

# IT Analytics

for the Windows operating system

Software Version: 2.00

---

## ETL User Guide

Document Release Date: February 2010  
Software Release Date: July 2009

*Last published date: July 2009*



## Legal Notices

### Warranty

The only warranties for HP products and services are set forth in the express warranty statements accompanying such products and services. Nothing herein should be construed as constituting an additional warranty. HP shall not be liable for technical or editorial errors or omissions contained herein.

The information contained herein is subject to change without notice.

### Restricted Rights Legend

Confidential computer software. Valid license from HP required for possession, use or copying. Consistent with FAR 12.211 and 12.212, Commercial Computer Software, Computer Software Documentation, and Technical Data for Commercial Items are licensed to the U.S. Government under vendor's standard commercial license.

### Copyright Notices

© Copyright 2008–2010 Hewlett-Packard Development Company, L.P.

### Trademark Notices

Adobe® is a trademark of Adobe Systems Incorporated.

Microsoft® and Windows® are U.S. registered trademarks of Microsoft Corporation.

Oracle is a registered trademark of Oracle Corporation and/or its affiliates.

### Acknowledgements

This product includes ANTLR 3 software, which is Copyright © 2003-2008, Terence Parr All rights reserved. ([www.antlr.org/](http://www.antlr.org/))

This product includes software developed by the Apache Software Foundation. ([www.apache.org/](http://www.apache.org/))

This product includes software developed by the OpenSymphony Group. ([www.opensymphony.com/](http://www.opensymphony.com/))

This product includes Sentry Spelling-Checker Engine, which is Copyright © 2000 Wintertree Software Inc. ([www.wintertree-software.com](http://www.wintertree-software.com))

## Documentation Updates

This guide's title page contains the following identifying information:

- Software Version number, which indicates the software version.
- Document Release Date, which changes each time the document is updated.
- Software Release Date, which indicates the release date of this version of the software.

To check for recent updates, or to verify that you are using the most recent edition of a document, go to:

**<http://h20230.www2.hp.com/selfsolve/manuals>**

This site requires that you register for an HP Passport and sign-in. To register for an HP Passport ID, go to:

**<http://h20229.www2.hp.com/passport-registration.html>**

Or click the **New users - please register** link on the HP Passport login page.

You will also receive updated or new editions if you subscribe to the appropriate product support service. Contact your HP sales representative for details.

## Support

You can visit the HP Software Support web site at:

**<http://www.hp.com/go/hpsoftwaresupport>**

This web site provides contact information and details about the products, services, and support that HP Software offers.

HP Software Support Online provides customer self-solve capabilities. It provides a fast and efficient way to access interactive technical support tools needed to manage your business. As a valued support customer, you can benefit by using the HP Software Support web site to:

- Search for knowledge documents of interest
- Submit and track support cases and enhancement requests
- Download software patches
- Manage support contracts
- Look up HP support contacts
- Review information about available services
- Enter into discussions with other software customers
- Research and register for software training

Most of the support areas require that you register as an HP Passport user and sign in. Many also require a support contract.

To find more information about access levels, go to:

**[http://h20230.www2.hp.com/new\\_access\\_levels.jsp](http://h20230.www2.hp.com/new_access_levels.jsp)**

To register for an HP Passport ID, go to:

**<http://h20229.www2.hp.com/passport-registration.html>**

# Contents

- 1 Getting Started ..... 9
  - ETL User Guide Overview ..... 9
  - Other Documentation Resources ..... 10
- 2 Working With a Data Warehouse ..... 11
  - ETL ..... 12
  - Data Warehouse Model ..... 13
    - Source Data Models ..... 14
    - Data Integration Interface ..... 14
    - SAP BusinessObjects Data Services ..... 14
    - Audit, Balance, and Control ..... 14
    - Staging Database ..... 15
    - Target Database ..... 16
    - Queries ..... 16
    - Presentation ..... 16
  - Other Data Warehouse Components ..... 17
    - Data Models ..... 17
    - Schemas ..... 17
    - External Sources ..... 17
    - Consolidation ..... 17
      - Single Source Level 0 (S0) Consolidation ..... 18
      - Multiple Source Level 2 (M2) Consolidation ..... 18
      - Cross Reference Consolidation ..... 19
    - Metadata ..... 19
    - Historic Data ..... 19
    - Universe and Analytic Views ..... 20
    - Analytics and Dashboards ..... 20
- 3 ETL Process ..... 21
  - ETL Process Overview ..... 21
  - ETL Staging Layers ..... 22
  - ETL Stages ..... 23
    - ETL Processes ..... 24
  - Stage 1: Extract Source Data, Update Cross Reference Metadata, and Process External Sources ..... 25
    - Process 1: Check ETL Preconditions ..... 25
    - Processes 2 and 3: Extract the Source Data ..... 25
      - Extraction Criteria ..... 26
      - Flat Files ..... 26
    - Process 4: Update Cross Reference Metadata ..... 26
    - Process 5: Process External Sources ..... 26

Stage 2: Load the Source Data (EXT) and Process Data Matching .....	27
Processes 6 and 7: Load the Extraction Tables .....	27
Pre-fill the Extraction Tables .....	27
Data Validation .....	28
Load the Extraction Tables .....	28
Process 8: Data Matching .....	28
Source Interface Reference Views .....	28
Stage 3: Standardize the Source Data (SSI) and Generate Enterprise Keys .....	30
Processes 9 and 10: Single Source Identity .....	30
Process 11: Load Enterprise Cross Reference Tables .....	31
Processes 12 and 13: Generate Enterprise Keys .....	31
Enterprise Key .....	31
Stage 4: Populate Enterprise Keys (MSI) .....	32
Processes 14 and 15: Multiple Source Identity .....	32
Stage 5: Consolidate the Source Data (MSC) .....	33
Source Data Priorities .....	33
Processes 16 and 17: Single Source Consolidation .....	33
Backfill the MSI Tables .....	34
Processes 18 and 19: Backfill Keys .....	34
Processes 20 and 21: Backfill MSI Tables .....	34
Process 22: Multiple Source Consolidation .....	34
Stage 6: Populate Load-Ready Tables (XFR) .....	35
Process 23: Prepare Dimensions for Load to Target .....	35
Process 24: Prepare Fact Tables for Load to Target .....	36
Stage 7: Load the Data Warehouse (Target) .....	37
Process 25: Load from Staging to Target .....	38
Process 26: Life Cycle Management .....	38
Stage 8: Update the Snapshot Tables .....	39
Process 27: Consolidated Snapshot .....	39
Process 28: Single Source Snapshot .....	39
Process 29: Transactional Snapshot .....	39
Process 30: Check ETL Postconditions .....	39
<b>4 Data .....</b>	<b>41</b>
Source Systems .....	41
Pre-defined Source Views .....	41
External Sources .....	42
External Source Data Path .....	42
External Source Entities .....	43
Application .....	43
Budget .....	43
Cost Category .....	43
Currency .....	43
Exchange .....	43
IT Function .....	43
Period .....	44
Program .....	44

Consolidating Data .....	45
Data Priorities .....	45
Table Join Types .....	45
Keys Used for Consolidation .....	46
Business Key .....	46
Composite Key .....	46
Durable Key .....	46
Enterprise Key .....	46
Foreign Key .....	46
Primary Key .....	46
Surrogate Key .....	46
Data Validation .....	47
Administrative and Operational Data Reports .....	49
Historical Data in Ad Hoc Queries .....	49
<b>Glossary</b> .....	<b>51</b>
<b>A Naming Conventions</b> .....	<b>67</b>
Field Naming Conventions .....	67
Default Records and Values .....	68
Default Field Values .....	69
<b>B Data Models</b> .....	<b>71</b>
Target Data Models .....	71
Actual Cost .....	72
Allocated Actual Cost .....	73
Allocated Planned Cost .....	74
Planned Cost .....	75
Universe Data Model .....	76
IT Financial Analysis Universe .....	77
Cross Reference Consolidation Data Model .....	78
Cross Reference Consolidation .....	79





# 1 Getting Started

The Extract, Transform, and Load (ETL) process is an end-to-end transfer of external source data through several staging layers and into the target layer of the data warehouse. The source data can reside in a single database or multiple databases. The data warehouse uses the ETL process to extract, consolidate, and transform the source data into a meaningful target model that populates relevant business analytics.

## ETL User Guide Overview

This guide describes the ETL process in detail and provides information about where data originates and how the ETL process converts the data so that it can be used for reporting purposes. Table 1 summarizes the contents of each chapter.

**Table 1 Document Map**

Chapter	Describes
Chapter 1, Getting Started	ETL User Guide overview.
Chapter 2, Working With a Data Warehouse	General information about ETL, consolidation, and data warehouse components.
Chapter 3, ETL Process	Details about the end-to-end data warehouse ETL process.
Chapter 4, Data	Information about sources and source views, data consolidation, data validation, and the use of historic data for ad hoc queries.
Glossary	Data warehouse terminology and definitions.
Appendix A, Naming Conventions	Naming conventions for fields, default field values, and a list of default records.
Appendix B, Data Models	Entity relationship diagrams (ERDs) for the target data models, the IT Financial Analysis universe, and the cross reference consolidation data model.

## Other Documentation Resources

For data management and administration information, see the *HP IT Analytics Data Warehouse Administration Guide*.

For all data warehouse installation and configuration information, see the *HP IT Analytics Installation and Configuration Guide*.

For all SAP BusinessObjects Enterprise XI 3.1 or SAP BusinessObjects Data Services tasks and troubleshooting, refer to the SAP BusinessObjects Enterprise XI 3.1 documentation that accompanies the BusinessObjects installation software.

---

## 2 Working With a Data Warehouse

The data warehouse is the repository for key business data that support queries, reports, and populates business analytics. The data warehouse acquires, standardizes, enriches, and consolidates data based on the target model that is designed to satisfy the business analytic requirements. The data warehouse uses SAP BusinessObjects Data Services to complete the process that extracts data from one or more source applications.

The content of the data warehouse depends on the related HP applications that generate business data and IT Analytics, the analytic application that presents the data. Data can originate with external applications such as Asset Manager and Project and Portfolio Management. The data warehouse can consolidate similar data from these applications into a meaningful target model that populates relevant business analytics. Enabling multiple source applications to provide input data ensures that the application can produce robust and comprehensive results.

The data warehouse enables data integration and promotes data consolidation by using consistent data models regardless of the disparate data models among different data sources. The data warehouse software relies on common data models that support each phase of ETL. As the data is progressively aligned into a single target model, transformation processes assign common data names and attributes for each entity that appears in the target model.

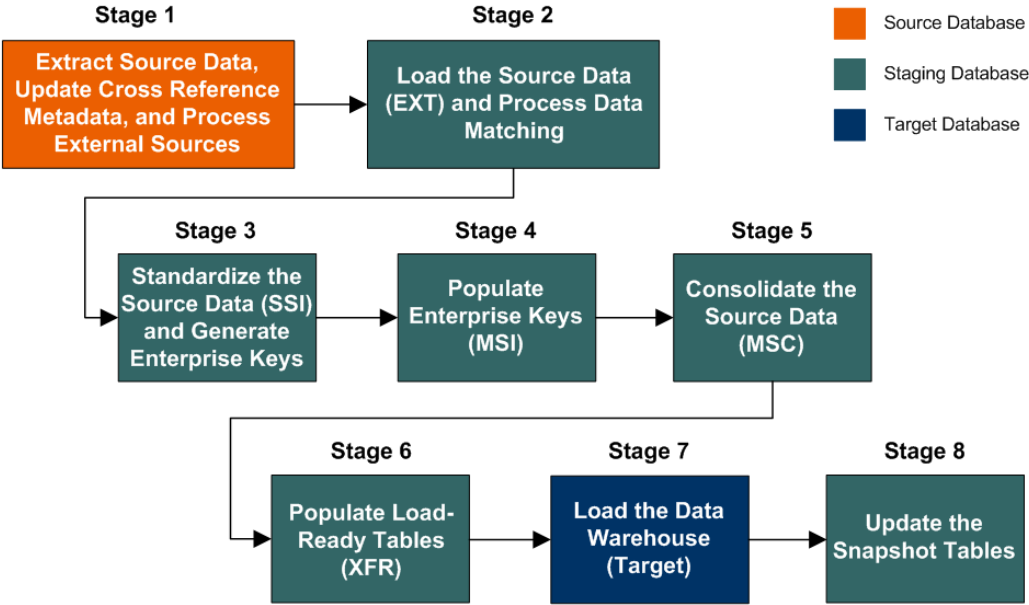
The data warehouse is scalable, extensible, and modular. It can grow and respond to added processes and an increasing volume of data. It is maintainable with a set of tools that monitor processes, troubleshoot problems, and produce reports.

# ETL

The Extract, Transform, and Load (ETL) process is an end-to-end transfer of external source data through several staging layers and into the target layer of the data warehouse. The source data can reside in a single database or multiple databases. The data warehouse uses the ETL process to extract, consolidate, and transform the source data into a meaningful target model that populates relevant business analytics. The ETL engine is SAP BusinessObjects Data Services.

The ETL process comprises eight primary stages. The stages that process depend on metadata and the nature of the sources. [Figure 1](#) shows the overall ETL process.

**Figure 1 ETL process**



For detailed information about each stage of the ETL process, see [ETL Process](#) on page 21.

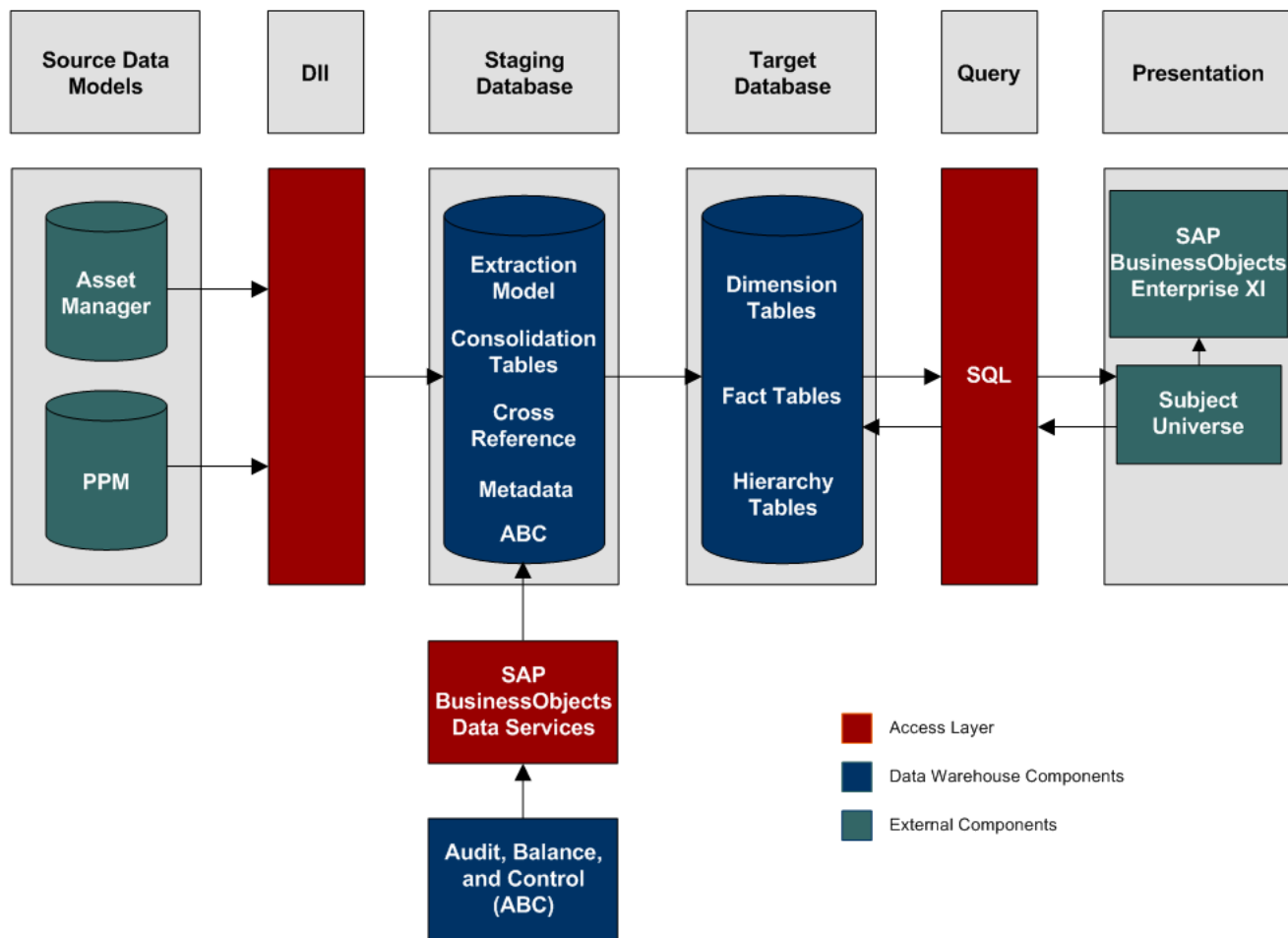
# Data Warehouse Model

The data warehouse model is an end-to-end solution for extracting source data and consolidating disparate source data models into a consolidated data model. The data organized to reflect this model resides in target tables that become the source for IT Analytics.

The data warehouse relies on an external Relational Database Management System (RDBMS) to provide daily maintenance and backup support, SAP BusinessObjects Data Services to assist with ETL, an Audit, Balance, and Control (ABC) infrastructure of utilities and processes to govern regularly scheduled ETL jobs, and internal data warehouse utilities to create the data warehouse and maintain data integrity.

Figure 2 shows the principal data warehouse components and how they interact. Read the following sections for more information about each component, its role, and its responsibilities.

**Figure 2 Data warehouse overview**



## Source Data Models

Source data originates in external application repositories. Data generated by an external application can be rich in business information, but unless you can integrate it with equally important business information generated by other applications, it is difficult to get a complete view of operational information.

The source data models describe the source and extraction format so that the data warehouse can consume this data and integrate it into a common view that you can use to populate comprehensive business intelligence reports and dashboards.



## Data Integration Interface

The data integration interface (DII) is the access layer that stands between the source database models and the extraction models in the staging database. The DII uses a mapping model that transforms the source data to fit into extraction work tables. The DII can accommodate minor variations in the source model schemas from one version to another.

## SAP BusinessObjects Data Services

SAP BusinessObjects Data Services is the access mechanism that performs the physical data extraction from the source repository to the staging database.

## Audit, Balance, and Control

Audit, Balance, and Control (ABC) processes ensure that source data is not corrupted in the extraction process, and that the integrity of the extracted data is maintained as it progresses from the initial source extraction through the logical and physical transformation into target tables. SAP BusinessObjects Data Services can move the data but does not evaluate whether the final result is correct. ABC establishes checkpoints and provides feedback on each step to guarantee that each step is valid before the next step begins.

The Audit component ensures consistency during ETL processing. It measures the number of records in and the number of records out for each step and displays these runtime statistics in a collection of audit reports.

Balance verifies that data in the data warehouse matches data in the source system. For example, if the data warehouse stores project dollar amounts then the balance process verifies that the aggregate project dollars in the data warehouse matches the aggregate project dollars in the source application data.

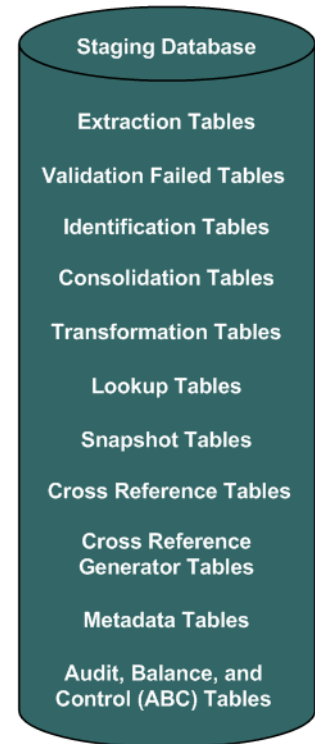
Control governs ETL processes. Control makes sure that there is a proper restart and recovery when a system error occurs. Control also manages job dependencies at runtime.

ABC utilities are applications that you can invoke automatically or manually to run ETL, solve problems, and perform ETL-related tasks.

## Staging Database

The staging database contains tables and snapshots that are critical to the ETL process. Metadata describes these tables. IT Analytics processes data in the staging database before loading it into the final target tables in the target database.

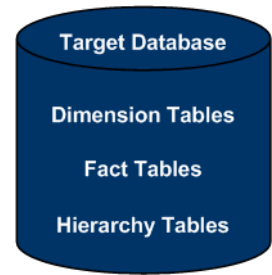
- Extraction (EXT) tables contain the initial version of source data. EXT tables are used during the stage of ETL when IT Analytics loads data from the extract flat files into the staging database.
- Validation Failed (VALF) tables are used during the data validation process that ensures all of the data is usable. VALF tables store source data that is rejected by the ETL process due to validation failure. Data validation is done as part of the process to load the EXT tables.
- Identification (SSI and MSI) tables restructure the diverse source system data into a conformed structure. IT Analytics uses SSI and MSI tables to standardize the data and add an enterprise key.
- Consolidation (MSC) tables integrate disparate source data into a consolidated entity. The MSC tables contain the uniform structure that IT Analytics uses to load the data into the target tables.
- Transformation (XFR) tables are load-ready tables that match the internal structure of the target tables. XFR tables are used during the stage of ETL when IT Analytics prepares the data for loading into the target tables in the target database.
- Lookup (LOOKUP) tables store durable keys and surrogate keys. IT Analytics uses these tables to establish dimension-to-dimension references in dimension tables and foreign key references in fact tables.
- Snapshot (TSNP, SSNP, and CSNP) tables are the persistent storage for ETL staging data. These snapshot tables capture the changed data and support the data warehouse recovery process. The data warehouse uses TSNP tables for source extraction staging, SSNP tables for single source integration staging, and CSNP tables for multiple source consolidation staging. The final ETL stage captures the snapshot data.
- Cross Reference (XREF) tables support the data integration processes that combine records from multiple sources into a single record in the consolidation model.
- Cross Reference Generator (XREFGEN) tables support data matching and cross reference inputs for the consolidation models.
- Metadata tables contain a logical model and describe how to translate that model into the physical implementation of tables and views.
- ABC tables contain the ETL workflow and enable job control and sequencing. They also host audit information to ensure data quality.



## Target Database

The target data models describe the format of the data that produces business intelligence analytics. The instantiated data models are dimension, fact, and hierarchy tables populated with data obtained through ETL processes.

- Dimension tables contain the data model to be queried and presented by IT Analytics in a variety of reports and analytics.
- Fact tables contain data that describes events, transactions, or other granular information.
- Hierarchy tables capture hierarchical relationships among levels of data.



## Queries

You can use structured query language (SQL) to create a primary abstraction layer of data from the target tables to populate reports, dashboards, or resolve business questions. The SAP BusinessObjects Enterprise XI 3.1 suite of business tools simplify accessing this information and storing it in business-oriented views.

## Presentation

SAP BusinessObjects Enterprise XI 3.1 uses a universe abstraction layer to organize data for presentation in out-of-box or user-designed analytics. The universe provides a business oriented view of the data that resides in the data warehouse target tables.



# Other Data Warehouse Components

The language of the data warehouse includes other database concepts.

## Data Models

Data models describe the source, extraction, consolidation, and target components for a single entity:

- **Source entity models:** The source entity models are the data produced by an external application, such as Asset Manager or Project and Portfolio Management.
- **Extraction models:** ETL processes obtain the data in native format and stage it in intermediate tables that conform to the extraction data model for that entity. Extraction also validates the data to ensure that it is usable.
- **Consolidation models:** Consolidation models transform the data further by merging extracted data from diverse sources into meaningful data required by IT Analytics.
- **Target models:** The target models describe the final version of the data that adheres to the design required by IT Analytics.

For more information and to view entity relationship diagrams (ERDs) of the target data models, see [Data Models](#) on page 71.

## Schemas

Logical schemas define entities and attributes. Along with the metadata, schemas become the blueprint for organizing data. This predictable organization enables IT Analytics to access and display the data that it provides to end users.

## External Sources

External sources use Excel spreadsheets as a platform. The spreadsheets are used when the data needs to be manually consolidated or must be taken from a source other than Asset Manager or Project and Portfolio Management.

For more information, see [External Sources](#) on page 42.

## Consolidation

There are two primary levels of source data consolidations: Single Source Level 0 (S0) consolidations and Multiple Source Level 2 (M2) consolidations. Some M2 level entities also pass through cross reference consolidation.

## Single Source Level 0 (S0) Consolidation

For Single Source Level 0 (S0) consolidations, the data comes from a single table in a single source or from multiple tables in a single source. These are the S0 level entities:

- Application
- Budget
- Contract
- Cost Category
- Cost Center
- Currency
- Discretionary Flags
- Exchange
- IT Function
- Period
- Program
- Opex/Capex Flags
- Scenario

## Multiple Source Level 2 (M2) Consolidation

For Multiple Source Level 2 (M2) consolidations, the data comes from multiple sources. For example, the Person and Project entities contain content from both Asset Manager and Project and Portfolio Management. In these cases, IT Analytics uses priority rules to resolve and assign values for the common attributes. Some M2 entities also go through the cross reference consolidation process. These are the M2 level entities:

- Actual Cost
- Budgetline
- Configuration Item
- Location
- Model
- Organization
- Person
- Planned Cost
- Project

## Cross Reference Consolidation

Cross reference is the process for consolidating Multiple Source Level 2 (M2) entities. The cross reference consolidation process integrates records with similar information to produce a single version of data in the data warehouse. The cross reference process can consolidate data from up to three sources. These are the cross reference entities:

Actual Cost

Budgetline

Location

Organization

Person

Project

## Metadata

Metadata determines the order of execution of the ETL processes. Metadata defines for each process the nature of the data that is collected and how it is transferred from one step to another. The cross reference tables also use metadata as a source.

## Historic Data

The source system tables contain only a current view of the source data. No historic version of data is input to the ETL process. The data warehouse does manage historical tracking, however, after IT Analytics loads the data into the data warehouse.

The source database should update the last modification date field value when records are inserted and updated. IT Analytics uses this value to extract delta records.

For more information about using historic data in ad hoc queries, see [Historical Data in Ad Hoc Queries](#) on page 49.

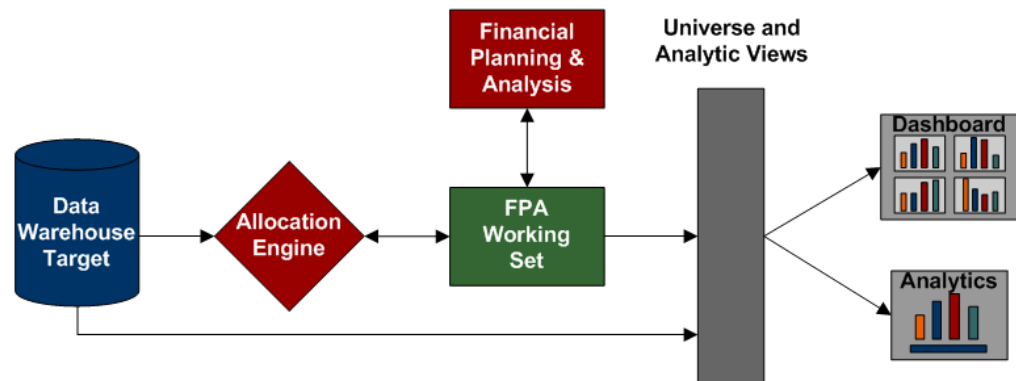
## Universe and Analytic Views

The universe is the result of a semantic layer of metadata that creates a business oriented view of the data. The universe contains a schema of the tables and joins that make up the universe and dimension and measurement objects. The universe is an interface between the data warehouse and the analytics that display the data.

The universe contains analytic views that enable you to see published and up-to-date allocation cost and planned cost data as well as unpublished data or published data that is not up-to-date. The views also ensure the consistency of data warehouse quality by showing only data that is successfully extracted, transformed, and loaded. Analytic views provide the mechanism to reduce the dependencies of analytic semantic models to data warehouse physical models.

For more information and to view an ERD of the universe, see [IT Financial Analysis Universe](#) on page 77.

**Figure 3 Universe and Analytic Views**



## Analytics and Dashboards

Analytics and dashboards are reports that summarize or aggregate data to help managers identify trends, analyze volumes of data, plan, forecast, and make business or strategic decisions.

## 3 ETL Process

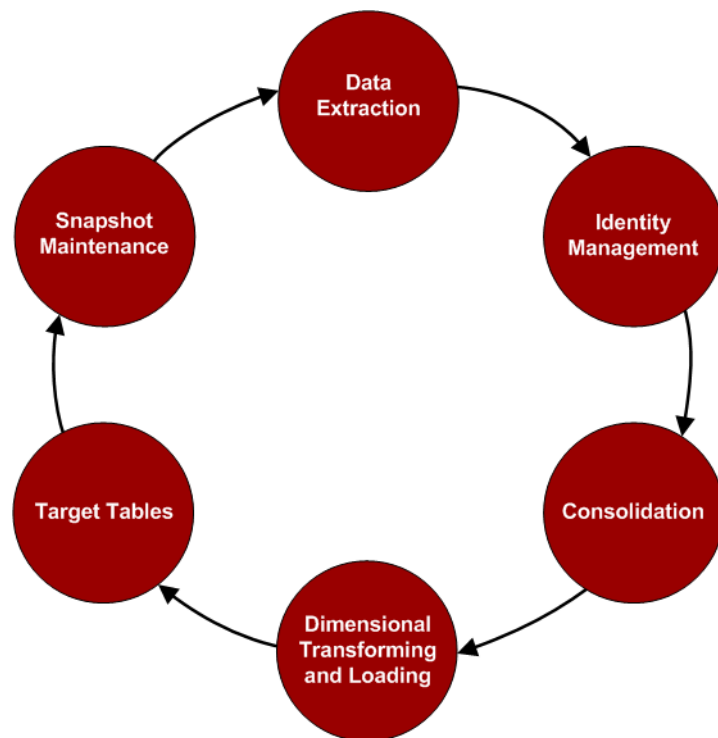
The ETL process is an end-to-end transfer of external source data through several staging layers and into the target layer of the data warehouse. The source data can reside in a single database or multiple databases. The data warehouse uses the ETL process to extract, consolidate, and transform the source data into a meaningful target model that populates relevant business analytics. The ETL engine is SAP BusinessObjects Data Services. Read this chapter for descriptions of the stages and processes that compose the ETL process.

For information about running ETL and about operations reports that provide measurements and metrics logged by the ETL processes as well as historical data and statistics for ETL runs, see the *HP IT Analytics Data Warehouse Administration Guide*.

### ETL Process Overview

Figure 4 shows an overview of the sequence IT Analytics follows during the ETL process. SAP BusinessObjects Data Services extracts external application data, then processes and consolidates the data before loading it into the target database. Data gets stored in data warehouse target tables for subsequent application processing.

**Figure 4 ETL Process Overview**

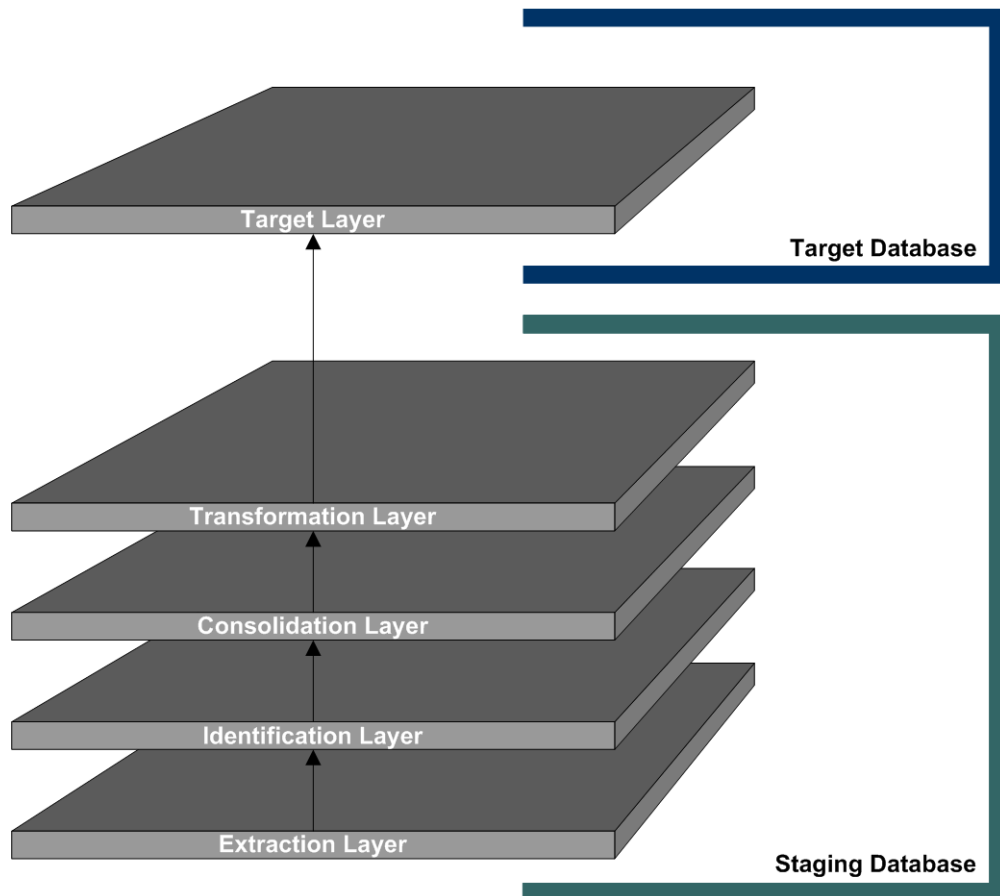


## ETL Staging Layers

IT Analytics moves data through four staging layers in the staging database before loading the data into the target layer of the target database. Each staging layer is a subsystem that functions independently. Data passes through the Extraction, Identification, Consolidation, and Transformation staging layers.

Figure 5 shows how data moves through the staging layers and into the target layer.

**Figure 5 ETL Staging Layers**



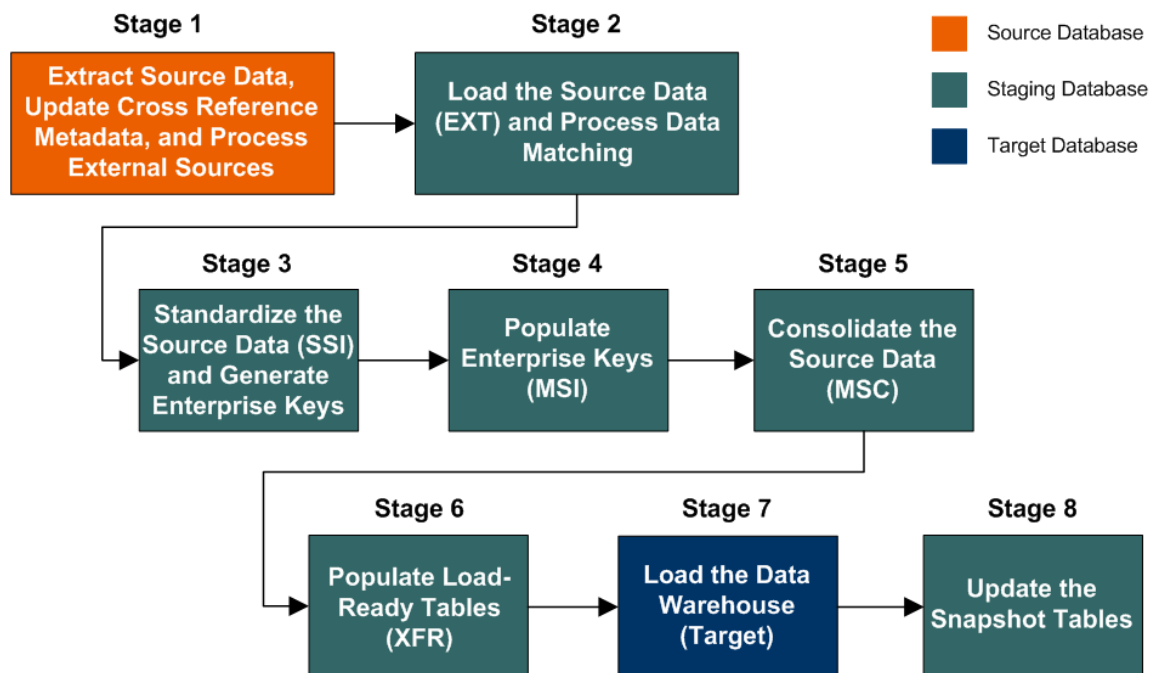
# ETL Stages

There are eight stages of the ETL process. The stages that process depend on metadata and the entity being transformed.

- Stage 1: Extract Source Data, Update Cross Reference Metadata, and Process External Sources on page 25
- Stage 2: Load the Source Data (EXT) and Process Data Matching on page 27
- Stage 3: Standardize the Source Data (SSI) and Generate Enterprise Keys on page 30
- Stage 4: Populate Enterprise Keys (MSI) on page 32
- Stage 5: Consolidate the Source Data (MSC) on page 33
- Stage 6: Populate Load-Ready Tables (XFR) on page 35
- Stage 7: Load the Data Warehouse (Target) on page 37
- Stage 8: Update the Snapshot Tables on page 39

Figure 6 shows the eight stages of ETL processing.

**Figure 6 ETL Stages**



## ETL Processes

Each stage of ETL comprises one or more processes.

Stage 1: Extract Source Data, Update Cross Reference Metadata, and Process External Sources:

- [Process 1: Check ETL Preconditions](#) on page 25
- [Processes 2 and 3: Extract the Source Data](#) on page 25
- [Process 4: Update Cross Reference Metadata](#) on page 26
- [Process 5: Process External Sources](#) on page 26

Stage 2: Load the Source Data (EXT) and Process Data Matching:

- [Processes 6 and 7: Load the Extraction Tables](#) on page 27
- [Process 8: Data Matching](#) on page 28

Stage 3: Standardize the Source Data (SSI) and Generate Enterprise Keys:

- [Processes 9 and 10: Single Source Identity](#) on page 30
- [Process 11: Load Enterprise Cross Reference Tables](#) on page 31
- [Processes 12 and 13: Generate Enterprise Keys](#) on page 31

Stage 4: Populate Enterprise Keys (MSI):

- [Processes 14 and 15: Multiple Source Identity](#) on page 32

Stage 5: Consolidate the Source Data (MSC):

- [Processes 16 and 17: Single Source Consolidation](#) on page 33
- [Processes 18 and 19: Backfill Keys](#) on page 34
- [Processes 20 and 21: Backfill MSI Tables](#) on page 34
- [Process 22: Multiple Source Consolidation](#) on page 34

Stage 6: Populate Load-Ready Tables (XFR):

- [Process 23: Prepare Dimensions for Load to Target](#) on page 35
- [Process 24: Prepare Fact Tables for Load to Target](#) on page 36

Stage 7: Load the Data Warehouse (Target):

- [Process 25: Load from Staging to Target](#) on page 38
- [Process 26: Life Cycle Management](#) on page 38

Stage 8: Update the Snapshot Tables:

- [Process 27: Consolidated Snapshot](#) on page 39
- [Process 28: Single Source Snapshot](#) on page 39
- [Process 29: Transactional Snapshot](#) on page 39
- [Process 30: Check ETL Postconditions](#) on page 39



# Stage 1: Extract Source Data, Update Cross Reference Metadata, and Process External Sources

Significant processing occurs in the first stage of the ETL process. In Stage 1, the system runs a check to determine whether any ETL preconditions are met, and IT Analytics extracts the source data, updates the cross reference metadata, and processes the external source spreadsheets.

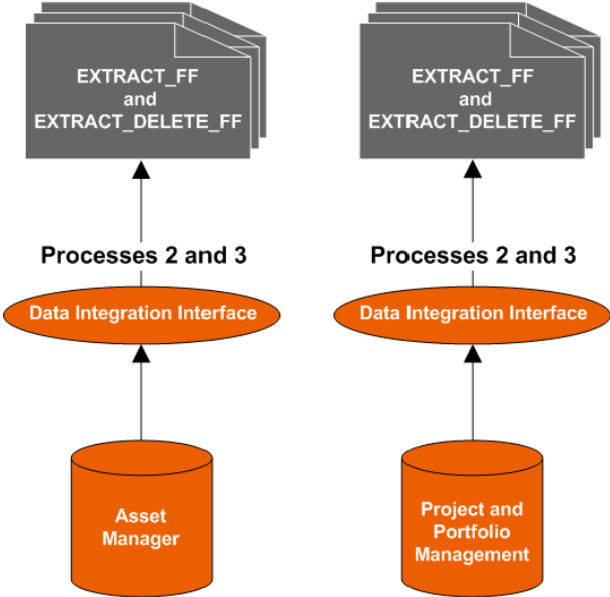
## Process 1: Check ETL Preconditions

The Check ETL Preconditions (SYS\_READY) process is the first process to run. This process marks the start of ETL and can be customized to check whether any preconditions are met before ETL begins.

## Processes 2 and 3: Extract the Source Data

The Extract PPM to Flat Files (PPM\_SOURCE\_EXTRACT) and Extract AM to Flat Files (AM\_SOURCE\_EXTRACT) processes extract delta data from the source systems and write it to delimited flat files on the file system. IT Analytics accesses data through pre-defined database views.

**Figure 7 Data Extraction Processing**



For information about the database views, see [Pre-defined Source Views](#) on page 41.

## Extraction Criteria

The data warehouse uses the Change Data Capture (CDC) technique for extracting data from the source tables. IT Analytics extracts data if the last modification date is greater than or equal to the maximum last modification date value.

For the initial extraction, the default beginning date is *1/1/1900*. Because the last modification date is later than 1/1/1900, SAP BusinessObjects Data Services extracts all data from the source tables. After the initial extraction, the date range depends on the maximum last modification date value in the Transactional Snapshot (TSNP) tables for the most recent successful ETL run. By using the maximum last modification date value, IT Analytics extracts only data inserted or updated since the last successful ETL run.

## Flat Files

Flat files are used to decouple the source system database from the target database. By decoupling the two databases, no process has an open connection to the source system at the same time as it has an open connection to the target system. This ensures that processing on the target system does not affect the source system.

Using flat files also enables the data warehouse to access the source system history in the form of delta records. Using flat files also enables the data warehouse to recover data to a point in time after the initial load.

There are two types of flat files:

### Extract Flat Files

Extract Flat Files (EXTRACT\_FF) house records that were updated or inserted since the last time ETL ran. These flat files contain copies of the source view structure.

### Extract Delete Flat Files

Extract Delete Flat Files (EXTRACT\_DELETE\_FF) are used to identify source system deletes. These flat files contain the complete list of business keys from the source view. This list gets compared to the Transactional Snapshot (TSNP) to determine which records have been deleted from the source. If the key existed in the previous run but not in the current extraction flat file, IT Analytics flags the record for that key as a deleted record.

## Process 4: Update Cross Reference Metadata

The Cross Reference Initial Synchronization (ENTERPRISE\_CONF) process updates the metadata, such as consolidation level, priority, and source system information, that is stored in the Cross Reference (XREF) tables. IT Analytics uses this information to complete data matching in [Process 8: Data Matching](#) on page 28.

## Process 5: Process External Sources

The Extract External Files (EXTERNAL\_FILE) process converts the data from the Budget, Cost Category, Currency, Exchange, IT Function, Period, and Program external source spreadsheets into dimension format and then loads the data into the target layer of the data warehouse.

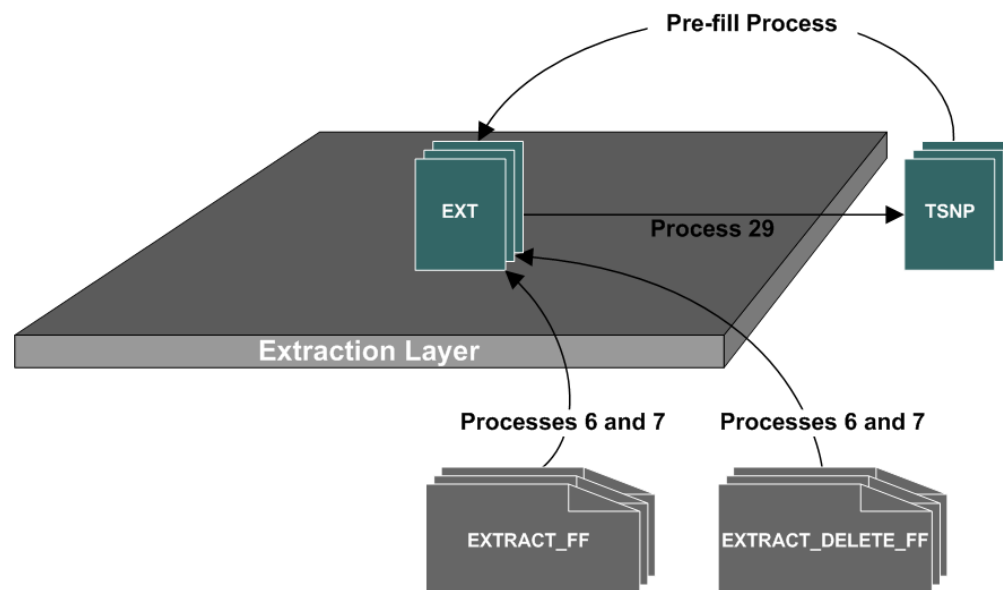
## Stage 2: Load the Source Data (EXT) and Process Data Matching

In Stage 2, SAP BusinessObjects Data Services loads the source data from the flat files generated during Stage 1 into the Extraction (EXT) tables in the data warehouse staging database. In the Extraction layer, records exist in database form.

### Processes 6 and 7: Load the Extraction Tables

The Load PPM Flat Files (PPM\_EXT) and Load AM Flat Files (AM\_EXT) processes extract all of the data from the flat files created in Stage 1 and load the data into the EXT tables in the data warehouse staging database. These processes also pre-populate or pre-fill the EXT tables and perform data validation.

**Figure 8 Extraction Layer Processing**



### Pre-fill the Extraction Tables

Because the ETL process is driven from delta records, some tables that are involved in joins may have incomplete join criteria. For example, if there is a two table join and only one table has a delta record for that day, the join cannot be completed. To ensure that tables have all of the data available to complete joins, the pre-fill process loads Transactional Snapshot (TSNP) records into the EXT tables that are involved in joins. The TSNP tables contain a version of the source data as it existed at the beginning of the previous ETL cycle. The data that SAP BusinessObjects Data Services loads in the EXT tables during the pre-fill process is flagged as pre-fill records.

## Data Validation

IT Analytics processes the data from the flat files through a validation transformation to ensure that the data is usable. As part of the data validation, the system looks for issues such as null business keys, invalid numbers, and invalid date formats.

Some of the records with invalid data may have default values that allow the data to be processed. For example, IT Analytics may set an invalid date format to an early date, such as 01/01/1900, if it is not vital for the integrity of the data. Records that have errors that prevent SAP BusinessObjects Data Services from processing the data are written to the Validation Failed (VALF) tables.

For more information, see [Data Validation](#) on page 47.

## Load the Extraction Tables

After the pre-fill process and data validation completes, SAP BusinessObjects Data Services loads the source data flat files (EXTRACT\_FF and EXTRACT\_DELETE\_FF) into the Extraction (EXT) tables.

IT Analytics compares the business keys in EXTRACT\_DELETE\_FF to the business keys in the Transactional Snapshot (TSNP) tables. If a key exists in the TSNP tables but not in EXTRACT\_DELETE\_FF, the associated record is identified as a deleted record.

## Process 8: Data Matching

The Cross Reference Generator (XREF\_GENERATOR) process completes data matching. It relates similar entities that exist in different source systems. For example, employee *1000*, defined as *John Doe* in the amEmplDept table in AM is the same as user *5200*, defined as *jdoe* in the KNTA\_USERS table in PPM. In order to identify John Doe as the same entity in the two applications, the entities must be cross referenced.

XREF\_GENERATOR populates the Cross Reference Input (XXX\_XREF\_INPUT, where XXX is the entity name) tables with the relationship information. IT Analytics uses this data in [Process 11: Load Enterprise Cross Reference Tables](#) on page 31.

## Source Interface Reference Views

The entities are cross referenced using a source interface reference view. This view comprises the source table name, business key, last modification date, and a reference qualifier column. Each cross reference entity has two reference views, one for AM and one for PPM. The views are defined in SQL scripts.

### Reference Qualifiers

The definition of the reference qualifier column varies by view and can be built using data from multiple tables in a source. The reference qualifier column links an entity in AM with its corresponding entity in PPM. Once the entities are linked, the XREF\_GENERATOR process assigns both records the same enterprise key. In subsequent ETL processing, IT Analytics consolidates the data from these records based on the consolidation priorities.

[Table 2](#) lists the data that is used as a reference qualifier for each source for each cross reference entity.

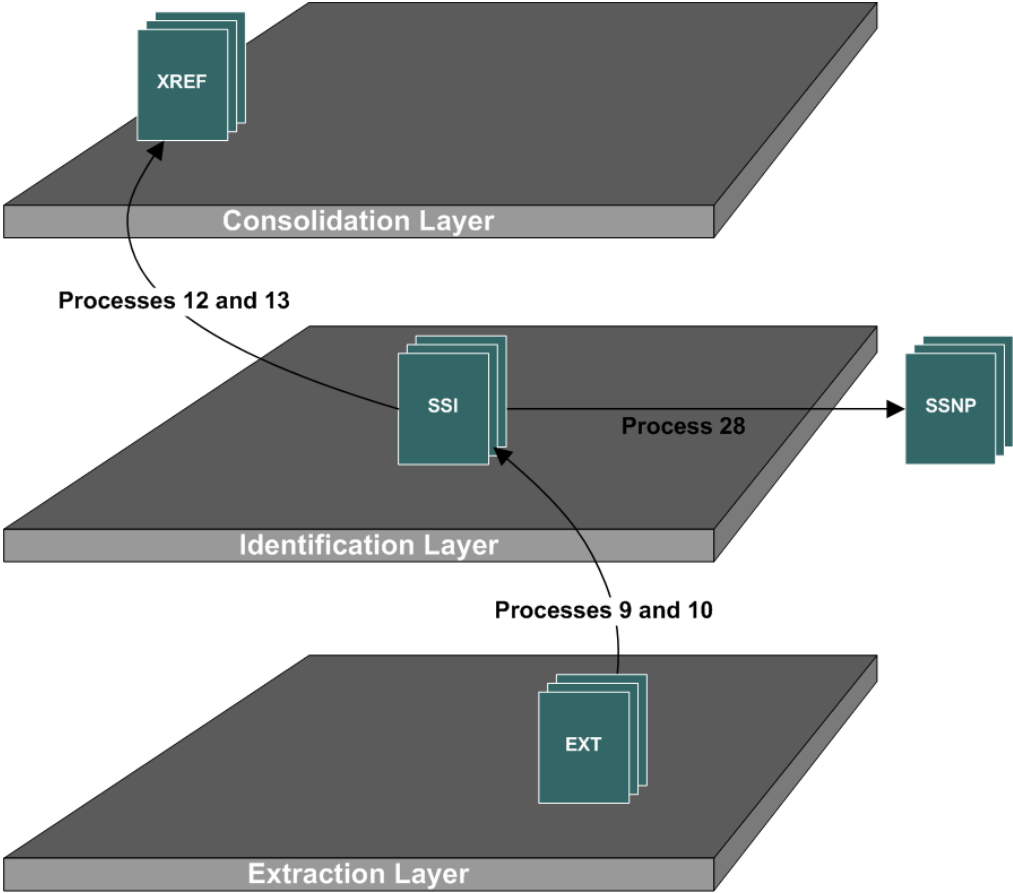
**Table 2 Reference Qualifier Definitions**

Entity	Source	View Name	Table and Column
Actual Cost	AM	AM_AMEXPENSELINE_REF_V	amExpenseLine.Title amExpenseLine.dBilling
	PPM	PPM_ACTUALCOST_REF_V	KCRT_FG_PFM_PROJECT.PRJ_PROJECT_ID KCRT_FG_PFM_ASSET.REQUEST_ID KCST_BUDGET_PERIOD_SUM.PERIOD_ID (KNTA_PERIODS.START_DATE)
Budgetline	AM	AM_AMBUDGLINE_REF_V	amBudget.Name amBudgetCategory.Name
	PPM	PPM_KCST_BUDGET_LINES_REF_V	KCST_BUDGETS.BUDGET_NAME KCST_BUDGET_LINES.EXPENSE_TYPE_CODE KCST_BUDGET_LINES.CATEGORY_CODE
Location	AM	AM_AMLOCATION_REF_V	amLocation.Name
	PPM	PPM_KNTA_REGIONS_REF_V	KNTA_REGIONS.REGION_NAME
Organization	AM	AM_AMORG_REF_V	amEmplDept.Name amCompany.Name
	PPM	PPM_KRSC_ORG_UNITS_REF_V	KRSC_ORG_UNITS.ORG_UNIT_NAME
Person	AM	AM_AMPERSON_REF_V	amEmplDept.Name amContact.Name
	PPM	PPM_KNTA_USERS_REF_V	KNTA_USERS.USERNAME
Project	AM	AM_AMPROJECT_REF_V	amProject.Code
	PPM	PPM_PM_PROJECTS_REF_V	PM_PROJECTS.PROJECT_ID

# Stage 3: Standardize the Source Data (SSI) and Generate Enterprise Keys

Because source system data structures vary among application sources, the differences in the data structures must be resolved before IT Analytics can load the data into the target database. In Stage 3, IT Analytics transforms these diverse data structures into standard structures and generates enterprise keys.

**Figure 9 SSI Identification Layer Processing**



## Processes 9 and 10: Single Source Identity

The PPM Single Source Identity (PPM\_SSI) and AM Single Source Identity (AM\_SSI) processes restructure the diverse source system data into a conformed structure. The result is a set of Single Source Identity (SSI) tables that contain the conformed data.

## Process 11: Load Enterprise Cross Reference Tables

The Load Cross Reference Table (ENTERPRISE\_XREF) process updates the Enterprise Cross Reference (XXX\_ENTERPRISE\_XREF, where XXX is the entity name) tables with the information for records with relationships that were defined in [Process 8: Data Matching](#) on page 28. ENTERPRISE\_XREF loads data from the XXX\_XREF\_INPUT tables to the XXX\_ENTERPRISE\_XREF tables.

## Processes 12 and 13: Generate Enterprise Keys

The PPM Single Source Identity Cross Reference (PPM\_SSI\_XREF) and AM Single Source Identity Cross Reference (AM\_SSI\_XREF) processes add to the XXX\_ENTERPRISE\_XREF tables the business keys for any new records that were created in [Processes 9 and 10: Single Source Identity](#) on page 30. These new records will have business keys but no associated enterprise key in the Enterprise XREF table. The SSI\_XREF processes examine records in the SSI tables to determine which business keys do not have an associated enterprise key. IT Analytics then creates a record in the XXX\_ENTERPRISE\_XREF table to generate the enterprise key.

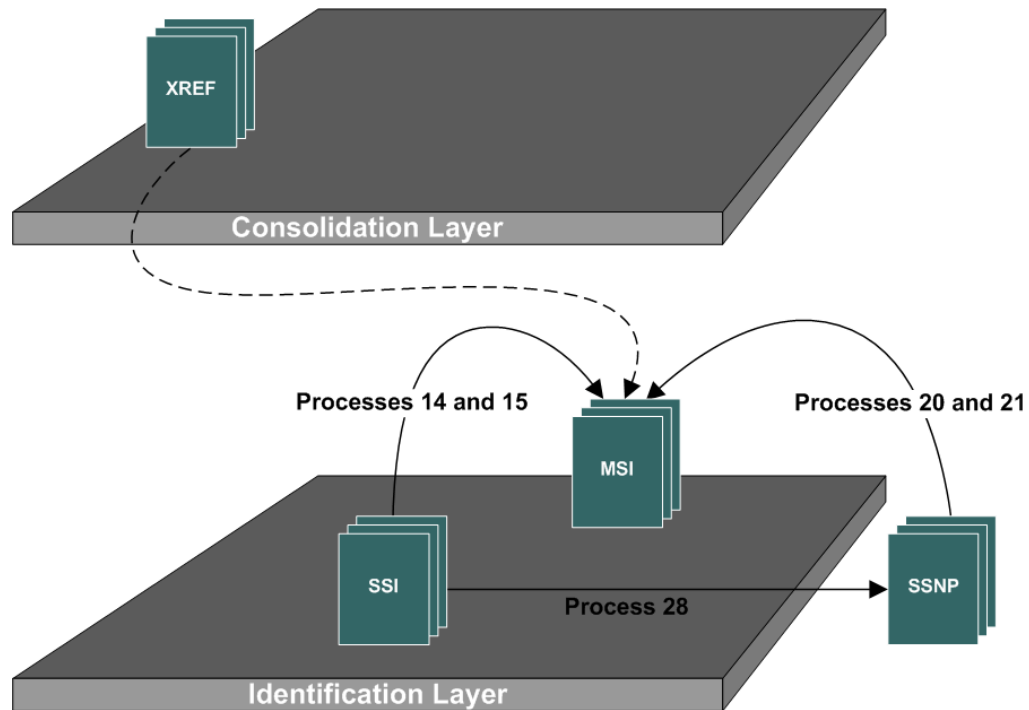
### Enterprise Key

Since the data warehouse uses data from multiple source systems that are not required to use the same business keys for the same entities, IT Analytics must create an internal key to link entity records. This internal key is the enterprise key.

## Stage 4: Populate Enterprise Keys (MSI)

During Stage 4, IT Analytics takes the data from [Stage 3: Standardize the Source Data \(SSI\)](#) and [Generate Enterprise Keys](#) on page 30 and populates it with the enterprise keys to form the Multiple Source Identity (MSI) tables.

**Figure 10 MSI Identification Layer Processing**



### Processes 14 and 15: Multiple Source Identity

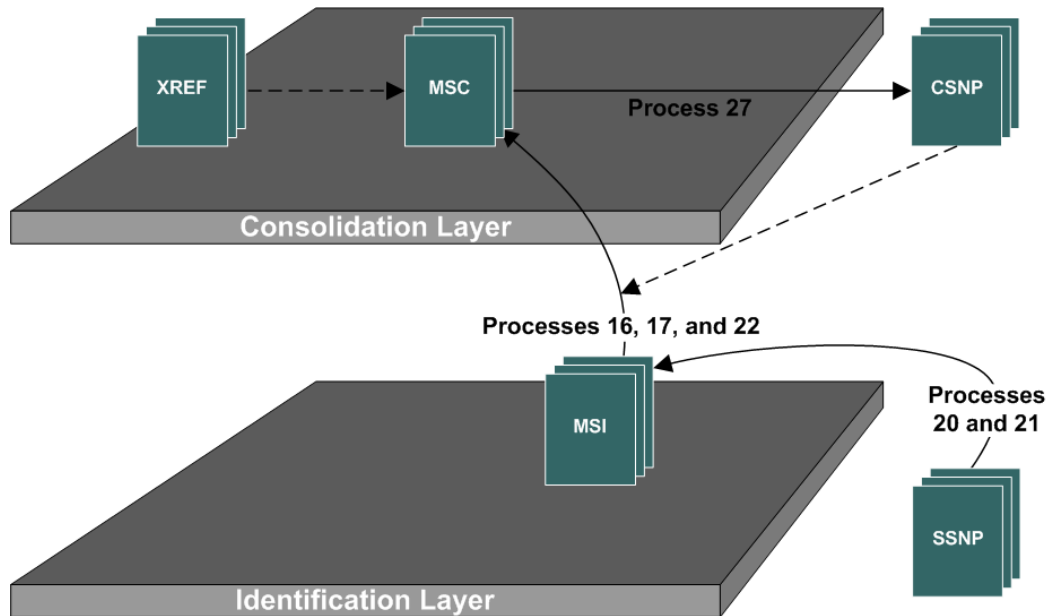
The PPM Multiple Source Identity (PPM\_MSI) and AM Multiple Source Identity (AM\_MSI) processes copy the data in the SSI tables and add it to the MSI tables. The MSI processes then look up the enterprise keys from the XXX\_ENTERPRISE\_XREF tables and populate the enterprise key columns in the MSI tables. The enterprise key enables IT Analytics to consolidate records in [Stage 5: Consolidate the Source Data \(MSC\)](#) on page 33. The MSI tables match the format of the SSI tables with the addition of the enterprise key.



## Stage 5: Consolidate the Source Data (MSC)

In Stage 5, the Multiple Source Consolidation (MSC) process combines MSI tables into a single consolidated entity. IT Analytics accesses priority information in the cross reference metadata tables to determine which data should take precedence in the MSC tables. The MSC tables contain the uniform structure that IT Analytics uses to load the data into the target tables.

**Figure 11 Consolidation Layer Processing**



### Source Data Priorities

Data priorities identify which source system value has precedence during multiple source data consolidation if more than one source has the same information. By default, data from AM takes precedence over data from PPM. Priorities can be defined at two levels:

- Entity level: All attributes in an entity inherit the priorities of the entity.
- Attribute level: Enables you to set the priority for each attribute separately. Attribute-level priorities override the entity-level priorities.

### Processes 16 and 17: Single Source Consolidation

The PPM Consolidation (PPM\_CON) and AM Consolidation (AM\_CON) processes extract data from the MSI tables and load it into the MSC tables.

IT Analytics flags records as *inserted*, *updated*, or *deleted*. The flag type is determined by comparing each record with the previous version of the record stored in the Consolidated Snapshot (CSNP) tables.

## Backfill the MSI Tables

If one source for an MSI record is updated during a cycle, and another related source does not get updated, only the updated record from the first source can be inserted. If the other source is left unaccounted for, the join between source MSI tables remains incomplete. To ensure that all joins can be completed, IT Analytics backfills the MSI tables using data from a view of the enterprise keys and MSI tables.

### Processes 18 and 19: Backfill Keys

The PPM Backfill Keys (PPM\_BACKFILL\_CTRL) and AM Backfill Keys (AM\_BACKFILL\_CTRL) processes prepare the data so that the backfill can be completed.

IT Analytics reads all of the enterprise keys from the MSI tables from a backfill view. The view that contains the list of keys obtains from the Enterprise XREF table the business keys for all of the missing sources. IT Analytics then loads the business key and enterprise key pairs into control tables.

### Processes 20 and 21: Backfill MSI Tables

The PPM Backfill (PPM\_BACKFILL) and AM Backfill (AM\_BACKFILL) processes complete the backfill of the MSI tables.

IT Analytics joins the control tables created in [Processes 18 and 19: Backfill Keys](#) to the Single Source Snapshot (SSNP) tables. The SSNP tables are snapshots of the complete SSI data set as of the previous ETL cycle. All of the missing records identified in [Processes 18 and 19: Backfill Keys](#) will be pulled from the SSNP tables and loaded into the MSI tables. This completes the backfill process and guarantees that the system has a complete set of data between all sources regardless of which source changed.

## Process 22: Multiple Source Consolidation

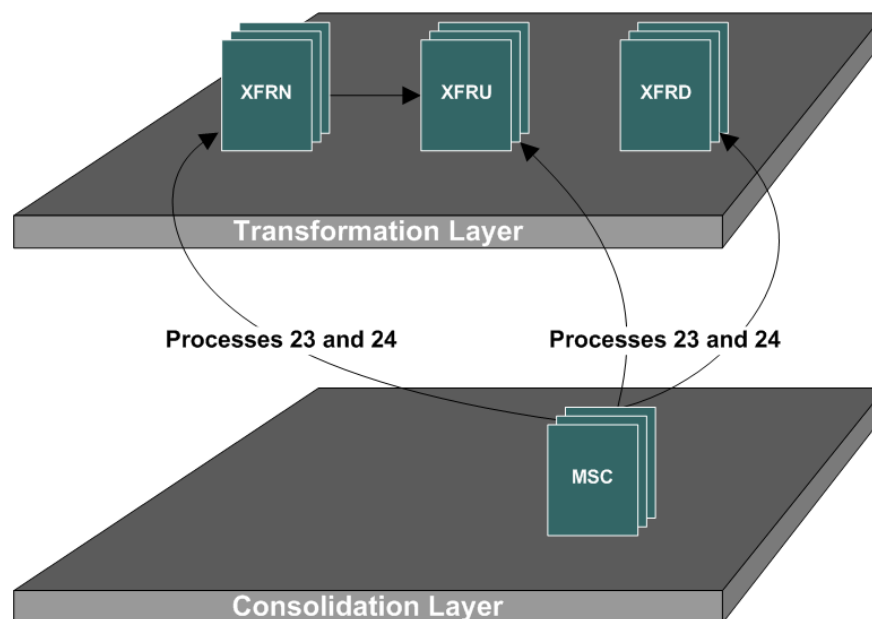
The Multiple Source Consolidation (MS\_CON) process combines MSI tables into a single consolidated entity with information from all source system inputs.

To ease ETL maintenance, the MS\_CON process uses the consolidated database view of the MSI tables. As part of the view, a flag attribute (MD\_FLAG) determines if a record is a deleted, updated, or new record. The records that IT Analytics identifies as delete records are loaded directly to the MSC tables flagged as deleted records. IT Analytics compares records that are not flagged as delete to the Consolidated Snapshot (CSNP) tables to determine if they are updated records or new records.

## Stage 6: Populate Load-Ready Tables (XFR)

The Transformation layer contains load-ready tables that match the internal structure of the data warehouse target tables. In Stage 6, IT Analytics transforms data into the format of the target tables and populates the load-ready target staging tables.

**Figure 12 Transformation Layer Processing**



### Process 23: Prepare Dimensions for Load to Target

The Prepare Dimension Load (XFR\_DIM) process extracts data from the MSC tables, converts it into a dimension format that is compatible with the target database, and loads it into the load-ready target staging tables.

IT Analytics separates records by their flagged value: deleted, inserted, or updated. It does the following:

- Stores deleted records in the Load-Ready Delete (XFRD) tables. The XFRD tables are used to apply a logical delete to the dimension table in the target layer.
- Stores insert records in the Load-Ready New (XFRN) tables.
- Compares the update records to the Consolidated Snapshot (CSNP) tables to determine if the update is a slowly changing dimension (SCD) type 1 or type 2 change.
  - A type 1 SCD record goes directly into the XFRU table.
  - A type 2 SCD record becomes two separate records. The first record goes into the XFRN tables. The second record goes to the XFRU tables. This XFRU record will expire the current active record in the target table. The XFRN record will become the active record for that durable key.

The data warehouse layer tracks the history of both type 1 and type 2 SCDs.

## Process 24: Prepare Fact Tables for Load to Target

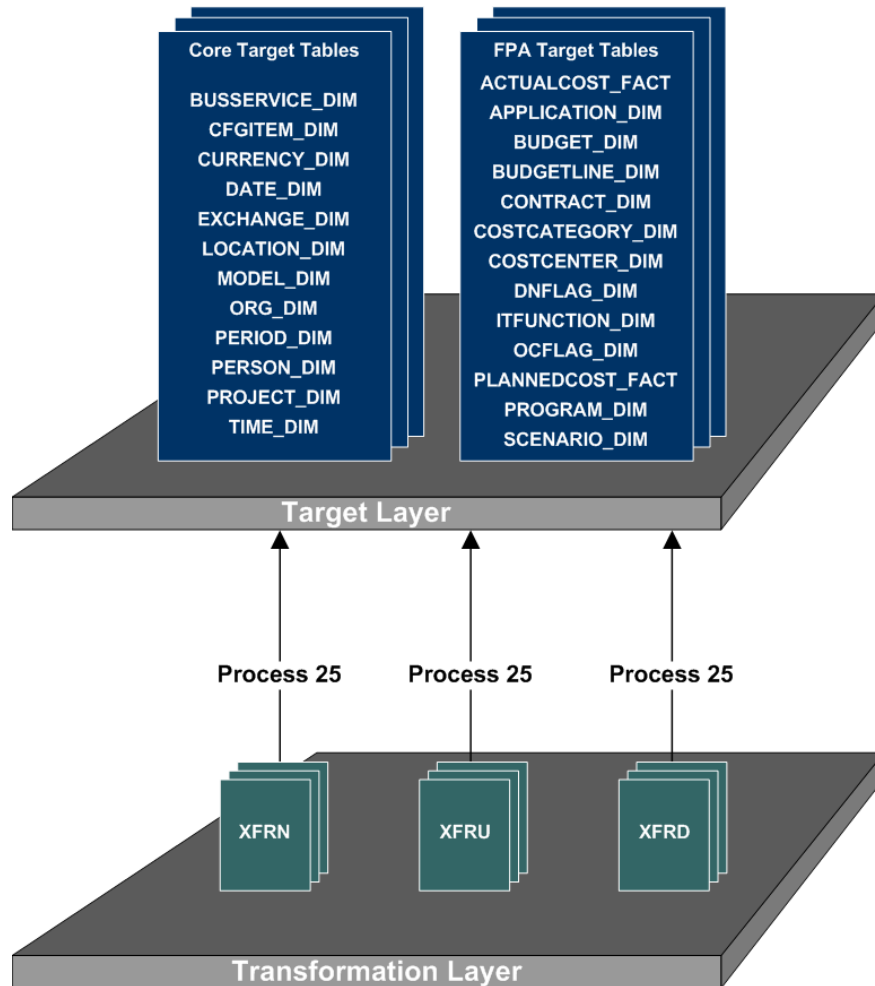
The Prepare Fact Load (XFR\_FACT) process extracts data from the MSC tables, converts it into a fact format that is compatible with the target database, and loads it into the load-ready target staging tables. IT Analytics separates records by their flagged value: deleted, inserted, or updated. IT Analytics stores deleted records in the XFRD tables, insert records in the XFRN tables, and update records in the XFRU tables. XFR\_FACT also calculates fact measures and populates foreign keys.

## Stage 7: Load the Data Warehouse (Target)

The data warehouse target layer contains the data that IT Analytics uses for reporting purposes. The data resides in standard fact and dimension tables. The data warehouse tracks the history of the dimension table contents for administrative reports. In Stage 7, IT Analytics adds new records, updates existing records, and flags deleted records in the target layer fact and dimension tables.

For more information about the fact and dimension target tables, see [Data Models](#) on page 71.

**Figure 13 Load Data to the Target Layer**



## Process 25: Load from Staging to Target

The Load from Staging to Target (DW) process extracts data from the load-ready target staging tables and loads it into the target dimension and fact tables. IT Analytics adds expired flags to deleted records and inserts new records with active flags.

To maintain data integrity, deletes are processed first. For deleted records, IT Analytics takes data from the XFRD tables, locates the record in the fact and dimension tables, applies a delete for that record, and sets the active flag to *N*.

After processing the deletes, IT Analytics processes the SCD type 1 updates to ensure that new records are not marked non-active. IT Analytics loads the records from the XRFU tables into the target fact and dimension tables as normal updates. The updates are SCD type 1 changes, and the expired active records are SCD type 2 changes.

For inserted records, IT Analytics loads the records from the XFRN tables into the target fact and dimension tables. These records are the new records from normal inserts and the new SCD type 2 records.

## Process 26: Life Cycle Management

The Target Durable Key Updates (DW\_DK\_LFCL) process performs the life cycle management updates. If there are changes such as updates to cross reference relationships, priorities, or records that need to be re-consolidated, this process updates the appropriate consolidation and target models.

IT Analytics updates or deletes all of the consolidation models without history. IT Analytics updates any target models as SCD type 1 or SCD type 2, depending on the changed attributes and target table types.

## Stage 8: Update the Snapshot Tables

Updating the snapshot tables is part of a maintenance process that occurs at the end of an ETL cycle. After loading new records or updates to the data warehouse, IT Analytics must update the snapshot tables for use during the next ETL cycle. There are three processes that run in Stage 8 to update the snapshot tables.

### Process 27: Consolidated Snapshot

The Consolidated Snapshot (CSNP) process extracts data from the MSC tables and loads it into the CSNP tables.

IT Analytics flags records in the consolidation layer as inserts, updates, and deletes. In order to determine what flag should be associated with which record, IT Analytics must compare the current data to a previous version of the data. The CSNP tables contain the previous version of the data and are used to do the comparison. CSNP tables are complete sets of data for all MSC tables. The data exists in the state it was in as of the end of the previous ETL run. The CSNP tables do not hold logical deletes. Deletes are applied physically to the snapshots.

### Process 28: Single Source Snapshot

The Single Source Snapshot (SSNP) process extracts data from the SSI tables and loads it into the Single Source Snapshot (SSNP) tables.

The SSNP tables contain data in the same structure as the SSI tables. It is the entire set of the data in the state it existed at the end of the previous ETL cycle. IT Analytics uses the SSNP tables as part of the process to backfill the MSI tables.

For more information, see [Backfill the MSI Tables](#) on page 34.

### Process 29: Transactional Snapshot

The Transactional Snapshot (TSNP) process extracts data from the EXT tables into the TSNP tables.

The TSNP tables contain a version of the source data as it existed at the beginning of the previous ETL cycle. The TSNP tables are used to identify deleted records. The business keys in the EXTRACT\_DELETE\_FF flat files are compared to the business keys in the TSNP tables. If a key exists in the TSNP tables but not in EXTRACT\_DELETE\_FF, the associated record is identified as a deleted record and loaded into the EXT tables. The TSNP tables are also used as part of the pre-fill process.

For more information, see [Pre-fill the Extraction Tables](#) on page 27.

### Process 30: Check ETL Postconditions

The Check ETL Postconditions (SYS\_CLOSE) process is the last process to run. This process marks the end of ETL and can be customized to check whether any ETL postconditions have been met. This process also signals to the allocation engine that ETL is complete.

For information about the allocation engine, see the *HP IT Analytics Applications User Guide*.





# 4 Data

This chapter contains information about sources and source views, the table connection types and keys used for data consolidation, data validation, and the use of historic data for ad hoc queries.

## Source Systems

The source data originates in external applications. You can extract data from HP Asset Manager (AM) and HP Project and Portfolio Management (PPM). These source systems may reside in remote databases managed by an Oracle or Microsoft SQL Server RDBMS. IT Analytics does not connect directly to source systems, but accesses data through pre-defined source views.

### Pre-defined Source Views

A view is created for each source table. In addition to dimension and measure data, each view has a business key and a foreign key.

- The business key is the unique key for the table. If the source table also has a composite key, it will be concatenated with the business key. For example, the source application might have a department table that contains values for both a company and a department. If the company is *HP* and the department is *Sales*, the business key becomes *HP|Sales*.
- If the view is for a child table for a single source, the foreign key contains a value that maps to the business key of the internal parent table. For example, the business key for the *amAsset* view is *lAstId*, and the foreign key is *lPortfolioItemId*.

For more information about keys, see [Keys Used for Consolidation](#) on page 46.

# External Sources

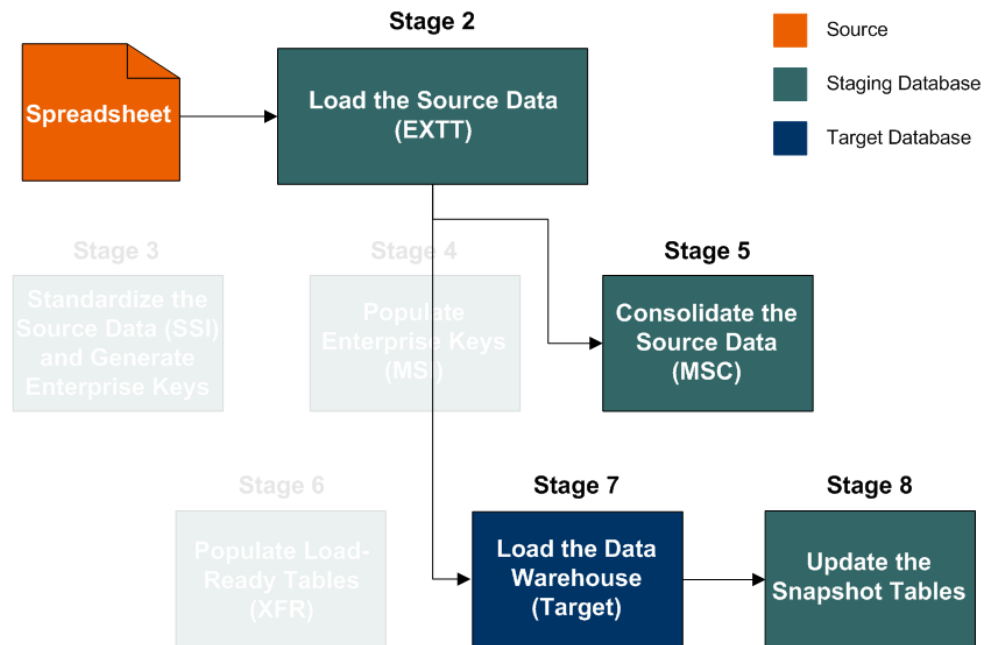
External sources use Excel spreadsheets as a platform. The spreadsheets are used when the data needs to be manually consolidated or must be taken from a source other than AM or PPM. The external file format is based on the consolidation model that it supports. Each record in the spreadsheet contains source and consolidation information. The source information includes the source primary key, source name, and source attributes used to match the consolidated records. The consolidation information includes the consolidated attributes for the consolidation model.

For information about editing external source spreadsheets, see the *HP IT Analytics Data Warehouse Administration Guide*.

## External Source Data Path

In the ETL process for external source spreadsheets, IT Analytics imports some spreadsheets into the Multiple Source Consolidation (MSC) stage and some spreadsheets directly into the target tables. The data in external source spreadsheets comes from a manually created file that matches the target structure.

**Figure 14 External Source ETL Data Flow**



## External Source Entities

The Application, Budget, Cost Category, Currency, Exchange, IT Function, Period, and Program entities require external source spreadsheets.

### Application

Applications represent large scale implementations of software that incur costs and generate dependencies. They are used exclusively for allocation scenarios. Since AM and PPM do not contain applications, the application external source spreadsheet, APPLICATIONS.xls, does not have source columns. Applications are manually entered by name.

For information about allocation scenarios, see the *HP IT Analytics Applications User Guide*.

### Budget

Budgets represent high-level containers for planned costs as defined in the source applications. The Budget external source spreadsheet, BUDGETS.xls, consolidates the data from the Budget table *amBudgets* in AM and the Budget table *KCST\_BUDGETS* in PPM.

### Cost Category

Cost categories are descriptive groups used to sort planned and actual costs. The cost category external source spreadsheet, COSTCATEGORY.xls, consolidates the data from the Cost Category table *amCostcategory* and the Budget Category table *amBudgetCategory* in AM and the Budget Category definitions *KNTA\_LOOKUPS* in PPM. *KNTA\_LOOKUPS* is a multi-purpose storage area. PPM does not have a cost category concept.

### Currency

Currencies represent all supported national currencies used for display and conversion. The currency external source spreadsheet, CURRENCY.xls, is populated with the currencies the data warehouse supports. CURRENCY.xls determines the base currency as well as the display names for the currencies.

### Exchange

The exchange external source spreadsheet, EXCHANGE.xls, represents conversions between currencies. The conversions are valid for a particular date range, and they may be actual or projected. Exchange rates can be updated through a subscription service provided by a number of financial sites.

### IT Function

IT functional domains are used exclusively for allocation scenarios and represent high-level divisions of your IT operations that may differ from traditional divisions, such as departments or locations. Since AM and PPM do not contain IT functional domains, the IT function external source spreadsheet, ITFUNCTIONS.xls, does not have source columns. IT functions are manually entered by name.

For information about allocation scenarios, see the *HP IT Analytics Applications User Guide*.

## Period

Periods represent fiscal calendar months used in financial planning. The period external source spreadsheet, PERIOD.xls, consolidates the data from the Period Definition tables *amPeriod*, *amFYDivision*, and *amFinancialYear* in AM and the Periods table *KNTA\_PERIODS* in PPM.

## Program

Programs are used exclusively for allocation scenarios and represent large-scale groups of resources and projects. Since AM and PPM do not contain programs, the program external source spreadsheet, PROGRAMS.xls, does not have source columns. Programs are manually entered by name.

For information about allocation scenarios, see the *HP IT Analytics Applications User Guide*.

# Consolidating Data

During the ETL process, IT Analytics can consolidate the data that comes from multiple tables within a source as well as data that comes from multiple sources. To accomplish the consolidation, IT Analytics follows data priorities and uses different types of table connections and keys.

## Data Priorities

Data priorities identify which source system value has precedence during multiple source data consolidation if more than one source has the same information. By default, data from AM takes precedence over data from PPM. Priorities can be defined at two levels:

- Entity level: All attributes in an entity inherit the priorities of the entity.
- Attribute level: Enables you to set the priority for each attribute separately. Attribute-level priorities override the entity-level priorities.

## Table Join Types

There are three types of connections the data warehouse uses to consolidate tables. The type of connection depends on the characteristics of the tables and whether it is a single source or multiple source consolidation.

- Tables from a single data source where one table is considered the parent or primary table are consolidated with a join. In a join, all records from the parent table are imported, and records that have a matching foreign key are imported from child tables.
- Tables that exist across multiple data sources and where reconciliation is done are consolidated with a merge. In a merge, all records from all tables are imported, and reconciliation is done where appropriate. Merges join tables where they overlap and also import all other records.
- Tables that occur in either single or multiple data sources and have records that are assumed to be equal and independent are consolidated with a union. In a union, all records from all sources are imported but are not reconciled. They are simply added to the rowset.

# Keys Used for Consolidation

Several types of keys get stored in the data warehouse for use during consolidation.

## Business Key

Business keys are the field or fields that make up the unique value for each record in the source application.

## Composite Key

A composite key is a primary key that is made up of two or more columns. Composite keys are also known as concatenated keys or aggregate keys.

## Durable Key

The durable key is a key that remains constant throughout the life of an entity. It is a simple, sequentially assigned integer with no structure or semantics embedded in the key value.

## Enterprise Key

The enterprise key is used to generate the durable key in the target entities.

## Foreign Key

A foreign key is the field in a table that refers to the primary key of another table in the same data source. Every relationship in the model must be supported by a foreign key. Foreign keys enforce referential integrity by completing an association between two entities.

## Primary Key

Primary keys are the field or fields that make up the unique value for each record. In the source application, this is represented by the business key. In the data warehouse, it is the surrogate key.

## Surrogate Key

A surrogate key is a primary key for the dimension tables. It is the field that makes up the unique value for each record in the data warehouse. Surrogate keys are numeric and are controlled by the ETL process.

## Data Validation

To ensure all of the data is usable, data validation is done as part of the process to load the Extract (EXT) tables. As part of the data validation, IT Analytics looks for issues such as null business keys, invalid numbers, and invalid date formats.

Some of the records with invalid data may have default values that allow the data to be processed. For example, IT Analytics may set an invalid date format to an early date, such as *01/01/1900*, if it is not vital for the integrity of the data. Records that have errors that prevent SAP BusinessObjects Data Services from processing the data are written to the Validation Failed (VALF) tables.

By default, IT Analytics validates the source tables that contain the data that is used to populate the cost fact tables, ACTUALCOST\_FACT and PLANNEDCOST\_FACT. The source tables are AM\_AMBUDGLINE, AM\_AMCOSTCENTER, and AM\_AMEXPENSELINE from Asset Manager and PPM\_KCST\_BUDG\_LINE\_CELLS, PPM\_KCST\_BUDG\_PERIOD\_SUM, PPM\_KCST\_CURRENCY\_LINES, and PPM\_KCST\_CURRENCY\_CELLS from Project and Portfolio Management. [Table 3](#) lists the validation rules for these tables. Using the VALIDATION\_REF file, data validation can be turned on and off for each source table.

**Table 3 Default Validation Rules**

Source and Target Table	Column	Validation Rule
Source: AM_AMBUDGLINE Target: PLANNEDCOST_FACT	BUDGETCUR	Must be one of the currencies in CURRENCY_DIM.
	LPERIODID	Must be in the SRC_PRIMARY_KEY column in the PERIOD.xls external source spreadsheet when the source (SOURCE_NAME) is AM.
	MBUDGET	Must not be zero.
Source: AM_AMCOSTCENTER Target: ACTUALCOST_FACT	MD_BUSINESS_KEY	Must not be NULL.
Source: AM_AMEXPENSELINE Target: ACTUALCOST_FACT	DBILLING_LOC	Must not be NULL and must be a valid year (in the range of AM timezone year).
	DEBITCUR	Must be one of the currencies in CURRENCY_DIM.
	MDEBIT	The value in this field minus the value in MCREDIT must be greater than zero.
Source: PPM_KCST_BUDG_LINE_CELLS Target: PLANNEDCOST_FACT, ACTUALCOST_FACT	PERIOD_ID	Must be in the SRC_PRIMARY_KEY column in the PERIOD.xls external source spreadsheet when the source (SOURCE_NAME) is PPM.

**Table 3 Default Validation Rules**

Source and Target Table	Column	Validation Rule
Source: PPM_KCST_BUDG_PERIOD_SUM Target: ACTUALCOST_FACT	PERIOD_ID	Must be in the SRC_PRIMARY_KEY column in the PERIOD.xls external source spreadsheet when the source (SOURCE_NAME) is PPM.
Source: PPM_KCST_CURRENCY_CELLS Target: PLANNEDCOST_FACT, ACTUALCOST_FACT	ACTUAL_VALUE_LCL	Must be greater than zero, or PLAN_VALUE_LCL must not be zero.
	CURRENCY_CODE	Must be one of the currencies in CURRENCY_DIM.
Source: PPM_KCST_CURRENCY_LINES Target: ACTUALCOST_FACT	ACTUAL_CAPEX_LABOR_LCL	Must be greater than zero, or ACTUAL_CAPEX_NONLABOR_LCL, ACTUAL_OPEX_LABOR_LCL, or ACTUAL_OPEX_NONLABOR_LCL must be greater than zero.
	CURRENCY_CODE	Must be one of the currencies in CURRENCY_DIM.



**Note:** Some of the out-of-box demonstration data for Asset Manager does not pass the default validation rules.



## Administrative and Operational Data Reports

The data warehouse provides administrative and operational reports that enable your system administrator to view the model definitions, get insights into the metadata, and audit and troubleshoot generated schemas. The data warehouse also includes reports such as the following that provide insight into the ETL process and the data:

- Data model overview reports show an overview of the entities in the data warehouse and their relationships. These reports indicate the lineage of data from source to target.
- The Schema report diagnoses whether the ETL staging and target layers are aligned with the metadata.
- The Details report shows source, extraction, consolidation, dimension, and fact details for the entities.
- Operations reports provide measurements and metrics logged by the ETL processes as well as historical data and statistics for ETL runs.

For more information, see the *HP IT Analytics Data Warehouse Administration Guide*.

## Historical Data in Ad Hoc Queries

Several dimension tables in the data warehouse track significant historical changes to data. IT Analytics tracks these changes by using durable keys and surrogate keys. The durable key remains constant throughout the life of an entity, giving it a consistent identity over time. Surrogate keys change. They are unique to each table and allow each record to have a unique identity.

When a slowly changing dimension (SCD) type 2 field is updated, IT Analytics does not overwrite the original record in the table. Instead, each SCD type 2 change generates a new record. IT Analytics makes a copy of the record and flags the original record as historical. Both the new and old record share the same durable key, indicating that they are the same piece of data. In addition, IT Analytics gives the new record a unique surrogate key and flags the record as current.

You can view the history of a specific record by displaying all records with the same durable key. Create and transition dates denote the time when each record is valid. And the *Is Current?* flag denotes which record is the latest incarnation. By default, all reports filter on *Is Current? = 'Y'*.

Historical tracking also impacts foreign keys. All fact-to-dimension or dimension-to-dimension links are represented by both the durable key and the surrogate key. This enables you to view the related record as it exists currently or as it existed at the time the primary record was created.

By default, all joins in the IT Financial Analysis universe aim for the current version of the data. However, to enable you to view the historical version of the data for trend analysis and comparisons, the universe also contains flexible joins. By editing the report query to include the *Show Historical Information* object for the foreign key in a Web Intelligence Report, the universe links to the relevant history record.



---

# Glossary

IT Analytics uses a variety of terms to describe its functionality and objectives. Click any term to view the definition.

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [L](#) [M](#) [N](#) [O](#) [P](#) [R](#) [S](#) [T](#) [U](#) [W](#) [X](#)

## A

ABC

ABC Batch

ABC Job stream

ABC Job stream step

Ad hoc query

Aggregation

Allocation engine

Allocation scenario

Allocation stage

Analytic

Analytic application

Assignment groups

## B

Backfill record

Business driver

Business impact analysis report

Business intelligence

Business process

Business service

Business service downtime impact

## C

Catalog

Change data capture

Composite key

Configuration item

Configuration Management Database

Conformed dimension  
Consolidated model  
Core model  
Corporate information factory  
CSNP  
Crosstab  
Cube

## D

Dashboard  
Data architecture  
Data flow  
Data integration  
Data quality  
Data mart  
Data mining  
Data model  
Data quality  
Data warehouse  
Dimension  
Downstream

## E

Efficiency and effectiveness  
Enterprise key  
ETL  
Executable  
EXT  
Extract (ETL)

## F

Fact table  
Financial Planning & Analysis  
Foreign key

## G

Governance

Granularity

## H

Hierarchy

Historical data

## I

Impact analysis

Impact rules

ITIL best practices

## J

Join

## L

Load (ETL)

Log

## M

Measure

Merge

Metadata

Metric

MSC

MSI

Multi-dimensional

## N

Near-real-time

## O

OLAP

Organization

## P

Predictive analysis

Primary key

Prioritization

Process

## R

Real-time

Reconciliation

Report

Report library

Resource allocation

Risk

## S

Slowly changing dimension

Snowflake schema

SSI

SSNP

Staging

Star schema

Stream step detail

Surrogate key

## T

Transform (ETL)

TSNP

## U

Union

Universe

Upstream

## W

Work flow

X

XFR

XREF

## **ABC**

Audit, Balance, and Control (ABC). ABC is an application that triggers the ETL processes.

Audit ensures consistency during ETL processing. It measures the number of records in and the number of records out for each step and displays these runtime statistics in a collection of audit reports.

Balance verifies that data in the data warehouse matches data in the source system. For example, if the data warehouse stores project dollar amounts then the balance process verifies that the aggregate project dollars in the data warehouse matches the aggregate project dollars in the source application data.

Control governs ETL processes. Control makes sure that there is a proper restart and recovery when a system error occurs. Control also manages job dependencies at runtime.

## **ABC Batch**

An instantiated ABC job stream built from the job stream definition. The execution is sequential: one batch must finish execution before a new batch for the same job stream can start. In the case where there is both upstream and downstream ETL, an upstream batch and a downstream batch can run concurrently. Each executable batch has a unique ID.

## **ABC Job stream**

A flow of individual execution (job stream) steps and a description of related dependencies. A job stream can have one or more start steps, one or more end steps, and execute several parallel branches. A stream ID value identifies the individual stream. The ABC job stream contains one-to-many job stream steps.

## **ABC Job stream step**

An individual step that invokes a command line utility or executes an SAP BusinessObjects Data Services job. An ABC job stream step is unique within an ABC job stream.

## **Ad hoc query**

Ad hoc queries are spontaneous requests for information that are normally formulated and run a single time. Ad hoc queries are constructed with a SQL statement or by using a data access tool.

## **Aggregation**

Aggregation is the accumulation of related data into a single table.

## **Allocation engine**

The allocation engine is a process that communicates with the Financial Planning & Analysis (FPA) web application. It provides status information about allocations when changes are made to stages or rules. It selects costs, updates relationships, and splits costs across the data model.

By using allocation scenarios, stages, and rules, financial analysts can allocate the costs to compare planned versus actual costs and use the results for proper cost and budget management.

### **Allocation scenario**

An allocation scenario is a tool that enables you to analyze planned costs and actual costs by categorizing them. You can use categorized costs to compare planned (budgeted) costs to the actual costs.

### **Allocation stage**

Allocation stages are a required component of the allocation engine. They identify a dimension that is associated with a Cost fact table. You can define Actual Cost and Planned Cost stages.

### **Analytic**

Analytics are reports that summarize or aggregate data. Analytics help managers to identify trends, analyze volumes of data, plan, forecast, and make business or strategic decisions.

### **Analytic application**

Analytic applications produce a series of logically integrated interactive reports, including dashboards and scorecards. These reports support management decisions that usually involve demographic analysis, trend analysis, pattern recognition, drill-down analysis, and profiling. For example, Financial Planning & Analysis (FPA), IT Performance Analytics (ITPA), and IT Financial Management (ITFM) are analytic applications.

### **Assignment groups**

Assignment groups are lists of users who are responsible for an incident. When an incident opens or escalates, a notification mechanism alerts the users in the group.

**Example:** A service desk operator receives a request to fix a disabled workstation. The operator creates an incident and assigns it to the IT assignment group. The IT technician determines that the hard drive must be replaced. Because the drive must be purchased, the technician updates the incident and assigns it to a Materials Management assignment group for acquisition. If the hard drive purchase is delayed, and the incident cannot be closed in a reasonable amount of time, a pre-defined process can escalate the incident to the operations manager for intervention.

Incident category definitions can include a default assignment group for all incidents in that category. For example, the hardware category can list IT as the assignment group if the IT department is always the first group to handle hardware incidents.

### **Backfill record**

Backfill records are used to complete partial data entities in the change data capture (CDC) records for a given cycle. Backfill records are used to ensure that completed data is loaded into the target environment.

### **Business driver**

- 1 Business drivers are the tasks, information, and people that support and promote the goals of the enterprise.



- 2 Business drivers are the requirements that describe the business goals, for example, more quality data and faster response to queries.
- 3 Business drivers are problems that are important enough to mean the difference between success and failure for an organization.

### **Business impact analysis report**

A business impact analysis report describes potential risks to the organization. A good business impact analysis report should identify the cost of outages in terms of lost revenue, replacement of equipment, salaries, and other opportunity costs.

### **Business intelligence**

Business intelligence describes the result of in-depth analysis of detailed business data, including database and application technologies and analysis practices. Business intelligence is sometimes used synonymously with decision support but is much broader, potentially encompassing knowledge management, enterprise resource planning, data mining, and other practices.

### **Business process**

Business processes are a set of business services that support a single corporate purpose.

### **Business service**

A business service is the combination of one or more business applications plus the supporting configuration items that enable delivery of the business service to the customer. The products that IT offers its customers are internal and external business services. A business service supports zero-to-many organizations, and a business service incident usually impacts associated organizations. An organization can rely on one-to-many business services; however, an incident that affects the organization may not always affect the associated business service.

Enterprise resource planning (ERP) is a good example of a business service. ERP is a business management system that integrates all facets of the business, including planning, manufacturing, sales, and marketing. There can be many component business applications such as inventory control, order tracking, customer service, finance, and human resources.

### **Business service downtime impact**

Business service downtime impact (BSDI) is a metric that quantifies the cost of service downtime in monetary terms, including productivity and business opportunity impact.

### **Catalog**

A list of declared available executables.

### **Change data capture**

IT Analytics uses the change data capture (CDC) technique to detect inserts and updates from source data based on the previously extracted change timestamp for each entity. Delete records are detected using a full key compare against the Transactional Snapshot (TSNP).

### **Composite key**

Composite keys are primary keys that consist of more than one column. Composite keys are also known as concatenated keys or aggregate keys.

### **Configuration item**

Configuration items (CI) are components of your infrastructure or a business service. CIs are catalogued in your Configuration Management Database (CMDB). A CI can be a hardware component in a larger hardware system, a software application, or any other identifiable item that supports a business service. For example, a CI can be items such as operating systems, monitors, accounting applications, routers, and telephones.

### **Configuration Management Database**

A well-designed Configuration Management Database (CMDB) is a critical building block that stores configuration, asset, and incident data. The CMDB integrates information about assets, incidents, problems, change requests, and configuration items (CIs).

IT Analytics uses the data in the CMDB to run scenarios and produce analytics that suggest how to tune your business services and IT infrastructure. Historical information about assets and CIs, and how well they meet performance commitments, helps you make decisions about future IT investments and cost-saving strategies. The more robust your data, the more precise these scenarios and analytics will be.

A well-tuned CMDB can relate assets to CIs, provide information about asset class availability and CI performance cost, and relate asset and service availability to business service performance.

### **Conformed dimension**

A conformed dimension is used in more than one cube and has the same meaning and data regardless of the fact table that references it.

### **Consolidated model**

IT Analytics uses the consolidated model. During the ETL process, IT Analytics gathers information from one or more source systems and conforms the data into a single consolidated entity before loading it into the data warehouse.

### **Core model**

In the core model, data marts are architected to share the same master set of conformed dimensions and standardized facts across the enterprise.

### **Corporate information factory**

Corporate information factories are the framework that surrounds the data warehouse. Corporate information factories typically contain an operational data store, a data warehouse, data marts, decision support systems applications, exploration warehouses, data mining warehouses, alternate storage, and other data management systems. IT Analytics uses the Ralph Kimball methodology.

### **CSNP**

Consolidated Snapshot.

### **Crosstab**

Crosstab (cross tabulation), also known as a contingency table, is a combination of two or more variable tables. A contingency table analyzes whether the value of one variable is associated with or contingent on another.

## **Cube**

Cubes are groups of related dimensions and measures that are used for data analysis.

## **Dashboard**

Dashboards are reporting tools that consolidate, aggregate, and arrange measurements and metrics on a single screen so information can be monitored at-a-glance.

## **Data architecture**

- 1 Data architecture is the framework design used to organize data resources for collection and future access.
- 2 Data architecture is the data, processes, and technologies that an organization uses to create and operate information systems.

## **Data flow**

Data flow is a reusable object that contains steps to define the transformation of data from source to target. Data flows are called from inside a work flow or job. You can pass information into or out of data flows using parameters.

You can store data flows inside work flows, but you cannot store data flows inside other data flows.

**Source:** SAP® Business Objects. SAP Help Portal (<http://help.sap.com>).

## **Data integration**

Data integration consolidates records with similar information to produce a single version of data in the data warehouse.

Data integration includes the steps from acquisition through publication, such as extraction, cleansing, consolidation, removing duplication, key generation, transformation, and aggregation, of source data before loading to the target table.

## **Data quality**

Data quality is the degree of data excellence. For example, the data is stored according to data type, it is consistent, complete, not redundant, follows business rules, and corresponds to established domains. The data should be timely, well understood, and satisfy the needs of the business. The user should be satisfied with the validity of the data and the information derived from that data.

## **Data mart**

Data marts are highly focused versions of a data warehouse. They are data stores that contain data for a specific business function within an organization, for example, finance or sales.

## **Data mining**

Data mining is the process of analyzing data to identify relationships, trends, or patterns in the data.

## **Data model**

Data models describe the source, extraction, consolidation, and target components for a single entity.

## **Data quality**

Data quality is the degree of data excellence. For example, the data is stored according to data type, it is consistent, complete, not redundant, follows business rules, and corresponds to established domains. The data should be timely, well understood, and satisfy the needs of the business. The user should be satisfied with the validity of the data and the information derived from that data.

## **Data warehouse**

Data warehouses are a repository of integrated information that is available for queries and analysis. Users can run queries against a database without having to know anything about the underlying data structures in the database.

## **Dimension**

- 1 Dimensions are entities that describe, qualify, or otherwise add meaning to facts in a star schema fact table. Dimensions are the items in the analysis of facts by product, market, time, and period, for example.
- 2 Dimensions describe the measurements (facts) that business users want to analyze.

## **Downstream**

Downstream ETL is a process that retrieves data from the data warehouse and passes it to a consumer application.

## **Efficiency and effectiveness**

Efficiency describes how well IT meets customer expectation with minimal resources, expense, or unnecessary effort. Effectiveness measures the degree of success in achieving a given goal.

Unlike efficiency, effectiveness does not measure the resources expended, the expense incurred, or the effort, but only the degree of success.

IT can be effective without being efficient. IT Analytics enables you to construct scenarios that alter resources, expenses, or other efficiency components to improve effectiveness.

## **Enterprise key**

The enterprise key is used to generate the durable key in the target entities.

## **ETL**

The Extract, Transform, and Load (ETL) process is an end-to-end transfer of external source data through several staging layers and into the target layer of the data warehouse. The source data can reside in a single database or multiple databases. The data warehouse uses the ETL process to extract, consolidate, and transform the source data into a meaningful target model that populates relevant business analytics.

## **Executable**

A program that executes an ETL task:

- Any custom script or program that complies with the data warehouse ABC API.
- An SAP BusinessObjects Data Services job that complies with the data warehouse ABC utilities.

## **EXT**

Extract Staging.

### **Extract (ETL)**

The extract process extracts delta data from the source systems and writes it to delimited flat files on the file system. The data warehouse uses the Change Data Capture (CDC) technique for extracting data from the source tables. IT Analytics extracts data if the last modification date is greater than or equal to the maximum last modification date value.

### **Fact table**

Fact tables are the central tables in star schemas. Facts are information about the business that are typically numeric and additive. Fact tables contain the measures that the business users wish to analyze to find new trends or to understand the success or failure of the organization.

### **Financial Planning & Analysis**

Financial Planning & Analysis (FPA) is a component of IT Analytics that helps organizations make better IT financial decisions by reducing the effort involved with managing IT finance and giving business the financial transparency it demands. FPA does this by consolidating and allocating planned and actual costs from HP Asset Manager and HP Project and Portfolio Management to help IT manage those costs more effectively.

### **Foreign key**

A foreign key is the field in a table that refers to the primary key of another table in the same data source. Every relationship in the model must be supported by a foreign key. Foreign keys enforce referential integrity by completing an association between two entities.

### **Governance**

Governance is a structures process or group used to control change and provide oversight. Governance groups are commonly used to control changes in data models, data modeling practices, and architecture standards.

### **Granularity**

Granularity is the level of the measures within a fact table represented by the lowest level of the dimensions.

### **Hierarchy**

Hierarchies use ordered levels to organize data. For example, in a time dimension, a hierarchy might be used to aggregate data from the month level to the quarter level to the year level.

### **Historical data**

Historical data is data from previous time periods. It is used for trend analysis and for comparisons to other time periods.

### **Impact analysis**

Impact analysis is the process of analyzing the effect if a business process fails or loses efficiency.

## **Impact rules**

Impact rules define the effect of an incident on a business service, application, or the organization. Impact rules are associated with these characteristics:

- Incident type, such as degraded performance, limited capacity, corruption, an outage, or a request for information.
- Scope of the impact. For example, the impact might affect one user, an entire department, a business unit, or the entire enterprise.
- A schedule that defines when the rule applies. An outage at midnight might not be as expensive as an outage during normal business hours.
- Fixed amount, fixed rate, or stepped rate impact calculation.

These variables make each rule unique. As you expand the number of impact rules to cover different scenarios, your impact calculations become more precise. For example, if all variables are the same except for the scope, the impact cost is exponentially greater as a larger segment of the organization is affected. Different rules also enable you to calculate the impact of an outage at these levels:

- Business service level, which has immediate impact.
- Organization level, which extends the impact to include customer good will, lost revenue, and more.

IT Analytics uses impact rules to calculate the impact of historical events as well as hypothetical scenarios.

## **ITIL best practices**

IT Analytics users need to adhere to the best practices advocated by the IT Infrastructure Library (ITIL). When you adhere to the ITIL best practices and processes for problem, change, and service level management, you have a business maturity model that is compatible with IT Analytics processes and objectives. To validate the benefits of ITIL best practices, you can evaluate comparative industry process metrics to confirm the improved efficiency and effectiveness of your internal processes.

IT Analytics follows ITIL version 3 guidelines for naming conventions and to categorize analytics.

## **Join**

Tables from a single data source where one table is considered the parent or primary table are consolidated with a join. In a join, all records from the parent table are imported, and records that have a matching foreign key are imported from child tables.

## **Load (ETL)**

The load process adds new records, updates existing records, and flags deleted records in the data warehouse target layer fact and dimension tables.

## **Log**

A log displays an alert that an issue occurred during execution. The information in the execution log can become an alert when configured by the alert system.

## **Measure**

A value collected by the executable during execution, such as the number of rows processed during an ETL job, or an amount extracted from a table that describes expenses in a source application.

The ABC application does not make any assumptions about the business tasks performed by the executable or impose any semantic requirements on computed measure values. The ABC application provides an interface for the executable to store measures. The ABC reports provide drill-down capabilities into the stored measures.

## **Merge**

Tables that exist across multiple data sources and where reconciliation is done are consolidated with a merge. In a merge, all records from all tables are imported, and reconciliation is done where appropriate. Merges join tables where they overlap and also import all other records.

## **Metadata**

- 1 Metadata includes a description of data models, a description of the layouts used in database design, the definition of the system of record, the mapping of data from the system of record to other places in the environment, and specific database design definitions.
- 2 Metadata usually refers to agreed-upon definitions and business rules stored in a centralized repository to ensure that business users across departments and systems use common terminology for key business terms. Metadata can include information about data currency, ownership, source system, derivation (such as profit equals revenues minus costs), or usage rules. Metadata prevents data misinterpretation and poor decision making due to limited understanding of the true meaning and use of corporate data.

**Source:** SAP® Business Objects. SAP Help Portal (<http://help.sap.com>).

## **Metric**

- 1 Metrics measure quality. They establish and collect measurements of success or failure.
- 2 The executable that collects the measures and generates the status. The ABC application provides an interface for the executable to store metrics.

## **MSC**

Multiple Source Consolidated.

## **MSI**

Multiple Source Identity.

## **Multi-dimensional**

Multi-dimensional is the aggregation of data by the dimensions of the business. For example, sales by region by product by time.

## **Near-real-time**

Near-real-time provides users with rapid updates from transactional systems.

## **OLAP**

Online analytical processing (OLAP) is an application that enables multidimensional analysis and visualization of data.

## **Organization**

Organizations are parts of your business that provide a product or service to an internal or external customer. It can be an entire business unit, a cost or profit center, or a functional department. It is an entity that adds value directly or indirectly to your customer community.

## **Predictive analysis**

Predictive analysis uses data mining techniques to predict future behavior, probability, and trends. The more historical data that you have, the more accurately you can predict the likelihood of future events.

For example, auto insurance companies use safety predictors such as age, gender, and driving record when they issue car insurance policies. Mortgage underwriters use financial predictors such as income, debt ratio, and payment history when they issue home mortgages.

## **Primary key**

Primary keys are the field or fields that make up the unique value for each record. In the source application, this is represented by the business key. In the data warehouse, it is the surrogate key.

## **Prioritization**

When the number of services required exceeds your capacity, you must establish rules to organize open incidents into a work queue. The most urgent incidents are at the beginning of the queue; incidents with lower priority are at the bottom of queue.

## **Process**

A process is an execution of a job stream step in a specific batch. A single job stream step may lead to multiple processes in cases of retries.

## **Real-time**

Real-time reporting is on demand reporting versus scheduled, prebuilt reporting.

## **Reconciliation**

Data reconciliation helps ensure the quality of the data.

## **Report**

Reports are pre-defined documents that target a specific audience or user.

## **Report library**

Report libraries are sets of programs that have been created, fully tested, quality assured, documented, and made available to the user community. The programs in these libraries are called canned, predefined, parameterized, or skeleton queries. They are launched by a user who enters only a variable such as a date, region number, range of activity, or some other set or sets of values the program needs to generate a query or report.



## **Resource allocation**

The challenge for IT is to reduce costs without compromising services. Reducing personnel can reduce costs, but may seriously affect services that require high service levels and minimal outages. IT Analytics enables you to plan for resource needs, the cost of reducing head count, and validate that resources levels are correct for your service requirements.

For example, increased revenues may mean increased demand for services and more unplanned outages. How much budget do you need to assign enough technicians to cover expected demand?

## **Risk**

Risk is the possibility of loss that is measurable in monetary terms. Risk can also be an intangible opportunity cost. The business service owner must define the level of risks, such as the amount of acceptable downtime or economic impact. IT must set expectations for how it will support the applications and infrastructure of business service and mitigate risks.

Examples of risks include a corrupt database, a power outage, or a natural disaster. The list of business risks is large. Some are common to all enterprises, like a natural disaster; some are specific to your organization or line of business.

## **Slowly changing dimension**

A slowly changing dimension (SCD) is a dimension that contains data that changes over time. For SCD type 1 changes, the new data overwrites the original data. For SCD type 2 changes a new record is generated. IT Analytics makes a copy of the record and flags the original record as historical. Both the new and old record share the same durable key, indicating that they are the same piece of data.

## **Snowflake schema**

A snowflake schema is a data model that has a centralized fact table that references multiple dimension tables, and the dimension tables reference multiple related tables.

## **SSI**

Single Source Identity.

## **SSNP**

Single Source Snapshot.

## **Staging**

Staging areas are collection areas where data is transformed for presentation, integration, or enhancement. For example, data might be staged until all relevant components are in place.

## **Star schema**

Star schemas are simpler than snowflake schemas. A star schema is a data model that has one or more centralized fact tables that reference a single layer of dimension tables. The dimension tables do not reference other tables.

## **Stream step detail**

A sub-task of the executable job stream step. There are no assumptions about the content of the executable; therefore, the stream step detail is not modeled. The ETL designer is responsible to define the job stream step details.

**Surrogate key**

A surrogate key is a primary key for the dimension tables. It is the field that makes up the unique value for each record in the data warehouse. Surrogate keys are numeric and are controlled by the ETL process.

**Transform (ETL)**

During the transformation process, IT Analytics transforms data into the format of the target tables and populates the load-ready target staging tables. The Transformation layer contains load-ready tables that match the internal structure of the data warehouse target tables.

**TSNP**

Transactional Snapshot.

**Union**

Tables that occur in either single or multiple data sources and have records that are assumed to be equal and independent are consolidated with a union. In a union, all records from all sources are imported but are not reconciled. They are simply added to the rowset.

**Universe**

The universe is the result of a semantic layer of metadata that creates a business oriented view of the data. The universe contains a schema of the tables and joins that make up the universe and dimension and measurement objects. The universe is an interface between the data warehouse and the analytics that display the data.

The universe contains analytic views that enable you to see published and up-to-date allocation cost and planned cost data as well as unpublished data or published data that is not up-to-date. The views also ensure the consistency of data warehouse quality by showing only data that is successfully extracted, transformed, and loaded. Analytic views provide the mechanism to reduce the dependencies of analytic semantic models to data warehouse physical models.

**Upstream**

Upstream ETL retrieves data from one or more external applications, consolidates it into a target data model, and stores it in the data warehouse.

**Work flow**

A work flow is a reusable object containing steps to define the order of job execution. Work flows call data flows but cannot manipulate data. You can call work flows from inside other work flows or jobs and can pass information into or out of work flows using parameters. You can reuse work flows by dragging existing ones from the object library.

**Source:** SAP® Business Objects SAP Help Portal (<http://help.sap.com>)

**XFR**

Dimensional Transform.

**XREF**

Cross Reference.

# A Naming Conventions

This appendix lists naming conventions for fields as well as the default records and default field values.

## Field Naming Conventions

**Table 4** Field Naming Conventions

<b>Field</b>	<b>Naming Convention</b>
Metadata fields (Primary Business Key)	MD_BUSINESS_KEY
Metadata fields (Primary Enterprise Key)	MD_ENTERPRISE_KEY
Metadata fields (Primary Durable Key)	MD_DURABLE_KEY
Metadata fields (Other)	MD_<entity>
Source field used only for ETL	SRC_<entity>
Business Key	<entity>_BUSINESS_KEY
Enterprise Key	<entity>_ENTERPRISE_KEY
Durable Key	<entity>_DURABLE_KEY
Surrogate Key (Primary)	PK_<entity>_ID
Surrogate Key (Foreign)	<entity>_ID
Boolean	FLAG_<entity>
Date or Date-Time (Source Localized)	DATE_<entity>_LOC
Date or Date-Time (Data Warehouse Standardized)	DATE_<entity>_BASE
Date or Date-Time (Universal Time)	DATE_<entity>_UTC
Currency or Exchange Rate Foreign Key (Source Localized)	<entity>_LOC_ID <entity>_LOC_BUSINESS_KEY <entity>_LOC_DURABLE_KEY <entity>_LOC_ENTERPRISE_KEY

**Table 4 Field Naming Conventions**

Field	Naming Convention
Currency or Exchange Rate Foreign Key (DW Standardized)	<entity>_BASE_ID <entity>_BASE_BUSINESS_KEY <entity>_BASE_DURABLE_KEY, <entity>_BASE_ENTERPRISE_KEY
Money (Source Localized Currency)	<entity>_LOC
Money (DW Standardized Currency)	<entity>_BASE

**Notes:**

- If adding affixes makes the field name too long, the <entity> portion becomes truncated.
- Local (LOC) refers to the time zone or currency. In the staging Tables, LOC refers to the source system locale. In the target tables, LOC refers to the data warehouse locale.
- Base (BASE) refers to the default currency of the data warehouse. This value may or may not match the default value of the source applications. The ETL process standardizes the base so that it becomes consistent across all records. BASE versions of any field remain NULL until the MSC stage.

## Default Records and Values

The ETL process creates three default records in each table (Primary Key (PK) values -2 and -1).

**Table 5 Default Records**

Data Type	Unknown	No Match
Primary SK	-1	-2
FK	-1	-2
String /Text (length dependent)	UNKNOWN, UNK, U	NO MATCH, NM, M
Flag	M	M
Date	12/31/9999	12/31/9999

## Default Field Values

**Table 6 Default MD Field Values**

Field	New	SDC1	New SDC2	Old SDC2 (History Record)	Delete
MD_TRANSENDDATE	Null	Null	Null	Same as MD_CREATEDDATE on NEW SCD2	Null
MD_TRANSLASTIND	Y	Y	Y	N	Y
MD_ACTIVESTATUSIND	Y	Y	Y	Y	N
MD_DELETEDDATE	Null	Null	Null	Null	Sysdate
MD_CREATEDDATE	Sysdate	No Update	Sysdate	No Update	
MD_LASTMODDATE	Null	Sysdate	Null	Sysdate	Sysdate

### MD attribute population rules

- MD\_BATCH\_ID, MD\_PROCESS\_ID: Retrieved from an Audit, Balance, and Control (ABC) function call. Hard coded as a variable.
- MD\_FLAG: Based on a comparison with the snapshot. Valid values: NEW, UPD, DEL.
- MD\_PF\_FLAG: Valid values: D (Delta), P (Pre-fill).

### FLAG Defaults

FLAG\_ fields always use the following values:

- Y (Yes)
- N (No)
- U (Unknown)
- M (Metadata)



## B Data Models

This appendix contains entity relationship diagrams (ERDs) for the target data models, the IT Financial Analysis universe, and the cross reference consolidation data model.

### Target Data Models

This section contains ERDs for the cost entities: Actual Cost, Allocated Actual Cost, Allocated Planned Cost, and Planned Cost. The ERDs are logical views of the cost Fact tables and associated Dimension tables. View the target database for physical views of the target data models.



To view interactive data models that provide column descriptions and data types for the tables that make up the model, download the *IT Analytics 2.00 Entity Relationship Diagrams* from the HP Software Product Manuals web site:

**<http://h20230.www2.hp.com/selfsolve/manuals>**

This site requires that you register for an HP Passport and sign-in. To register for an HP Passport ID, go to:

**<http://h20229.www2.hp.com/passport-registration.html>**





This entity relationship diagram (ERD) is a logical view of the target data model. It is interactive and contains details about the tables used in the model. To take advantage of the interactivity and be able to view additional details, open the PDF using Adobe® Reader® 9. Free upgrades are available at [www.adobe.com](http://www.adobe.com).

If you are viewing this data model using Windows Internet Explorer or Mozilla Firefox, go to **File > Save As** (or **Save Page As**), and save the PDF to your computer. Then open it using Adobe Reader 9.

**To view table details using the Model Tree**

1. From the **Tools** menu, select **Analysis > Object Data Tool**.

2. Click this Model Tree icon.



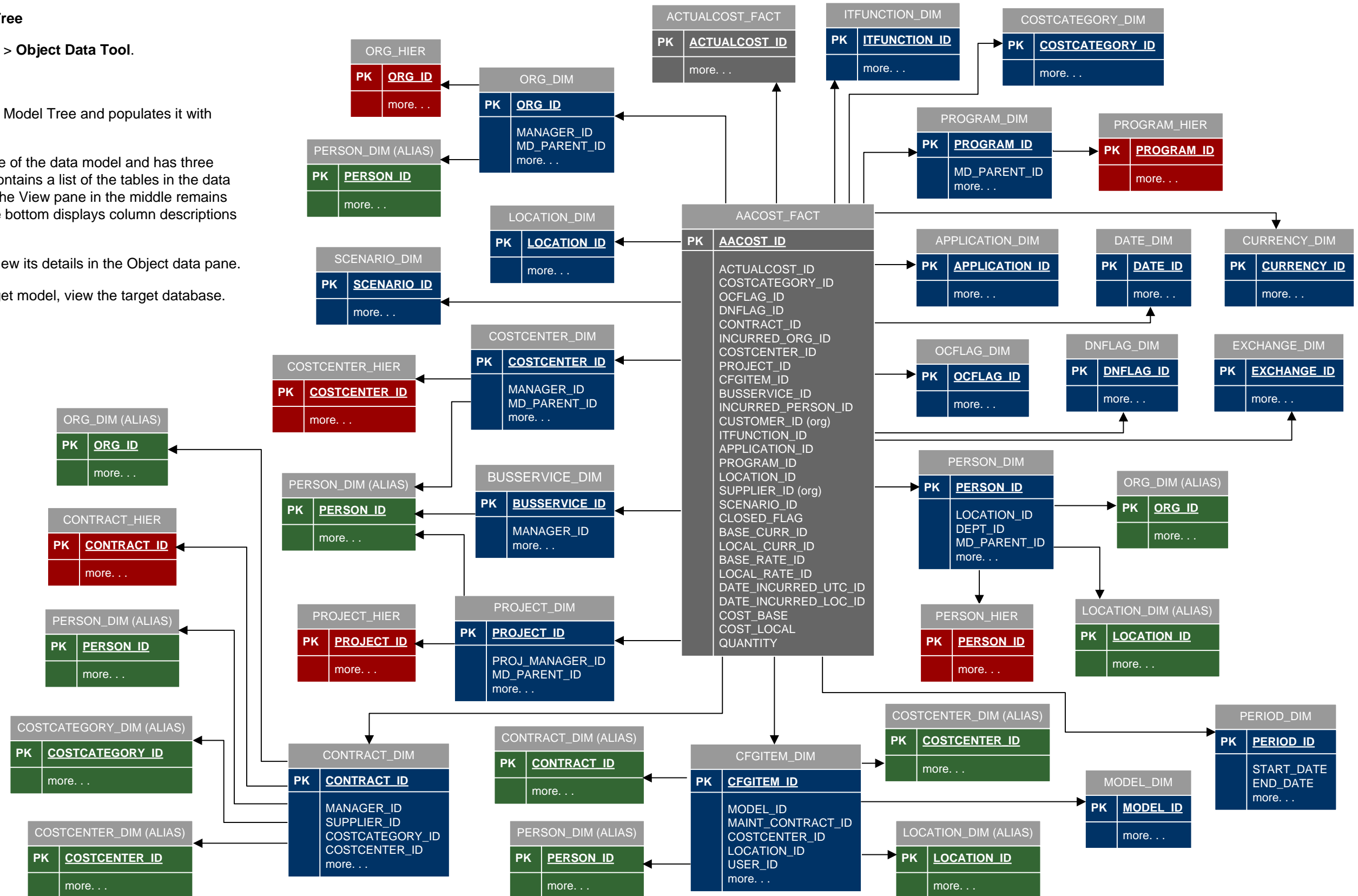
Clicking the Model Tree icon opens the Model Tree and populates it with tables from the data model.

The Model Tree appears on the left side of the data model and has three panes: The Structure pane at the top contains a list of the tables in the data model that have detailed object data. The View pane in the middle remains empty. And the Object data pane at the bottom displays column descriptions and data types for the selected table.

3. Click an item in the Structure pane to view its details in the Object data pane.

**Note:** To see the physical view of the target model, view the target database.

# Allocated Actual Cost Target Data Model



# Allocated Planned Cost Target Data Model

## View the ERD using Adobe Reader 9

This entity relationship diagram (ERD) is a logical view of the target data model. It is interactive and contains details about the tables used in the model. To take advantage of the interactivity and be able to view additional details, open the PDF using Adobe® Reader® 9. Free upgrades are available at [www.adobe.com](http://www.adobe.com).

If you are viewing this data model using Windows Internet Explorer or Mozilla Firefox, go to **File > Save As** (or **Save Page As**), and save the PDF to your computer. Then open it using Adobe Reader 9.

## To view table details using the Model Tree

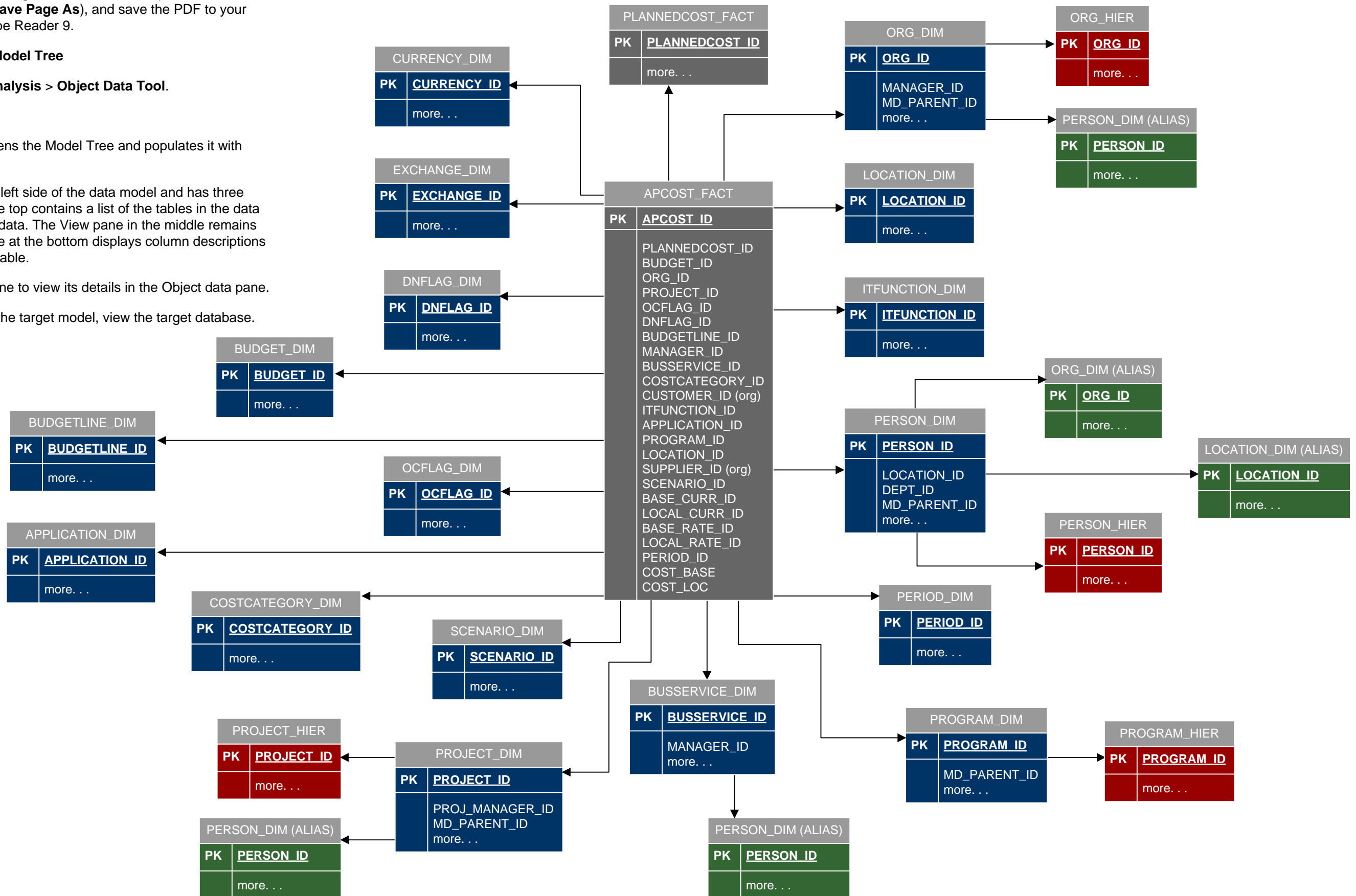
1. From the **Tools** menu, select **Analysis > Object Data Tool**.
2. Click this Model Tree icon. 

Clicking the Model Tree icon opens the Model Tree and populates it with tables from the data model.

The Model Tree appears on the left side of the data model and has three panes: The Structure pane at the top contains a list of the tables in the data model that have detailed object data. The View pane in the middle remains empty. And the Object data pane at the bottom displays column descriptions and data types for the selected table.

3. Click an item in the Structure pane to view its details in the Object data pane.

**Note:** To see the physical view of the target model, view the target database.



# Planned Cost Target Data Model

## View the ERD using Adobe Reader 9

This entity relationship diagram (ERD) is a logical view of the target data model. It is interactive and contains details about the tables used in the model. To take advantage of the interactivity and be able to view additional details, open the PDF using Adobe® Reader® 9. Free upgrades are available at [www.adobe.com](http://www.adobe.com).

If you are viewing this data model using Windows Internet Explorer or Mozilla Firefox, go to **File > Save As** (or **Save Page As**), and save the PDF to your computer. Then open it using Adobe Reader 9.

## To view table details using the Model Tree

1. From the **Tools** menu, select **Analysis > Object Data Tool**.

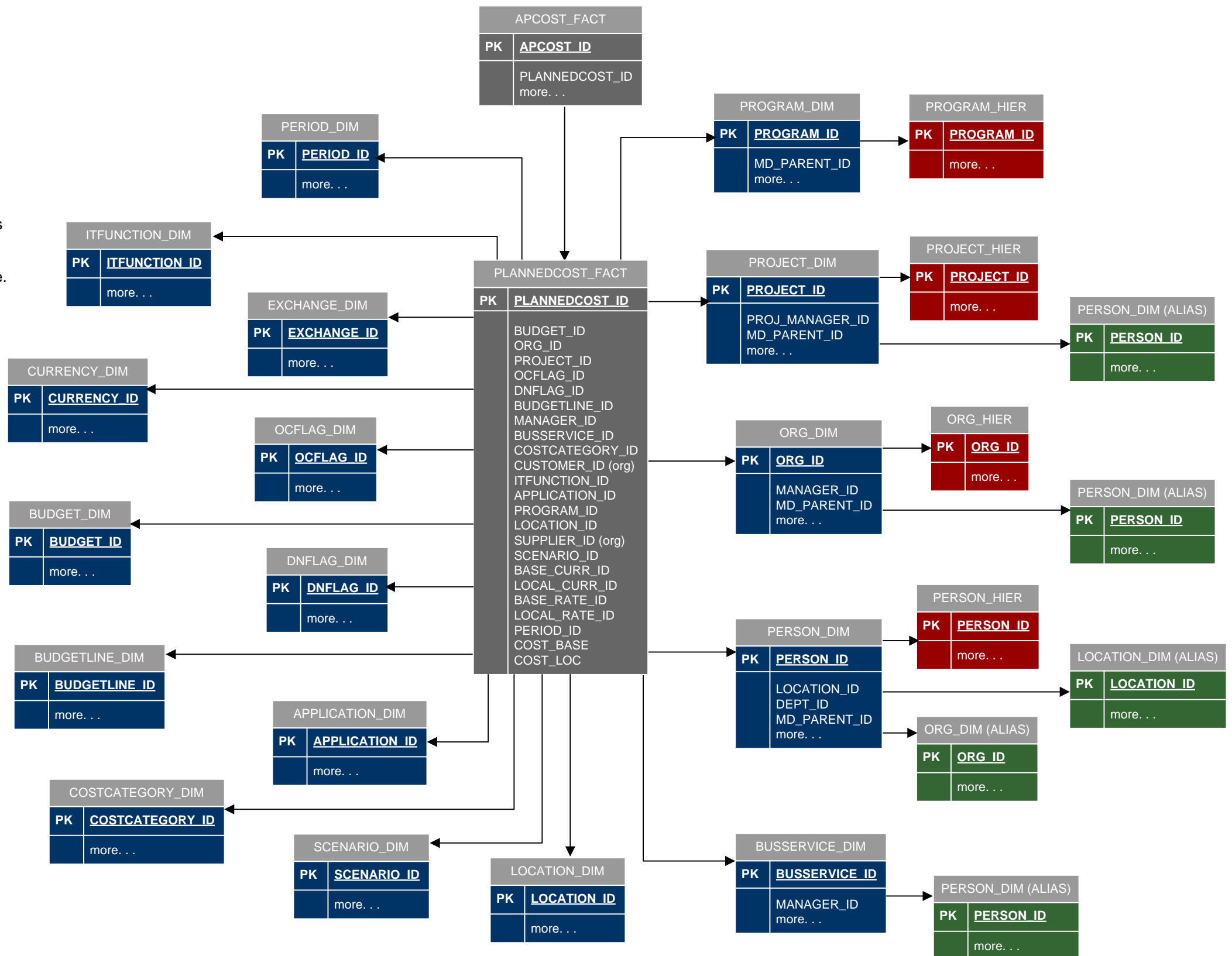
2. Click this Model Tree icon. 

Clicking the Model Tree icon opens the Model Tree and populates it with tables from the data model.

The Model Tree appears on the left side of the data model and has three panes: The Structure pane at the top contains a list of the tables in the data model that have detailed object data. The View pane in the middle remains empty. And the Object data pane at the bottom displays column descriptions and data types for the selected table.

3. Click an item in the Structure pane to view its details in the Object data pane.

**Note:** To see the physical view of the target model, view the target database.



## Universe Data Model

This section contains an ERD for the IT Financial Analysis universe file. The ERD is interactive and provides details about the cubes, hierarchies, tables, dimensions, measures, and analytic views used in the universe.



# IT Financial Analysis universe

## View the universe data model using Adobe Reader 9

This data model is interactive and contains details about the cubes, hierarchies, and tables used in the universe. To take advantage of the interactivity and be able to view additional details, open the PDF using Adobe® Reader® 9. Free upgrades are available at [www.adobe.com](http://www.adobe.com).

If you are viewing this data model using Windows Internet Explorer or Mozilla Firefox, go to **File > Save As** (or **Save Page As**), and save the PDF to your computer. Then open it using Adobe Reader 9.

### To view cube, hierarchy, and table details using the Model Tree

1. From the **Tools** menu, select **Analysis > Object Data Tool**.

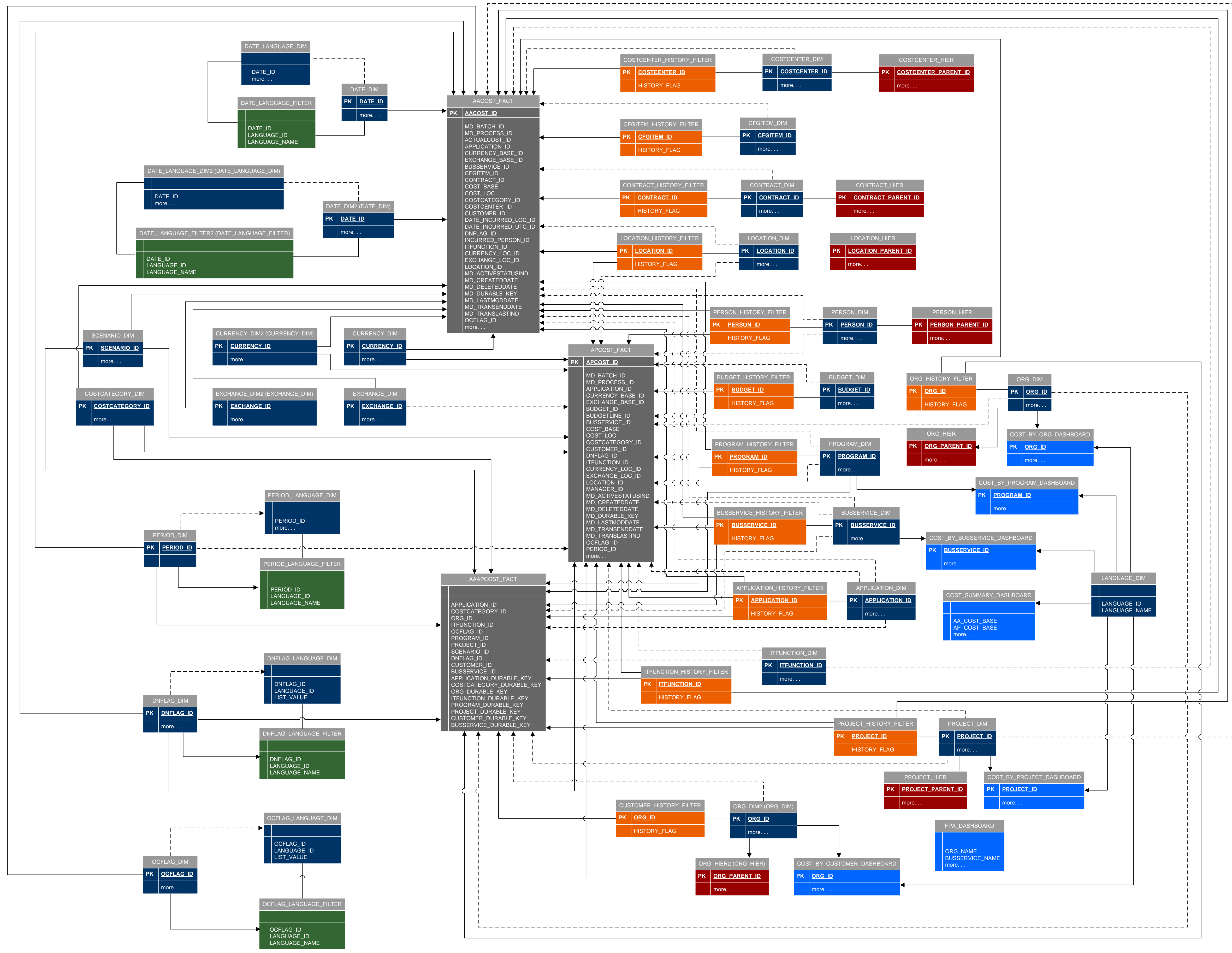
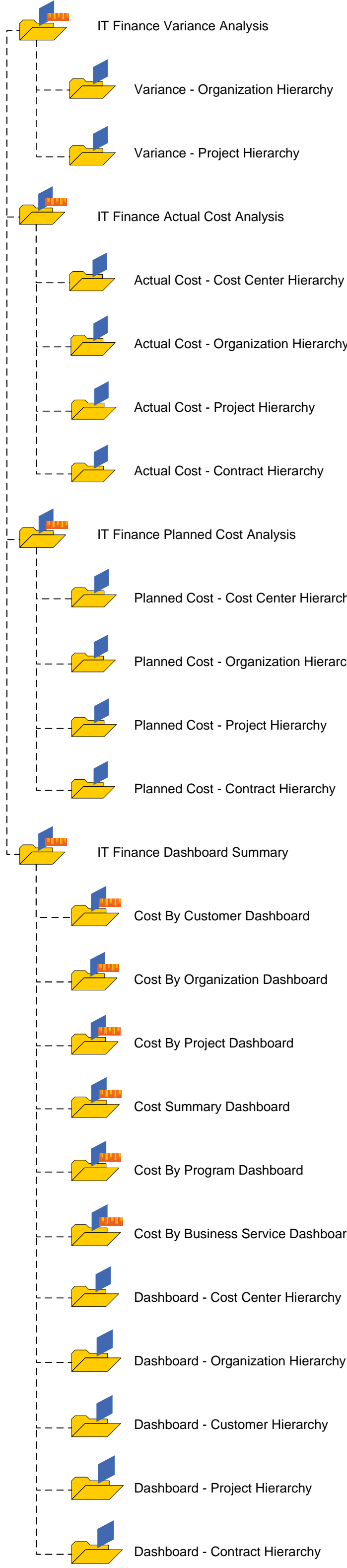
2. Click this Model Tree icon.

Clicking the Model Tree icon opens the Model Tree and populates it with cubes, hierarchies, and tables from the data model.

The Model Tree appears on the left side of the data model and has three panes:

- The Structure pane at the top lists the cubes, hierarchies, and tables in the model.
- The View pane in the middle remains empty.
- The Object data pane at the bottom displays these details for the selected object:
  - column names and any associated dimensions and measures in the selected table.
  - dimensions and measures in the selected cube.
  - dimensions in the selected hierarchy.

3. Click an item in the Structure pane to view its details in the Object data pane.



## Cross Reference Consolidation Data Model

This section contains an ERD for the cross reference consolidation data model. The ERD is interactive and provides details about the tables used in the cross reference consolidation.

## View the ERD using Adobe Reader 9

This entity relationship diagram (ERD) is interactive and contains details about the tables used in the cross reference consolidation. To take advantage of the interactivity and be able to view additional details, open the PDF using Adobe® Reader® 9. Free upgrades are available at [www.adobe.com](http://www.adobe.com).

If you are viewing this ERD using Windows Internet Explorer or Mozilla Firefox, go to **File > Save As** (or **Save Page As**), and save the PDF to your computer. Then open it using Adobe Reader 9.

## To view table details using the Model Tree

1. From the **Tools** menu, select **Analysis > Object Data Tool**.

2. Click this Model Tree icon. 

Clicking the Model Tree icon opens the Model Tree and populates it with tables from the data model.

The Model Tree appears on the left side of the ERD and has three panes: The Structure pane at the top contains a list of the tables in the ERD that have detailed object data. The View pane in the middle remains empty. And the Object data pane at the bottom displays the table and column definitions for the selected table.

3. Click an item in the Structure pane to view its details in the Object data pane.

# Cross Reference Consolidation

 Cross Reference Table

**XXX** Consolidated entity name  
Example: ORG\_ENTERPRISE\_XREF

