California Franchise Tax Board (FTB)

Enterprise Architecture
Business Intelligence (BI) Definition

Version No. 1.2

April 11, 2008

Author:
Enterprise Architecture Council

Document Information

Document Source

This document is controlled through Document and Deliverable Management. To verify that this document is the latest version, contact Enterprise Architecture.

Revision History

Version No.	Date	Summary of Changes	Revision Marks
1.1	02/05/2009	Removed references	JZ
1.2	02/05/2009	Title change	John R.

Table of Contents

List of Figures

1.0 Executive Summary and Charter

1.1 Overview

Business intelligence (**BI**) is a business management term, which refers to applications and technologies that are used to gather, provide access to, and analyze data and information about company operations and performance. Business intelligence systems help companies have a more comprehensive knowledge of the factors affecting their business, such as metrics on sales, production, internal operations, and they can help companies to make better business decisions. Three main components are reporting, data mining, and predictive analytics.

An aspect of Business Intelligence that is of special interest to FTB is modeling and scoring. The desire for improved metrics on efficiencies and outcomes for modeling requires data services that will provide access to enterprise-wide Business Intelligence information. Access to this critical information will provide a wider array of information to model on and will lead to better modeling results. It will also provide for:

- Less data redundancy
- Improved data inaccuracy
- Increased access to FTB Data
- More efficient use of taxpayer and third party information
- More consistent treatment across debt types.

1.2 Scope

The scope of the Business Intelligence Architecture Definition concerns analytic processing technologies that depend upon Data & Delivery Management Architecture Definition for the quality of data in both transactional (OLTP) and analytical data (OLAP) stores. The BI Architecture Definition is comprised of technologies and solutions that include data warehouse, data marts and federated data solutions that cater to reporting, data mining or predictive analysis needs. Tools and technologies to deliver business intelligence solutions that are applicable to FTB will be addressed. Since this Architecture Definition deals with the capture, architecture, storage and analysis of data for OLAP purposes, included in the scope are architecture, methodology and processes to create a enterprise data warehouse (EDW) and data marts.

1.3 High-level Requirements

The following table outlines the high-level requirements of the Business Intelligence Architecture Definition.

Figure 1.3-1: Business Intelligence Architecture Definition – High Level Requirements

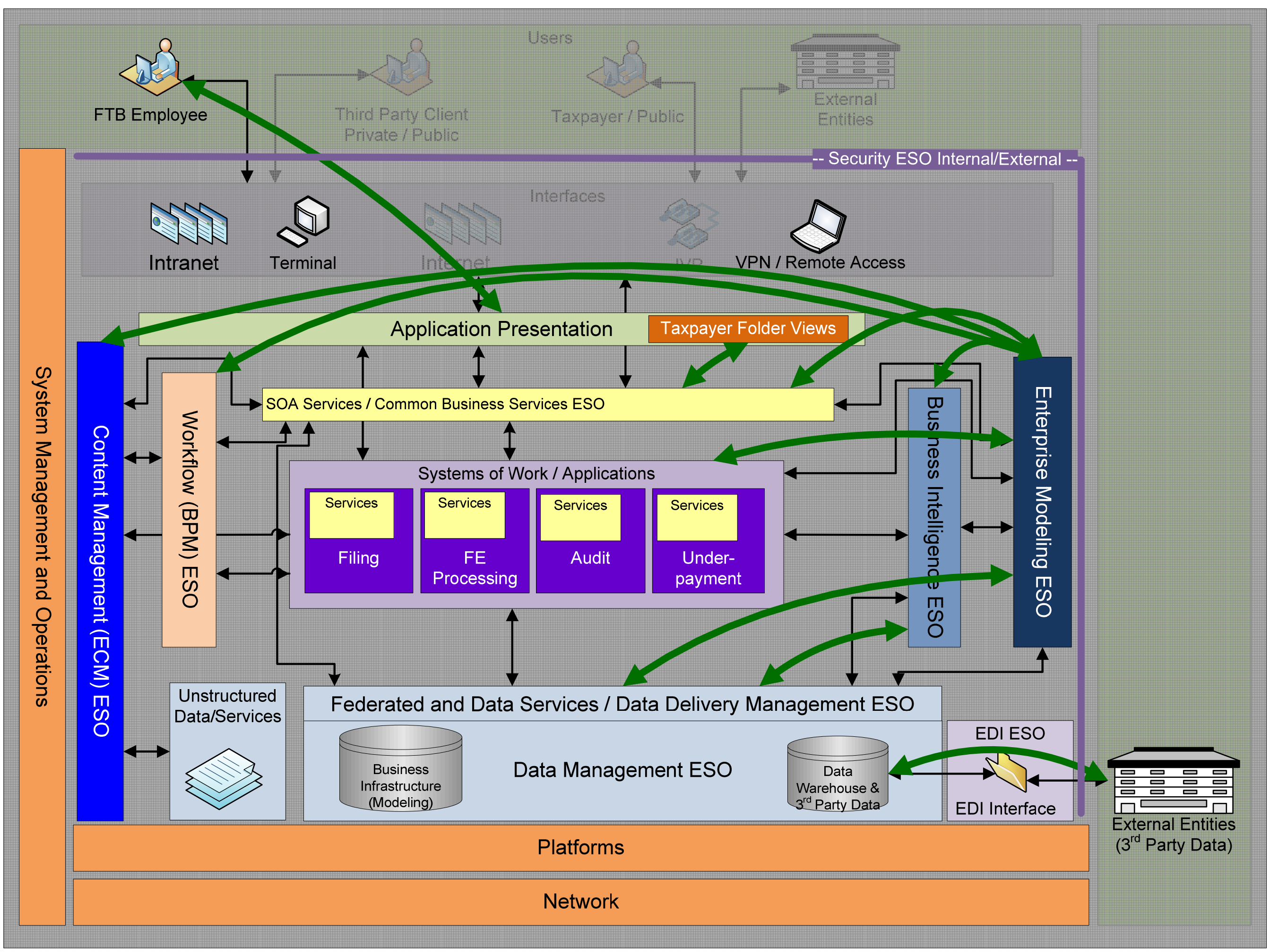
Requirement	Description
Capture historical data from all relevant OLTP data repositories	Business Intelligence is enhanced by the capture of historical OLTP data. OLTP data will be developed to include time values so that the OLAP data is prepared for historical data reporting and intelligence.
Provide suitable architecture for analytical processing of data	Business Intelligence is enhanced by strategies to enhance analytical processing. First is the data warehouse as the “trusted” data source for OLTP and OLAP needs. De-normalized relational data from the warehouse will be used in multi-dimensional structures for use in data marts etc. Also the use of ‘dashboards’ for proper business decision-making and management, and the use of federation. This also implies that a prudent mixture of ETL solutions and Federated solutions is used, and the criteria for using one or the other be specified.
Provide a process methodology for capturing, analyzing and delivering user requirements for business intelligence	A process for capturing, analyzing and then delivering the user requirements for business intelligence will be defined. This, along with the data architecture and models as well as tools help to define and implement the total business intelligence solutions as desired by the users.
Ability to address new requirements for reporting, data mining, modeling or analytical processing with agility and promptness	It is important for business intelligence solutions to address any requirements for analytical processing with agility and promptness as they arise, whether for modeling, reporting, data mining etc. This necessitates that a architecture be built that captures all relevant historical data, and has furthermore layers or tiers of data storage, from mere capture, to de-normalized structures to data marts, which can address these needs.
Provide Usage Analysis	Collecting and reviewing usage of the information provided by the BI platform is critical to determine the success the BI Architecture Definition. FTB will track usage on any exposed object, such as a report or dashboard and who is using it. This will allow BI to inform users of upcoming changes as well as inform management if certain areas of information delivery are not being utilized. This information will help determine the ROI to create and maintain the environment. Both IT and business partners will get a view of where the organization currently is delivering information well and areas that need improvement

1.4 Conceptual Model

The following diagram illustrates the portion of the Enterprise System Architecture that will interact with, supply information for, or is a dependency of the Business Intelligence Architecture Definition. The green arrows represent information flows between the Business Intelligence system, other systems, and to the FTB employee.

The diagram below illustrates how FTB employees interface with applications to perform Business Intelligence activities and how the business intelligence system interacts with common business services, System of Work services, and data services to not only acquire information for Business Intelligence, but to provide BI results to systems within the enterprise. Unstructured data from the Content Management System that does not exist within system databases, such as non-captured tax form fields, will provide additional information for Business Intelligence. Information generated through the Business Intelligence System, such as analyses of third party data will also contribute data to the enterprise. All of this information provides a more comprehensive view of the Systems of Work (SOW) lifecycle that can then result in more effective workflow processes. However, Business Intelligence at an enterprise level will require more data exchange across the enterprise and may require additional network and hardware resources in order to transmit the additional data.

Figure 1.4-1: Business Intelligence Architecture Definition – "To Be" Technical Architecture Model



2.0 Current Architecture

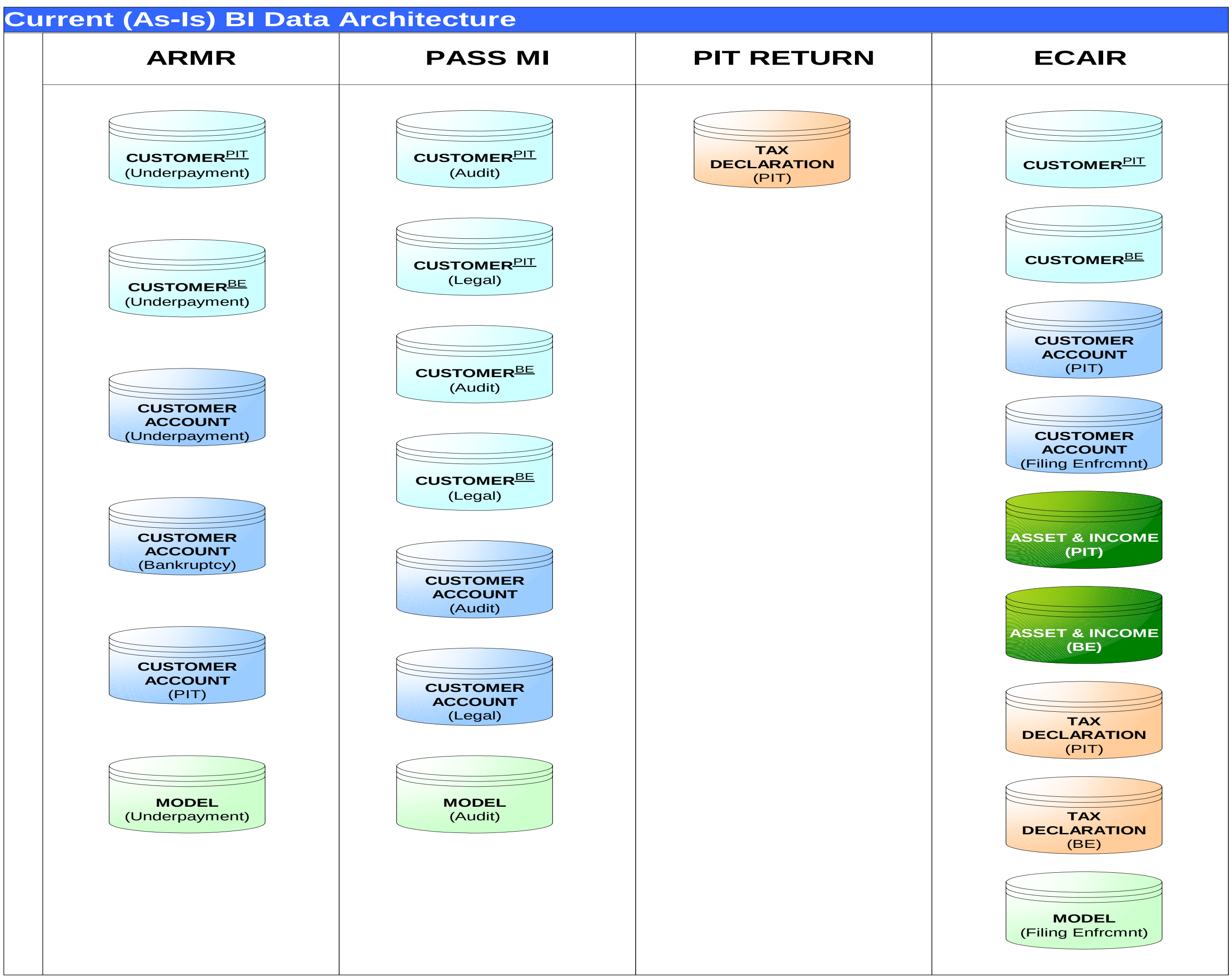
2.1 Current Capabilities and Components

BI is on a continuum of change from the monolithic or distributed data warehouses that defined BI's first generation to emerging trends in BI. The trends in BI are concerned with the analytics of data whether the data is in the operational space (Operational BI) or the reporting, analytical space (Traditional BI).

2.1.1 Current Data Architecture

Figure 2.1-1 illustrates the current BI data architecture environment at FTB. This data architecture employs the architectural subject areas, but stores the data redundantly for each of the BI systems. The data and the BI applications are siloed.

Figure 2.1-1: Current BI Data Architecture

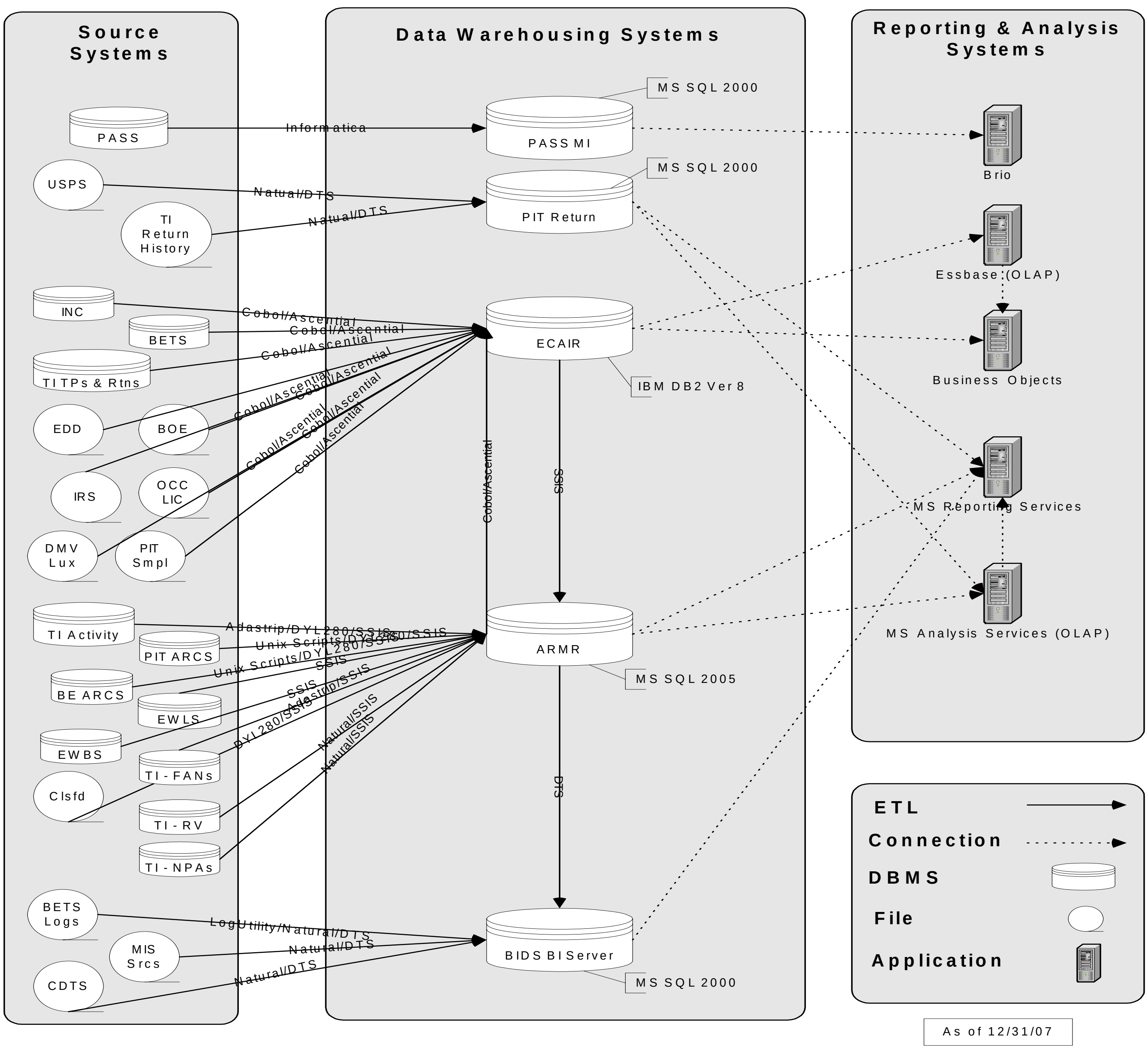


2.1.2 Current BI Technical Architecture and System Inventory

The current technical architecture for Business Intelligence parallels the data architecture and adds the process and delivery mechanisms for BI focused data.

Figure 2.1-2 illustrates the current technical architecture and environment of FTB's BI systems. The diagrams following the BI technical architecture depict the current technical systems inventory of the major BI systems at FTB. The diagrams show System Identity, User Community, Level of Usage, Data Content, Data Sources, Staff Resources, Hardware Resources and Software Resources.

Figure 2.1-2: Current BI Technical Architecture



2.2 Current Governance and Organization

The current BI organization and system architectures are tightly connected to the evolution of overall information technology services in the department since 1995. Prior to 1995, information technology services at FTB were "centralized" in one organization that serviced all FTB business areas. Until this time, information needs were primarily provided through mainframe applications that lent themselves to central control and management. The increasing use of personal computers and the emergence of distributed computing, coupled with a concern about the ability of the department to respond to rapid changes in technology for meeting business needs, prompted FTB to reorganize technology services into a "decentralized" model. Under this model, each major business area was responsible for developing and maintaining the specific technologies that supported their business activity.

Business activity influenced the emergence of BI at FTB as efforts to provide for reporting and analytical needs were tightly integrated with the deployment of each project. Just as each initiative was developed at different points in time, under different management, and in partnership with different vendors, so were corresponding BI efforts, for example:

- **ARMR** (Accounts Receivable Management Reporting System) implemented as a separate, parallel project, to ARCS in 2000.
- **PASS MI** (Professional Audit Support System Management Information) implemented as separate project, subsequent to PASS in 2004.
- **BETS Revenue Reports** implemented as a separate effort, subsequent to BETS, in 1997.
- **ECAIR** (Enterprise Customer, Asset, Income, & Returns) was implemented as the last phase of INC, in 2002.

These are not all of the BI systems developed during this period, but these examples illustrate the driving forces behind the existence the current "stove-pipe" BI systems and the efforts FTB is taking to transition to a more coordinated and cohesive architecture through organization and governance.

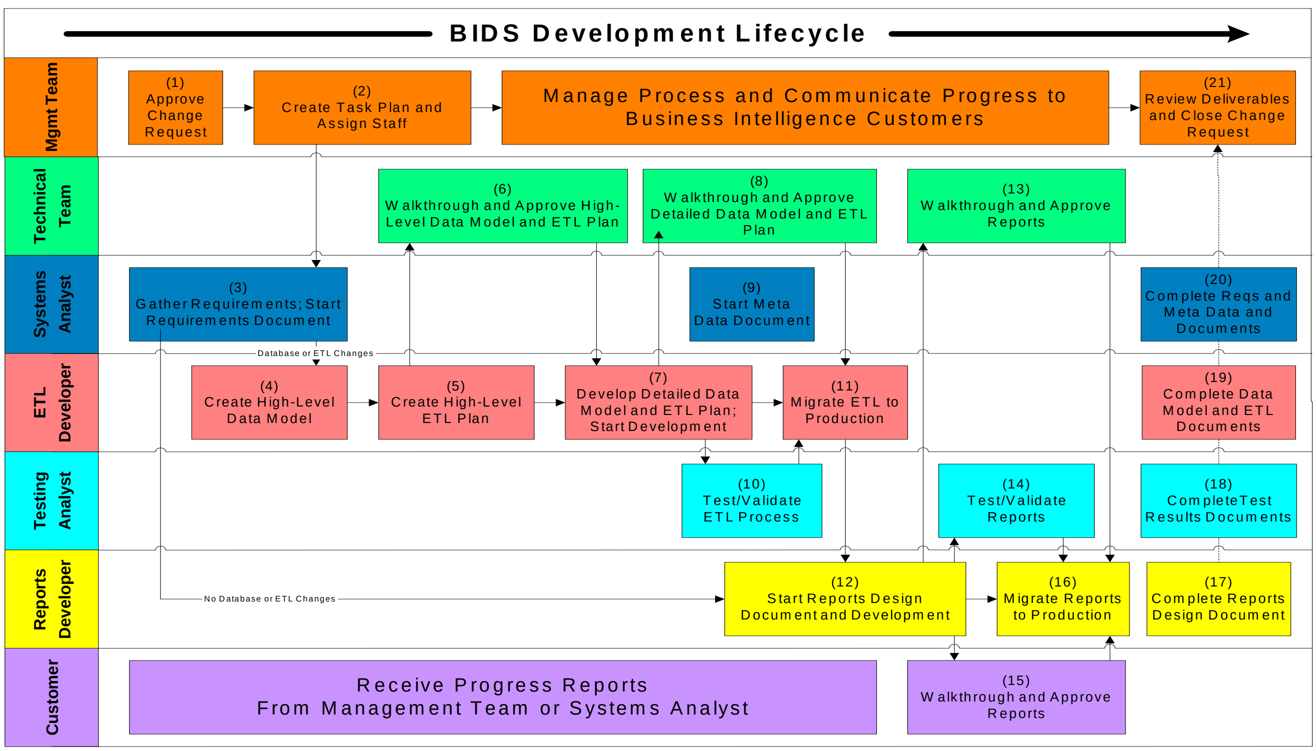
2.2.1 Business Intelligence and Data Services (BIDS)

The Business Intelligence and Data Services (BIDS) group was formed in early 2006 by consolidating most of the staff involved in supporting the analytical and reporting needs of FTB's business customers. One of the key goals of this consolidation was to examine the BI activities and practices and the technologies and tools for opportunities to reduce redundancies, standardize, and implement industry best practices. To minimize potential negative impact to customers, the staff was organized around the systems or activities they supported:

- **ARMR/PASS MI Unit**: This unit consists of analysts, developers, and testers that support ARMR and PASS MI.
- **ECAIR Unit**: This unit consists of analysts, developers, and testers that support the ECAIR data warehouse.
- **ADM BI Unit**: This unit consists of analysts and developers who support the PIT Return Data Mart, MIS Reports, and ad hoc reporting and other data services.

BIDS Development Lifecycle BIDS implemented a consolidated development lifecycle applicable to all BIDS development activity. BIDS considers a standardized development lifecycle key to improving their ability to effectively meet BI users' needs, increase the quality of data and products, and reduce potential redundancy and development costs. As part of this lifecycle, the BIDS Technical Team is empowered to approve or deny all proposed changes to the BI systems. This has helped the organization move away from the prior practices of development or system changes that ignored the impact to other BI systems or other development activities – practices the perpetuated "stove-pipe" systems. The following is a diagram of the BIDS Development Lifecycle:

Figure 2.2-1: BIDS Development Lifecycle



2.2.2 Other FTB Business Intelligence Activities

While BIDS is the primary provider of business intelligence reporting and analytics, other organizations within the department also engage in providing reports, analytics, and ad hoc support. These areas include:

Economic and Statistical Research Bureau – provides a wide range of reports and analytics relating to corporate and individual filing activity and demographics. The organization is responsible for overseeing the creation of annual samples of filing data that are critical for determining the potential impacts of changes in tax policy or trends in economic activity. The information is used for a variety of purposes, such as the department's Annual Report, news releases, revenue impact analyses, a variety of special studies, and providing answers to questions posed by other departmental units, other state departments, the Legislature, and the general public. Although the organization does not represent itself as a "provider of BI," many of their activities are clearly BI in nature.

Privacy, Security and Disclosure Bureau – develops security policies and procedures, including security measures for the protection of FTB's facilities, and to prevent, detect, and track unauthorized access to information technology systems, networks, and data. Some activity related to tracking data exchange agreements with third parties and analyzing system usage activity for potential violations of security policy are BI in nature.

Financial Management Bureau – in addition to other fiscal responsibilities, provides reports and analytical support relating to CALSTARS and Activity Based Costing (ABC). These activities are often considered to be within the BI umbrella.

Compliance Systems Bureau – provides some reporting and analytics related to various applications developed and maintained for audit and collection activity, including ad hoc support. BIDS often refers ad hoc support to this organization when requests actually need operational data associated to systems supported by this group. The organization is currently developing a BI application associated with the Court Ordered Debt Expansion (CODE) project.

Tax Systems & Applications Bureau – provides reporting and analytics relating to tax systems and web-based systems that support taxpayer activities through FTB's public website. Web statistics are collected using a commercial-off-the-shelf application and provided to users through FTB's intranet or through ad hoc support. BIDS also works with this organization when requests for data services need operational data associated to systems supported by this group.

Network Management Bureau – has responsibility for the design, implementation and day-to-day operation of the voice and data network infrastructure. Most notably, this organization has deployed a new call center system, Enterprise-Wide Customer Service Platform 2 (ECSP2), which includes a "Reporting Info Mart" that is a BI application focused on call center and IVR data.

2.2.3 Current Business Intelligence Users

Figure 2.2-2: Illustrates the type of user base for FTB BI systems

	Type of Users			
BI System	**Casual Users**	**Power Users**	**Developers**	**Total**
ARMR	500	12	5	**517**
PIT Return DM	50	11	4	**65**
ECAIR DW	20	10	5	**35**
PASS MI	100	3	3	**106**
Outside of BIDS	195	8	20	**223**
Total	**670**	**36**	**17**	**946**

3.0 Target Capabilities and Components

3.1 Target BI Data Architecture

A collaborative and cooperative target environment for Business Intelligence with no unplanned duplication and efficient use of licenses, staff and other resources requires three major requirements:

1. ***FTB will use single data inputs for key subject areas***
   a. *Supports the business rule to prevent unplanned redundancy. Data will not be replicated unless there is a business need to do so (such as performance or security needs).*
   b. *Requires each feed be assigned to a specific BI group. Responsibilities include: primary extraction of data, understanding and representing the data accurately, consulting as many stakeholders as possible to make sure that all needs of the enterprise are considered and met.*
   c. *Requires a Governance Structure*

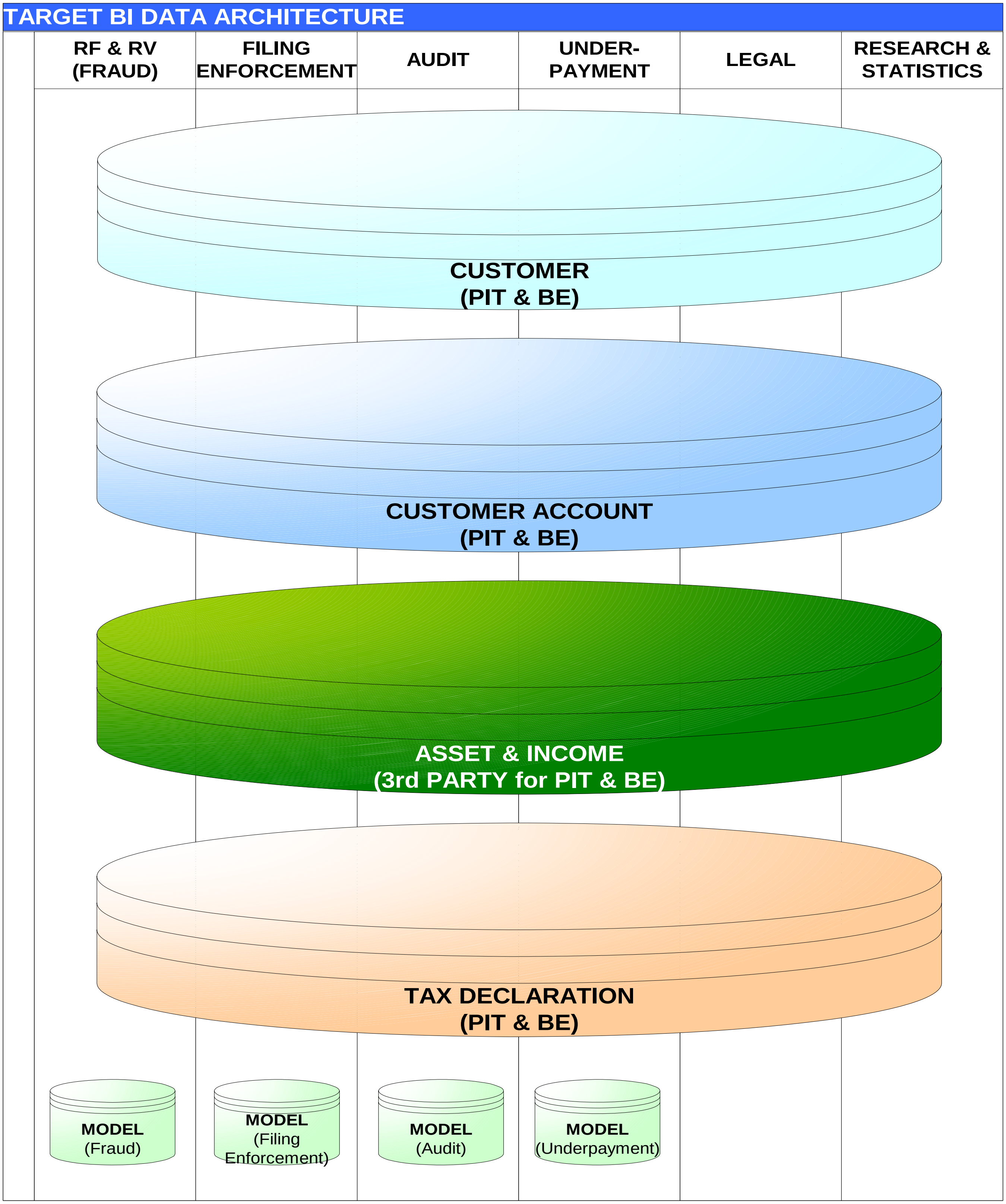
2. ***FTB shall use a Common Data Architecture***
   a. *Data must be compatible for distributed structure to work*
   b. *Implies common keys and surrogate keys exist to ensure that data is being gathered consistently and integrated correctly for the same party*
   c. *Requires standard definitions be developed*
   d. *Implies data quality issues are addressed (preferably in the operational systems)*

3. ***Compatible business intelligence technical and application architecture***
   a. *An initial focus on a distributed structure that acknowledges the current mixed (heterogeneous) business intelligence environment and seeks to minimize short-term disruptions.*
   b. *Level One data marts reflect major categories of data, consistent with the department's Enterprise Data Architecture. Level One data marts receive the enterprise data feeds*
   c. *Level Two data marts use data sourced from level one data marts. May actually extract data for their use or simply provide direct reporting.*
   d. *Implies common platforms and tools*
   e. *FTB will incorporate Traditional and Operational BI to meet business needs*

The Figure below represents the target BI data architecture illustrating the reuse of data from common data subject areas defined by the Enterprise Data Architecture and demonstrates no data redundancy is necessary. Uniqueness is inherent in the MODEL subject area to each of the business functions, yet the models will be able to relate to one another via the CUSTOMER subject area. This is for identifying a CUSTOMER with multiple problems across business functions.

.

Figure 3.1-1: Target BI Data Architecture



3.1.1 Target Enterprise Data Warehouse Architecture

The data warehouse is an environment that must support three primary roles:

1. **Data Acquisition or Collection -** This is the intake role. Taking data from the operational support systems and placing it into the data warehouse environment.
2. **Data Distribution -** This is the role of making the data available for distribution to the end-user
3. **Data Access -** This is the role of providing easy/optimized access to information.

These three roles must be supported by a data architecture that is comprised of three physical tiers.

1. **Staging Area -** A Staging Area is defined as "any data store designed primarily to receive data into a warehousing environment". Other applications shall not access these datastores for any other purpose than loading the warehouse. Access to the staging shall be restricted for timeliness of data reporting. Once the data in concern is stored in the warehouse, accessing that data in staging is forbidden.
2. **Data Warehouse -** A Data Warehouse is defined as "a data structure that is optimized for distribution. It collects & stores integrated sets of data from staging and provides feeds to data marts." It The Data Warehouse is the *"trusted"* source of data to the enterprise. is *subject-oriented*, *integrated*, *time-variant* and *non-volatile*.

   Subject-oriented: A data warehouse is organized around major subjects, such as party/customer, tax declaration, customer account, and *asset & income. Rather than concentrating on the day-to-day operations and transaction processing of an organization, a data warehouse focuses on the modeling and analysis of data for decision makers. Hence, data warehouses typically provide a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.*

   Integrated: A data warehouse is usually constructed by integrating multiple heterogeneous sources, such as relational databases, flat files, and on-line transaction records. Data cleaning and data integration techniques are applied to ensure consistency in naming conventions, encoding structures, attribute measures, and so on.

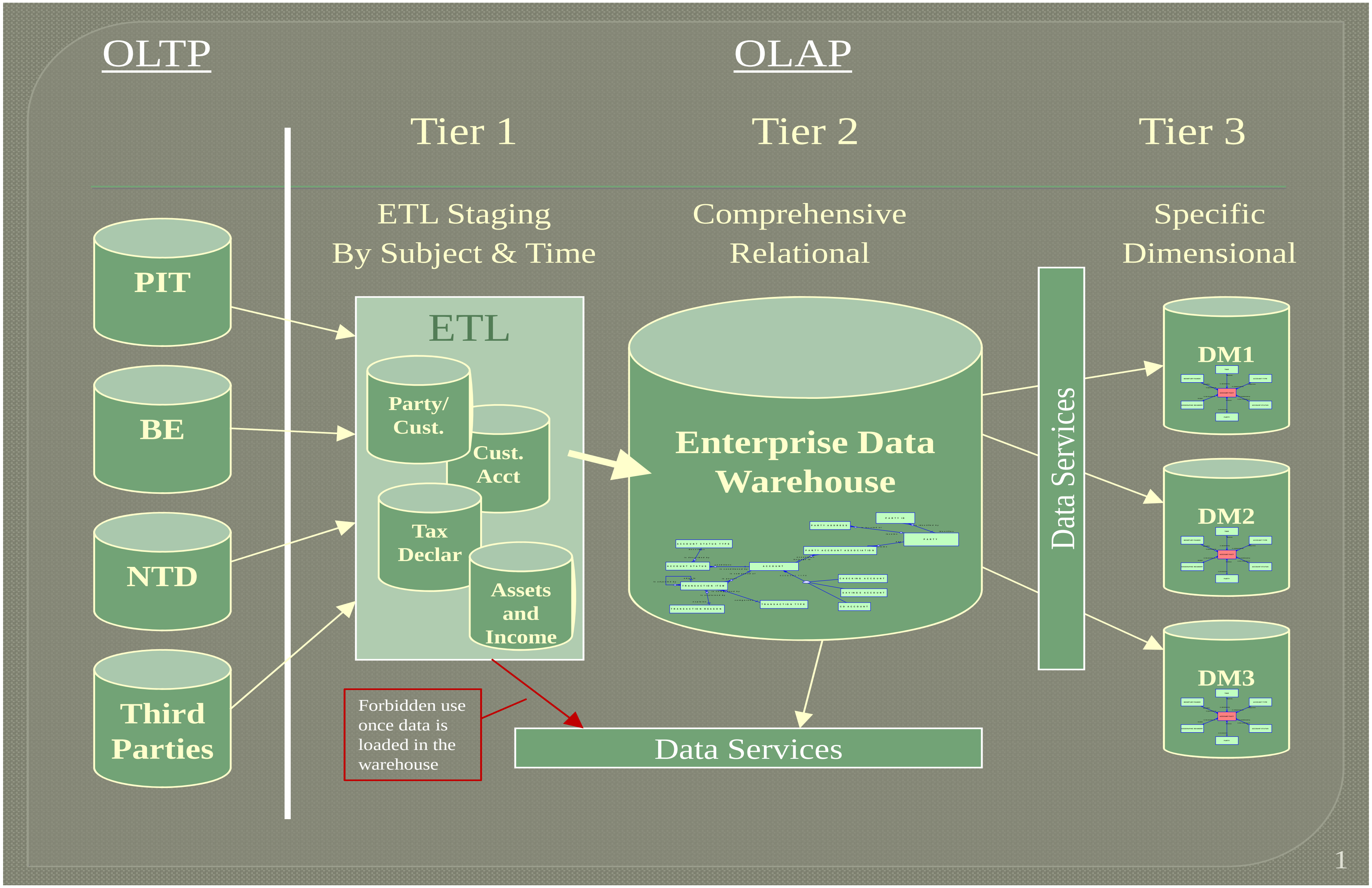
   Time-variant: Data are stored to provide information from a historical perspective (e.g., the past 5-10 years). Every key structure in the data warehouse contains, either implicitly or explicitly, an element of time.

   Nonvolatile: A data warehouse is always a physically separate store of data transformed from the application data found in the operational environment. Due to this separation, a data warehouse does not require transaction processing, recovery, and concurrency control mechanisms. It usually requires only two operations in data accessing: initial loading of data and access of data

3. **Data Marts -** A Data Mart is defined as "a data structure that is optimized for access. It is designed to facilitate end-user analysis of data. It typically supports a single analytic application for a distinct set of consumers."

   Figure 3.1-2 illustrates the three-tier architecture

Figure 3.1-2: Target Data Warehouse Architecture



3.2 Target BI Technical Architecture

This section discusses and illustrates traditional and operational BI in FTB's environment.

3.2.1 Target Traditional BI

Traditional BI will not be replaced by real-time or near real- time Operational BI. FTB has started the move toward the future in transitioning the ECAIR data warehouse to the enterprise level data warehouse. FTB will create standardized data inputs to the warehouse and create a staging area for stable and useable data (trusted data). The conformed data dimensions and data marts will be configured for reuse, performance and predictive analytics.

3.2.2 Target Operational BI

Operational BI will incorporate what may be called data sensors, data sensor networks and knowledge discovery persistence. Advanced analytics will allow for decision-making in real time. Operational BI focuses on providing real-time monitoring of business processes and activities as they are executed within computer systems. Data is captured at a point in time using logic and queries and saved by persistent data logging for future identity matching of situations and patterns.

*3.2.2.1 **Data Sensors explained***

- **Data Sensors**
  A Data Sensor is a software instrument (logic and/or query) that records specific parameters of a data stream and/or events. The sensors are constructed to record, measure, and then analyze the data stream against its history or other conditions as set by business rules (also see BPM Architecture Definition). A metaphor for a data sensor is much like an anemometer records wind speed. The wind speed is the selected data element that can be tested against performance parameters and/or its history. Wind data is then used to determine what construction methods and materials can be used to build a house in a high wind or low wind area. Data sensors at FTB could record the 1040 AGI and 540 AGI along with a TPs ID or other identifying data. A sensor can be a counter of blue-path returns. Added to the blue-path counter can be a blue path refund amount sensor and blue path remit amount sensor. Blue-path refund vs. remit amounts can be compared to counts. Date sensors may be added as well. Advanced implementations allow threshold detection, alerting and providing feedback to the process execution systems. PIT Return Data Mart would be a good candidate for Operational BI architecture.

  The data sensor knows what the needle looks and feels like, so the rest of the haystack of data can be ignored.

- **Data Sensor Networks**
  Data sensor networks consist of a set of distributed data sensors to cooperatively monitor operational conditions, such as return counts vs. amount, which returns impede operations, which business functions have faster turnaround time, and what patterns identify fraud or other threats. Data sensor networks can be business function specific. The following analogy will help one understand this concept. One pixel in a graphic will not let you envision the entire picture, as a set of pixels will allow you the possibility to see the entire picture. Likewise one sensor equals one pixel where a sensor network equals the entire picture.

An example of Operational BI at FTB may occur with return filing and verification (RF&V) activities for fraud threat detection. A Data Sensor Network is created for discovery and recording of the difference between what an employer reports on the employee's W-2 compared to the employee's filed return. This will immediately flag the return for fraud for either identifying the fraudulent employer or that employee. For employer fraud, patterns will need to be discovered over time. These patterns are stored and then future referenced as data passes through the operational systems. This is an example of data learning from data and queries learning from queries. The benefit of Operational BI for fraud activities is identification of a problem before refunds are issued. A Sensor Network can be created specifically for fraud activities or other modeling needs where data is available to perform these tasks.

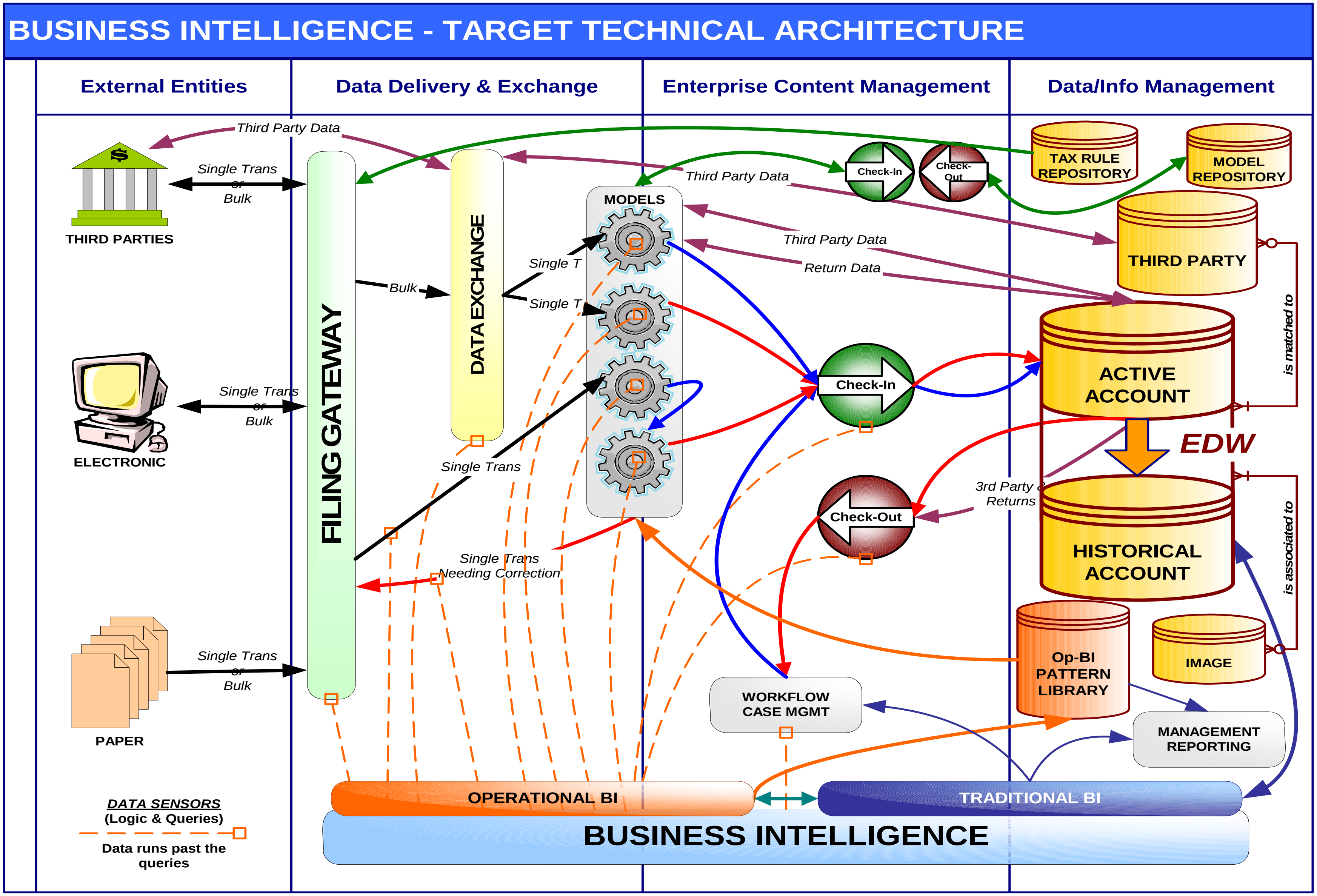
The target data warehousing/BI environment will reduce or eliminate the snapshot concept and the batch extract, transform and load (ETL) that has dominated since the very beginning. The majority of developmental dollars and a massive amount of processing time is used retrieving data from operational databases.

Additionally, the target data warehousing/BI environment at FTB will:

- eliminate write then detect extract process;
- read the same data stream that flow into and between the operational system modules;
- create data that is meaningful to the data warehouse environment by the operational system to a queue as it was created.

In Figure 3.2-1 on the next page, the architecture data delivery, content management, and data management are shown as vertical functions where business intelligence traverses these functions. The details of data delivery, content management and data management are addressed in their appropriate architecture definition documents. Note the data sensors testing selected values that are of interest to the enterprise throughout the entire life cycle of a tax declaration.

Figure 3.2-1: Target BI Technical Architecture (Operational with Traditional)



3.3 Target Enterprise Governance

3.3.1 Business Intelligence and Performance Management Maturity Curve

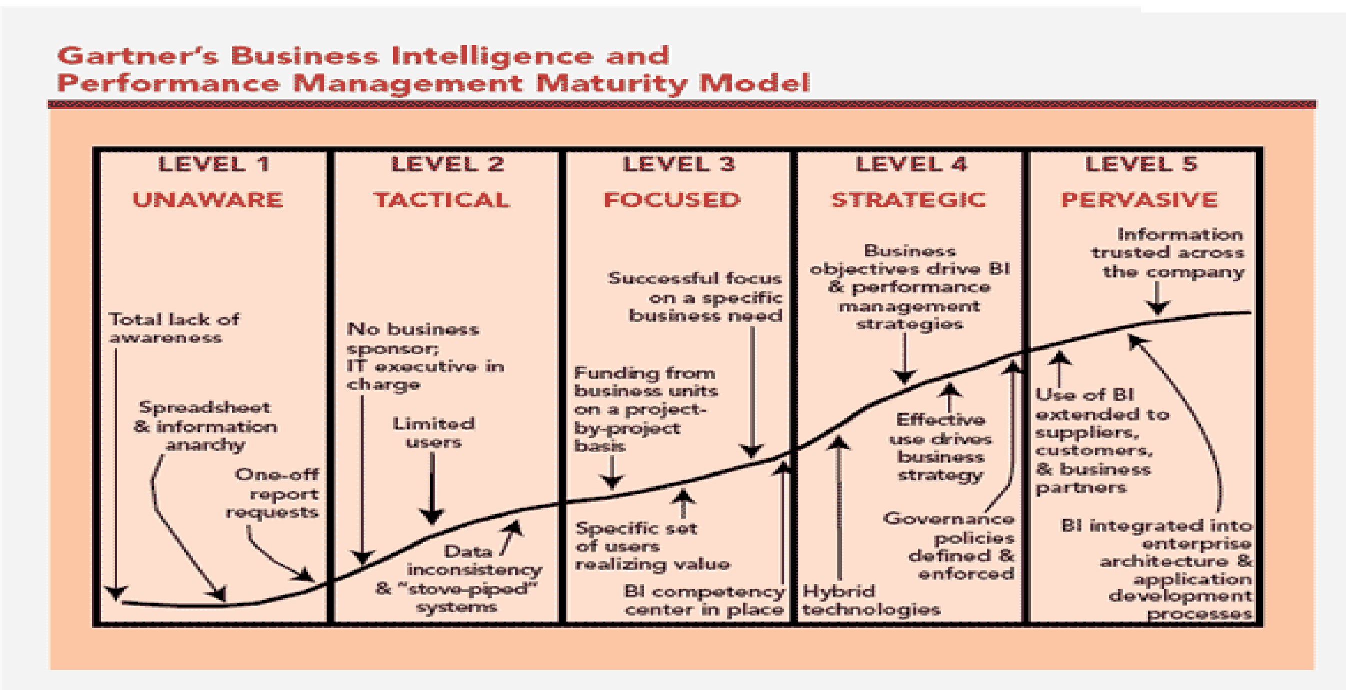
FTB's target BI organization must be tightly connected to the business goals and objectives of the enterprise. The Gartner Group offers a useful tool for understanding where an organization is with regard to BI and what it needs to do to move to the next level. Gartner refers to this tool as the "Business Intelligence and Performance Management Maturity Curve." The curve is based on the real-world phenomenon that organizational change is usually incremental over time.

FTB is currently "**Level 3 – Focused**" category. The recommendations for moving to the **Level 4 – Strategic** level include:

- *Meet with current users to understand how they are leveraging the data and determine if there is additional information you can provide to improve their decision-making processes and their ability to impact the business objectives.*
- *Work with other business units that are key to achieving corporate goals.*
- *Establish a BI Competency Center (BICC) under the Data and Business Centers of Excellence, which is a group of business, IT and information analysts working together, virtually or actively, to define business intelligence strategies and requirements.*
- *Look for opportunities to move to a more overall business-driven vision, by focusing on an architectural approach to support the appropriate tools and applications needed by a broad range of users.*

FTB will use the Gartner Group Business Intelligence and Performance Management Maturity Model to move to Level 4 and to Level 5 to measure BI's effectiveness to the enterprise.

Figure 3.3-1: Gartner's BI & PM Maturity Model



As illustrated in the above chart, the BI & PM Maturity Curve is based on a model that describes the transition in five levels. For the purposes of this document, FTB's current level of 3, and target levels of 4, & 5 are explained:

Level 3 – Focused: Successful BI initiatives deployed, supporting defined business initiatives; one or more business sponsors; metrics are formally defined to analyze specific departmental or functional performances; analytic applications deployed with ongoing user training for a larger number of users across different departments; information often still resides in stove-piped databases, reports, applications, and spreadsheets.

Level 4 – Strategic: BI Competency Center (BICC – discussed later in this document) is in place; dynamic and proactive effort to use BI to meet strategic business goals; funding considered strategic, not just a cost center; well defined repeatable processes in place; effort to reduce unnecessary redundant tools and away from stove-piped systems.

Level 5 – Pervasive: BI is part of corporate culture; information is trusted and valued; scope of BI initiatives extended, sometimes to include external customers/suppliers; BI integrated into business processes; BI is pervasive across systems, applications, technologies and devices. There are real-time, online systems that are broadly, pervasively and effectively used.

3.3.2 Business Intelligence Competency Center (BICC)

One of the key recommendations for moving an organization's BI effort to the next level is establishing a Business Intelligence Competency Center (BICC). Establishing competency centers is a best practice for promoting the effective use of technologies or processes that have widespread impact on organizations and would benefit from cross-functional communication and direction, or governance. The BICC will be a part of the Data Center of Excellence and the Business Center of Excellence.

BICC Defined: A BICC consists of a small team of BI experts consisting of a balanced mix of business users, IT providers, and information analysts. Their basic charter is linking FTB's business-driven goals and objectives with the information, applications, processes, training, policies, and technology necessary to promote the effective use of BI across the organization.

4.0 Gap Analysis

4.1 BI Gap Analysis Defined

A major issue with BI at FTB is the massive amount of data that needs to be extracted, transformed and loaded (ETL) to a data warehouse system. The operations and analysis performed on this data and the ETL process, is performing at a decreasing rate. Performance of our BI systems is a primary complaint from the business community.

Although the Enterprise Customer, Asset, Income and Return (ECAIR) data warehouse is becoming enterprise focused, there are no rules or standards to pull the 3rd Party data into FTB's data stores. Currently, 3rd Party data is not managed and is not using a stable front-end to load the data. This results in data that is unusable. Many of the data feeds are hand coded resulting in reduced data quality.

FTB will develop and maintain a central metadata repository for the enterprise, and eliminate multiple silo metadata repositories around FTB. Metadata should not be hand coded/entered.

4.2 Gap Analysis Table

Figure 4.2-1 discusses the changes required to close the gap between FTB's current BI architecture and the target BI architecture.

Figure 4.2-1: Gap Analysis Table

DATA & TECHNICAL DOMAIN	GAP	CHANGE REQUIRED	BENEFITS	RISKS	IMPACT
Performance degradation will become increasingly problematic as more Third Party Data is added.	The addition of more data for modeling is slowing performance of the database.	Restructure DW Partition DW Use best practice query algorithms. Use Decision Analytics. Employ or train staff in higher mathematics and statistics to create algorithms for performance	Data is to be either restructured/normalized and/or partitioned to support future quantities of data. More attention to query algorithms must be developed to support more data. Making modeling faster will allow RF&V to	If data is not restructured either by normalization, partitioning, and/or query algorithm improvements, then the data warehouse for modeling will continue to take longer result times. This will contribute for the need of more computing and personnel resources.	**High**

DATA & TECHNICAL DOMAIN	GAP	CHANGE REQUIRED	BENEFITS	RISKS	IMPACT
		for ETL and Intelligence algorithms for Analytics.	generate notices more quickly and accelerate revenue to FTB. The restructure would include better structures for OLAP processing as well as better time slicing of OLTP data		
Real time Tax Rule Creation and Modification	Business Rules are hard coded in software applications.	Remove Tax Business Rules from software application code. Create Tax Business Rule Repository and Tax Business Rule Management engine. BI needs to support and/or work with the Business Process Management area.	Real time Tax Rule Creation and Modification will allow FTB to: • Identify potential trends, strategies, and adjustments • Fine tune and adjust the validation and verification processes	Need to separate the tax validation rules from the software application code, and store these rules and business knowledge facts into separate agile data stores. If this is not accomplished, annual changes costs will continue to be expensive as IT staff will always need to change the business rules in the application code.	**High**
Single enterprise feeds for key subject areas, or single data warehouse.	BI at FTB has Awareness of the key subject areas, but they are not implemented. 3rd Party data is not loaded in any one place or method. Incomes in many forms, times, and areas. Some ETL logic is hand coded. ETL should be only automated.	Projects must comply BI requirements, and be responsible for their processes to feed the warehouse. This must be automated as much as possible via direct feeds from the business	Reduces cost of now capturing the data in one place, and not having to recapture the same data for each database or application that uses the data. Supports the business rule to prevent unplanned redundancy. Data will not be replicated unless there is a business need to do so (such as performance or security needs). Creates a stable front	Governance needs to be a prime concern for feeding the data warehouse from return and third party data. Multiple non-standard or redundant data feeds will add to the cost of doing business in the area of cleansing the data after the fact to conform to the DW.	**High**

DATA & TECHNICAL DOMAIN	GAP	CHANGE REQUIRED	BENEFITS	RISKS	IMPACT
			end giving way to useable data.		
Compatible data architecture and BI Methodology. & Compatible business intelligence technical and application architecture	FTB does have a target data architecture that is generally accepted or is completely ignored. A consistent BI methodology is required along with the target data, technical and application architectures. Before an OLTP system is created, BI must be considered in the planning stage or FSR stage.	Establish the BI Competency Center Enforce the target architecture, through Enterprise Architecture and the BI Competency Center (BICC). Accept or develop a compatible methodology for BI and FTB business.	Consistency in the development and deployment of BI systems.	The risk not to have a compatible architectures and BI methodology will continue to result in poor data quality, complex ETL, higher resource costs, and poor performance.	**High**
Enterprise Metadata Repository for OLTP and OLAP	Metadata Repositories are silo and hand code. No metadata standards are adhered to.	FTB needs to govern and create a central enterprise metadata repository that is compatible with industry metadata standards	Consistent and Contextual meaning of data reduces the cost of staff trying to analyze each piece of data for meaning and use each time it is used. If it is defined upfront and made useful, then time and cost is minimized for delivery of the BI product.	Governance and creation of a central or federated metadata repository is imperative in BI. If this is not done costs will remain high in developing BI products to the business.	**High**

4.3 Strategies

- **Continue organizational consolidation of BI functions**. The effort to consolidate BI functions, which began in 2006, has helped to reduce redundant activities and initiate standardization of processes and practices. Some additional BI activities continue in the department, outside of the organization with primary responsibility for BI. Continued consolidation of these activities will help to prevent the proliferation of siloed stovepipes and promote a single vision and clear responsibility toward meeting enterprise goals and objectives.
- **Strengthen communication between BI users and BI providers**. Establish the BI Competency Center (BICC). BI providers must have a clear understanding of those expectations and have a communication process in place that continually updates those expectations, as conditions change. Conversely, BI users must be fully aware of the BI products and services available to them to meet their business goals.
- **Strengthen the link between BI initiatives and corporate planning efforts**. While BI at FTB has successfully met some business areas' tactical needs, moving to the next level, in the BI Maturity Curve, requires a stronger focus on the strategic direction of the organization. A stronger link to corporate planning efforts will help propel BI to be a strategic enterprise resource.
- **Utilize BI maturity assessment tools to identify additional opportunities for growth**. A number of assessment tools are available for uncovering potentially hidden opportunities or risks specific to BI organizational efforts. One of the most recognized sources for these tools is The Data Warehouse Institute and available on their website: http://www.tdwi.org/
- **Implement the Business Intelligence Competency Center**. Deployment of a BICC is a widely accepted best practice for organizations interested in ensuring their BI efforts are in line with business goals and objectives and that their BI customers effectively use the products and services available from their BI providers.

5.0 Roadmap

Target BI at FTB will focus on the analytics and reporting both in real-time operational and traditional analytical. Figure 5-1 identifies the phases and sub-phases to reach the target BI architecture.

Figure 4.3-1: Business Intelligence Architecture Definition Phases

Task Name	Scheduled Work (in hours)
Business Intelligence	**8,204**
Business Intelligence - PHASE I	**2856**
Business Intelligence Planning	
Coordinate BI Planning with Enterprise Data Management & Data Services	160
Plan and Document Metadata Repository	240
Document Business Intelligence Plan	160
Review Business Intelligence Plan	40
Enterprise Data Warehouse (EDW) Analysis	
Determine EDW Requirements	120
Develop EDW Logical Data Model	120
Analyze Requirements and Identify EDW Tables	112
Milestone - Agreement on EDW Data Contents	
Enterprise Data Warehouse Design	
Determine Design Approach	96
Document Design Approach	80
Review Design Approach	40
Enterprise Data Warehouse Data Model Design	
Analyze Requirements and Identify EDW Tables	80
Design EDW Views Template	80
Design EDW Physical Data Model	40
Enterprise Data Warehouse Population Design	
Validate Data Replication	24
Design EDW Population Approach	80
Document Data Warehouse Population Approach	40
Review EDW Population Approach	16
Design EDW Population Template	80
Review EDW Population Template	16
Enterprise Data Warehouse Data Services Design	
Design EDW Data Service Approach	80
Document EDW Data Service Approach	40
Review EDW Data Service Approach	16
Design EDW Data Service Template	80
Review EDW Data Service Template	16
Milestone - EDW Design Complete	
Enterprise Data Warehouse Build & Test	
Build Automation Script for generating View code	80
Confirm EDW Environment	40
Enterprise Data Warehouse Data Structures Build & Test	
Build EDW Database	80
Build EDW Views	80
Enterprise Data Warehouse Data Services Build & Test	
Develop, Test & Review Data Services	360
Enterprise Data Warehouse Population Build & Test	
Develop, Test & Review Replication Process	360
Develop, Test & Review Population Processes	360

Milestone - EDW Build & Test Complete	0
Business Intelligence - PHASE II	**1100**
Underpayment - Replicate/Decouple Workflow Rules Analysis	
Determine Workflow Rule Requirements	80
Create Workflow Rule Model	40
Document Workflow Rule Requirements	20
Review Workflow Rule Requirements	16
Milestone - Agreement on Workflow Rule Requirements	0
Underpayment Modeling - Data Mart Analysis	
Determine Data Mart Requirements	40
Create Data Mart Logical Data Model	40
Document Data Mart Requirements	20
Review Data Mart Requirements	16
Milestone - Agreement on Data Mart Requirements	0
Underpayament Modeling - Data Mart Design	
Design Data Mart (Facts and Dimensions)	40
Design Data Mart - Cube Implementation	40
Create Data Mart Physical Data Model	40
Document Data Mart Design	20
Review Design of Financial and Performance Data Marts	16
Milestone - Agreement on Design of Financial and Performance Data Marts	0
Data Mart ETL & Data Services Design	
Design Data Mart ETL & Data Services Approach	80
Document Data Mart ETL & Data Services Approach	40
Review Data Mart ETL & Data Services Approach	16
Data Mart Build & Test	
Data Mart Data Structures Build	120
Confirm Data Mart Environment	20
Build Data Mart Data Structure	40
Milestone - Data Mart Data Structures Build Complete	0
Data Mart ETL & Data Services Build	
Develop, Test, Review Data Mart ETL & Data Services	176
Data Mart Build - Cube Implementation	
Confirm Financial Data Mart - Cube Implementation Environment	20
Build Data Mart - Cube Implementation	160
Milestone - Data Mart Build Complete	0
Business Intelligence - PHASE III	**1856**
Business Intelligence Filing Enforcement (FE) Modeling - PHASE III-A	**928**
FE Modeling - Data Mart Analysis	
Determine Data Mart Requirements	40
Create Data Mart Logical Data Model	40
Document Data Mart Requirements	20
Review Data Mart Requirements	16
Milestone - Agreement on Data Mart Requirements	
FE Modeling - Data Mart Design	
Design Data Mart (Facts and Dimensions)	40
Design Data Mart - Cube Implementation	40
Create Data Mart Physical Data Model	40

Document Data Mart Design	20
Review Data Mart Design	0
Milestone - Agreement on Data Mart Design	
FE Data Mart ETL & Data Services Design	
Design Data Mart ETL & Data Services Approach	80
Document Data Mart ETL & Data Services Approach	40
Review Data Mart ETL & Data Services Approach	0
FE Data Mart Build & Test	
Data Mart Data Structures Build	120
Confirm Data Mart Environment	20
Build Data Mart Data Structure	40
Milestone - Data Mart Data Structures Build Complete	0
Data Mart ETL & Data Services Build	
Develop, Test, Review Data Mart ETL & Data Services	176
Data Mart Build - Cube Implementation	
Confirm Cube Implementation Environment	20
Build Data Mart - Cube Implementation	160
Milestone - Data Mart Build Complete	0
Business Intelligence Audit Modeling - PHASE III-A	**928**
Audit Modeling - Data Mart Analysis	
Determine Data Mart Requirements	40
Create Data Mart Logical Data Model	40
Document Data Mart Requirements	20
Review Data Mart Requirements	16
Milestone - Agreement on Data Mart Requirements	
Audit Modeling - Data Mart Design	
Design Data Mart (Facts and Dimensions)	40
Design Data Mart - Cube Implementation	40
Create Data Mart Physical Data Model	40
Document Data Mart Design	20
Review Data Mart Design	16
Milestone - Agreement on Data Mart Design	
Audit Data Mart ETL & Data Services Design	
Design Data Mart ETL & Data Services Approach	80
Document Data Mart ETL & Data Services Approach	40
Review Data Mart ETL & Data Services Approach	0
Audit Data Mart Build & Test	
Data Mart Data Structures Build	120
Confirm Data Mart Environment	20
Build Data Mart Data Structure	40
Milestone - Data Mart Data Structures Build Complete	0
Data Mart ETL & Data Services Build	
Develop, Test, Review Data Mart ETL & Data Services	176
Data Mart Build - Cube Implementation	
Confirm Financial Data Mart - Cube Implementation Environment	20
Build Data Mart - Cube Implementation	160
Milestone - Data Mart Build Complete	0
Business Intelligence - PHASE IV	**2392**

Fraud Modeling - Operational BI Analysis	
Determine Operation BI Requirements	80
Create Operational BI Logical Data Model	120
Create Operational BI Rules for Fraud Modeling	160
Document Fraud Operational BI Requirements	40
Review Fraud Operational BI Requirements	32
Milestone - Agreement on Fraud Modeling - Operational BI Requirements	
Fraud Modeling - Data Mart Analysis	
Determine Data Mart Requirements	60
Create Data Mart Logical Data Model	60
Document Data Mart Requirements	32
Review Data Mart Requirements	16
Milestone - Agreement on Data Mart Requirements	
Fraud Modeling - Data Mart Design and OpBI Datastore Design	
Design Data Mart (Facts and Dimensions)	40
Design OpBI Datastore	80
Create Data Mart Physical Data Model	40
Create OpBI Physical Data Model	60
Document Physical Data Model Designs	40
Review Design of Data Mart and OpBI Datastore	32
Milestone - Agreement on Data Mart OpBI Datastore Design	
Fraud Data Mart ETL & Data Services Design	
Design Data Mart ETL & Data Services Approach	120
Document Data Mart ETL & Data Services Approach	40
Review Data Mart ETL & Data Services Approach	32
Fraud Data Mart Build & Test	
Data Mart Data Structures Build	120
Confirm Data Mart Environment	20
Build Data Mart Data Structure	40
Milestone - Data Mart Data Structures Build Complete	0
Data Mart ETL & Data Services Build	
Develop, Test, Review Data Mart ETL & Data Services	176
Data Mart Build - Cube Implementation	
Confirm Data Mart - Cube Implementation Environment	20
Build Data Mart - Cube Implementation	160
Milestone - Data Mart Build Complete	0
Fraud OpBI Data Services Design	
Design OpBI Data Services Approach	120
Document OpBI Data Services Approach	40
Review OpBI Data Services Approach	32
Fraud OpBI Build & Test	
Build OpBI Data Structure	80
Build OpBI Data Services	120
Confirm OpBI Environment	20
Develop, Test, Review OpBI Data Structures & Data Services	360
Milestone - OpBI Data Structures & Data Services Build Complete	

6.0 Appendix

6.1 Definitions

BICC – Business Intelligence Competency Center

CONFORMED DIMENSION - A dimension shared across fact tables. This guarantees that the same terminology and values will be used throughout all reporting in the organization, which eliminates departmental conflicts over terminology definitions and relates one truth to all users.

DATA QUALITY - Data Quality is accurate and consistent data resulting in useable information and knowledge.

DIMENSION - A table that identifies a major perspective of a business data set such as Customers

ECAIR - Enterprise Customer, Asset, Income and Return data warehouse

FACT - A table that holds the metrics for a given subject area such as units sold

OPERATIONAL BUSINESS INTELLIGENCE – See Section 3.3.2

PREDICIVE ANALYTICS and **PREDICTIVE TAX COMPLIANCE MANAGEMENT** – See Section 3.3.4

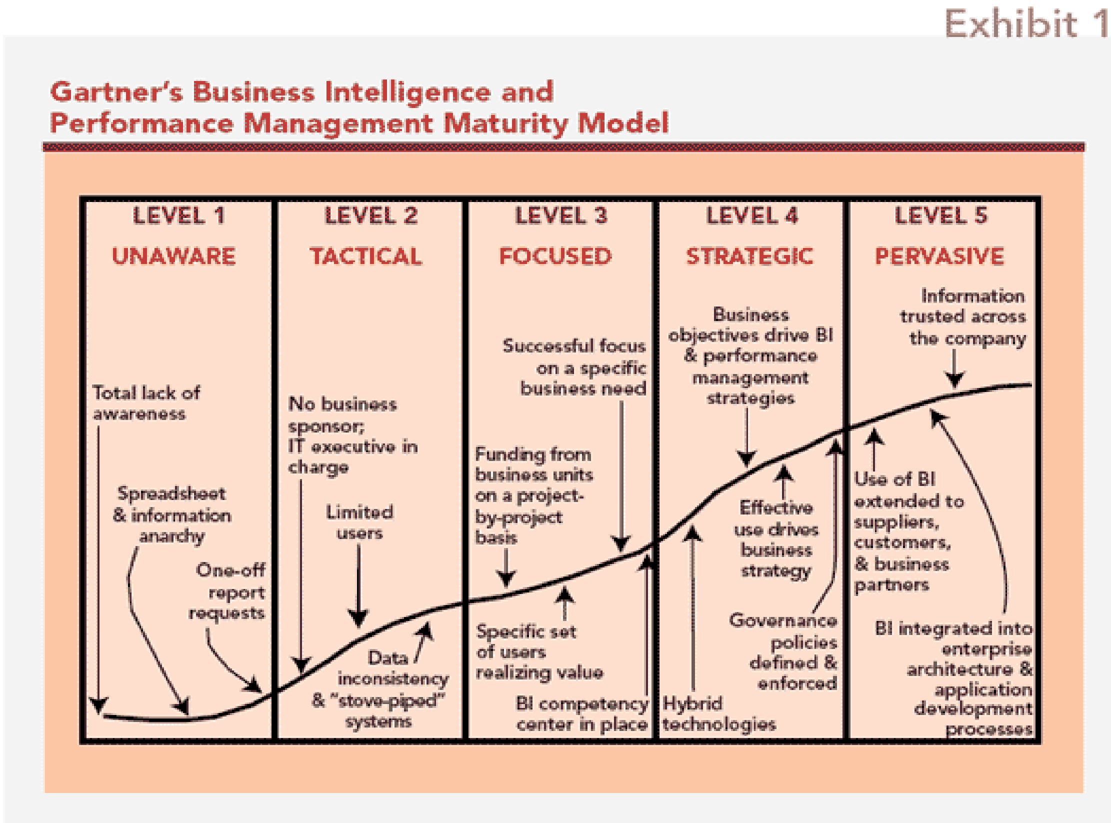
TRADITIONAL BUSINESS INTELLIGENCE – See Section 3.3.1

TRUSTED DATA – Trusted Data comes from a central source (i.e., Enterprise Data Warehouse) where the data has been formatted and verified for quality and use for all downstream processes and datastores.

6.2 Industry Best Practices

As illustrated in section 2.2.2.1, the composition of the target BI organization must be tightly connected to the business goals and objectives of the overall organization. The Gartner Group offers a useful tool for understanding where an organization is with regard to BI and what it needs to do to move to the next level. Gartner refers to this tool as the “Business Intelligence and Performance Management Maturity Curve.” The curve is based on the real-world phenomenon that organizational change is usually incremental over time.

Figure 6.2-1: Gartner's BI & PM Maturity Model



6.3 Industry Trends

The current industry trend for BI is the move toward adding operational intelligence analytics to the current BI toolbox. Operational BI stems from the ability to be proactive running the business versus reactive.

FTB can benefit by Operational BI analytics as seen with such BI systems in the PIT Return Data Mart, some of the ARMR and PASS MI, as well as INC. Not all BI would require operational real-time data, as there will always be the need for historical financials and other traditional BI reporting. The greatest value using Operational BI would be in the area of Fraud Detection Modeling before a refund is sent to the customer.

6.3.1 Traditional Business Intelligence – (also called Static BI)

Traditionally, reporting drives BI. At the simplest level, a report is the rendering of information requested from existing data, with at least some level of formatting and usually some added calculations, such as subtotals and totals at a minimum. An interactive pivot table doesn't really change the nature of a report. The ability to select parameters, for instance, is actually a reporting application surrounding just another report. Latency is another issue for BI where most of the data analysis is after the fact reporting on events that have already passed. Offline knowledge discovery for tax compliance functions include but not limited to:

- Predictive analysis from distributed heterogeneous data
- Mining unusual patterns from massive and disparate return and financial data

6.3.2 Operational Business Intelligence - (also called Dynamic BI)

Real time operational business intelligence is the process of delivering information about business operations without any latency. In this context, real time means delivering information in a range from milliseconds to a few seconds after the business event. While traditional business intelligence presents historical information to users for analysis, real time business intelligence compares current business events with historical patterns to detect problems or opportunities automatically. This automated analysis capability enables corrective actions to be initiated and or business rules to be adjusted to optimize business processes. Additionally, operational BI gives the organization the ability to identify a threat such as fraud so that action can be taken before a refund or any monies are in an unlikely customer's possession.

Gartner Reports:

"Real-time analytics demand real-time data warehousing. Know the critical issues facing the real-time warehouse and understand both your options and the decision criteria that drive your architectural choices.

Key Findings

- *Real-time data warehouses are defined by the presence of a single real-time updated data attribute as required by a business analysis.*
- *There are actually three classes of data warehouse refresh or update rates: periodic, intra-day and real-time.*
- *Business users tend to insist on real-time updates even when they are unable to affect business outcomes.*

Recommendations

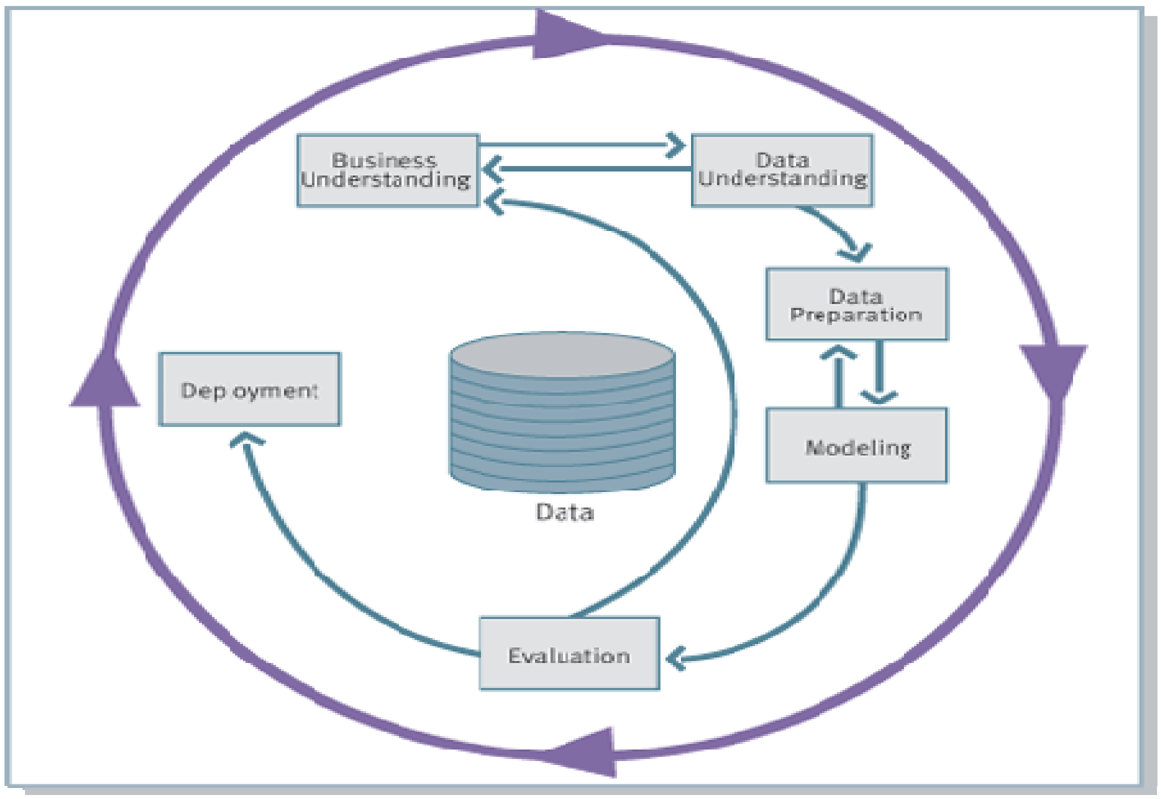
- *Create real-time data attributes in your data warehouse that are attached to real-time business drivers that influence business operations in real-time.*
- *Avoid the false assumption that a real-time data warehouse requires real-time feeds for all sources.*
- *Utilize data staging areas such as operational data stores, data integration hash files and near-real-time partitions to manage real-time data to periodic data relationships.*

6.3.3 Methodology - CRoss Industry Standard Process for Data Mining (CRISP-DM)

CRoss Industry Standard Process for Data Mining (CRISP-DM) is a data mining methodology that has been developed as an industry- and tool-neutral data mining process model. CRISP-DM is a methodology that makes data mining and predictive analytics projects more efficient, better organized, more reproducible, more manageable and more likely to yield business success. Even small-scale data mining investigations benefit from using this methodology.

The CRISP-DM process model for data mining provides an overview of the life cycle of a data mining project. It contains the corresponding phases of a project, their respective tasks and relationships between these tasks. At this description level, it is not possible to identify all relationships. An electronic copy of the CRISP-DM Version 1.0 Process Guide and User Manual is available at http://www.crisp-dm.org/index.htm .

Figure 6.3-1: Phases of the CRISP-DM Process Model



The life cycle of a data mining project consists of six phases. The sequence of the phases is not strict. Moving back and forth between different phases is always required. It depends on the outcome of each phase which phase, or which particular task of a phase, that has to be performed next. The arrows indicate the most important and frequent dependencies between phases.

The outer circle in the figure symbolizes the cyclic nature of data mining itself. A data mining process continues after a solution has been deployed. The lessons learned during the process can trigger new, often more focused business questions. Subsequent data mining processes will benefit from the experiences of previous ones.

Below follows a brief outline of the phases:

Business Understanding
This initial phase focuses on understanding the project objectives and requirements from

a business perspective, and then converting this knowledge into a data mining problem definition, and a preliminary plan designed to achieve the objectives.

Data Understanding
The data understanding phase starts with an initial data collection and proceeds with activities in order to get familiar with the data, to identify data quality problems, to discover first insights into the data, or to detect interesting subsets to form hypotheses for hidden information.

Data Preparation
The data preparation phase covers all activities to construct the final dataset (data that will be fed into the modeling tool(s)) from the initial raw data. Data preparation tasks are likely to be performed multiple times, and not in any prescribed order. Tasks include table, record and attribute selection as well as transformation and cleaning of data for modeling tools.

Modeling
In this phase, various modeling techniques are selected and applied, and their parameters are calibrated to optimal values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. Therefore, stepping back to the data preparation phase is often needed.

Evaluation
At this stage in the project you have built a model (or models) that appears to have high quality, from a data analysis perspective. Before proceeding to final deployment of the model, it is important to more thoroughly evaluate the model, and review the steps executed to construct the model, to be certain it properly achieves the business objectives. A key objective is to determine if important business issues have not been sufficiently considered. At the end of this phase, a decision on the use of the data mining results should be reached.

Deployment
Creation of the model is generally not the end of the project. Even if the purpose of the model is to increase knowledge of the data, the knowledge gained will need to be organized and presented in a way that the customer can use it. Depending on the requirements, the deployment phase can be as simple as generating a report or as complex as implementing a repeatable data mining process. In many cases it will be the customer, not the data analyst, who will carry out the deployment steps. However, even if the analyst will not carry out the deployment effort it is important for the customer to understand up front what actions will need to be carried out in order to actually make use of the created models.

CRISP-2.0: Updating the Methodology

Many changes have occurred in the business application of data mining since CRISP-DM 1.0 was published. Emerging issues and requirements include:

- *The availability of new types of data—text, Web, and attitudinal data, for example—along with new techniques for pre-processing, analyzing, and combining them with related case data*

- *Integration and deployment of results with operational systems such as call centers and Web sites*
- *Far more demanding requirements for scalability and for deployment into real-time environments*
- *The need to package analytical tasks for non-analytical end users and integrate these tasks in business workflows*
- *The need to seamlessly integrate the deployment of results and closed-loop feedback with existing business processes*
- *The need to mine large-scale databases in "situ", rather than exporting an analytical dataset*
- *Organizations' increasing reliance on teams, making it important to educate greater numbers of people on the processes and best practices associated with data mining and predictive analytics*

Figure 6.3-1 above is supported by the industry direction and methodology as prescribed by the CRISP-DM processes both in versions 1.0 and CRISP-DM 2.0. CRISP—DM 2.0 supports the need to mine large-scale databases in "situ" meaning operational databases, rather than exporting large-scale analytical data stores. FTB will have the need for real time analytics as well as logging storage of these results.

6.3.4 Business Intelligence Predictive Tax Compliance Management

FTB manages large, disparate sources of data that contain the hidden knowledge needed to improve compliance operations. This hidden knowledge that can be found in Third Party data, has enormous potential for improving decision making, but cannot be manually extracted and put to use. Most of this data is in house at FTB, but not accessible or available at the enterprise level. Advanced analytics, is the process of uncovering patterns and relationships in data. For tax administrators, these patterns improve decision making by uncovering areas for compliance process improvement—helping FTB make better, timelier decisions to achieve their goals.

BI with data mining produces predictive models to improve non-filer discovery, audit selection and collections management. Based on the results from using advanced analytics, FTB can determine which actions will drive the best outcome, and then deliver those recommended actions to the systems or people that can take appropriate action. In tax compliance management, the goal of data mining is to discover "knowledge" that enables a tax agency to optimally focus limited resources to detect non-compliance, and ultimately enhance voluntary compliance.

In order to achieve compliance, advanced analytics such as data mining are used to examine the way in which tax compliance management issues relate to data on past, present, and projected future actions. Advanced analytics include statistical, mathematical and other algorithmic techniques, and are more complex than the basic analytics used to compute frequencies, cross-tabs, and query and reporting cubes. Advanced analytics complement basic analytic environments such as OLAP and reporting—providing deeper levels of insight. Advanced Predictive Analytics is more than running SQL against a data warehouse. The figure below illustrates a predictive analytics framework for tax compliance management.

Figure 6.3-2: Predictive Analytics Framework for Tax Compliance Management