

PROJECT REPORT
(SEMESTER TRAINING)

Informatica PowerExchange for Oracle Essbase

Submitted by

Atishay Jain

Roll No 10783017

Department Of Computer Science and Engineering
THAPAR UNIVERSITY, PATIALA
(Deemed University)



Jan-May/Jun 2010

ANNEXURE – VIII

DECLARATION

I hereby declare that the project work entitled (“Title of the project”) is an authentic record of my own work carried out at (Place of work) as requirements of semester project term for the award of degree of B.E. (Computer Science & Engineering), Thapar University, Patiala, under the guidance of (Name of Industry coordinator/Name of Faculty coordinator), during 10 Jan to 25 Jun, 2010).

(Signature of student)
Atishay Jain
10783017

Date: _____



ATTENTION : BE – IVth Year (CSE)

CONTENTS OF THE REPORT

(Binding: Hard Bound)

(Number of Copies = 2)

1. Cover page – on hard paper
2. Inner page – same as cover page but on the soft paper
3. Declaration
4. Acknowledgement (if any)
5. Certificate (place of training)
6. Contents
 - Summary of the Project (1-2 Pages)
 - Introduction to the programming/development environment
 - Project
 - Design Phase should include DFD/E-R diagrams
 - Details of the work including work program
 - Testing – state the different test cases taken to test the project
 - Results, Conclusions and Future Scope of Work
 - References (if any)
 - CD including presentation and text file (*.doc) of the report

Please note:

1. The upper case of letters in the cover page. The 3rd line is 16 pt bold and other lines are 12 pt. The page is centered. Department and Institute names are bold.
2. The matter contained in the report should be typed in MS word (1.5 spacing) Times New Roman, 12 pt or equivalent with other software. The text should be properly justified.
3. Figures and tables may be inserted in the text as they appear or may be appended in order.
4. Subject matter should be typed on single side.
5. Text should be properly justified and Pages should be numbered.
6. Source-code will not be a part of the report.

IMPORTANT:

The reports are to be kept ready & submitted at the time of presentations also include the screen shots of the project in the presentations too. Make sure the Report is signed by your mentor/coordinator.

(IAP Coordinator)

(Head, CSED)



Informatica Corporate Overview

Informatica Corporation (NASDAQ: INFA) is the world's number one independent provider of data integration software. Informatica Corporation provides enterprise data integration and data quality software and services in the United States and internationally. Its software handles various enterprise-wide data integration initiatives, including data warehousing, data migration, data consolidation, data synchronization, and data quality, as well as the establishment of data hubs, data services, cross-enterprise data exchange, and integration competency centers. The company primarily offers PowerCenter, which accesses, discovers, and integrates data from a business system and delivers that data throughout the enterprise; PowerExchange that enables IT organizations to access all sources of enterprise data without having to develop custom data access programs; and Data Quality, which delivers data quality to stakeholders, projects, and data domains. It also provides B2B Data Exchange, a software product for multi-enterprise data integration; Application Information Lifecycle Management, which helps IT organizations to manage every phase of the data lifecycle; Complex Event Processing that enables enterprises to detect, correlate, analyze, and respond to data-driven events; and Cloud, which consists of data integration cloud services and data integration cloud platform. In addition, the company offers product-related customer support, consulting, and education services. Informatica Corporation serves energy and utilities, financial services, government and public sector, healthcare, high technology, insurance, manufacturing, retail, services, telecommunications, and transportation sectors. The company distributes its products through direct sales, systems integrators, resellers, distributors, and original equipment manufacturers. Its strategic partners principally include Accenture, Affecto, Hewlett-Packard, IPI Grammttech, Infosys, Tata Consultancy Services, Teradata, and Wipro. The company was founded in 1993 and is headquartered in Redwood City, California.

Introduction to Programming and Development Environment

Development Environment:

I) Informatica Platform:

Concept: As the volume, variety, and need to share data increases, more and more enterprise applications require data integration capabilities. Data integration requirements can include the following capabilities:

- The ability to access and update a variety of data sources in different platforms
- The ability to process data in batches or in real time
- The ability to process data in different formats and transform data from one format to another, including complex data formats and industry specific data formats
- The ability to apply business rules to the data according to the data format
- The ability to cleanse data to ensure the quality and reliability of information

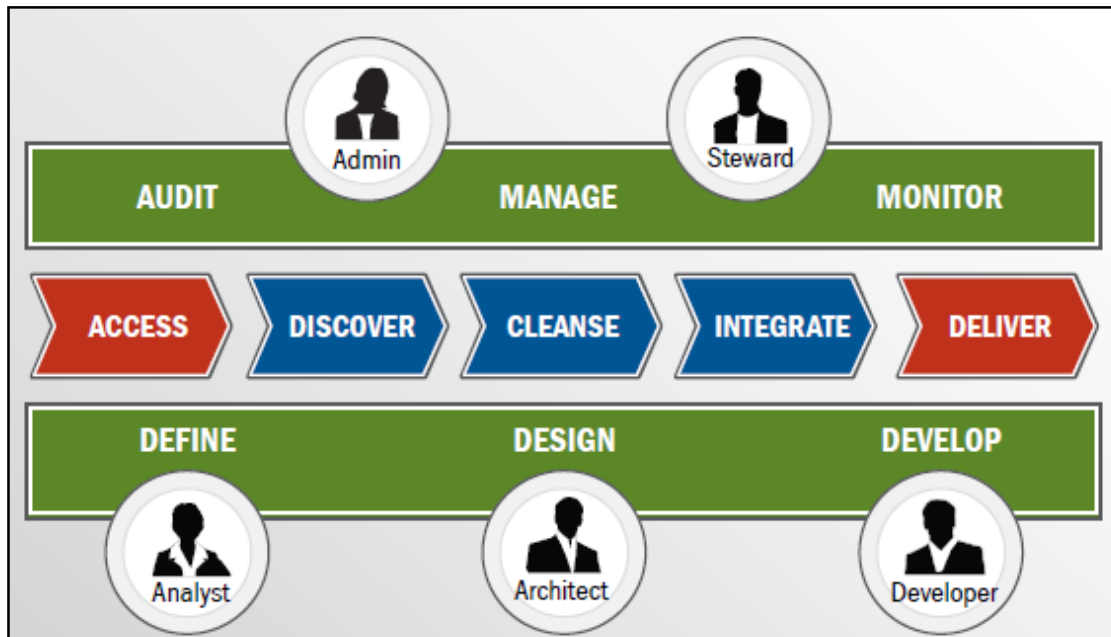


Figure 1: Informatica Platform

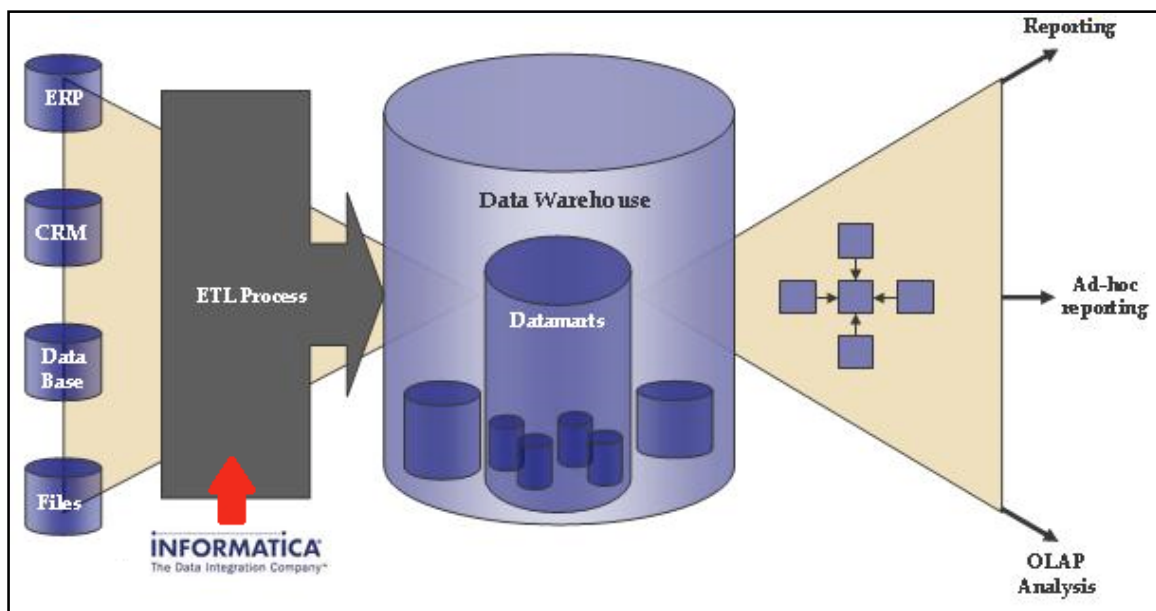
Application development can become complex and expensive when you add data integration capabilities to an application. Data issues such as performance and scalability, data quality, and transformation are difficult to implement. Applications fail if these complex data issues are not addressed appropriately.

PowerCenter data integration provides application programming interfaces (APIs) that enable you to embed data integration capabilities in an enterprise application. When you leverage the data processing capabilities of PowerCenter data integration, you can

focus development efforts on application specific components and develop the application within a shorter time frame.

PowerCenter data integration provides all the data integration capabilities that might be required in an application. In addition, it provides a highly scalable and highly available environment that can process large volumes of data. An integrated advanced load balancer ensures optimal distribution of processing load. PowerCenter data integration also provides an environment that ensures that access to enterprise data is secure and controlled by the application. It uses a metadata repository that allows you to reuse processing logic and can be audited to meet governance and compliance standards.

ETL: ETL stands for extract, transform, and load. ETL is software that enables businesses to consolidate their disparate data while moving it from place to place, and it doesn't really matter that that data is in different forms or formats. The data can come from any source. ETL is powerful enough to handle such data disparities. For example, a financial institution might have information on a customer in several departments and each department might have that customer's information listed in a different way. The membership department might list the customer by name, whereas the accounting department might list the customer by number. ETL can bundle all this data and consolidate it into a uniform presentation, such as for storing in a database or data



warehouse.

Figure 2: Role of ETL and Informatica in Business Intelligence

Another way that companies use ETL is to move information to another application permanently. For instance, word-processing data might be translated into numbers and letters, which are easier to track in a spreadsheet or database program. This

is particularly useful in backing up information as companies transition to new software altogether.

One important function of ETL is "cleansing" data. The ETL consolidation protocols also include the elimination of duplicate or fragmentary data, so that what passes from the E portion of the process to the L portion is easier to assimilate and/or store. Such cleansing operations can also include eliminating certain kinds of data from the process. If you don't want to include certain information, you can customize your ETL to eliminate that kind of information from your transformation.

The T portion of the equation, of course, is the most powerful. ETL can transform not only data from different departments but also data from different sources altogether. For example, data in an email program such as Microsoft Outlook could be transformed right along with data from an SAP manufacturing application, with the result being data of a common thread in the end.

Key Entities:

1. **Repository Manager:** To create and manage repository, create and manage users
2. **Repository server admin console:** To manage the Repository Server
3. **Designer Tools:** Use the Designer to create mappings that contain transformation instructions for the Integration Service.
 - a. **Source Analyzer:** to import source definitions from third party software.
 - b. **Target designer:** to import target definitions, to write to third party applications.
 - c. **Transformation developer:** a rep and parses data. This object can be r
 - d. **Mapping:** a mapping is nothing but a set of source and target definitions linked by transformation objects. A mapping is what represents the data flow from source to target. Mapping Designer is used to create Mappings.
 - e. **Mapplet designer:** A mapplet is a set of objects that you use to build reusable transformation logic.
4. **Workflow:** A workflow is a set of instructions that tells the Power Center Server how to

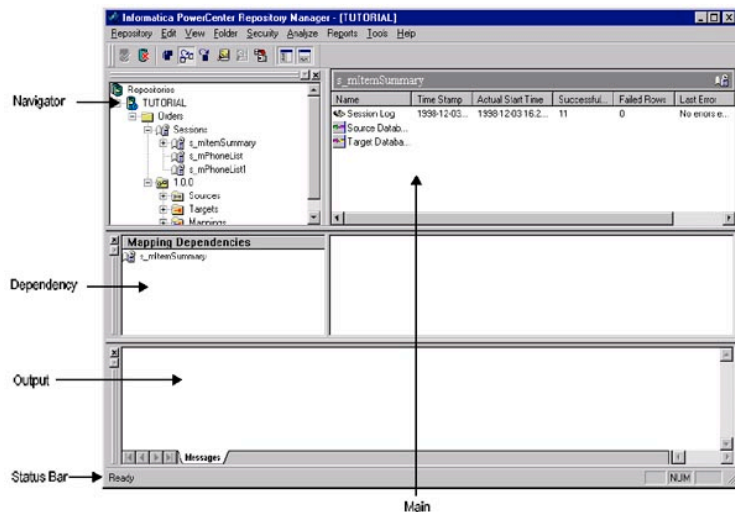


Figure 3: Repository Manager

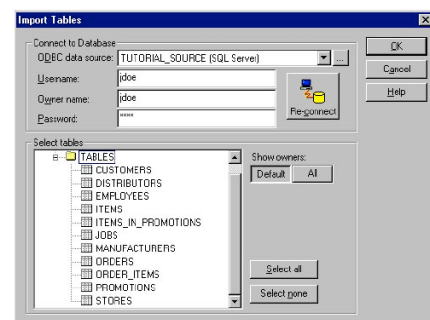


Figure 4: Import Sources

execute a task. Workflow Manager can be used to create and run workflows. Workflow Monitor can be used to monitor the running of workflows

- a. **WorkFlow Manager :** Use the Workflow Manager to create, schedule, and run workflows. A workflow is a set of instructions that describes how and when to run tasks related to extracting, transforming, and loading data.

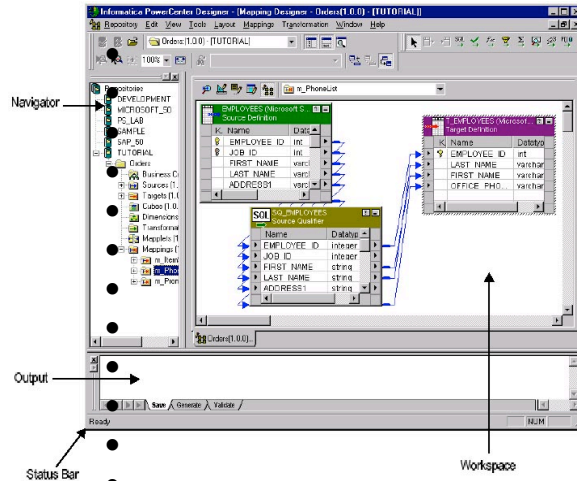


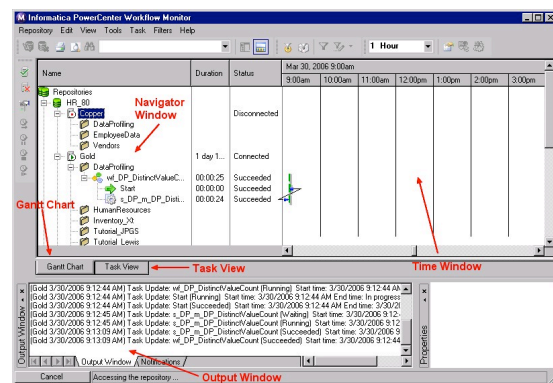
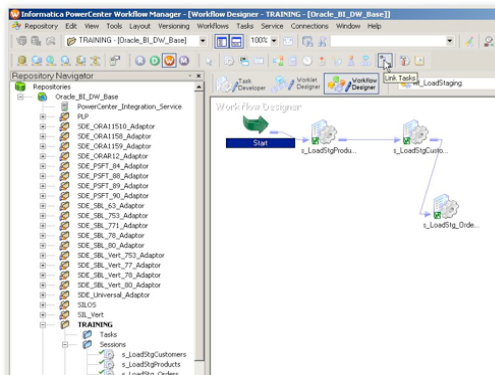
Figure 5: Designer with Mapping

Column	Column Name	Data type	Prec	Scale	Not null	Key type
1	EMPLOYEE_ID	int	10	0	<input checked="" type="checkbox"/>	PRIMARY KEY
2	JOB_ID	int	10	0	<input checked="" type="checkbox"/>	FOREIGN KEY
3	FIRST_NAME	varchar	30	0	<input checked="" type="checkbox"/>	NOT A KEY
4	LAST_NAME	varchar	30	0	<input checked="" type="checkbox"/>	NOT A KEY
5	ADDRESS1	varchar	72	0	<input checked="" type="checkbox"/>	NOT A KEY
6	ADDRESS2	varchar	72	0	<input checked="" type="checkbox"/>	NOT A KEY
7	CITY	varchar	32	0	<input checked="" type="checkbox"/>	NOT A KEY
8	STATE	varchar	2	0	<input checked="" type="checkbox"/>	NOT A KEY
9	POSTAL_CODE	varchar	10	0	<input checked="" type="checkbox"/>	NOT A KEY

Figure 6: Edit Table Properties

- Monitor, add, edit, delete Informatica Server info in the repository
- Stop the Informatica Server
- Configure database, external loader, and FTP connections
- Manage sessions and batches - create, edit, delete, copy/move within a folder, start/stop, abort sessions, view session logs, details, session performance details.

- b. **WorkFlow Monitor:** Use the Workflow Monitor to monitor



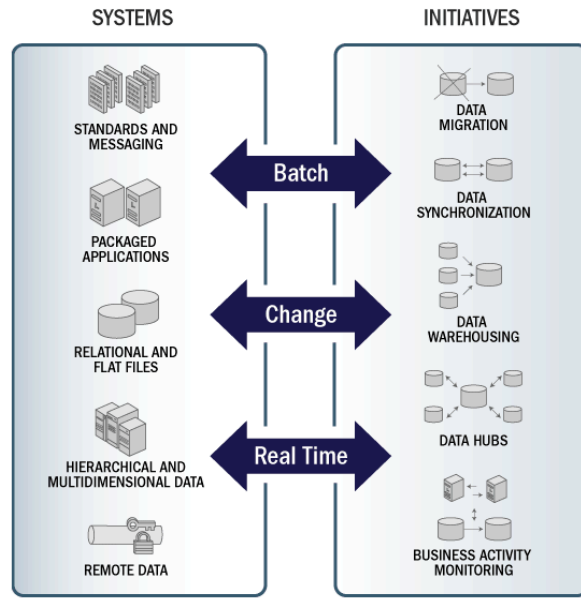
scheduled and running workflows for each Integration Service.

II) Informatica PowerExchange:

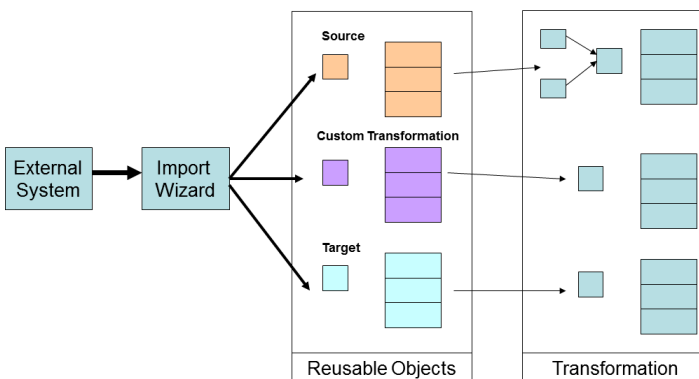
Informatica PowerExchange is a comprehensive family of data access products that enable the IT team to access mission-critical operational data seamlessly and readily deliver it to people and processes across your organization, wherever and whenever it's needed. No hand coding is necessary, which increases the efficiency of limited IT resources while dramatically reducing the cost of delivering data.

PowerExchange lets one access all major enterprise data formats—from the mainframe to the desktop—without having to develop custom data access programs. You can push data to data marts and warehouses, pull data from real-time message queues, or expose data as a service. One can also write data into third-party systems, such as SAP and Siebel, when migrating and synchronizing

Figure 9: Informatica PowerExchange



Informatica PowerExchange, based on a services-oriented architecture (SOA), provides on-demand access to data in all critical enterprise data systems, including mainframe, midrange, and file-based systems. Available as a standalone service or tightly integrated with Informatica PowerCenter, PowerExchange helps organizations leverage mission-critical operational data by making it available to people and processes without requiring manual coding of data extraction programs. Its SQL access to native database APIs provides high-performance extraction, conversion, and filtering of data without intermediary staging and programming. Shared services offer data delivery options that enable IT organizations to flexibly and efficiently manage processing demands.



Organizations today demand immediate access to accurate information for quick decision making and high-speed operations. At the same time, the volume and variety of data is exploding, stretching the capacity of IT resources and infrastructures. To leverage the full value of their information,

Figure 10: PowerExchange (Adapter) working

organizations must be able to integrate data from a wide variety of transactional applications and systems for easy access and "right time" delivery.

III) Oracle Hyperion Essbase:

Essbase stands for Extended Spread Sheet database. It is a multidimensional database management system (MDBMS) that provides a multidimensional database platform upon which to build analytic applications (OLAP). Online analytical processing (OLAP) is a multidimensional, multi-user, client-server computing environment for users who need to analyze enterprise data.

OLAP applications span a variety of organizational functions. Finance departments use OLAP for applications such as budgeting, activity-based costing (allocations), financial performance analysis, and financial modeling. Sales departments use OLAP for sales analysis and forecasting. Among other applications, marketing departments use OLAP for market research analysis, sales forecasting, promotions analysis, customer analysis, and market/customer segmentation. Typical manufacturing OLAP applications include production planning and defect analysis.

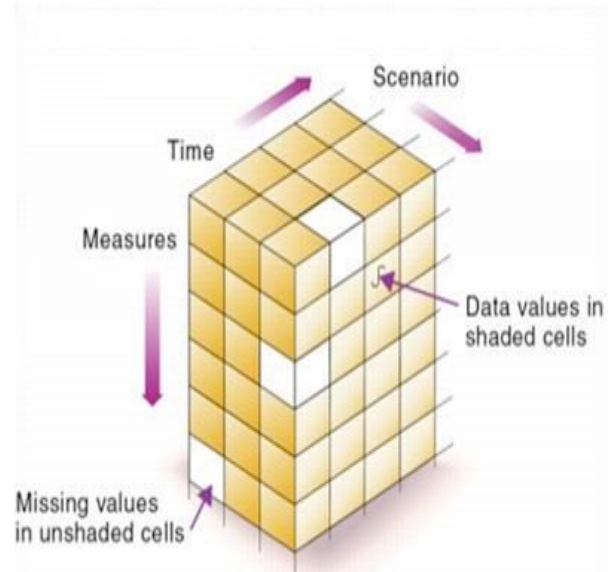


Figure 11: OLAP Cube

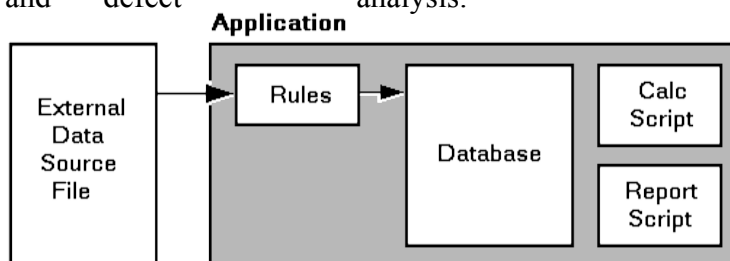


Figure 12: Essbase Application

Essbase **Analytic Services** – Essbase OLAP Server: Analytic Services—a multi-threaded OLAP database software that takes advantage of symmetric multi processing hardware platforms—is

based upon Web-deployable, thin-client architecture.

The server acts as a shared resource, handling all data storage, caching, calculations, and data security. The Analytic Server client needs only to retrieve and view data that resides on a server. All Analytic Services application components, including database outlines and calculation scripts, application control, and multi-dimensional database information, reside on a server.

A Hyperion Essbase application is a collection of one or more of the following types of files:

- Multidimensional databases
- Rules for loading data
- Scripts to calculate your data
- Scripts to generate reports on your data
- Security information and application security

Key Entities:

1. **Outline:** The database structure of a multidimensional database, including all dimensions, members, tags, types, consolidations, and mathematical relationships. Data is stored in the database according to the structure defined in the outline.

2. **Dimensions:** A data category that is used to organize business data for retrieval and preservation of values. Each dimension usually contains a hierarchy of related members grouped within it. For example, a Year dimension often includes members for each time period, such as quarters and months. Other common business dimensions may be measures, natural accounts, products, and markets.

- Standard dimensions:** These are core components of a business plan and often relate to departmental functions. For example, Time, accounts, product line
- Attribute dimensions:** These dimensions are associated with standard dimensions and are used to group and analyze members of your standard dimensions. For example, You can compare the profitability of your non-caffeinated products

3. **Members:** A discrete component within a dimension. For example, a time dimension might include such members as Jan, Feb, and Qtr1.

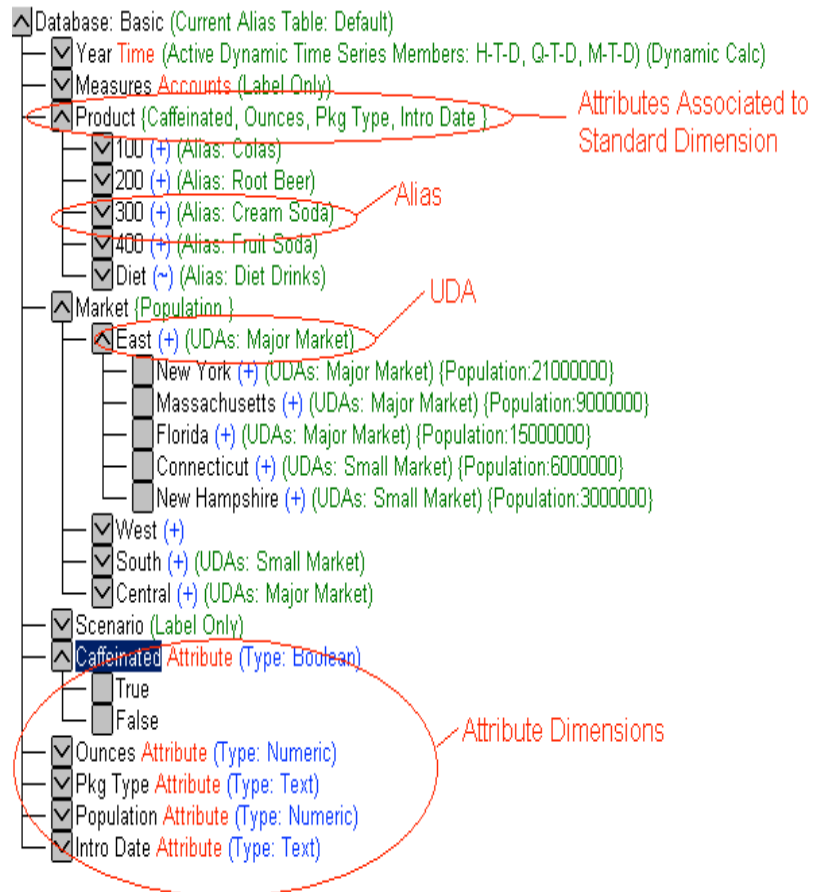


Figure 13: Essbase Outline

4. **Alias:** Aliases are alternative name for a dimension, member, or description. They can improve the readability of an outline or a report. For example, members in Product dimension may be identified both by product codes, such as 100, and by more descriptive aliases, such as Colas.
5. **Base Dimension:** A standard dimension that is associated with an attribute dimension. To classify a member of base dimension, you can associate it with a member of an attribute dimension that describes the classification, such as a specific flavor. For example, assuming products have flavors, the Product dimension is the base dimension for the Flavors attribute dimension.
6. **Dynamic Time Series:** A process that is used to perform dynamic period-to-date reporting.
7. **UDA:** A UDA or user-defined attribute is a term associated with members of an outline to describe a particular characteristic of the members. Users can specify UDAs within calc scripts and reports to return lists of members that have the specified UDA associated with them. UDAs can be applied to dense as well as sparse dimensions.
8. **Report:** An Analytic Services report script enables you retrieve formatted summaries from an Analytic Services database. It consists of a text file containing Essbase Analytic Services Report Writer commands that generate one or more production reports. The report script is a text file that contains data retrieval, formatting, and output instructions.

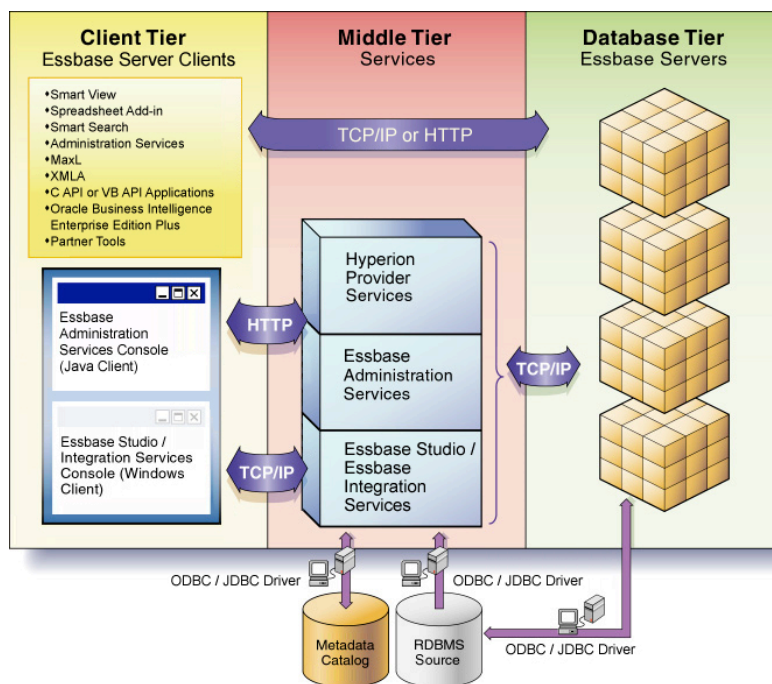


Figure 14: Essbase Architecture

Essbase Administration Services. The database tier is made up of Essbase Servers. Communication between the client and middle tiers, and the middle and database tiers, is

Essbase Architecture:

Essbase products incorporate powerful architectural features to handle a wide range of analytic applications across large multi-user environments. The client tier includes Essbase Server clients, such as Oracle Hyperion Smart View for Office, Fusion Edition, Administration Services Console and Oracle Hyperion Smart Search, Fusion Edition. The middle tier includes services, such as Oracle Hyperion Provider Services and Oracle

through HTTP. Communication between the client and database tiers is through TCP/IP or HTTP. Communication between data sources and the metadata catalog with the middle and database tiers is through ODBC and JDBC drivers.



Programming Environment:

I) Java Programming Language

A high-level programming language developed by Sun Microsystems. The language derives much of its syntax from C and C++ but has a simpler object model and fewer low-level facilities. Java applications are typically compiled to byte-code (class file) that can run on any Java Virtual Machine (JVM) regardless of computer architecture. Java is general-purpose, concurrent, class-based, and object-oriented, and is specifically designed to have as few implementation dependencies as possible. It is intended to let application developers "write once, run anywhere". Java is considered by many as one of the most influential programming languages of the 20th century, and is widely used from application software to web applications

II) Sun Java J2SE 1.6

Java++ Platform, Standard Edition or Java SE is a widely used platform for programming in the Java language. It is the Java Platform used to deploy portable applications for general use. In practical terms, Java SE consists of a virtual machine, which must be used to run Java programs, together with a set of libraries (or "packages") needed to allow the use of file systems, networks, graphical interfaces, and so on, from within those programs.

III) Eclipse IDE

Eclipse is a multi-language software development environment comprising an integrated development environment (IDE) and an extensible plug-in system. It is written primarily in Java and can be used to develop applications in Java and, by means of various plug-ins, other languages including C, C++, COBOL, Python, Perl and PHP.

IV) SWT Framework

The Standard Widget Toolkit (SWT) is a graphical widget toolkit for use with the Java platform. It was originally developed by IBM and is now maintained by the Eclipse Foundation in tandem with the Eclipse IDE. It is an alternative to the AWT and Swing Java GUI toolkits provided by Sun Microsystems as part of the Java Platform, Standard Edition.

SWT is written in Java. To display GUI elements, the SWT implementation accesses the native GUI libraries of the operating system using JNI (Java Native Interface) in a manner that is similar to those programs written using operating system-specific APIs. Programs that call SWT are portable, but the implementation of the toolkit, despite the fact that part of it is written in Java, is unique for each platform.

The toolkit is licensed under the Eclipse Public License, an Open Source Initiative approved open source license