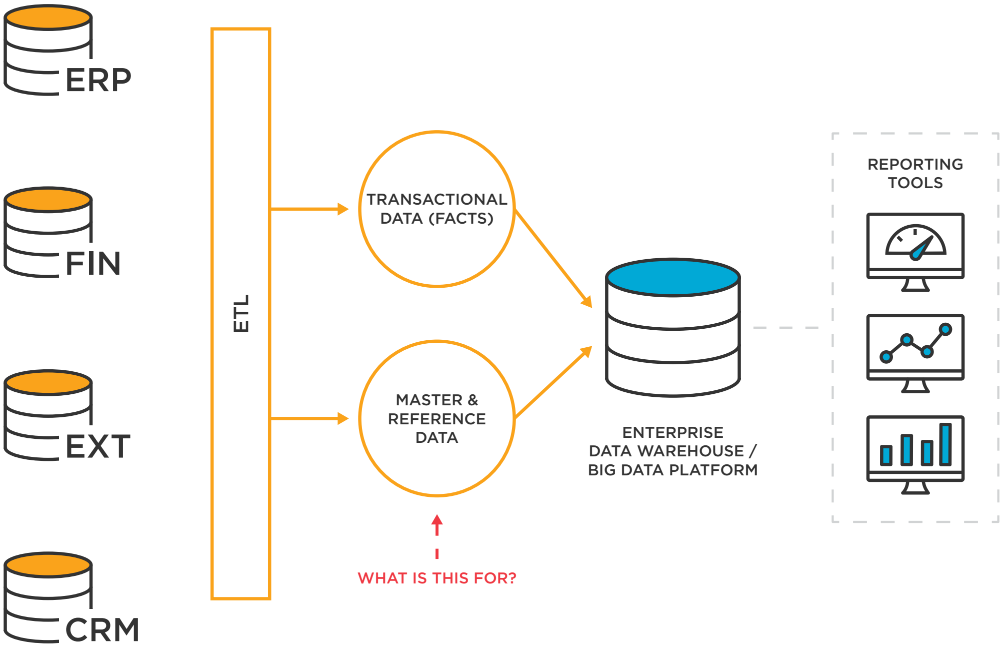
Enterprise Dimension and Hierarchy Management for BI and Big Data

INTRODUCTION

The goal of business intelligence and big data programs is to provide more timely analysis to help organizations improve their understanding, planning, and performance evaluation processes.

The architecture that supports these processes is generally understood by the teams that use it. They recognize that enterprise data warehouses (EDW) or big data platforms, such as Hadoop or DFS, will hold all the data. Business and operational intelligence tools will process that information and deliver the analysis in the form of visualizations, real-time dashboards, reports, and drilldowns to business users. Lastly, data integration technologies, such as ETL, will move data from your organization's systems of record into a data warehouse or big data platform.

However, when it comes to understanding the rationale behind managing enterprise dimensions, attributes, and hierarchies, many people are at a loss. This is a problem because neglecting the management of enterprise dimensions, attributes, and hierarchies will result in erroneous analysis. Fostering a better understanding of this problem, its issues and solution approaches, is the purpose of this paper. It in two parts: first, we highlight the importance of major challenges in enterprise dimensions and hierarchy management. And then, we describe where our solution, TIBCO EBX software version 5 can help.


ERP
FIN
EXT
CRM
ETL
TRANSACTIONAL DATA (FACTS)
MASTER & REFERENCE DATA
WHAT IS THIS FOR?
ENTERPRISE DATA WAREHOUSE / BIG DATA PLATFORM
REPORTING TOOLS

ENTERPRISE DIMENSION MANAGEMENT

One of the classic issues in data warehousing and business intelligence is how to deal with non-conformed (or inconsistent) dimensions, such as customers, products, locations.

Non-conformance exists because the underlying systems of record define or use dimensions in different ways; or the common dimensions are missing necessary classifications and attributes.

MANAGING MISMATCHING (NON-CONFORMED) DIMENSIONS

With mismatching dimensions, the crux of the problem is that your systems of record have defined your dimensions and attributes differently. Before the collected facts can be used in analysis, all dimensions and attributes need to be standardized and conformed. Using the facts before standardization runs the risk of producing inaccurate or erroneous analysis because you're comparing apples and oranges. Complicating the problem is that there may be a business reason behind the mismatch.

Here's a simple example from marketing. Sales and marketing want to understand how effective your industry-focused marketing programs have been at generating new opportunities. You need to be able to connect the industry-coded events to your prospects and closed deals. The problem is that while all those dimensions may have an industry attribute, marketing is using a public standard such as standard industry codes (SICs), while sales is using an internally created industry classification scheme.

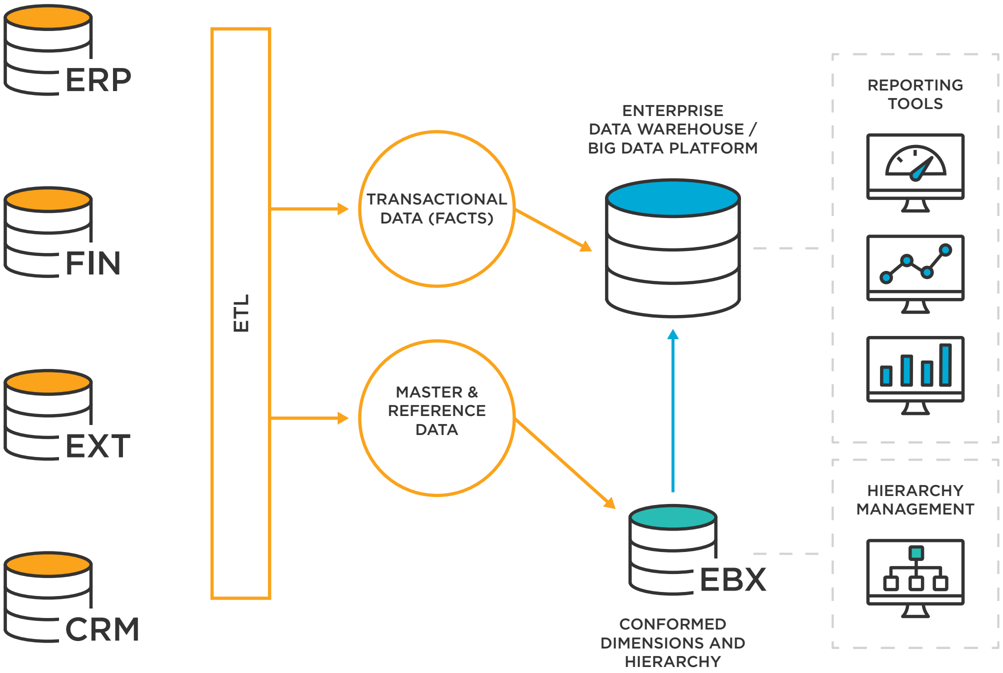
A common reaction, especially among data governance enthusiasts, is to urge standardization, but it may very well be that the attributes and dimensions are different for important reasons. In one trivial example, the marketing team may be using SIC because that's how all their vendors segment audiences. While the sales team may have decided to define their own classification because SIC was too complex.

Another less trivial illustration occurs in organizations that sell through distributors. Often enterprise applications define customers using different levels of granularity. For example, the CRM and customer service system may define customers by the sold-to, ship-to, or installed location address, with one record for each site. Financial systems may define customers by a bill-to contact or the payer at a site. This creates a situation where a single sold-to customer in the CRM maps to multiple payer customers in the ERP simply because the end-user may have purchased products directly and through several distributors over time.

What this means is that the analytical team will require a bridging mechanism, such as crosswalks or mappings that standardize between the different definitions so that each system of record can maintain its local definitions while supporting cross-divisional efforts. While it is technically possible to encode these mappings outside of a governed process (Excel/spreadsheet mappings), without any sort of auditable review and approval process you run the risk of quietly introducing mischaracterizations that are virtually impossible to find at the consolidated level.

ENRICHING DIMENSIONS WITH CLASSIFICATIONS AND EXTENDED ATTRIBUTES

In this situation, your systems of record may not persist all the classifications and attributes the business requires. Most of the time this happens because the system of record does not require maintenance of those attributes to meet its objectives. Take a product dimension maintained in an ERP. How critical are the brand management attributes in a supplier management context?


ERP
FIN
EXT
CRM
ETL
TRANSACTIONAL DATA (FACTS)
MASTER & REFERENCE DATA
ENTERPRISE DATA WAREHOUSE / BIG DATA PLATFORM
EBX
CONFORMED DIMENSIONS AND HIERARCHY
REPORTING TOOLS
HIERARCHY MANAGEMENT

The system of record may not have all the classifications and attributes you need.

It may seem like the simplest solution is to extend and manage attributes in the primary system of record, but that won't work for your users. Depending on the context, they may need different attribute values altogether. For example, market segmentation attributes—market, segment, sub-segment—when used in a performance context will reflect your products' current segmentation. Whereas, the very same attributes may hold different values if brand management is engaged in planning and analysis and they'd like to examine reorganizations or variations of the product lineup.

This illustrates the need for a separate system that business users can utilize to enrich dimensions before they're loaded into the EDW or big data store. One key requirement will be a mechanism to support the planning and analysis contexts with not just new attributes, but multiple alternate attribute value sets and temporary extensions.

HIERARCHY MANAGEMENT

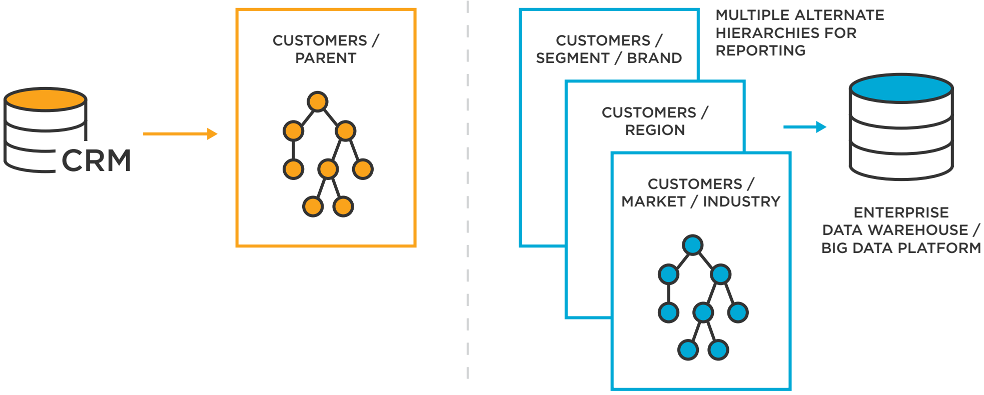
Hierarchies are often described as a system or organization in which elements are ranked, one above the other, according to status or authority. In general, hierarchy management requires that the levels and nodes are both accurate and in the right place.

However, in analytical context, hierarchies have unique requirements depending on their intended purpose. For example, in reporting, each level in the hierarchy represents a reporting level. Higher levels in the hierarchy are rollups or consolidations of their children. You can see this clearly in a P&L or income statement where all the cost and revenue lines, organized by type, roll up to each immediate parent, eventually summing to total cost and total revenue.

On the other hand, especially in research and planning contexts, hierarchies are used for classification. Your marketing organization may build multiple versions of the product hierarchy that represents the products by industry, by target customer size, by geography. All that said, the challenge in hierarchy management—reporting or classification—is not to generate "the hierarchy" but to support the need to create, maintain, (and in some cases distribute) alternate versions, or alternate hierarchies.

ALTERNATE HIERARCHIES IN REPORTING

In a performance reporting context, where the hierarchies represent rollups, alternate hierarchies are a different set of roll ups. By aggregating the data differently (and as long as each level or dimension is conformed), the user gets a different perspective on the same information.


CRM
CUSTOMERS / PARENT
CUSTOMERS / SEGMENT / BRAND
MULTIPLE ALTERNATE HIERARCHIES FOR REPORTING
CUSTOMERS / REGION
CUSTOMERS / MARKET / INDUSTRY
ENTERPRISE DATA WAREHOUSE / BIG DATA PLATFORM

Business teams need alternate points of view.

For example, a business user might define hierarchies to analyze revenues and growth by customer, by brand, by industry, and by region to determine where investment should rise, remain stable, or decline from a product line, industry, or geographic point of view.

In this context, each level within the hierarchy is a conformed dimension. The ability to enrich each dimension with new attributes and classification schemes, as described in the previous section, provides additional perspectives for analysis. And the lowest level, or leaf nodes, would be atomic elements such as a ledger account.

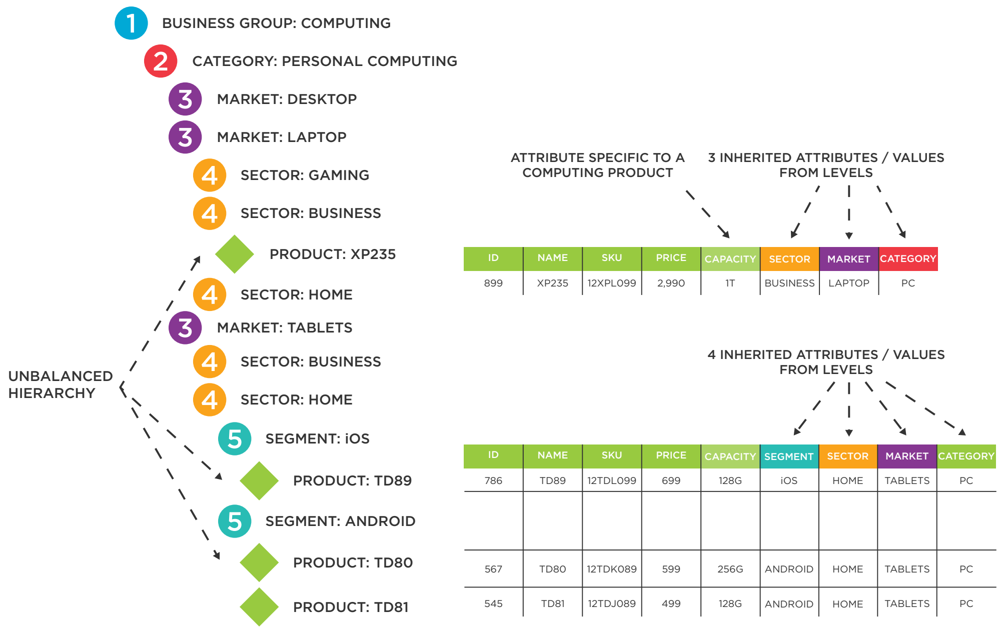
Under these assumptions you can see a couple of interesting issues unique to the reporting context. First, because these hierarchies represent costs, revenues, profits, there's a mutual exclusivity constraint on the leaf nodes. Unadjusted repeating leaf nodes that appear under multiple dimensions create double counting problems. Second, a reporting hierarchy and its alternate hierarchies will need to roll up to the same account. This means there is a comprehensiveness requirement. Every ledger account used in the original reporting hierarchy must be accounted for in every alternate; otherwise, it's not an alternate hierarchy but a hierarchy that represents something entirely different. And this brings us to our final point: how do the alternate versions deal with changes in their source hierarchy? For example, in many organizations, the finance department owns the performance hierarchies that are the source of accounts for tax and risk. What happens to those downstream hierarchies as finance adds and removes accounts? How do they synchronize?

Supporting alternate reporting hierarchies means that the team must have a system that facilitates the creation of user-defined, or explicit hierarchies, that can be organized and reorganized to create different rollups and consolidations. At the same time, the program team will need a mechanism to enforce those business rules around mutual exclusivity and comprehensiveness to ensure that the hierarchies are consistent with their alternate cousins. And finally, there will be a need to govern (and audit) the process of maintaining consistency among all the various alternate hierarchies.

ALTERNATE HIERARCHIES IN CLASSIFICATION

Classification hierarchies (also known as taxonomies) are used to divide elements into finer and finer subdivisions. In some classification exercises, elements are sorted into specific groups because users believe they have attributes in common with the other members of that group. For example, in classical systems such as the Linnaean taxonomy, Canidae and Felidae have been grouped together in Carnivora because they share the same sort of diet (meat eaters). Also peculiar to classification hierarchies is attribute inheritance. Attributes that belong to a parent are almost always inherited by their children. Because Canis lupus familiaris belongs to Canis lupus, which is a type of Canidae, and are themselves members of Carnivora, we know that c. l. familiaris eats meat.

We can see this kind of behavior in commercial classification schemes, such as product hierarchies. In the example beneath, the product inherits attributes from its parents higher up in the hierarchy, such as segment, sector, market, category, and business group. But what you'll notice in the example is that inheritable attributes may differ between groupings. Here, because the product TD89 is attached to the iOS segment, it inherits an additional attribute that does not exist for product XP235. Also our example hierarchy is unbalanced because products (the leaf nodes) can be attached to different levels of the hierarchy.

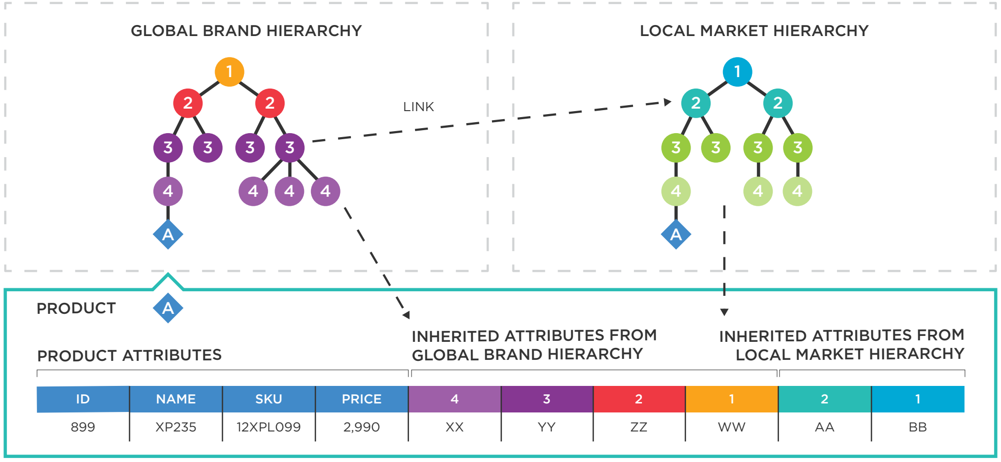


ID	NAME	SKU	PRICE	CAPACITY	SECTOR	MARKET	CATEGORY
899	XP235	12XPL099	2,990	1T	BUSINESS	LAPTOP	PC

ID	NAME	SKU	PRICE	CAPACITY	SEGMENT	SECTOR	MARKET	CATEGORY
786	TD89	12TDL099	699	128G	iOS	HOME	TABLETS	PC
567	TD80	12TDK089	599	256G	ANDROID	HOME	TABLETS	PC
545	TD81	12TDJ089	499	128G	ANDROID	HOME	TABLETS	PC

Attribute inheritance does not mean that all elements share the same attributes.

One challenge with classification hierarchies is that there really isn't a "right" set of attributes to underpin a grouping exercise (meaning that there can be an unlimited number of alternate hierarchies). While children in a classification hierarchy will always inherit attributes from their parents, if we link two hierarchies together, should the children inherit attributes from each set of hierarchy-specific parents? For example, if you have a global product classification hierarchy and your local teams "attach" their hierarchy to the global scheme, there may very well be a reason for your products to inherit attributes from both the global and local hierarchies. Having access to the union of all attributes from multiple classification hierarchies provides additional ways to slice and dice the information.


GLOBAL BRAND HIERARCHY
LOCAL MARKET HIERARCHY
LINK
1
2
3
4
A
PRODUCT
PRODUCT ATTRIBUTES
INHERITED ATTRIBUTES FROM GLOBAL BRAND HIERARCHY
INHERITED ATTRIBUTES FROM LOCAL MARKET HIERARCHY

ID	NAME	SKU	PRICE	4	3	2	1	2	1
899	XP235	12XPLO99	2,990	XX	YY	ZZ	WW	AA	BB

Linking hierarchies for consolidated reporting.

For the program team this means seeking out hierarchy management solutions that can support the development of multiple kinds of hierarchies, such as reporting and classification. There needs to be a mechanism to add business rules, such as leaf node exclusivity constraints and control comprehensiveness. Finally, inheritance is required to support any business requirement where classification is on order.

MANAGING MULTIPLE VERSIONS OF DIMENSIONS AND HIERARCHIES

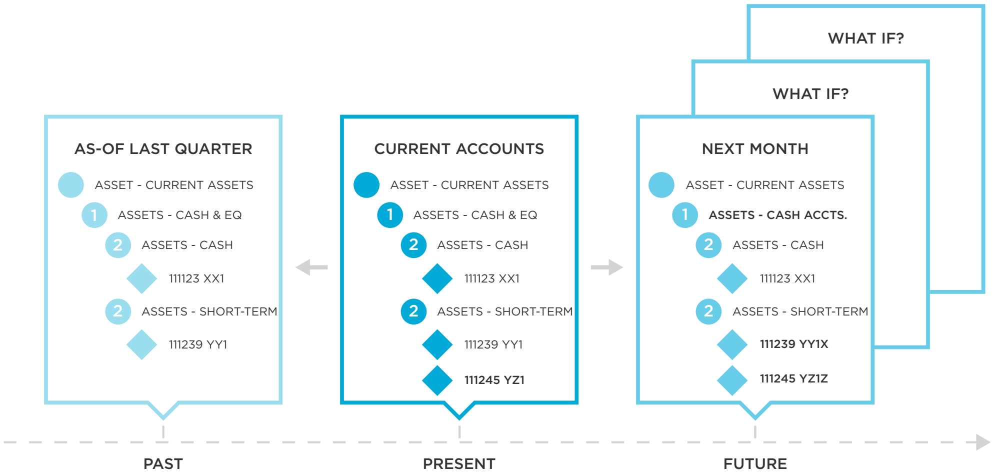
As we pointed out in the previous sections, your standard and alternate enterprise dimensions and hierarchies will need to be managed over time. However, this doesn't just mean support for prior, or as-of, versions, it also means providing the ability to craft hypothetical or what-if versions. Support for future versions transforms the BI and big data program from one of mostly performance analysis (looking back), to one that adds a data-driven approach for evaluating the fitness of new strategies and tactics.

AS-OF: MORE ACCURATE PERFORMANCE REPORTING

Business teams need to be able to rollback to the past and compare points in time for their dimensions and hierarchies. This is especially important in performance contexts since you want the same classifications and hierarchies applied to the facts being evaluated. For example, a user wants to compare revenues for the past five years. Because the chart of accounts (a hierarchy!) has changed considerably over the analysis period, the report needs to be based on a consistent version of the account hierarchy. It may also be important for the analyst to understand not just what has changed, but also, how those changes connect to each other across time.

When we consider your enterprise dimensions, the structure and values need to be versioned. On the structure side this includes keeping track of the user-defined extended attributes in addition to the core attributes from systems of record. From a historical, or as-of, point of view, having access to prior versions of your attributes and dimensions will provide not only an audit trail for governance purposes, but also information about how your facts were characterized, or classified, at a specific point in time. Understanding how attributes evolved is of critical importance in longitudinal, or time-series, analyses.

Focusing on hierarchies, it's also recommended that you version both the hierarchy structure and the values of the leaf-nodes and dimensions. The hierarchy structure represents both the ordering of the levels in your hierarchy (relationships between levels) and the placement of the leaf-nodes within the hierarchy itself (relationship between the elements and their parent level). From a values perspective, this is really no different from the versioning requirement of enterprise dimensions and attributes.



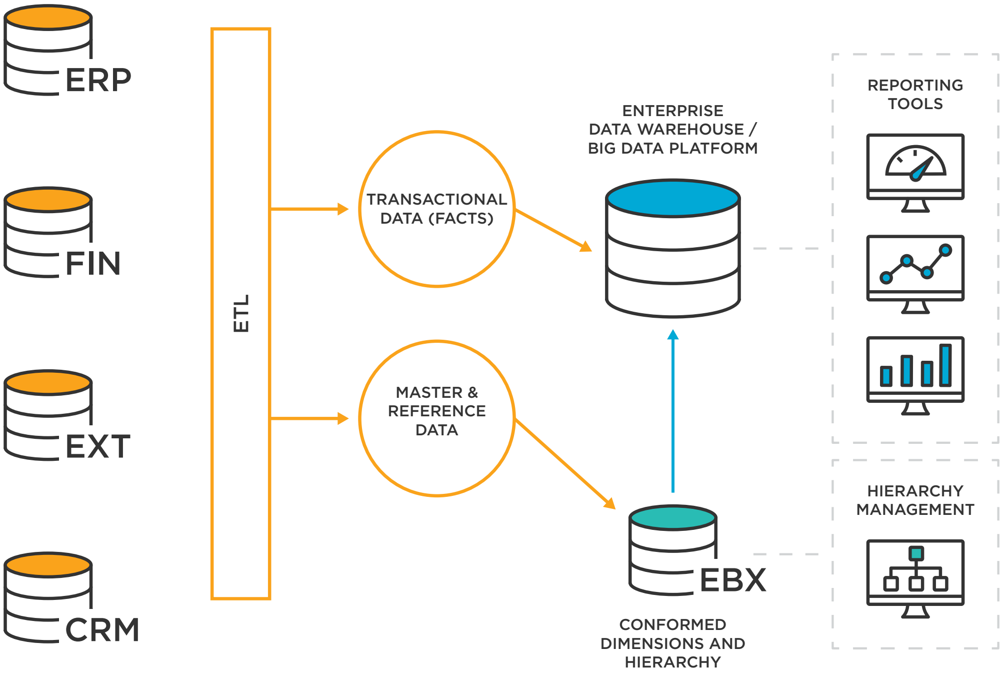
WHAT-IF: TESTING NEW BUSINESS SCENARIOS

Often BI and big data teams focus on as-of versions because of the time-sensitive nature of reporting performance to managers, regulators, and investors. However, we would argue that the need to support your business users with a sandbox for analyzing new business ideas and scenarios is equally important.

What this usually entails is providing the ability to create a new version from existing versions that can be modified by the users. Brand managers might apply a new organization of the product hierarchy or add new segmentations to see what a proposed strategy would look like using current fact data/reporting. In some cases, regulatory/government affairs or legal might apply their regulators' proposed definitions to the existing roster of asset/liabilities/products to understand the magnitude of change facing the company. Also, applying the what-if version of enterprise dimensions and hierarchies to existing data can be useful when establishing models that require historical data for calibration.

INTRODUCING TIBCO EBX SOFTWARE

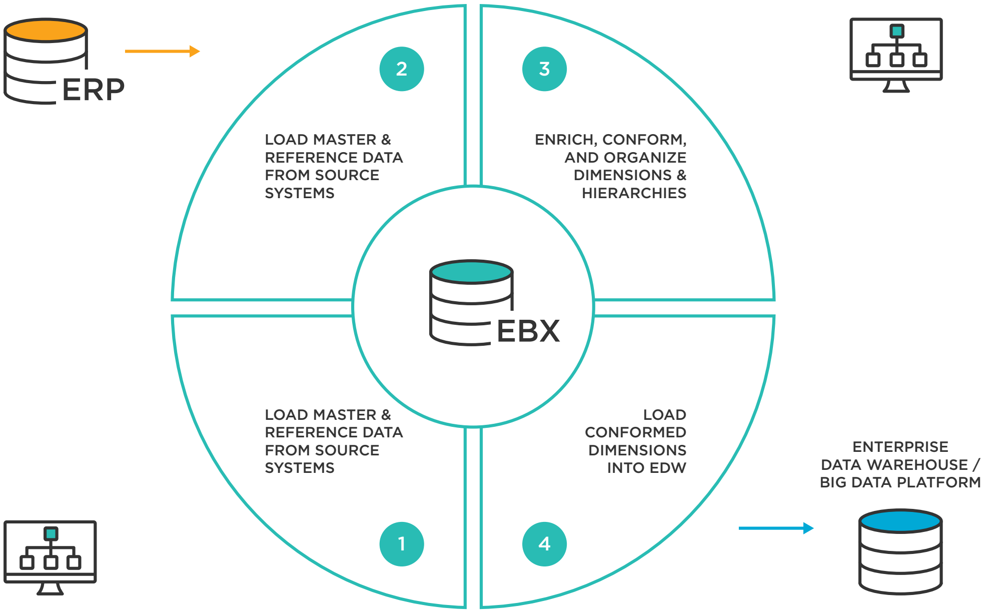
EBX software is a comprehensive and user-friendly solution for governing enterprise dimensions, attributes, and hierarchies—or analytical master data. It's comprehensive because it includes all the capabilities you need in one software solution. Our easy-to-use and easy-to-configure browser-based interfaces are designed to support business users who must work together to enrich, standardize, and create enterprise dimensions and hierarchies.


ERP
FIN
EXT
CRM
ETL
TRANSACTIONAL DATA (FACTS)
MASTER & REFERENCE DATA
ENTERPRISE DATA WAREHOUSE / BIG DATA PLATFORM
EBX
CONFORMED DIMENSIONS AND HIERARCHY
REPORTING TOOLS
HIERARCHY MANAGEMENT

HOW IT WORKS

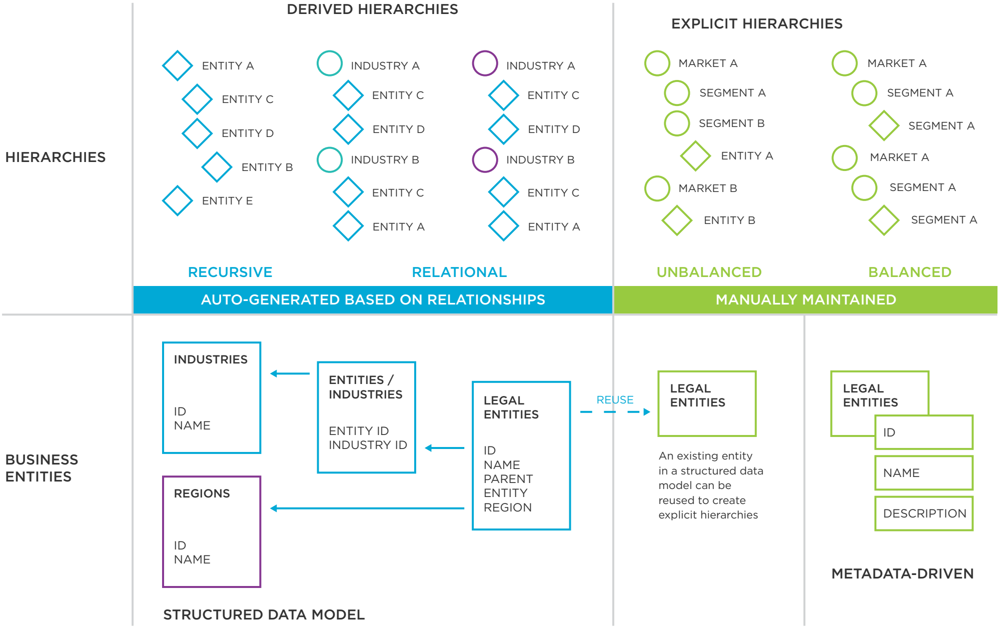
Our software fits seamlessly into your business intelligence architecture and sits between your data integration layer and data warehouse, or big data platform. Instead of manipulating your dimensions and hierarchies in the systems of record, data integration layer, or in the warehouse, your business teams adjust, enrich, or create dimensions and hierarchies in audited workflows inside the software.

All this helps ensure that your analytical master data is subjected to a high degree of data governance. Also, your teams can leverage our unique capabilities to design and manage complex hierarchies.


ERP
2
LOAD MASTER & REFERENCE DATA FROM SOURCE SYSTEMS
3
ENRICH, CONFORM, AND ORGANIZE DIMENSIONS & HIERARCHIES
EBX
LOAD MASTER & REFERENCE DATA FROM SOURCE SYSTEMS
1
LOAD CONFORMED DIMENSIONS INTO EDW
4
ENTERPRISE DATA WAREHOUSE / BIG DATA PLATFORM

COMPREHENSIVE HIERARCHY MANAGEMENT SUPPORT

As we described in the previous section on hierarchies, there are many different use cases for hierarchies ranging from data exploration to classification to reporting. The software supports all those use cases because, on a technical level, it supports both derived and explicit hierarchy types.



Support for derived and explicit hierarchies in one solution.

Derived Hierarchies

Derived hierarchies are generated from the relationships that exist in a data model. In the example above, legal entities are linked to industries and regions, as well as to themselves (recursive relationship). Based on those three relationships, EBX software can dynamically render three hierarchies that organize the entities by parent (recursive) or by region and industry.

Explicit Hierarchies

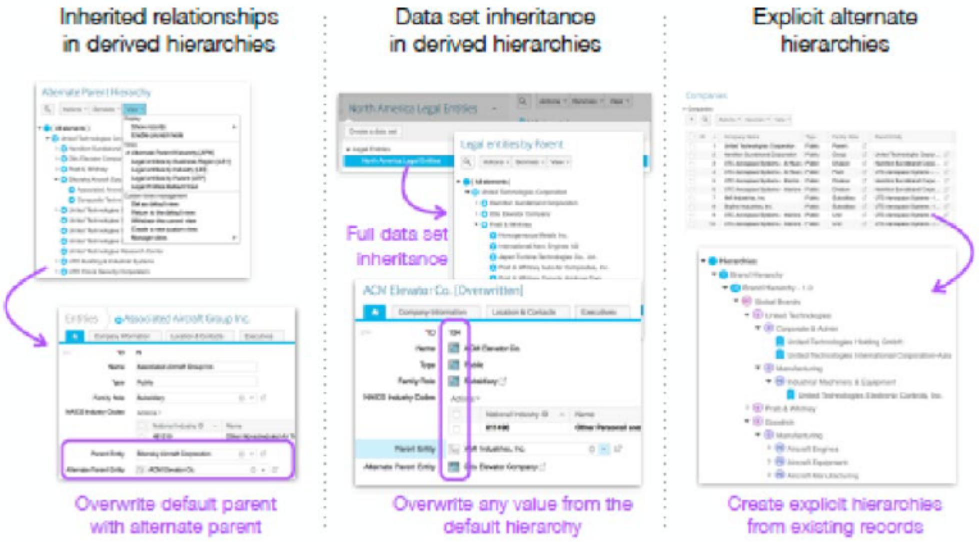
Explicit hierarchies are created by business users. With EBX software, users craft their hierarchy by explicitly defining levels and attributes. Explicit hierarchies can be balanced or unbalanced. We commonly see explicit hierarchies in classification and reporting contexts.

Connecting the Two Together

One unique aspect of EBX software is the ability to reuse elements from your derived hierarchies into your explicit hierarchies. In the example above, the legal entity dimension can be used as leaf nodes for your explicit hierarchies. This means that you can keep your hierarchies consistent with your evolving enterprise dimensions and each other.

ALTERNATE HIERARCHIES

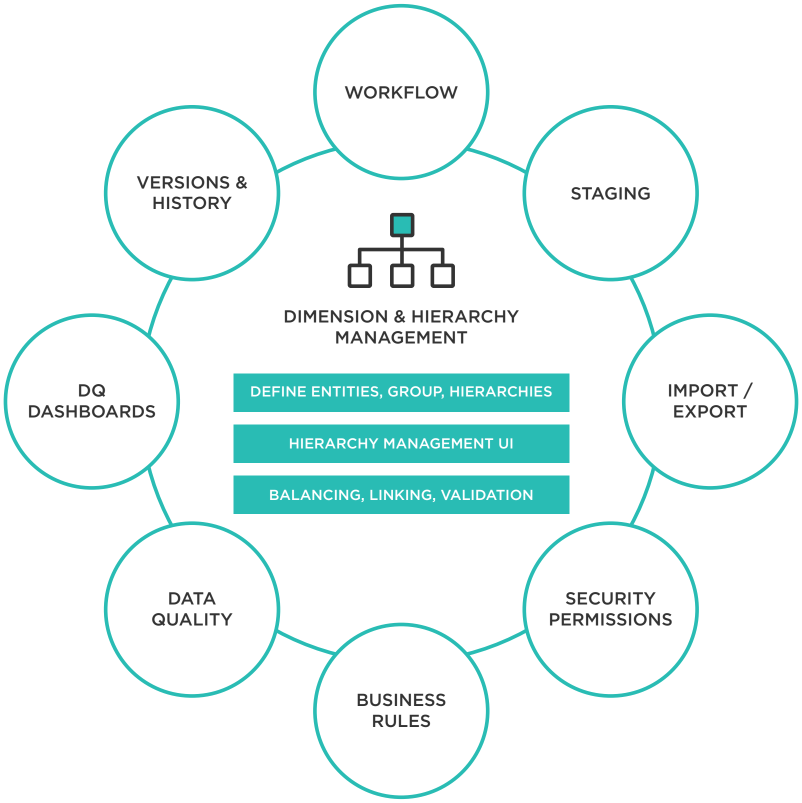
Finally, an important part of hierarchy management is the ability to support alternate hierarchies. It's a topic we covered at length in the previous sections. Today, one of the biggest issues is that, in most business intelligence programs, the end user must duplicate their hierarchies in order to define alternate versions. Creating copies means that hierarchies will be out of sync if there's change. For example, if a bank's risk management team creates the risk management hierarchy by duplicating the firm's chart of accounts, it will miss out on all the finance updates (hierarchy moves and account adds/updates/deletes). This means, rather quickly, the risk management hierarchy diverges from the financial hierarchy.


Inherited relationships in derived hierarchies
Data set inheritance in derived hierarchies
Explicit alternate hierarchies
Full data set inheritance
Overwrite default parent with alternate parent
Overwrite any value from the default hierarchy
Create explicit hierarchies from existing records

With EBX software, there are multiple ways to create alternate hierarchies without any data duplication. Here are two examples:

Using EBX software inheritance, you can create a child of your hierarchy. This child, by default, inherits all the relationships and values from its parent. The user selectively overrides the attributes and relationships they would like to change. For inherited elements, parent updates flow through to the child automatically, meaning the two are never out of sync. With explicit hierarchies it is possible to reuse your existing conformed dimensions to create a new hierarchy. These conformed dimensions can be new levels or new leaf nodes. In the example above, we're reusing the legal entity dimension as the leaf node of the hierarchy.

EBX FEATURES



EBX software is an enterprise-class solution for managing your dimensions and hierarchies. Key features for dimension and hierarchy management include:

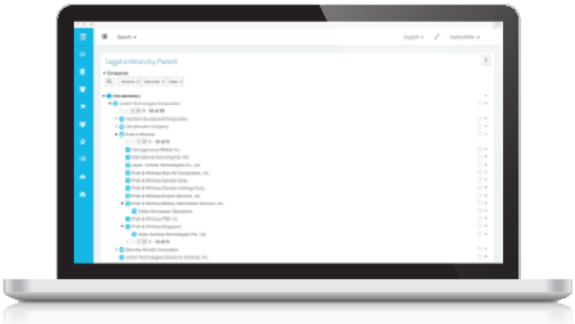
- Tools to define dimensions, attributes and hierarchies
- Support for all types: derived, explicit, balanced, and unbalanced
- Services to balance and link hierarchies
- Versioning to manage past, present, and future dimensions and hierarchies
- Inheritance to manage alternate hierarchies
- User interfaces to manage dimension, attributes, and hierarchies
- No-code, browser-based UIs
- Built-in full text and fuzzy searching and filtering
- Perspectives and custom layout to configure role-specific UIs
- Services for integration and distribution to downstream systems, ad-hoc analysis tools, data warehouses, and big data platforms

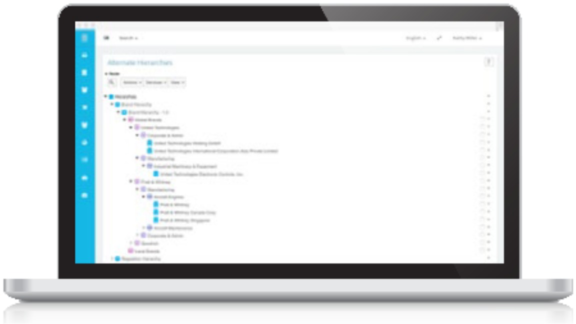
In addition to its dimension and hierarchy management capabilities, EBX software provides a set of advanced features to manage, share, and govern data, including:

- Collaborative workflow to support your governance and change management processes
- Business rules to enforce controls and validation
- Data quality engine to cleanse and match data
- Dashboard to track data quality and program performance
- Fine-grained security and permissions

CAPABILITY HIGHLIGHTS

HIERARCHY MANAGEMENT: VIEW AND MANAGE COMPLEX HIERARCHIES





Browser-based
EBX software provides end users with a browser-based user interface for hierarchy management. Hierarchies can be presented in multiple ways (tables, trees, graphs) where users can:

- Browse, search, or filter hierarchies
- Perform updates to the hierarchy with single and multi-node adds, updates, moves, and deletes
- Import and export content from/to various formats including MS Excel and text

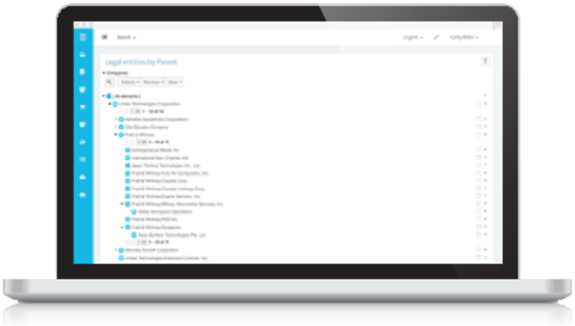
End-users can also craft sharable views that automatically apply pre-set filters and customizations of their on-screen display.

Support for complex hierarchies
EBX software has native support to manage complex hierarchies, like unbalanced and balanced hierarchies, with services to balance hierarchies, and customization tools to update level labels and icons.

In EBX software, especially for explicit hierarchies, classification attributes can be created at any level of a hierarchy. Each node can automatically inherit the attributes and values from upper levels in a hierarchy.

User control

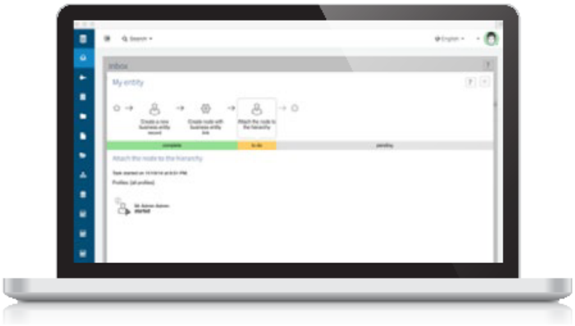
The user can define which inherited values are maintained or overwritten at any particular level.



COLLABORATIVE WORKFLOW: ENABLE DISTRIBUTED DATA GOVERNANCE ACROSS TEAMS

The built-in workflow in EBX software helps you design, execute, and monitor the governance processes for your enterprise dimension and hierarchies, for example, change management and approval. The built-in workflow includes a wide range of features including:

- Design of multi-step workflows
- User and server tasks
- Parallel or sequential approvals
- Notification, deadlines, reminders
- Workflow inbox for end users
- Dashboard for tracking workflow completion
- Reporting on active and completed workflow

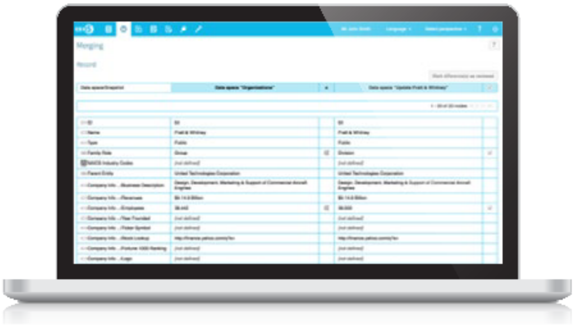


VERSION CONTROL: MANAGE AND CONNECT PAST, PRESENT, AND FUTURE EDITIONS

With EBX software, it is possible to manage and connect past, present, and future versions of your enterprise dimensions and hierarchies.

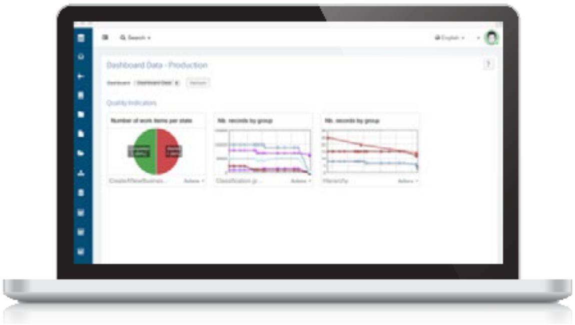
With snapshots, keep images of your dimensions and hierarchies at points in time in order to rollback or compare two points in time. In addition, EBX software keeps track of every single transaction, so it is possible to display or query a record-level history on each dimension.

Using data spaces, it is possible to create drafts or what-if versions of dimensions and hierarchies. What-if versions are completely isolated, meaning your finance team can work on the new legal entity hierarchy without affecting the existing legal entity hierarchy. These draft data spaces can also be used to create purely speculative or hypothetical hierarchies.



PERFORMANCE MEASUREMENT

EBX software also provides features that help teams engage in a process of continuous improvement. With reports and dashboards from our insight module, users can measure changes in quality and process performance over time. This data is an important piece of information when determining how and where to make improvements. Also, if you'd like to use your own business intelligence tools, all of the performance data is easily accessible.



Learn more about TIBCO EBX software at www.tibco.com/resources/demand-webinar/introducing-tibco-ebx

Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

TIBCO fuels digital business by enabling better decisions and faster, smarter actions through the TIBCO Connected Intelligence Cloud. From APIs and systems to devices and people, we interconnect everything, capture data in real time wherever it is, and augment the intelligence of your business through analytical insights. Thousands of customers around the globe rely on us to build compelling experiences, energize operations, and propel innovation. Learn how TIBCO makes digital smarter at www.tibco.com.