Individual may operate using additional Sole-Trader Business Names, one for each alternative Sole-Trader Business.

Appendix

III. A Business must be associated with one and only one Organization or one and only one Sole- Trader Individual.

2. Customer, Customer Account & Customer Associated Account

I. Customer Accounts can be 'logical', that is, defined for internal WIB management processes and not visible externally to the Customer. They are used for instance by Risk and Relationship Managers as aggregation points to determine product revenue measures at different levels.

II. An instance of a Customer Account must be associated with one and only one Primary Customer, however a Customer Account may be owned by more than one Customer based on rules defined by a Customer Account Ownership Type.

III. A Customer Associated Account represents the fact that a Customer may hold one or more than one Customer Account instances.

3. Party, Party Profile Type, Customer Account & Customer Associated Account

I. Additional Parties may be part of the Customer Associated Account under defined Party Profile Types. Party Profiles are used to nominate additional roles other than that of a Customer as an Account Holder required to manage the Customer Account, for example, a Guarantor and an Authorized Signatory.
II. A Party using Party Profile Type in relation to a Customer Account, may at the same time be a Customer for one or more other Customer Account instances, or may be a Non-Customer.

4. Customer, Customer Profile & Customer Account

I. A Customer Profile represents one or more roles that a Customer plays across one or more Customer Account instances or transactions that the Customer is involved in the bank; for example, a Customer as a Guarantor for a Loan Account or a Customer as a Counterparty for a Trade Transaction.

Appendix

3. Business Data Standards Sample

A Trustee/ Responsible Entity (collectively 'Trustees') • A trustee holds or manages cash/assets/ property for a beneficiary (i.e. investor). The trustee has a fiduciary duty to act in the best interest of the beneficiary and is responsible for providing oversight of the Fund's assets and appointing third party providers such as investment managers, fund administrators and custodians.

Superannuation/Pension Trustees are responsible for an investor's fund's pre and post-retirement savings to provide them with a lump sum or an income stream in retirement. This is typically done by a Trustee Board setting a target asset allocation and fund manager selection. The Board may use Asset/Investment Consultants.

ATF, ATF For, As Trustee Of, As Trustee For, In Its Capacity as Trustee For ARE, A.R.E.

FOR, ARE FOR, AREF, AS RE FOR, As Responsible Entity For

An Investment Manager or Fund Manager is appointed by the Trustee/ Responsible Entity of the Fund to implement a Fund's investment strategy in accordance with the stated investment guidelines. Their customers are the trustee/ responsible entity of the Fund on behalf of the investors. They are typically the banks counterparty rather than the Trustee

AIM, ACTING ON BEHALF OF ITS SUB-FUND, AIM For, AIMF, AS IM FOR, AS A & M FOR

An Agent is a party authorised to act on behalf of another party. Following keywords to be used to identify Agent:

As Agent For, As Manager and Agent For, On Behalf Of

Appendix

A Custodian Typically appointed by the trustee of the Fund, a Custodian is a financial institution that holds customers' securities for safekeeping to minimise the risk of their theft or loss. Most custodians offer other services such as fund administration (NAV calculation), transaction settlements, collection of dividends & interest payments, tax support, escrow services and FX. :
ACF, AS Custodian For, ANF, Nominee, As nominee for

A Guarantor is a party that agrees to be responsible for another party's debt or financial performance. Government guarantees were created during the GFC for credit purposes and such rules no longer used. Following keywords to be used to identify Guarantor:

GTEED, GOVT GTEED

Note that Business Data Standards evolve as you work through the Exceptions identified.

Appendix

4. Adherence to Data Quality Metrics

Data Quality	APRA Data Quality Metric	
Correctness	(a) accuracy: the degree to which data is error free and aligns with what it represents; **Correctness**	(f) **fitness for use**: the degree to which data is relevant, appropriate for the intended purpose and meets business specifications. APRA's assessment of fitness for use: "Data validation is the assessment of the data against business rules to determine its fitness for use prior to further processing."
Completeness	(b) **completeness**: the extent to which data is not missing and is of sufficient breadth and depth for the intended purpose;	
Consistency	(c) **consistency**: the degree to which related data is in alignment with respect to dimensions such as definition, value, range, type and format, as applicable;	
Data quality concepts based on process and not based on content of the data	(d) timeliness: the degree to which data is up-to-date;	
	(e) availability: accessibility and usability of data when required; and	

Appendix

5. Data Quality Rules: Nature and Alignment to Regulators' Metrics

	Exception Type Code	Exception Type Description	Warning Message Indicator	Exception Type Action	DQ Rule alignment
Internal Customer Master Data	001	Expected Primary Party not found in target entity Party - "Party" must exist prior to creating "Primary Party	N	Reject data	Completeness
	002	Expected Associated Party not found in target entity Party - "Party" must exist prior to creating "Associated Party"	N	Reject data	Completeness
	003	Party record found in RMM/MyClient not in master CIS	Y	Reject data	Accuracy, completeness, consistency
	004	Duplicate source records with same Candidate Key	N	Reject data	Accuracy, completeness, consistency
	005	Integrity violation, mandatory candidate key is NULL	N	Reject data	Accuracy, completeness, consistency
	006	Multiple records in internal master CIS matching the same External record	N	Reject data	Accuracy, completeness, consistency
Relationship Hierarchy Data	101	Ultimate Parent for Legal Hierarchy exists Externally, not internally	Y	Data not rejected	Accuracy, completeness, consistency
	102	Entity exists in internal Legal Hierarchy, not in External legal Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	103	Ultimate Parent for Legal Hierarchy exists internally, not Externally	Y	Data not rejected	Accuracy, completeness, consistency
	104	Entity exists in external Legal Hierarchy, not in internal legal Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	105	Immediate parent different between internal and external Legal Hierarchies	Y	Data not rejected	Accuracy, completeness, consistency
	106	Level of hierarchy different between internal and external Legal Hierarchies	Y	Data not rejected	Accuracy, completeness, consistency
	107	Ultimate Parent for External Legal Hierarchy exists, and not in internal RM Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	108	Entity exists in internal RM Hierarchy, not in External legal Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	109	Ultimate Parent for Internal RM Hierarchy exists, not for External Legal Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	110	Entity exists in external Legal Hierarchy, not in internal RM Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	111	Immediate parent different between internal RM Hierarchy and external Legal Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	112	Level of hierarchy different between internal and external Legal Hierarchies	Y	Data not rejected	Consistency
	113	Ultimate Parent for Internal Legal Hierarchy exists, and not in internal RM Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	114	Entity exists in internal RM Hierarchy, not in internal Legal Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	115	Ultimate Parent for Internal RM Hierarchy exists, not for Internal Legal Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	116	Entity exists in Internal Legal Hierarchy, not in internal RM Hierarchy	Y	Data not rejected	Accuracy, completeness, consistency
	117	Immediate parent different between internal RM Hierarchy and Internal Legal Hierarchy	Y	Data not rejected	Consistency
	118	Level of hierarchy different between internal RM Hierarchy and Internal Legal Hierarchy	Y	Data not rejected	Consistency
Product System Data	201	Mandatory Customer Attributes Missing in Product System	Y	Data not rejected	Completeness
	202	Mandatory Customer Attributes Invalid in Product System	Y	Data not rejected	Accuracy, completeness, consistency
	203	Customer Record Missing in Products system	Y	Data not rejected	Completeness
	204	Customer details not matching between Product System and Primary Customer Source	Y	Data not rejected	Consistency
	205	Duplicate customer record in Product System	Y	Data not rejected	Accuracy, completeness, consistency
	206	Customer Exists as Organisation and Individual in Product System	Y	Data not rejected	Accuracy, completeness, consistency

For further information visit:

zetaris.com