Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Capital Markets | the way we see it

Architecting Enterprise Data Management Systems

People matter, results count.

Contents

1 Introduction

Data management has been pushed to the forefront today by the multi-pronged squeeze of compliance, risk management, cost efficiency of operations, effectiveness of client relationships and marketing. All of these business functions rely on the accuracy of reference data, but multiple parties like credit risk, operations, trading and compliance executives all view the same information differently. This can lead to material disputes about how much information should be stored, in what form and who should control it.

An enterprise data management project brings all of this under its fold, holding responsibility to establish standards conformity, integrity, reliability and increase efficiency and throughput. To be successful, the enterprise data management team requires a deep understanding of both the data needs of various business processes and the building blocks for various asset classes.

2 Getting the Details Right

We know that the devil is in the details. To avoid being ensnared by the mammoth load of tons of data, a sound strategy is essential for a successful project. Understanding and addressing the "three Cs" is key to successfully harnessing the data. These three Cs are Classification of Data, Consolidation of Data and Consumption of Data: and they provide a solid framework to support your project and data models.

2.1. Classification of Data

The approach to the project and data modeling hinges to a large extent on the classification of data by asset classes. The highest level of classification is Equities, Fixed Income, Derivatives, Asset-Backed Instruments and Structured Products.

The asset coverage targeted in the project will determine the granularity of the data model. It also gives a view of future consolidation of vendor feeds and proprietary or common industry standard that can be applied to the data pumped in. Identification of the asset classes and the data handled in each category is essential to clearly understand the demands expected of the data model. Such a grid can be drawn up as shown in exhibit 1.

Exhibit 1: Asset Classes and Data Needs (Indicative Grid)

		Information				
Asset Class		Static Data	Pricing Data	Deal Terms & Conditions	Corporate Actions	Clearing Data
Fixed Income	Agencies	☐	☐	☐	☐	☐
	CMO/ABS	☐	☐	☐	☐	☐
	Corporates	☐	☐	☐	☐	☐
	MBS	☐	☐	☐	☐	☐
	Municipals	☐	☐	☐	☐	☐
	Treasuries	☐	☐	☐	☐	☐
Derivatives		☐	☐	☐	☐	☐
Equities		☐	☐	☐	☐	☐
Futures		☐	☐	☐	☐	☐
Indices		☐	☐	☐	☐	☐
Money Markets		☐	☐	☐	☐	☐
Options		☐	☐	☐	☐	☐
Structured Products		☐	☐	☐	☐	☐

2.2. Consolidation of Data

When it comes to consolidating data from various feeds for the same security or instrument type, the nature of challenges that have to be handled become even more varied. To be able to compare securities across different feeds the field formats, type and identification of a primary key set are critical. Identification of such a key set is necessary because there is no uniform standard that is followed by the numerous data vendors in the market. Some vendors may provide certain security details such as market lot while others do not. Depending on the pre-dominant geography of the data vendor certain data attributes that are essential for a particular market may be lacking in the feed. Ability to confidently fill in data from another feed after a strong match is established is an objective that must be fulfilled by the data model and associated business rules.

As these salient attributes of the design are addressed, the number of data feeds and their form can be rationalized; simultaneously bringing down the load on exception management. But this is not possible at the onset. Initial data loading, parsing and validation must occur within a critical time window to handle the large number of feeds typical of most organizations. Thus while the data model and the transformation layer must offer the capability to handle every form of data desired; they must also be optimized to perform with the least amount of overheads.

This is a mission critical operation and an enterprise data management system must be able to manage these volumes within the allotted time-frame. Slicing of the pre-data loading validations, loading, transformation to a standard schema, validation and consolidating to a single golden copy into independent asynchronous operations helps in achieving good operational efficiency.

2.3. Consumption of Data

The end business processes that will use the data and the manner in which it is needed is the third angle which must be taken into consideration during system design. Depending on the subscribing business process demands this can be classified as:

- Security master
- Client and counterparty
- Pricing
- Position and transactions
- Corporate actions

The storage demands of static data and market data are quite different. Relational databases are generally stretched to meet the needs of storing data that is in a state of continuous flux. Capacity handling, data archiving and data retrieval performance benchmarks also vary depending on the subscribing business process. It is different for trading systems which require real-time pricing information versus clearing and settlement processing which requires security master details, corporate actions and client/counterparty information. In respect to reference data the industry has repeatedly tried to address the lack of a common uniform system of referencing securities with little success. A security master thus has to build into itself the capability to store these multiple identification schemes without being too rigid so as to allow flexible matching and future rationalization.

While the challenges surrounding a security master is central to an enterprise data management project, the other sections like corporate actions also need close attention. They affect positions, payments and portfolios. Every year there are changes made to corporate actions and they have a significant impact on tax reporting duties that directly impact the bottom-line of the company. Correct analysis and automatic recalculation of strike price, contract sizes squarely depend on the accuracy, completeness and business rules associated with this data. Focus on operational risk, cost and regulatory compliance, all highlight the need for clean and reliable client and counterparty data. Research analytics and portfolio evaluation demand immediate market prices and complete historical price points. At the same time there must be adequate pricing tolerance built into the system so as to extrapolate missing pricing information.

Different types of data, their nature, transactional, transformational and storage demands must dictate not only the data model, processing rules and workflow but also the type of database. A golden source does not mean a single storehouse. Rather for an enterprise data management system this must translate into single storehouses for certain categories of data, where the distinction is made based on the data parameters such as type, usage, volume and consumption demands. This also allows for simplification of the data model, distribution and publication systems.

3 Under the Microscope

Amongst the various pieces of different asset classes, their issuers, account details, limits and positions, fees and charges, corporate actions, clearing and settlement, tax details and accounting in an enterprise data management project, asset classes is the biggest and most complex.

To address the various attribute sets that must be stored for different instruments, the data model must capture all asset classes in a hierarchy with defined domains, classes and sub-classes. A basic guideline for structuring the base model is Classification of Financial Instruments, ISO 10962. However, the depth of these instruments must be determined in far greater detail before setting out to identify attributes needed to completely define them for use. An example of such a classification in the domain of fixed income instruments is depicted in exhibit 2.

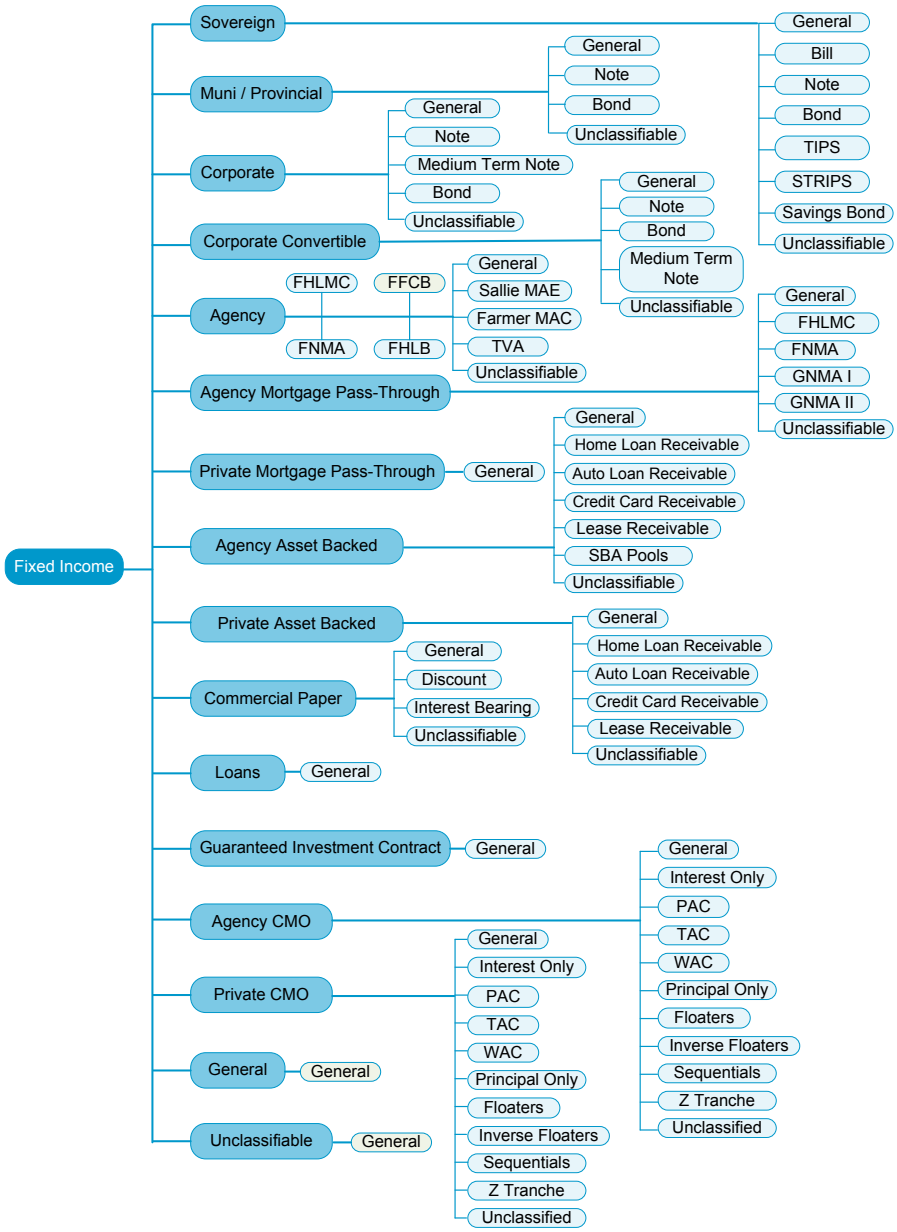
There has to be a single place that lists all the available domains in the data model and this is the Issue table. The issue table or related tables must be capable of storing information on the complete list of assets such as equity, debt instruments, options, futures, rights and others. For an enterprise data management system dealing with all types of asset classes, the issuer and issue relationship will have to be zero to many because certain assets such as metals or currencies do not have issuers. Consequently the issue table while storing an issuer reference cannot make it mandatory. A single table will not be able to completely define the issue. As complexity of securities increase a multitude of tables will be needed to uniquely reference an issue.

In the absence of an international issue identification scheme that is followed uniformly globally, a unique system generated identifier can be used for internal identification of the issue. Even then, all possible industry identifiers for the security must have corresponding table attributes as well as a means of storing identifiers in the old and new systems. Rather than store all issue information in a single table, they can be split into smaller tables such as Issue Type, Issue Description and Issue Terms. This segregation allows the Issue table to store information that is either uniform for all issues or is critical in uniquely referencing an issue.

Issue tables also can have information that is not necessarily in the public domain such as exchanges. These could be internal instruments such as real estate which must be cloaked from departments other than the one owning it. Building authorization over the information layer is a crucial step in architecting the data backbone of the enterprise.

To handle the different set of deal terms and conditions for various issues, a separate table such as Issue Terms and Characteristics can be modeled. This table will contain that set of parameters that lack any common ground among various securities. Identifying these set of attributes after taking out the common ones and containing them in a single table helps in targeted complexity management. For equities the attributes would include the likes of Dividend Type, Reinvestment

Exhibit 2: Charting Asset Class Hierarchy (fragment view)



Information and Voting Rights; while for Currencies they would include the likes of ISO Currency Code, Currency Symbol and Decimal Precision. For equities and debt instruments the details can be further segregated into separate tables that record issue underwriter or guarantor details.

With asset backed instruments the data needs are more complex. There are a host of different instruments and their data definition varies significantly from those of other asset class or domains. Descriptive data items for security types such as MBS, agency strips, CMO, ABS, SBA pools will have fields such as pre-payment date, payment date, post payment date, income history, ratings history, original and

current weighted average coupon, adjustable rate data for ARM pass-throughs and value added calculations such as historical pre-payment speeds. These fields require corresponding attributes in the data model.

To chart this complete dataset an iterative approach is essential. Using an iterative approach can help you target a select set of asset classes and further narrow it down to certain geographies. Given the sheer volume and details of data these are controls that can keep the volume at a manageable level and facilitate prototyping and refining at the early stages. Different feeds, internal and external, contain value added or legacy fields which may not be required by any business process in the organization.

To further trim the set down the mandatory and optional fields need to be identified. A sample classification snapshot is provided in exhibit 3. These then have to be mapped across all available feeds to see how the inbound data can be transformed to fit the data model of the central repository. Thus starts the journey to a common standard for all data across the enterprise and for a golden source.

Exhibit 3: Charting Asset Data Details (fragment view)

	Instrument		
Data	US pass-through securities	CMO and ABS	Fixed Income
Accrual begin date	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Amortization reset dates	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Amortization factor	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Previous amortization factor	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
CMO floating rate indicator	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Collateral Description	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Coupon	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Current pool balances	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Previous pool balances	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
First payment date	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Maturity date	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Reset dates	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
REMIC flag	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O
Redemption flag	☐ R ☐ O	☐ R ☐ O	☐ R ☐ O

4 Bringing it Together

While individual data needs have to be given attention, it is essential to keep an eye on the forest as well. From a system capability standpoint, the following exhibit provides a suggested guide to check for certain attributes of the enterprise data management system.

What is crucial for success in establishing an efficient enterprise data management system is that it is in tune with the organization's own specific and unique needs. Identifying intricacies of the building blocks, the data itself and the needs of the various business processes using the data is just the start.

Exhibit 4: High Level System Capability Checklist

Core Attributes

- Ability to maintain current data along with 5 years of price volume and 10 years of other historic data.
- Rule setup capabilities for consolidating vendor data into a single view and for specifying vendor data overrides.
- Available for more than 99% of the time in a year.
- Ability to process high volumes; process and standardize data feeds within a short timeframe from receiving these feeds on a daily basis.

External Factors

- Deal with changes in reference data that change infrequently, as well as data that changes daily, viz. price/volume data.
- Market data feeds have missing or incorrect data and feed vendors occasionally supply corrected restatements. The system must be able to detect and consolidate such data.
- Cater to different access patterns of client applications such as selection criteria, compliance rules etc.
- Ability to deliver during peak periods without failure. Adequate backup has to be in place as CPU's, servers and networks are liable to fail.

Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

www.capgemini.com/capitalmarkets

About Capgemini and the Collaborative Business Experience

Capgemini, one of the world's foremost providers of consulting, technology and outsourcing services, enables its clients to transform and perform through technologies.

Capgemini provides its clients with insights and capabilities that boost their freedom to achieve superior results through a unique way of working, the Collaborative Business Experience™.

The Group relies on its global delivery model called Rightshore®, which aims to get the right balance of the best talent from multiple locations, working as one team to create and deliver the optimum solution for clients.

Present in 40 countries, Capgemini reported 2010 global revenues of EUR 8.7 billion and employs around 112,000 people worldwide.

Capgemini's Global Financial Services Business Unit brings deep industry experience, innovative service offerings and next generation global delivery to serve the financial services industry.

With a network of 21,000 professionals serving over 900 clients worldwide, Capgemini collaborates with leading banks, insurers and capital market companies to deliver business and IT solutions and thought leadership which create tangible value.

For more information please visit **www.capgemini.com/financialservices**

Capgemini Financial Services
A6400 Shafer Court
Rosemont, Illinois 60018
Phone (847) 384 6100, Fax: (847) 384 0500