SAP Bods

tutorialspoint

SIMPLYEASYLEARNING

www.tutorialspoint.com

https://www.facebook.com/tutorialspointindia

https://twitter.com/tutorialspoint

About the Tutorial

SAP BO Data Services (BODS) is an ETL tool used for data integration, data quality, data profiling and data processing. It allows you to integrate, transform trusted data-to-data warehouse system for analytical reporting.

BO Data Services consists of a UI development interface, metadata repository, data connectivity to source and target system and management console for scheduling of jobs. This introductory tutorial gives a brief overview of the features of SAP BODS and how to use it in a systematic manner.

Audience

This tutorial will help all those readers who want to create their own local repository, configure a job server, start basic job development and execute the job to extract data from source systems and load the data to target systems after performing transformations, look-ups and validations.

Prerequisites

Before you start proceeding with this tutorial, we assume that you have basic knowledge of Data Warehousing and Business Intelligence. Basic knowledge of any RDBMS would be an added advantage.

Copyright & Disclaimer

© Copyright 2018 by Tutorials Point (I) Pvt. Ltd.

All the content and graphics published in this e-book are the property of Tutorials Point (I) Pvt. Ltd. The user of this e-book is prohibited to reuse, retain, copy, distribute or republish any contents or a part of contents of this e-book in any manner without written consent of the publisher.

We strive to update the contents of our website and tutorials as timely and as precisely as possible, however, the contents may contain inaccuracies or errors. Tutorials Point (I) Pvt. Ltd. provides no guarantee regarding the accuracy, timeliness or completeness of our website or its contents including this tutorial. If you discover any errors on our website or in this tutorial, please notify us at contact@tutorialspoint.com

Table of Contents

About the Tutorial i
Audience i
Prerequisites i
Copyright & Disclaimer i
Table of Contents ii

DW AND ETL – INTRODUCTION 1

1. **DW – Overview 2**
   Data Warehouse – Key Features 2
   Need of a DW System 2
   Difference between DW and ODB 2
   DW Architecture 3

2. **DW – Types 4**
   Data Mart 4
   Online Analytical Processing 4
   Online Transaction Processing 5
   Differences between OLTP and OLAP 5
   Predictive Analysis 6

3. **DW – Products & Vendors 7**
   SAP Business Warehouse 7
   Business Objects & Products 8

4. **ETL – Introduction 9**

SAP BO DATA SERVICES 11

5. **SAP BODS – Overview 12**
   Data Integration & Data Management 12

6. **SAP BODS – Architecture 13**
   Product Evolution – ATL, DI & DQ 14
   BODS – Objects 14
   BODS – Object Hierarchy 16
   BODS – Tools & Functions 17

7. **Data Services Designer 18**
   Repository 18
   BODS – Naming Standards 19

SAP BODS – REPOSITORY 20

8. **BODS Repository – Overview 21**
   BODS – Repository & Types 21

tutorialspoint
SIMPLYEASYLEARNING

9. Repository – Creating & Updating 22
Creating Repository 22
Updating the Repository 26

10. Data Services Management Console 28

11. DSMC – Modules 31
Administrator Module 31
Nodes 31

12. DS Designer – Introduction 36

13. ETL Flow in DS Designer 39
Create an ETL Flow 42

DATASTORES & FORMATS 45

14. Datastore – Overview 46
Create Datastore for a Database 47

15. Changing a Datastore 48

16. Memory Datastore 49
Creating a Memory Datastore 49
Memory Table as Source and Target 51

17. Linked Datastore 52

18. Adapter Datastore 53
Adapter Datastore – Definition 53

19. File Formats 54
Creating a File Format 55
Editing a File Format 56

20. COBOL Copybook File Format 58

21. Extracting Data from Database Tables 60
Importing Metadata 60

22. Data Extraction from Excel Workbook 62
Data Extraction from XML FILE DTD, XSD 63
Data Extraction from COBOL Copybooks 63

DATA FLOW AND WORK FLOW 64

23. Data Flow – Introduction 65
Example of Data Flow 65
Passing Parameters 67

tutorialspoint
SIMPLYEASYLEARNING

24. Data Flow – Changing Properties 68
Source and Target Objects 69

25. Workflow – Introduction 71
Example of Work Flow 71

26. Creating Workflows 72
Conditionals 73

SAP BODS – TRANSFORMS 75

27. Transforms – Types 76

28. Adding Transform to a Dataflow 77

29. Query Transform 79
Data Quality Transform 80
Text Data Processing Transform 81
Entity Extraction Transform 82
Differences between TDP and Data Cleansing 82

SAP BODS – ADMINISTRATION 83

30. Data Services – Overview 84
Real Time Jobs 84
Real Time vs Batch Jobs 84
Creating Real Time Jobs 84
Testing Real Time Jobs 86
Embedded Data Flows 86

31. Creating Embedded Data Flow 88
Variables and Parameters 89
Defining Local Variable 91

32. Debugging & Recovery Mechanism 93

33. Data Assessment – Data Profiling 95
Connecting to Profiler Server 96

34. Tuning Techniques 97

MULTIUSER DEVELOPMENT 98

35. Central vs Local Repository 99
Multiple Users 99

36. Central Repository – Security 101
Creating Non Secure Central Repository 101
Creating a Secure Central Repository 103

tutorialspoint
SIMPLYEASYLEARNING

37. Creating a Multiuser Environment 104
Migrating Multiuser Jobs 104
Central Repository Migration 104

tutorialspoint
SIMPLYEASYLEARNING

DW and ETL – Introduction

tutorialspoint
SIMPLYEASYLEARNING

1.DW – Overview

A Data warehouse is known as a central repository to store the data from one or multiple heterogeneous data sources. Data warehouse is used for reporting and analyzing of information and stores both historical and current data. The data in DW system is used for Analytical reporting, which is later used by Business Analysts, Sales Managers or Knowledge workers for decision-making.

The data in DW system is loaded from an operational transaction system like Sales, Marketing, HR, SCM, etc. It may pass through operational data store or other transformations before it is loaded to the DW system for information processing.

Data Warehouse – Key Features

The key features of a DW System are:

- It is central data repository where data is stored from one or more heterogeneous data sources.
- A DW system stores both current and historical data. Normally a DW system stores 5-10 years of historical data.
- A DW system is always kept separate from an operational transaction system.
- Data in DW system is used for different types of analytical reporting range from Quarterly to Annual comparison.

Need of a DW System

Suppose you have a home loan agency where data is coming from multiple applications like- marketing, sales, ERP, HRM, MM etc. This data is extracted, transformed and loaded in Data Warehouse.

For example, if you have to compare the Quarterly/Annual sales of a product, you cannot use an Operational transactional database, as this will hang the transaction system. Therefore, a Data Warehouse is used for this purpose.

Difference between DW and ODB

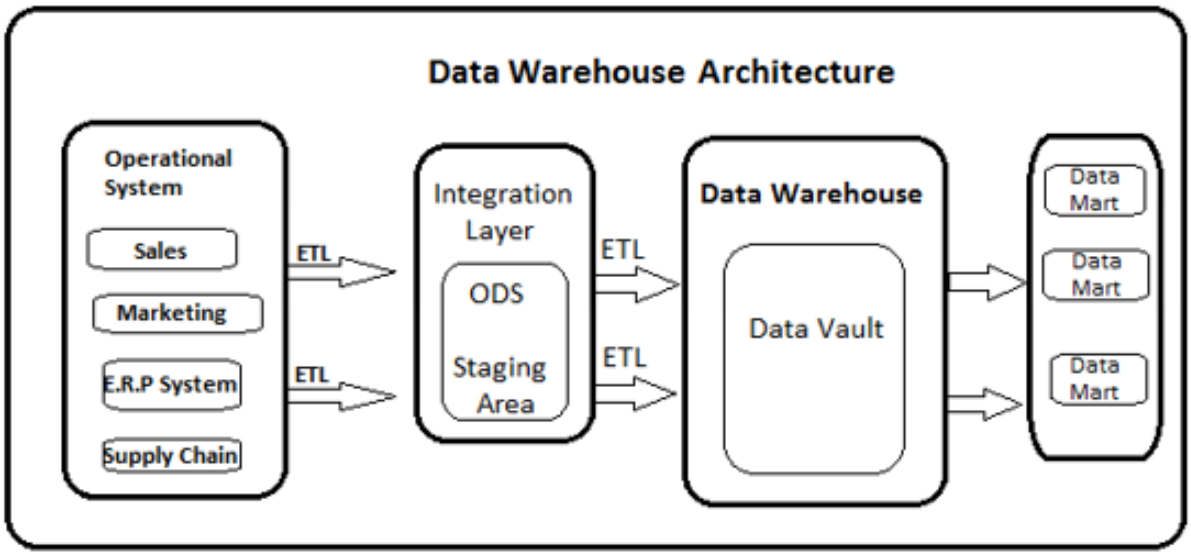
The differences between a Data Warehouse and Operational Database (Transactional Database) are as follows:

- A Transactional system is designed for known workloads and transactions like updating a user record, searching a record, etc. However, Data Warehouse transactions are more complex and present a general form of data.

- A Transactional system contains the current data of an organization and Data warehouse normally contains the historical data.
-
- Transactional system supports parallel processing of multiple transactions. Concurrency control and recovery mechanisms are required to maintain consistency of the database.
-
- An Operational database query allows to read and modify operations (delete and Update) while an OLAP query needs only read-only access of stored data (Select statement).

DW Architecture

Data Warehousing involves data cleaning, data integration, and data consolidations.



A Data Warehouse has a 3-layer architecture: **Data Source Layer, Integration Layer**, and **Presentation Layer**. The illustration given above shows the common architecture of a Data Warehouse system.

tutorialspoint
SIMPLYEASYLEARNING

2.DW – Types

There are four types of Data Warehousing system.

- Data Mart
- Online Analytical Processing (OLAP)
- Online Transactional Processing (OLTP)
- Predictive Analysis (PA)

Data Mart

A Data Mart is known as the simplest form of a Data Warehouse system and normally consists of a single functional area in an organization like sales, finance or marketing, etc.

Data Mart in an organization and is created and managed by a single department. As it belongs to a single department, the department usually gets data from only a few or one type of sources/applications. This source could be an internal operational system, a data warehouse or an external system.

Online Analytical Processing

In an OLAP system, there are less number of transactions as compared to a transactional system. The queries executed are complex in nature and involves data aggregations.

What is an Aggregation?

We save tables with aggregated data like yearly (1 row), quarterly (4 rows), monthly (12 rows) or so, if someone has to do a year to year comparison, only one row will be processed. However, in an un-aggregated table it will compare all rows.

```
SELECT SUM(salary)
FROM employee
WHERE title = 'Programmer';
```

Effective Measures in an OLAP system

Response time is known as one of the most effective and key measure in an **OLAP** system. Aggregated stored data is maintained in multi-dimensional schemas like star schemas (When data is arranged into hierarchical groups, often called dimensions and into facts and aggregate facts, it is called Schemas).

The latency of an OLAP system is of a few hours as compared to the data marts where latency is expected closer to a day.

Online Transaction Processing

In an OLTP system, there are a large number of short online transactions such as INSERT, UPDATE, and DELETE.

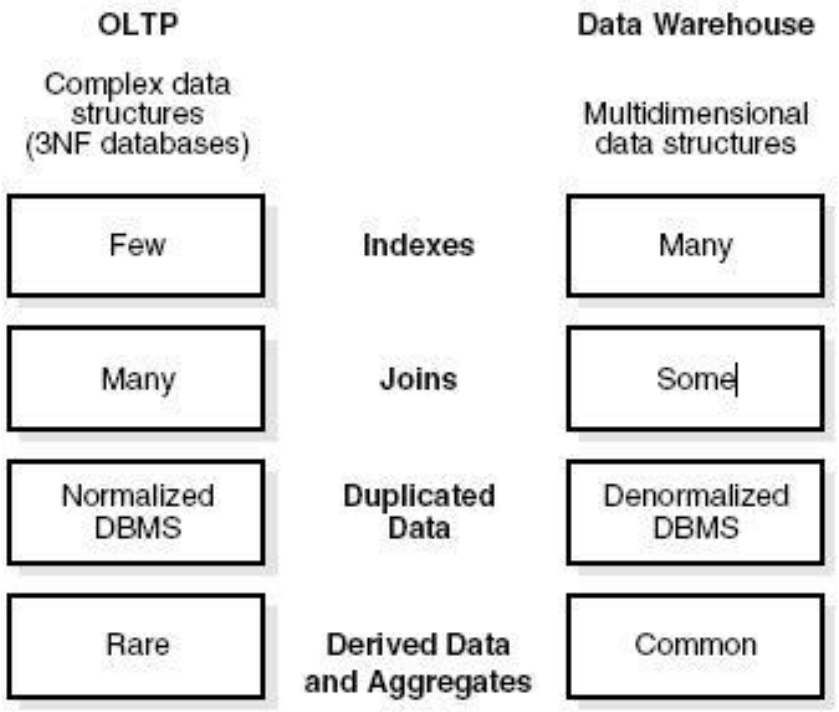
In an OLTP system, an effective measure is the processing time of short transactions and is very less. It controls data integrity in multi-access environments. For an OLTP system, the number of transactions per second measures the **effectiveness**. An OLTP data warehouse system contains current and detailed data and is maintained in the schemas in the entity model (3NF).

Example

Day-to-Day transaction system in a retail store, where the customer records are inserted, updated and deleted on a daily basis. It provides very fast query processing. OLTP databases contain detailed and current data. Schema used to store OLTP database is the Entity model.

Differences between OLTP and OLAP

The following illustrations shows the key differences between an **OLTP** and **OLAP** system.



- **Indexes:** OLTP system has only few indexes while in an OLAP system there are many indexes for performance optimization.

- **Joins:** In an OLTP system, large number of joins and data are normalized. However, in an OLAP system there are less joins and are de-normalized.

tutorialspoint
SIMPLYEASYLEARNING

- **Aggregation:** In an OLTP system, data is not aggregated while in an OLAP database more aggregations are used.

Predictive Analysis

Predictive analysis is known as finding the hidden patterns in data stored in DW system by using different mathematical functions to predict future outcomes.

Predictive Analysis system is different from an OLAP system in terms of its use. It is used to focus on future outcomes. An OALP system focuses on current and historical data processing for analytical reporting.

tutorialspoint
SIMPLYEASYLEARNING

3.DW – Products & Vendors

There are various Data Warehouse/database systems available in the market that meet the capabilities of a DW system. The most common vendors for data warehouse systems are:

- Microsoft SQL Server
- Oracle Exadata
- IBM Netezza
- Teradata
- Sybase IQ
- SAP Business Warehouse (SAP BW)

SAP Business Warehouse

SAP Business Warehouse is a part of SAP NetWeaver release platform. Prior to NetWeaver 7.4, it was referred to as SAP NetWeaver Business Warehouse.

Data warehousing in SAP BW means data integration, transformation, data cleansing, storing and data staging. DW process includes data modeling in BW system, staging and administration. The main tool, which is used to manage DW tasks in BW system, is the administration workbench.

Key Features

- SAP BW provides capabilities like Business Intelligence, which includes Analytical Services and Business Planning, Analytical Reporting, Query processing and information, and Enterprise data warehousing.

- It provides a combination of databases and database management tools that helps in making decision.

- Other key features of BW system include Business Application Programming Interface (BAPI) that supports connection to non-SAP R/3 applications, automated data extraction and loading, an integrated OLAP processor, metadata repository, administration tools, multi-language support, and a web enabled interface.

- SAP BW was first introduced in 1998 by SAP, a German company. SAP BW system was based on a model-driven approach to make Enterprise Data Warehouse easy, simple and more efficient for SAP R3 data.

- From last 16 years, SAP BW has evolved as one of the key system for many companies to manage their enterprise data warehousing needs.

- The Business Explorer **(BEx)** provides an option for flexible reporting, strategic analysis and operative reporting in the company.

- It is used to perform reporting, query execution and analysis function in BI system. You can also process current and historical data up to various degree of details over Web and in Excel format.

- Using **BEx** information broadcasting, BI content can be shared via email as document or in the form of links as live data or you can also publish using SAP EP functions.

Business Objects & Products

SAP Business Objects in known as the most common Business Intelligence tool and is used for manipulating data, user access, analyzing, formatting and publishing information on different platforms. It is a front-end based set of tools, which enables business users and decision makers to display, sort, and analyze business intelligence current and historical data.

It comprises of the following tools:

Web Intelligence

Web Intelligence (WebI) is called as the most common Business Objects detailed reporting tool that supports various features of data analysis like drill, hierarchies, charts, calculated measures, etc. It allows end-users to create ad-hoc queries in query panel and to perform data analysis both online and offline.

SAP Business Objects Xcelsius / Dashboards

Dashboards provide data visualization and dash-boarding capabilities to the end-users and you can create interactive dashboards using this tool.

You can also add various types of charts and graphs and create dynamic dashboards for data visualizations and these are mostly used in financial meetings in an organization.

Crystal Reports

Crystal Reports are used for pixel-perfect reporting. This enables the users to create and design reports and later use it for printing purpose.

Explorer

The Explorer allows a user to search the content in BI repository and best matches are shown in the form of charts. There is no need to write down the queries to perform search.

Various other components and tools introduced for detailed reporting, data visualization and dash-boarding purpose are Design Studio, Analysis edition for Microsoft Office, BI Repository and Business Objects Mobile platform.

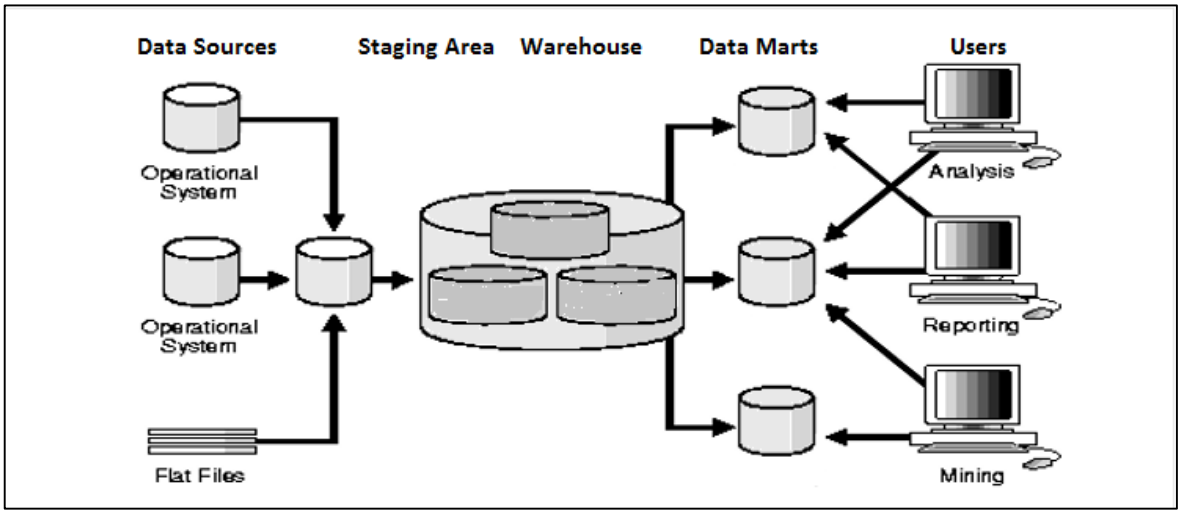
tutorialspoint
SIMPLYEASYLEARNING

4.ETL – Introduction

ETL stands for Extract, Transform and Load. An ETL tool extracts the data from different RDBMS source systems, transforms the data like applying calculations, concatenate, etc. and then load the data to Data Warehouse system. The data is loaded in the DW system in the form of dimension and fact tables.

Extraction

- A staging area is required during ETL load. There are various reasons why staging area is required.
- The source systems are only available for specific period of time to extract data. This period of time is less than the total data-load time. Therefore, staging area allows you to extract the data from the source system and keeps it in the staging area before the time slot ends.
- Staging area is required when you want to get the data from multiple data sources together or if you want to join two or more systems together. For example, you will not be able to perform a SQL query joining two tables from two physically different databases.
- Data extractions' time slot for different systems vary as per the time zone and operational hours.
- Data extracted from source systems can be used in multiple data warehouse system, Operation Data stores, etc.
- ETL allows you to perform complex transformations and requires extra area to store the data.



tutorialspoint
SIMPLYEASYLEARNING

Transform

In data transformation, you apply a set of functions on extracted data to load it into the target system. Data, which does not require any transformation is known as direct move or pass through data.

You can apply different transformations on extracted data from the source system. For example, you can perform customized calculations. If you want sum-of-sales revenue and this is not in database, you can apply the **SUM** formula during transformation and load the data.

For example, if you have the first name and the last name in a table in different columns, you can use concatenate before loading.

Load

During Load phase, data is loaded into the end-target system and it can be a flat file or a Data Warehouse system.

tutorialspoint
SIMPLYEASYLEARNING

SAP BO Data Services

tutorialspoint
SIMPLYEASYLEARNING

5.SAP BODS – Overview

SAP BO Data Services is an ETL tool used for Data integration, data quality, data profiling and data processing. It allows you to integrate, transform trusted data-to-data warehouse system for analytical reporting.

BO Data Services consists of a UI development interface, metadata repository, data connectivity to source and target system and management console for scheduling of jobs.

Data Integration & Data Management

SAP BO Data Services is a data integration and management tool and consists of Data Integrator Job Server and Data Integrator Designer.

Key Features

- You can apply various data transformations using Data Integrator language to apply complex data transformations and building customized functions.
- Data Integrator Designer is used to store real time and batch jobs and new projects in repository.
- DI Designer also provides an option for team based ETL development by providing a central repository with all basics functionality.
- Data Integrator job server is responsible to process jobs that are created using DI Designer.

Web Administrator

Data Integrator web administrator is used by system administrators and database administrator to maintain repositories in Data services. Data Services includes Metadata Repository, Central Repository for team-based development, Job Server and Web Services.

Key functions of DI Web Administrator:

- It is used to schedule, monitor and execute batch jobs.
- It is used for the configuration and start and stop real-time servers.
- It is used for configuring Job Server, Access Server, and repository usage.
- It is used for configuring adapters.
- It is used for configuring and controlling all the tools in BO Data Services.

Data Management function emphasizes on data quality. It involves data cleansing, enhancing and consolidating the data to get correct data in the DW system.

tutorialspoint
SIMPLYEASYLEARNING

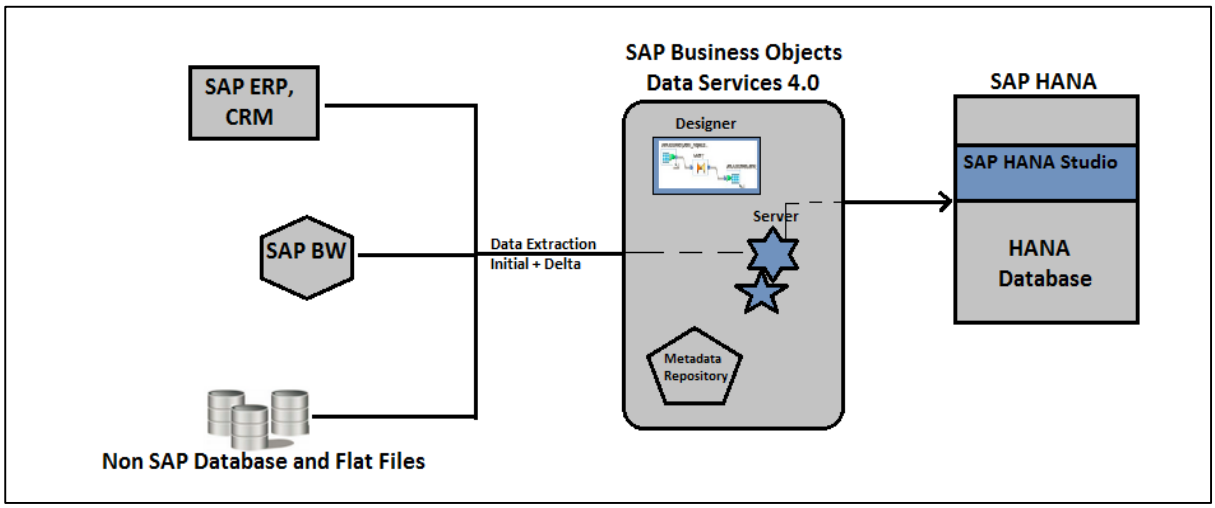
6.SAP BODS – Architecture

In this chapter, we will learn about the SAP BODS architecture. The illustration shows the architecture of BODS system with Staging area.

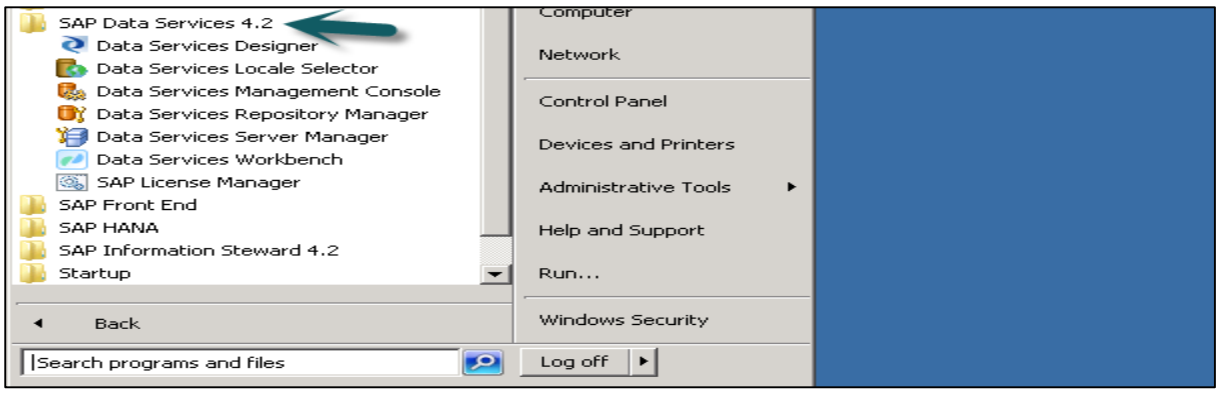
Source Layer

The source layer includes different data sources like SAP applications and non-SAP RDBMS system and data integration takes place in staging area.

SAP Business Objects Data Services includes different components like Data Service Designer, Data Services Management Console, Repository Manager, Data Services Server Manager, Work bench, etc. The target system can be a DW system like SAP HANA, SAP BW or a non-SAP Data warehouse system.



The following screenshot shows the different components of SAP BODS.

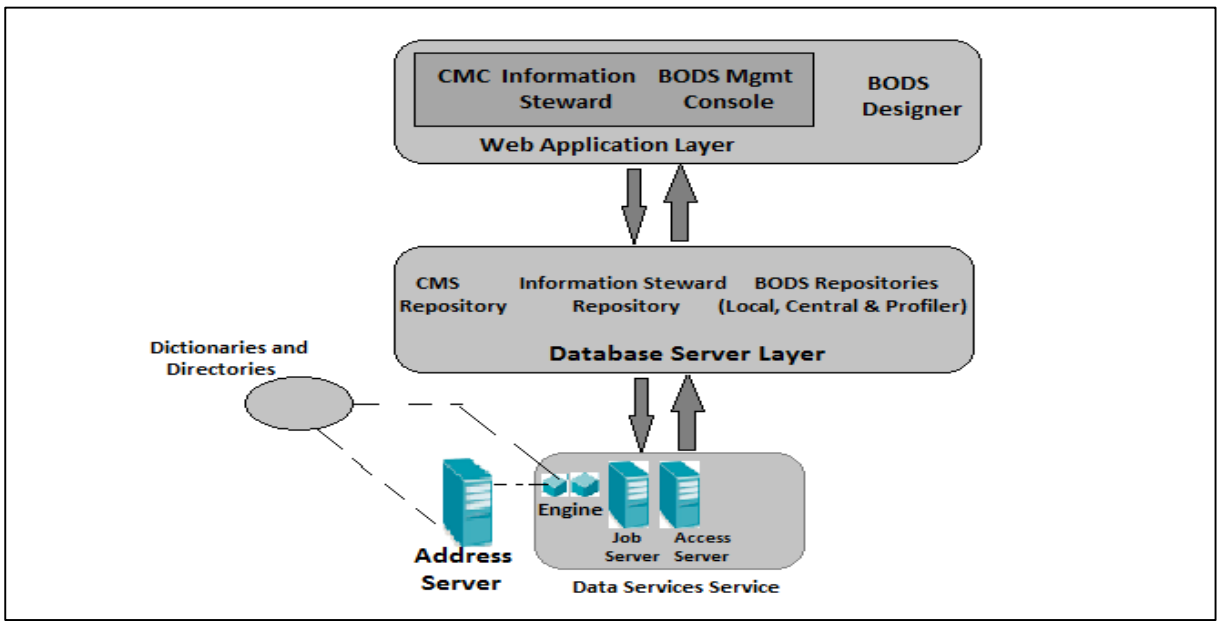


You can also divide BODS architecture in the following layers:

- Web Application Layer
- Database Server Layer
- Data Services Service Layer

tutorialspoint
SIMPLYEASYLEARNING

The following illustration shows the BODS architecture.



Product Evolution – ATL, DI & DQ

Acta Technology Inc. developed SAP Business Objects Data Services and later Business Objects Company acquired it. Acta Technology Inc. is a US based company and was responsible for development of first-data integration platform. The two ETL software products developed by Acta Inc. was the **Data Integration (DI)** tool and the **Data Management** or **Data Quality** (**DQ**) tool.

Business Objects, a French company acquired Acta Technology Inc. in 2002 and later, both the products were renamed as **Business Objects Data Integration (BODI)** tool and **Business Objects Data Quality (BODQ)** tool.

SAP acquired Business Objects in 2007 and both the products were renamed as SAP BODI and SAP BODQ. In 2008, SAP integrated both the products into single software product named as SAP Business Objects Data Services (BODS).

SAP BODS provides data integration and data management solution and in the earlier version of BODS, the text data-processing solution was included.

BODS – Objects

All the entities that are used in BO Data Services Designer are called **Objects**. All the objects like projects, jobs, metadata and system functions are stored in the local object library. All the objects are hierarchical in nature.

The objects mainly contain the following:

- **Properties** – They are used to describe an object and do not affect its operation. Example- Name of an object, Date when it is created, etc.

tutorialspoint
SIMPLYEASYLEARNING

- **Options –** Which control the operation of objects.

Types of Objects

There are two types of objects in the system – Reusable objects and Single Use objects. The type of object determines how that object is used and retrieved.

Reusable Objects

Most of the objects that are stored in the repository can be reused. When a reusable object is defined and saved in the local repository, you can reuse the object by creating Calls to the definition. Each reusable object has only one definition and all the calls to that object refer to that definition. Now, if definition of an object is changed at one place you are changing the object definition at all the places where that object appears.

An object library is used to contain object definition and when an object is dragged and dropped from library, a new reference to an existing object is created.

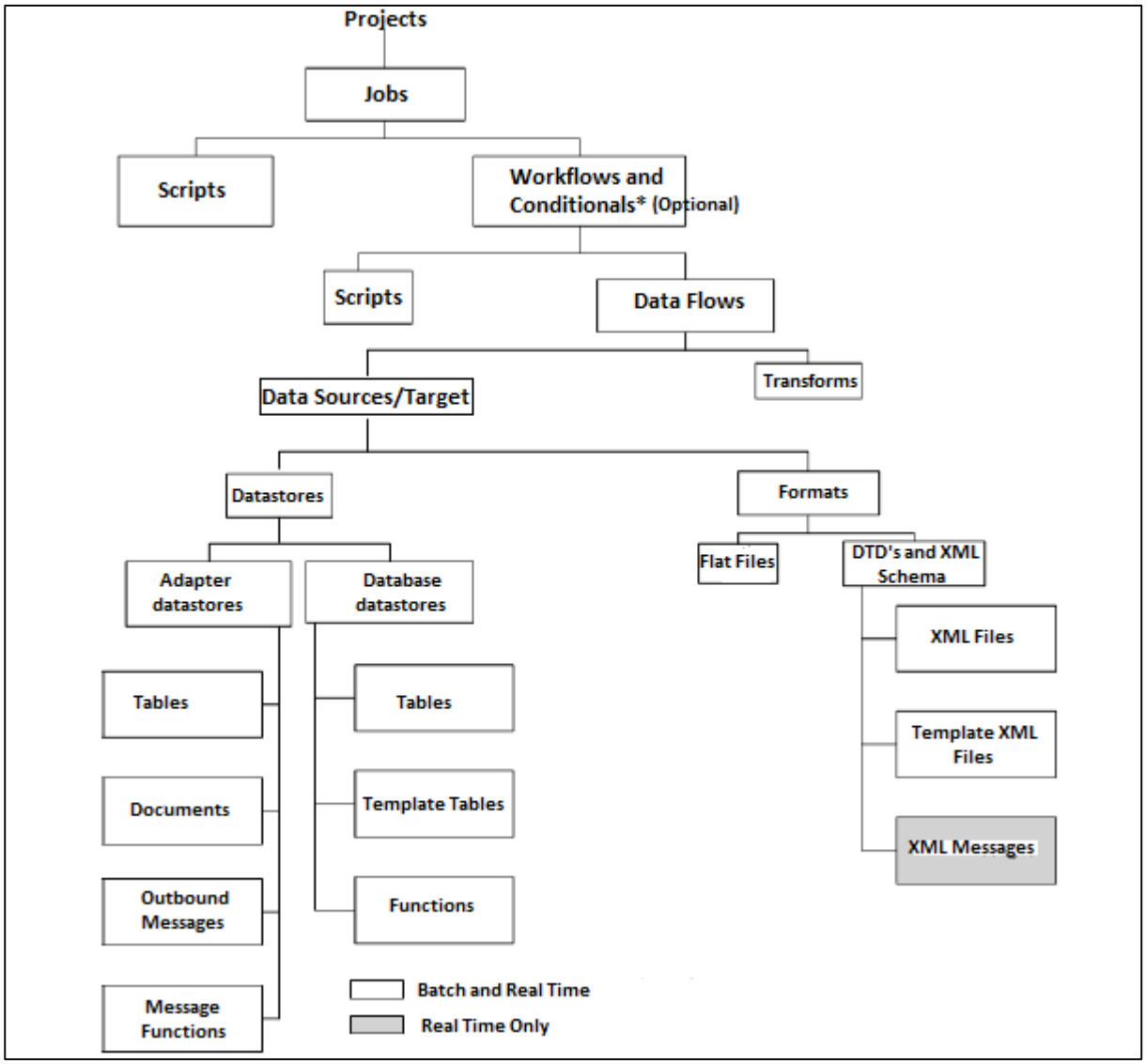
Single Use Objects

All the objects that are defined specifically to a job or data flow are known as single use objects. For example, specific transformation used in any data load.

tutorialspoint
SIMPLYEASYLEARNING

BODS – Object Hierarchy

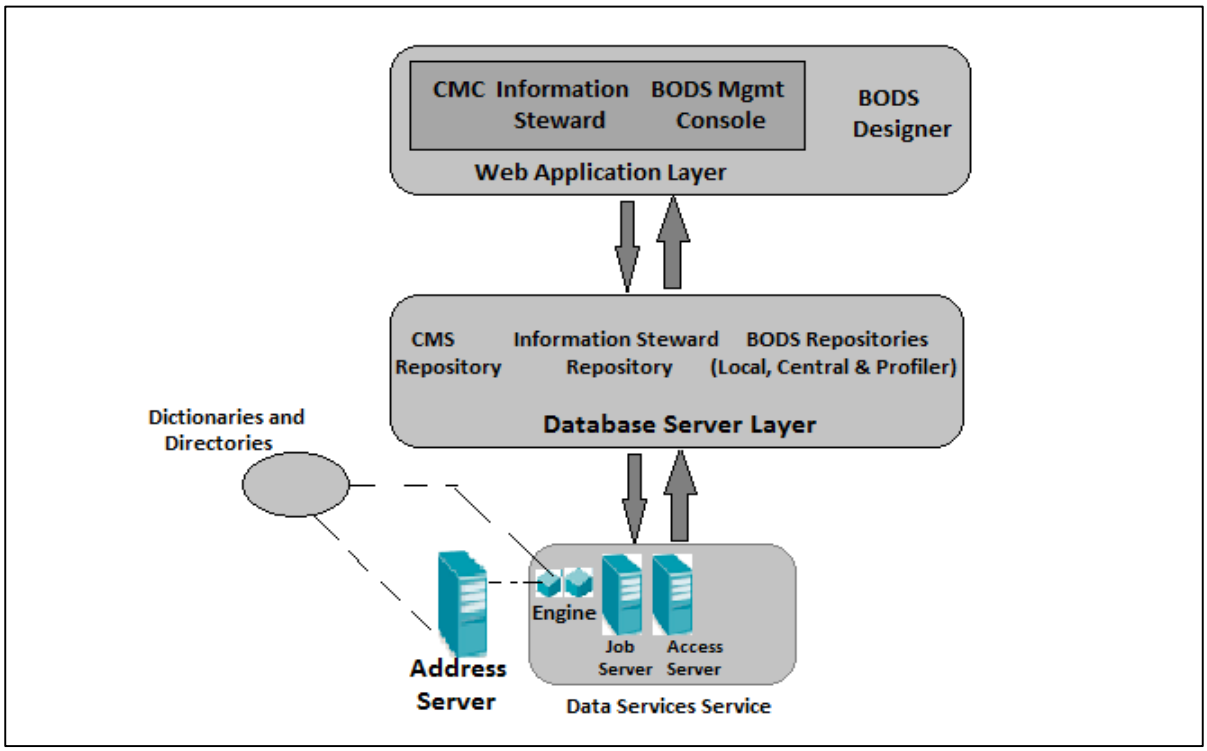
All the objects are hierarchical in nature. The following diagram shows the object hierarchy in SAP BODS system:



tutorialspoint
SIMPLYEASYLEARNING

BODS – Tools & Functions

Based on the architecture illustrated below, we have many tools defined in SAP Business Objects Data Services. Each tool has its own function as per system landscape.



At the top, you have Information Platform Services installed for users and rights security management. BODS depends on Central Management console **(CMC)** for user access and security feature. This is applicable to the 4.x version. In the previous version, it was done in Management Console.

tutorialspoint
SIMPLYEASYLEARNING

7.Data Services Designer

Data Services Designer is a developer tool, which is used to create objects consisting of data mapping, transformation, and logic. It is GUI based and works as a designer for Data Services.

Repository

Repository is used to store metadata of objects used in BO Data Services. Each Repository should be registered in Central Management Console and is linked with single or many job servers, which are responsible for executing jobs that are created by you.

Types of Repositories

There are three types of Repositories.

- **Local Repository:** It is used to store the metadata of all objects created in Data Services Designer like projects, jobs, data flow, work flow, etc.

- **Central Repository:** It is used to control the version management of the objects and is used for multiuse development. Central Repository stores all the versions of an application object. Hence, it allows you to move to previous versions.

- **Profiler Repository:** This is used to manage all the metadata related to profiler tasks performed in SAP BODS designer. CMS Repository stores metadata of all the tasks performed in CMC on BI platform. Information Steward Repository stores all the metadata of profiling tasks and objects created in information steward.

Job Server

Job server is used to execute the real time and batch jobs created by you. It gets the job information from respective repositories and initiates the data engine to execute the job. Job server can execute the real time or scheduled jobs and uses multithreading in memory caching, and parallel processing to provide performance optimization.

Access Server

Access Server in Data Services is known as real time message broker system, which takes the message requests, moves to real time service and displays a message in specific time frame.

Data Service Management Console

Data Service Management Console is used to perform administration activities like scheduling the jobs, generating the quality reports in DS system, data validation, documentation etc.

tutorialspoint
SIMPLYEASYLEARNING

BODS – Naming Standards

It is advisable to use standard naming conventions for all the objects in all systems as this allows you to identify objects in Repositories easily.

The table shows the list of recommended naming conventions that should be used for all jobs and other objects.

tutorialspoint
SIMPLYEASYLEARNING

End of ebook preview

If you liked what you saw…

Buy it from our store @ https://store.tutorialspoint.com

tutorialspoint
SIMPLYEASYLEARNING