

# **MS-BI**

**(SSIS,SSAS,SSRS)**

## **DATA WAREHOUSING TECHNOLOGIES**

## **sri raghavendra Xerox**

**All software language materials available**

**beside banglore iyyengars bakery opp:cdac balkampetroad ameerpet Hyderabad**

**cell :9951596199**



# MSBI INDEX

## SSIS

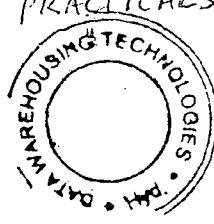
"Vinay Sir notes"

PAGE NOS

1. OPPORTUNITIES & MSBI PRESENCE	1
2. MSBI COMPETITIVE ADVANTAGES, DATA EVALUATION STAGES	2-3
3. DIFF BETWEEN 2000, 2005, 2008, 2008R2, 2011, DTS TO SSIS	4-7 14-16
4. INSTALLATION STEPS, REQUIREMENTS	7-8
5. WORKING MANAGEMENT STUDIO	9-11
6. SSIS ARCHITECTURE & COMPONENTS	12-13
7. SSIS PRACTICALS [ ARCHITECTURE, EXECUTION, DATAFLOW TASK OPERATIONS, SOURCES, DESTINATIONS, TRANSFORMATIONS]	16-45
8. CONTROL FLOW ITEMS [ CONTAINERS, TASKS, MAINTENANCE TASKS, EVENT HANDLING]	45-52 658
9. SCD [ WIZARD & MANUAL]	53-60
10. DEBUGGING, PRECEDENCE CONSTRAINTS	62-64
11. CHECK POINT, LOGGING, CONFIGURATIONS	64-68
12. DEPLOYMENT & SECURITY	68-72
13. INCREMENTAL LOADING, IMPORT EXPORT WIZARD	72-75
14. TRANSACTIONS, JOBS CREATION	75-79
15. PERFORMANCE TUNING	79-84

## OTHER CONTENTS:

1. REALTIME BUSINESS SPECIFICATION DOCUMENT  
[ SOURCEMAP, LLD ] - FOR PRACTICALS
2. DWH & BI FUNDAMENTALS
3. NORMALIZATION & T-SQL



# MICROSOFT BUSINESS INTELLIGENCE [MS-BI]

## Opportunities :-

1. MSBI Suite Developer.
2. SSIS " "
3. SSAS " "
4. SSRS " "
5. .NET + [SSIS (or) SSRS]
6. DWH [ETL (or) Reporting Tools] + MSBI.
7. ETL Testing.
8. SQL Server developer + MSBI.
9. SQL Server DBA + MSBI
- g. MSBI FRESHERS.

## MSBI Presence in the market :-

### DATAWAREHOUSING:-

IT is an RDBMS which has a huge volume of data in the support of Business decisions.

### ETL :-

E → Extract  
(Getting the data)

T → Transform  
(Performing intermediate operations).

Eg: currencies, <sup>Conversion</sup> etc.

L → Load (load to destination).

Tools:- Informatica, Data stage, ABINITIO, OWB, BODI etc.

Eg: Analytical operations / Business Decision.

### Transactional Sys (or) Traditional system

i.e., Day-to-Day Transactions.

(1900 - 2000)

Architects  
Founders  
Scientists

### DWH & BI

Approach.  
[Informational approach]

DATAWAREHOUSING / SSAS

.NET  
SQL server

JAVA  
Oracle

SAP  
Oracle

MAIN FRAMES  
DB2

2 dimensional operations

Analysis Tools:  
[SSAS]

DATAWAREHOUSING / SSAS

Reporting Tools / SSRS

[END USERS]

Business Analysts]

Analysis Tools: creates Multi Dimensional objects and provides multidimensional analysis e.g. cube ↓ i.e., Multi dimensional operations.

Tools: COGNOS, BO, HYPERION, MICROSTRATEGY, OBIEE.. etc.

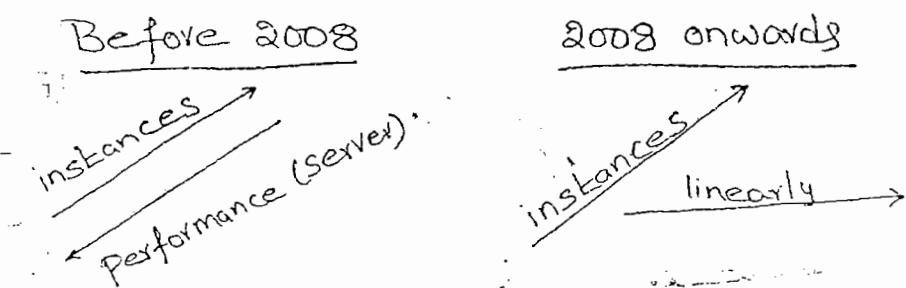
Reporting Tools: To represent the data in understandable format. To the end users/Analysts we require these tools. i.e., word, PDF, Excel sheets...etc.

Tools: COGNOS, BO, HYPERION, CRYSTAL REPORTS, PANORAMA, MICROSTRATEGY, OBIEE.. etc.

NOTE: The information we received in DW/H is in 2-dimensional data (i.e., rows and columns). FORMAT.

### MSBI Competitive Advantages

1. It is a single suite for end-to-end business solution.
2. It has extended capability for extraction, transformation and loading.
3. Low cost TCO [Total cost of Ownership]
  - Easy to Install
  - Easy to use
  - Easy to Maintain
  - cheaper price.
4. Highly Scalable – It supports multiple instances to connect to the server without sacrificing its performances.



- ⑥ → The BIMI is said to be "linear Scalable application" due to having CMS (Central Management Server) and PBM (Policy Based Management) servers. (2008 R<sub>2</sub>)

### Full Compatibility between the components of MSBI:-

→ All components runs in the single runtime called 'CLR' [Common Language Runtime]. i.e., it is having similar coding, Notation (naming).

→ Full Support to .Net, XML and webservices  
 ↓  
 (Universal language)

→ MSBI Provides very good support to the semiware house application (If we are reusing OLTP functionality in ware house).



### DATA Evaluation Stages

Traditional systems	SYSTEM TYPE	Time Taken	No of Rows it operates
OLTP (Online Transaction Processing) Eg: Immediate operations Booking Ticket, movie Tkt etc		Seconds	Less
DSS (Decision Support System) Eg: Minor Analysis The no of Saving Acc. opened in yr 2009 & 10		Minutes/hours	More
OLCP (Online complex processing) Eg: Loan Applications.		Min/hrs/Days	More
OLAP (Online Analytical Processing) main or complex Analytical System.		Min/hrs/Days	More

Eg:- The no of Saving accounts opened for the year 2009 & 2010 every month wise and every location of INDIA & SINGAPORE.

→ Differences between 2000 & 2005

	<u>2000</u>	<u>2005</u>
1.	DTS (Data Transformation Service) file ↔ file Database ↔ Database Database ↔ file (-DTS)	SSIS (SQL Server Integration services) 
	MSAS	SSAS [Analysis Services]
	No Reporting	SSRS [Reporting Services]
	No Notifications	SSNS [Notification Services]

2. XML Data Type Services
3. MDX [Multi Dimensional Extension] (or) Expressions.

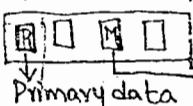
#### DBA features

- DISK mirroring
- ONLINE Restoring
- ONLINE indexing
- Dedicated DBA connection

#### Developer features

- Hosted CLR
- ADO.NET support
- Web services Support
- Advanced T-SQL etc.

Disk Mirroring:- specially designed for Administrators, Disk level one & only recovery mechanism is RAID:



Mirror image

Online Restoring :- DB operations, Restoring (Backing up) can be done simultaneously [database] in the same instance.

online Indexing :- Indexes creations & DB operations perform simultaneously in same database

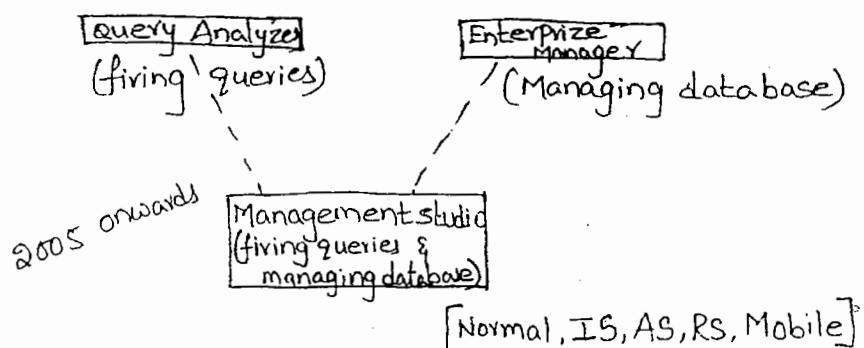
## Bussiness Intelligence features:-

1. SQL Server Integration Services (SSIS)
2. " " Analysis " (SSAS)
3. " " Reporting " (SSRS)
4. " " Notification " (SSNS)
5. Proactive caching
6. Report Builder.
7. Data mining.
8. MS-office full support.



## Management studio:-

Till 2000



Hosted CLR: However .Net languages running under a single runtime (CLR). MSBI runs under CLR.

Advantage: we can create tables (or) other objects in C-Sharp dot Net (or) VC# .Net later we can import it into SQL Server DB.

→ Debugging is more simpler than Solving in database.

## Advanced T-SQL:

There are many keywords added in 2005 to satisfy the demands of BI people.



SSIS- It is an High end platform which performs ETL (Extract, Transform, Load) operations and administrative tasks.

SSAS- It is an "MOLAP" tool to create, manipulate and provide multidimensional objects and analysis.

SSRS- It represents the data in understandable format i.e., word document, PDF etc.

SSNS- This is used to send (or) receive notifications based on the event of different applications, scripts (or) components.

Proactive Caching- used in analysis services to keep the cube and source databases in sync with each other (update)

Report Builder- Designed specially to create Adhoc reports.

Data Mining- It is knowledge analysis and discovery mechanism.

In terms of analysis we get the answers from questions like what happened? and why it has happened? but in datamining we get the answers for the questions like what will happen in the future?

Changes done (or) Made to the existing BI components:

SSIS Level:

- Data profile Task . . . . . Added
- Many Transfer Task . . . . . Added
- changes to @ lookup transformation
  - (b) for each loop container
  - (c) SCRIPT Task (C# .Net added)

SSRS Level:

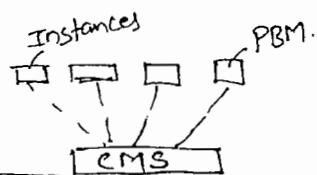
- Architecture changes
- IIS Independent
- Rich memory management
- DUNDAS Acquisition
- New rendering architecture etc...



⑦

- 0 SSAS Level → Design. : look & feel changed
- 0 → Default Attributes display changed.
- 0 → Added separate Tab "Design Aggregations" etc
- 2 → Processing options added/removed.
- 3 Differences between 2005 & 2008:-

- 0 1. Date, Time datatype added. (It supports 26 digits).
- 0 2. SPATIAL datatypes added.
- 0 3. GEO SPATIAL applications added.
- 0 4. Resource Governor Implemented.
  - 0 ↳ It does governance of various resources in 2008.
- 0 5. CMS added → All the instances managed under a single manager Server interface.
- 0 6. PBM → There are common policies implemented between various instances in CMS.
- 0 7. SSNS removed.



### Hardware & Software Requirements for MSBI Installation Steps :-

#### H/W requirements:-

1. 1 GHz Pentium-III compatible (or) faster processor (2 GHz or faster recommended).
2. 512 MB of RAM (or) more (2GB or higher recommended).
3. 2.1 GB free hard disk space for SQL Server installation files and samples.

#### S/W requirements:-

1. A Compatible operating system.

Many versions of Windows Server and desktop OS, including Win XP or later, Win Server 2003 (with SP2), Windows Vista, and Win Server 2008. → .NET Framework 2.5.

## Installation process :-

→ click "Setup"

↳ Select first option from the window opened.

i.e., New SQL Server stand-alone installation (or) added features

↳ Click on the required components

SSIS, AS, RS, BI Development Studio, management tool etc.

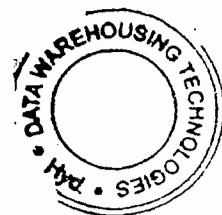
→ Instance configuration.

→ Server configuration.

Specify the same account, password for all the below services (or) different user id, password for each service.

### Services:-

- i. SQL Service Agent
- ii. SQL Service Database Engine
- iii. SQL Service Analysis Services.
- iv. SQL Service Reporting Services.
- v. SQL Service Integration Services 10.0.



Note:- Before Installation process we should have a valid user account and password.

→ Data base configuration:-

Authentication mode

↳ Select mixed mode (SQL Server authentication & windows authentication).

Built-in SQL Server system administrator account

User name:-

Password:-

→ Reporting services configuration

↳ Select "Install the native mode default configuration" option

→ After clicking the option it asks userid & password specify & click next..

(9)

## Tools and utilities required to work with MS-BI:

### ① BIDS (SQL Server Business Intelligence Development Studio):

This studio is required to work with IS, AS & RS applications.

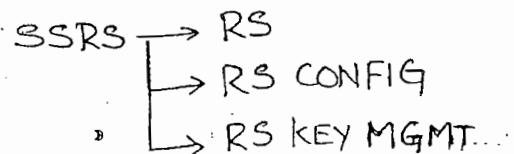
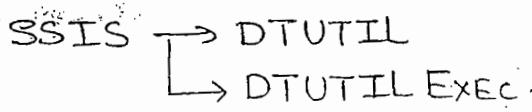
Navigation: Start → programs → SQL Server 2008 → SS BIDS

### ② SSMS (SQL Server Management Studio):

This is useful to work with IS, AS, RS, SQL Server mobile and normal databases (Managing and for firing queries on the databases, it is useful).

Navigation: Start → programs → SQL Server 2008 → SSMS.

### ③ Command Line Utilities:-



## Working with Management Studio:-



### ① open SSMS (SQL Server Management Studio)

### ② Select server type in the following

1. Database Engine → To connect and work with normal database.

2. Integration Service → To connect and work with <sup>Integration</sup> Service database.

3. Analysis Service → To connect and work with analysis services databases.

4. Reporting Services → To connect and work with Reporting Services databases.

5. SQL Server compact Edition → To connect and work with mobile applications.

(i) specify Server name → Either IP Address (or) actual Server name.

(ii) Select Authentication →

(i) Windows Authentication

↳ consider windows credentials to logon.

(ii) SQL Server Authentication

↳ server database credentials.

Note: select the option (ii) for the real time work process.

③ click connect ... Next ...

Creating Databases

↳ rt click on database

New database window

Database Name ← DBNEW.



→ click OK.

creating tables → DB New → Tables → rt click →

select 'New Table'

column Name      DataType

PARTY ID            Int

PARTY NAME        Varchar(50)

→ Next Save the table with a specified Name i.e., PARTY

→ click OK.

Adding data to the table of the database

→ rt click on TABLE → select the option "edit Top 200 rows".

PARTY ID      PARTY NAME

1  
2

X  
VINAY



## Firing Queries on database:-

- ① → Goto database (i.e., DBNew)
- ② → Right click → select new query
- ③ → SELECT \* FROM PARTY;
- ④ → click Execute (or) press "F5"
- ⑤ → Insert into party values (3, 'Madhu', 'Hyd', 5000)
- ⑥ → update party set party ID=20 where party ID=222.

## connecting to the other database:-

Two ways:-

- (i) By specifying <DatabaseName>.<Schema name>.<ObjectName>

Eg:- SELECT \* FROM VINAY\_DB.TEST

SELECT \* FROM SIVA\_DB.TEST

(no space required)

- (ii) Taking other database as current database

Syntax:- USE<DATABASE>;

USE SIVA\_DB ;

SELECT \* FROM <ObjectName>

SELECT \* FROM PARTY\_TEST;

Note:- Second way is recommandable if we are firing many queries on the same database (Eg:- SIVA\_DB)



# SSIS / SQL Server Integration Services

## ETL Operations

E → Extracting → Getting data.

T → Transform → performing Intermediate operations.

L → Load → Load to destination.

DB ↔ DB

DB ↔ File

File ↔ File.

## Administrative Tasks

→ Taking Backup of a database.

→ Shaving database

→ History cleanup

→ Transferring Database/Log/

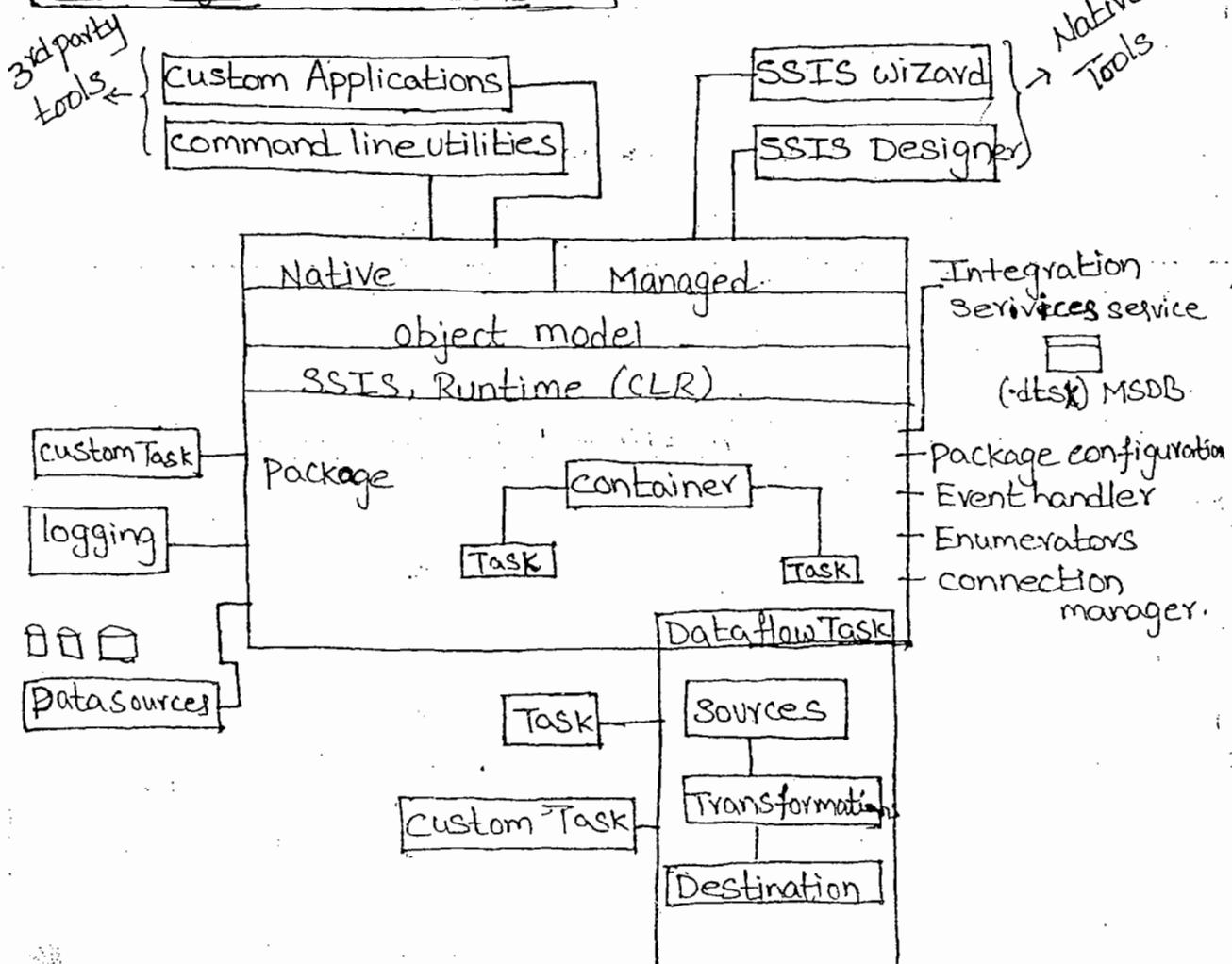
Error msg etc...

Databases: Oracle, SQL Server, Teradata etc...

Files: XML, Excel, Flatfile, Rawfile etc...



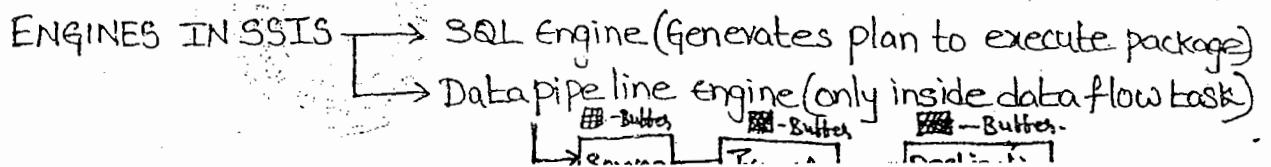
## SSIS logical Architecture



- ① There are 4 important components in SSIS Architecture. (13)
    - ② Object model
    - ③ SSIS Runtime
    - ④ Integration Services Service
    - ⑤ Dataflow Task.
  - ⑥ SSIS Designer- It's a native tool to create 'IS' packages and its components.
  - ⑦ Object model- It is an application programming interface which connects and understand custom tools & components.
  - ⑧ SSIS Runtime- This is 'CLR' which saves the layout of the packages (.dtx) runs the packages and manages the package components.
  - ⑨ Integration Services Service- This component helps us to store the packages in SQL Server database (MSDB), managing the packages and running the packages.
  - ⑩ Dataflow Task- To move the data between source to destination and perform different operations, dataflow task is required.  
It uses various "Inline buffers" while processing the data.  
It uses a "data pipeline engine" to move the data from source to destination & to manage buffers.
- \* Package and its Components-

Package is an important component in SSIS Architecture.

- It can be constructed through custom tools (or) native tools
- It performs all operations such as "ETL" & administrative task → It uses various other components as part of its processing. Eg: logging, event handling, package configuration etc.
- control flow task is mandatory for every package.
- To move the data from source to destination dataflow task is required.



## Differences between DTS & SSIS

### DTS

- SQL Server 7.0 introduced, available 2000 onwards
- Designed for ETS [Extract Transform Sources].
- It consists of single pane (i.e., pane means screen or frame) for all operations.  
It has data transformations  
Work flow etc.
- Data transformations available
- No DSV [Data Source View]  
No connection manager  
No Event handling  
No Looping through folders, files.
- Message boxes displayed in active-X Script
- Less Transformations.
- partial BI support (less)
- No Deployment Wizard
- Saved in  
① Enterprise manager  
(SQL server)
- ② File system  
(structured storage file)

### SSIS

- SQL Server 2005 onwards available.
- Designed for ETL [Extract Transform Load]
- It consists of multiple pane for multiple operations.  
It has control flow  
Data flow  
Package Explorer  
Event Handling.
- Data flow task introduced & transformation embedded.
- Available [Introduced]
- Message boxes displayed in Script Task.
- More Transformations.
- Full support to BI.
- Deployment wizards are there.
- Saved in local file system.  
Deployed to SQL Server.



## 2008 R2

- 2008 R2 is the second release of 2008.
- CODE NAME "KILIMANJARO"
- Released in middle of 2009.

### Supported features:



- Max 25 instances in CMS.
- Max 256 logical processors in CMS.
- Multi Server Administrator
- MDS [Master Data Services]
- "Data-Tier" applications
- PowerPivot for Virtualization
- Full support to
  - MS-Excel 2010.
  - MS visual studio 2010.
  - Full support to SharePoint server
  - Visualization etc.
- Data compression with UCS-2 code support.
- Available editions → Data centre edition
  - Huge data storage
  - Recommended for OLTP, ODS storage
- parallel DWH edition
  - Huge data storage (parallel)
  - Recommended for BI application



## 2008R2 to 2011:

- Its CODE NAME "DENALI"
- Multi sub-net failover clustering introduced.

### Programming Enhancements:

- Creating sequence introduced.

Syntax: Create Sequence <sequence name> START WITH

<value> INCREMENT BY <values>

(16)  
e.g:- Create sequence & start with 1 . increment by 1.

Insert int Test values (Next value for x, "vinay").

↳ empID  
↳ username

↳ Paging implemented in 2011:

→ It displays the required rows in page wise.

→ Full text search of Index introduced.

→ The usage of excel 'power pivot' enhanced so that reporting models are created easily.

→ Analysis service "BISM" (Business Intelligence semantic model) introduced. It is a 3-layer model.

→ WEB based visualization (Project cresend Introduced)  
It is a code name for representing applications for better visualizations.

### SSIS Practical Architecture:

Solution (collection of projects)

↳ project (collection of packages)

↳ package (A Discrete unit of work for doing ETL operations & Administrative tasks).

→ Control Flow (Tasks, Containers, planTasks, maintenance).

→ Data flow (Sources, Destinations, Transformations)

→ Package Explorer.

→ Event Handling.



### Navigation:

Start → programs → SQL Server 2008 → click BIDS → File  
→ New → project → select 'Integration Services' → project  
in templates window → Enter Name → click Ok.  
project name: Test\_Solution\_Project  
location: C:\TEST

→ Note: By default there is no solution is presented,  
So we can create a solution at the project creation time.  
→ click view menu → solution explorer (It describes the  
Projects, packages, data source views, data source information)

[  
    ↓  
    - It is used to connect to the DB  
    - It can be reusable "ACROSS packages"

**Data source view:** It is the logical object for the physical  
collection of tables (or) views in data sources.

**Connection Manager:**

For every connection we can take a name, the  
name can be reusable with in the program (or) package.  
→ In case of flat file: Folder & the File Name taken as connection  
string for connection manager.  
→ In case of relation, Source: Server name and database  
name taken as part of connection string in connection mana-  
ger.



**Various ways of Package Execution:**

- a) By pressing "F5"
- b) Solution Explorer → Packages → rightclick → Execute package.
- c) Debug Menu → Start debugging.
- d) SSMS → Integration Services → MSDB → Package → rightclick  
    → Run package.
- e) By using DTUTIL, DTUTIL Exec. facilities etc.

**Colors and their meanings:**

White → Ready to execute	Green → Success
Yellow → Running	Grey → disable
Red → Fail	

## Ensuring Package Success, Failure, Errors, Bottlenecks

We observe this information in "progress tab" (or) "Log provider".

(a) progress tab information: It describes how the package validated and executed step by step from starting to ending. Generally it displays

- (i) the No of rows operated
- (ii) Source and destination connections
- (iii) the amount of time taken b/w one statement & another statement etc.

(b) Log providers: DISCUSSED LATER IN THIS Book.

Variables: It is the value which is changeable within the package. There are two types of variables. They are:-

(i) System defined variables: These variables hold system information.

These variables store under SYSTEM<Namespace>

Syntax: SYSTEM::<Variable Name>

Eg: SYSTEM:: PACKAGE NAME

SYSTEM:: Execution ID.



User defined variables: These variables are created by the user only. These variables store under User Name space.

Syntax: USER::<Variable Name>

Eg: USER:: i

USER:: Name Var

USER:: Temp\_date etc.

Navigation: SSIS Menu → Variables → In the variables window click the top most left corner option to create New variable.

(19)

Name	Scope	Datatype	Value
i	Package	int32	1

Variable scope: The extent we use the variable is called the scope of variable. There are two scopes.

- (a) Package Level: Within the package, anywhere we can use the variable.
- (b) Task Level: Within the task only we can use the variable.

### Working with Data Flow Task

To move the data from source to destination and to perform intermediate operations, this task is mandatory.

Frequently used sources are flat file source and OLE DB sources and Frequently used destinations are flat file destination & OLE DB destination.

### Real Time Modes of Flat File Sources

<div style="border: 1px solid black; padding: 5px; display: inline-block;">123VINAY 456SIVAM</div> Fixed Width	<div style="border: 1px solid black; padding: 5px; display: inline-block;">123,VINAY 25,KISHOR</div> Delimited (Separated or) Variable Text
ID → 3 chars Name → 5 chars	ID → Max 10 char. Name → Max 30 char. Row Delimiter →   (Pipe) Column Delimiter → , (comma) Row Delimiter: CRLF (Enter char.) Column Delimiter → NO → [Carriage return line feed (or) forward]



Ex:- Moving data from File to File, (comma Delimited).

Sol:- (i) Take data flow task on control flow.

(ii) Go to data flow task → drag & drop flat file source & destination and do the below settings:

INPUT: party\_SRC → Notepad

File Name: party\_SRC

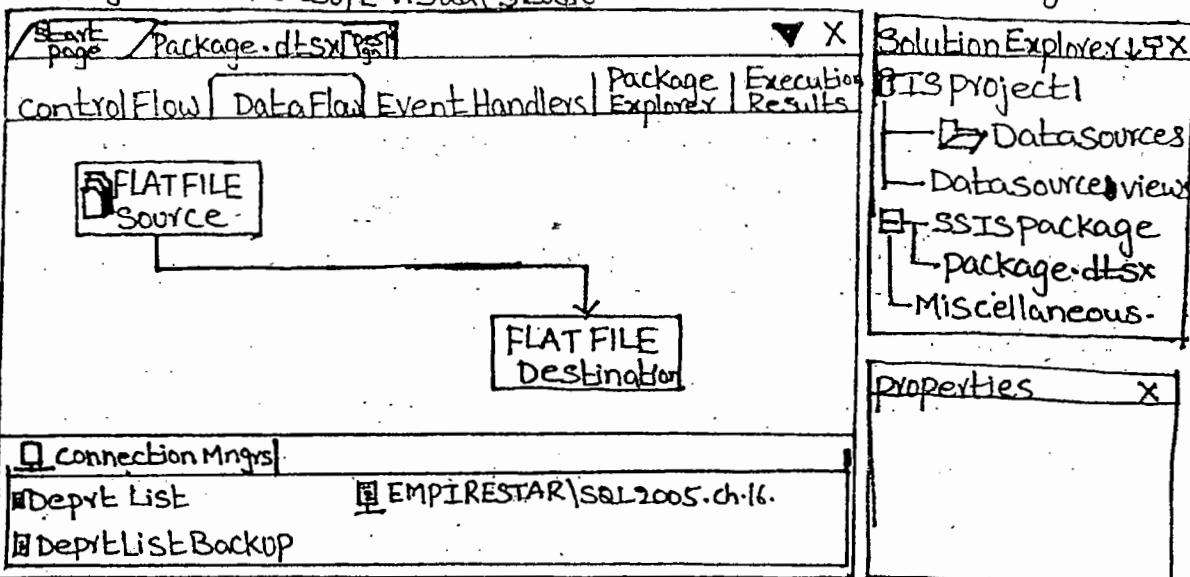
No of Rows: 13

FILE NAME: PARTY\_SRC.TXT NO OF ROWS: 2 CREATED BY: VINAY  
 Party ID, Party Name, PartyLoc, Party Income, party code

1 , SIVA , HYD , 30.000 , 30.  
 2 , MADHU , MUM , 40000 , 40

→ Connection Manager Name is reusable in the packages.

ISProject1 - Microsoft Visual Studio.



### FLAT FILE Source :-

↓  
 Select → rclick → Edit

Flat File connection Manager click

↪

connection Manager Name: SRC

File Name: Browse to the i/p folder party-SRC.Txt file.

Format: Delimited.

Header Row Delimited:  $\{CR\}-\{LF\}$

Header Row to Skip:

Column Names in the first data

Row: check it

click preview → click OK

↳ OK.

→

After assigning both Source and Destination Properties then

### FLAT FILE Destination :-

↓  
 Select → rclick → Edit

Flat File connection Manager →  
 click New → Select Delimited  
 → click OK.

connection Manager Name: TGT.

File Name: Browse to the o/p folder and specify party-output.txt

Format: Delimited.

Header Row Delimiter:  $\{CR\}-\{LF\}$

Column Names in the First data

Row: Check it.

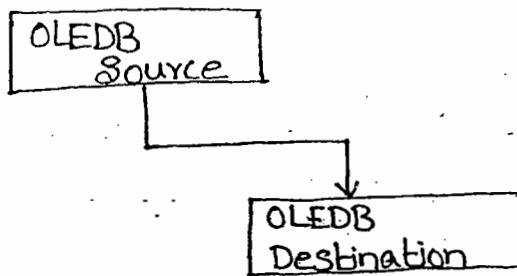
click preview → click OK → OK.

Go to mappings → check whether

Required source columns to

Required target columns → OK.

- ② OLEDB: Object Linking and Embedding DataBase. (21)
- Eg.: Moving data from table to table from source database DB-MSBI to MSBI-DB-destination database.
- Navigation: open BIDS → Take dataflow task <sup>in</sup> control flow



OLEDB → object Linking and embedding Database → universal provider to any database (or) application (Excel..etc)

OLEDB Source → Rt click → edit →

OLEDB connection Manager → New → New

Provider: Native OLEDB Native SQL Server client 10.0

Server name: Local Host

Select Authentication: Windows (or) SQL Server.

Select Database: DB-MSBI

Click Test connection → OK → OK → OK.

Name of the Table or view → Emp. → OK → OK.

OLEDB Destination → RC → edit

OLEDB connection Manager → New → New

Provider: Native OLEDB Native SQL Server client 10.0.

Server Name: Local Host

Select Authentication: Windows (or) SQL Server

Select Database: DB-VINAY

Click Test connection → OK → OK → OK.

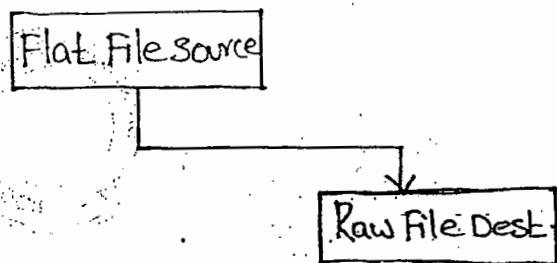
Name of the table or view → click New → change Table Name → OK.

Go Mappings → Connect required source columns to required.

Eg.: Moving data from one worksheet in Excel to another worksheet in another Excel.

Eg.: Moving data from flat file to Raw file.

Raw file contains binary information, so we are not able to read.



→ Raw File Destination → Edit → connection Manager → rclick  
Access Mode → File Name  
File Name → File path  
Write always → create always  
columns → Select columns → ok.

Eg.: Loading Data from XML file to a table in SQL Server database..

XML\_SRC.XML

<EMPS>

<student 1>

<EID> 001\_Vinay </EID>

<EName> Vinay </EName>

</student 1>

<student 1>

<EID> 002\_Siva </EID>

<EName> Siva </EName>

</student 1>

</EMPS>

XML Source

OLEDB destination



XML Source → rclick → Edit → Data access Mode → XML Location

OLEDB dest. → Generate XSD → ok.

Note: XML Schema definition must be specified (or) Generated to the corresponding XML file.

Transformations: These are intermediate operations performed between source and destination.

Eg: concatenation, Addition, Sorting, Merging etc.

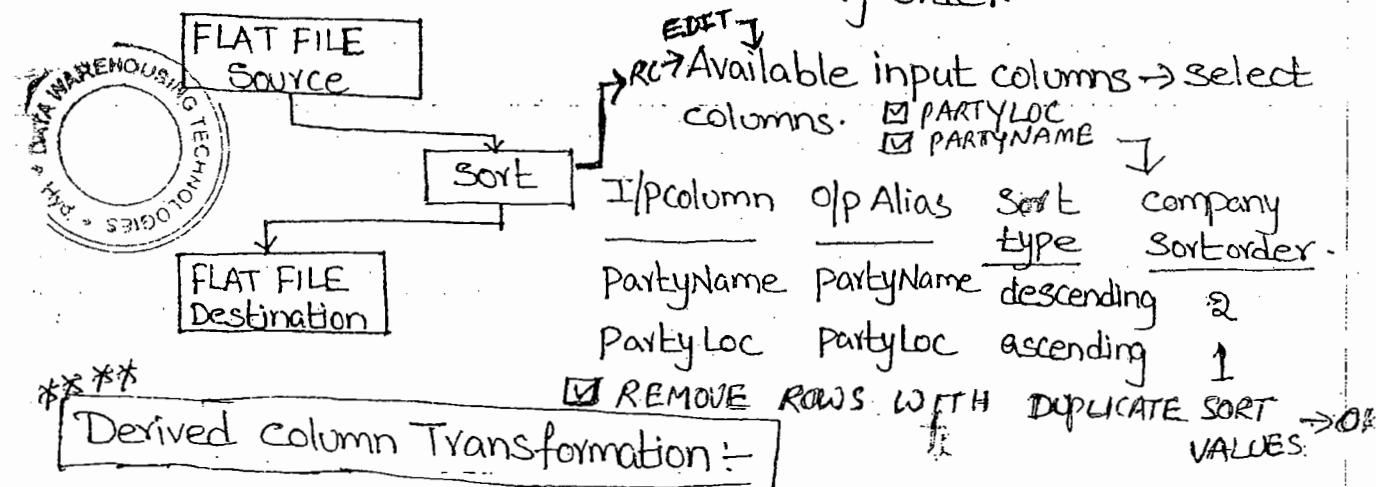
There are different dataflow transformations provided for different operations.

Sort Transformation: It sorts the data in the specified order (ascending or descending).

→ It has some flexibility to do sorting on multiple columns (by giving sort order).

→ There is a feature called remove rows with duplicates which helps us to display unique rows in sorting by eliminating duplicates.

Eg: Display the data by Locations in ascending order, within the locations names in descending order.



Derived Column Transformation:

It performs operations row by row: It does different calculations, aggregations, concatenations, transformations, conversions etc for the columns in the rows.

- Eg:   
① Display Name and Location by concatenation.  
② Display income, if it is Null, 99999.  
③ Display Income increment by 12%.

- ④ Display a new field with current date as business date  
 ⑤ Display default company code as 21000.



Derived column → rclick → Edit

### Derived Column Name

Address  
 PartyIncome-INC  
 Party Income  
 Business-DT  
 Source\_System\_code

### Derived Column

<add as new column>	(PartyName) + " - " + [PartyLoc]
<add as new column>	(DT_I4)[PartyIncome]*12/100;
Replace party income	[PartyIncome] == "999999" : [PartyIncome]
<add a new column>	GetDate()
<add a new column>	21000

### Expression

- Note :- (i) When we retrieve the data from FlatFile, all the columns belong to string datatype (DT-STR).  
 (ii) When we retrieve the data from Excel sheet, all the fields belongs to (DT-WSTR)

### Data Conversions :- Data conversions are done in three ways.

- ④ By using data conversion transformation.  
 ⑥ By using type cast operator in Expression.

Syntax :- <type cast operator> (column name)  
 (DT\_I4) (PartyIncome)  
 (DT\_DBDATE) ('2010-10-10')

- ⑤ Directly doing at flatfile source itself.

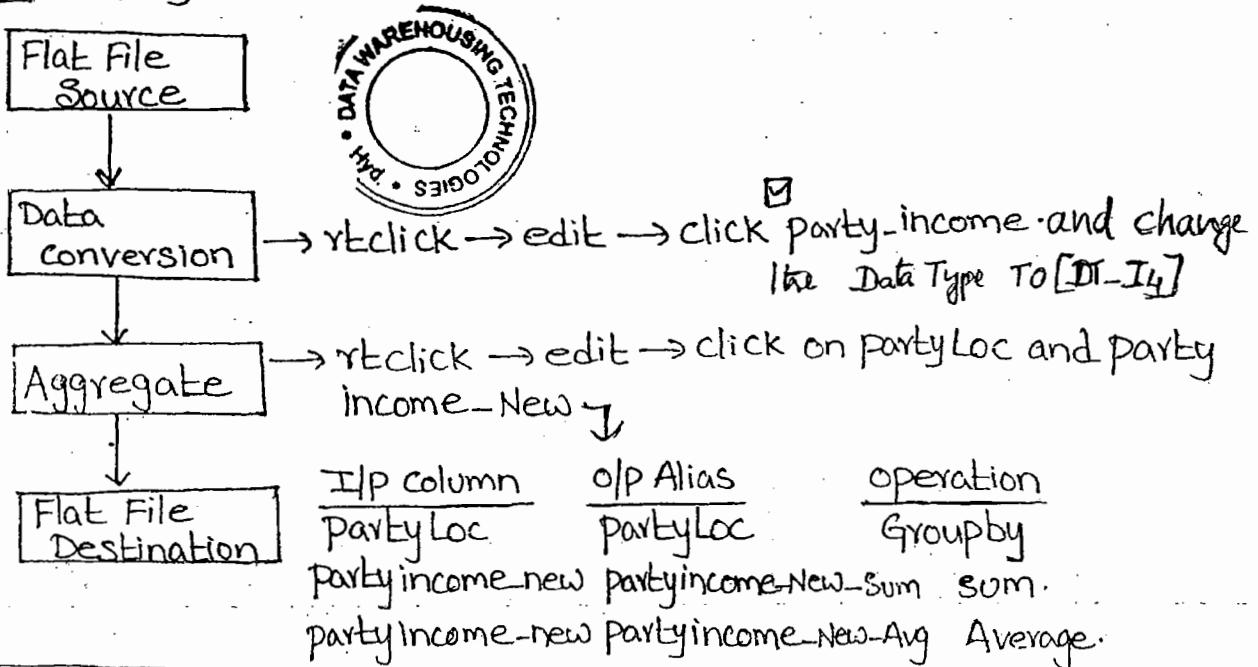
RClick on FlatFile source → Show advanced editor → Input and output properties → FlatFile source o/p → Output columns → PartyIncome → Data type properties → Data type → four-byte signed integer (DT\_I4)



## Aggregate Transformation :-

- IT performs aggregate operations such as
- Average, Sum, Count, Min, Max, Count Distinct, Group By etc
- If the field (or) Expression is of numeric data type we can perform the above all the operations.
- If the field is string (or) date, we perform limited operations like Group, count, Distinct count etc.

Eg:- Display Location wise Income sum and average



## Flat File Source additional properties:-

- Retain Null values from the source as null values.  
It helps us to treat Nulls coming from the source as Nulls only. (If we uncheck this option null treated as zero for integers, space for strings).

## Error output options:-

In case of error (or) truncated values coming from source we can use either of the below options.

- ① Ignore Failure → In case of error (or) truncations it ignores the failure.

3) Redirect Row → In case of error (or) truncation it redirect the row to the another destination. (26)

4) Fail component → In case of error (or) truncation it simply fails the component.

Navigations: FlatFile Source → rclick → edit → error op.

### Flat File Destination Files Format:-

1) Delimited → The columns are delimited by commas, except the last one which is delimited by the new line character.

2) Fixed width → The columns are defined by fixed widths.

3) Fixed width with row delimiters → The columns are defined by fixed widths. An extra column, delimited by the new line characters is added to define row delimiters.

4) Ragged Right → The columns are defined by fixed widths, except the last one which is delimited by the new line character.

### OLEDB SOURCE . PROPERTIES:-

OLEDB Source → rclick → Edit →

Data access mode:

- Table (or) View
- Table Name (or) View Name Variable
- SQL Command
- SQL Command from Variable.

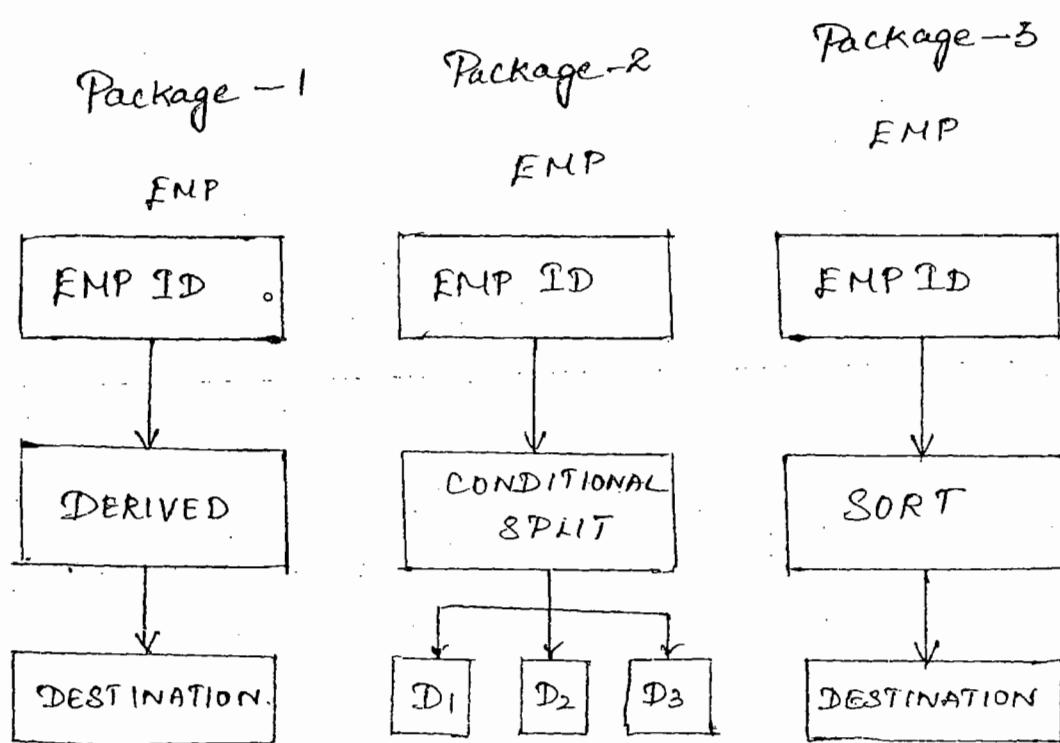
NOTE:- WE USE VARIABLES FOR DYNAMIC SOURCE RETRIEVAL.

## MULTICAST TRANSFORMATION

It creates multiple copies of same source.  
so that instead of doing multiple operations  
on the same source in multiple packages we  
can do in a single package by using Multicast.

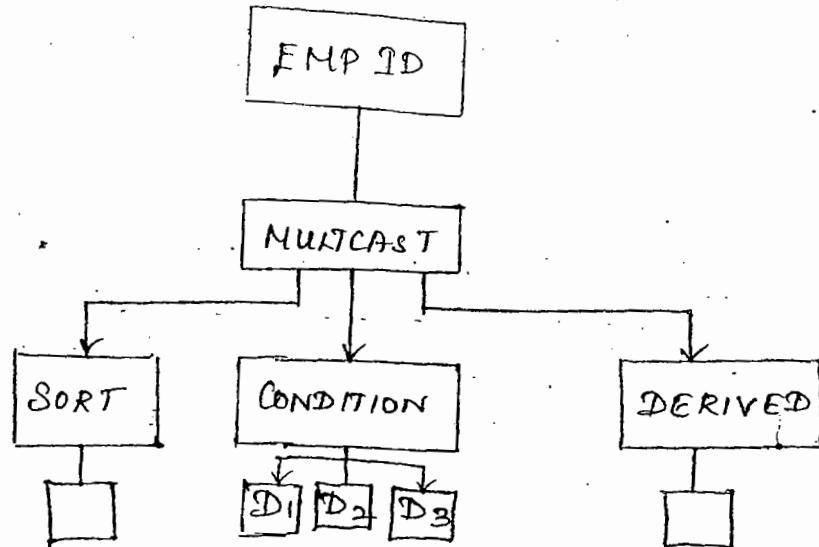
→ It improves the performance because it reads  
the data only one time from the source.

## DRAWBACKS WITH INDIVIDUAL PACKAGES.



- i. Three Buffers for Source
- ii. Multiple Reads on Source

## ADVANTAGES OF SINGLE PACKAGE SPLIT



- i) SOURCE READ ONLY ONE TIME
- ii) SINGLE TIME BUFFER OCCUPATION.

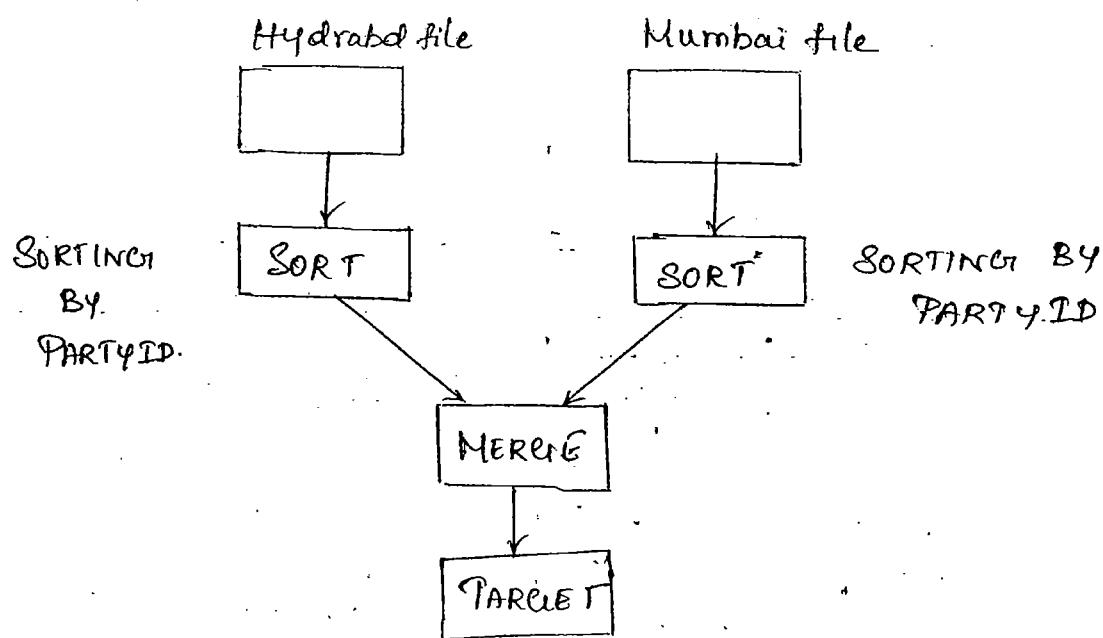
## MERGE TRANSFORMATION

It Merges Multiple Input data Sources.

Here the Restriction is, the Sources should be the Sorted Order so that the Output will also be in the Sorted Order.

There are two (2) ways to Implement.

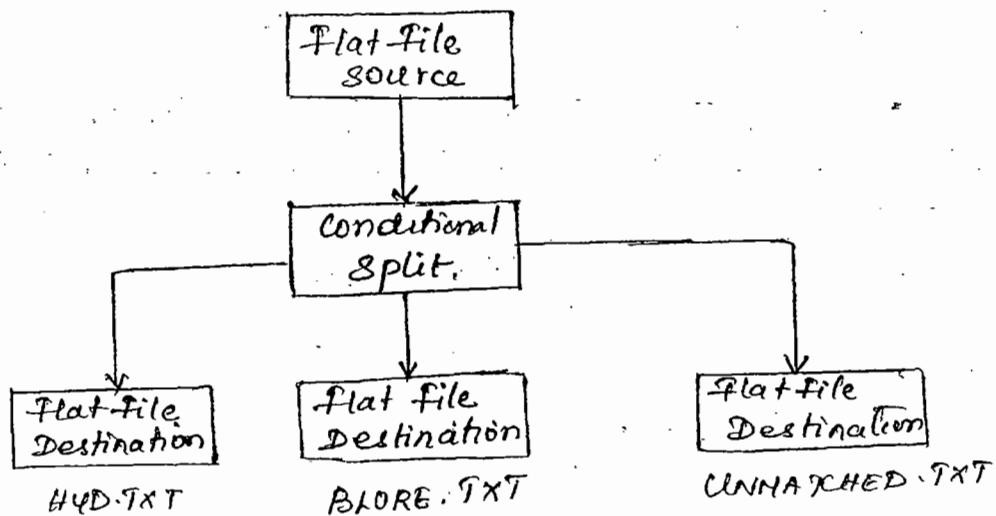
a. IF THE SOURCES ARE NOT IN SORTED ORDER  
DO THE BELOW PROCESS.



### CONDITIONAL SPLIT TRANSFORMATION

1. It splits the data based on the condition.
2. There are two types of output comes from this Transformation.
  - a. Conditions matched Output.
  - b. Conditions Unmatched Output (or) default Output.

Fig 1- Move Hyd , BANG , DATA To SEPERATE FILES AND UNMATCHED DATA TO ANOTHER FILE.



→ CONDITIONAL SPLIT → RT CLICK →

ORDER	OUTPUTNAME	CONDITION
1	HYD_DATA	[PARTYLOC] = "HYD"
2	BLORE_DATA	[PARTYLOC] = "BLORE"

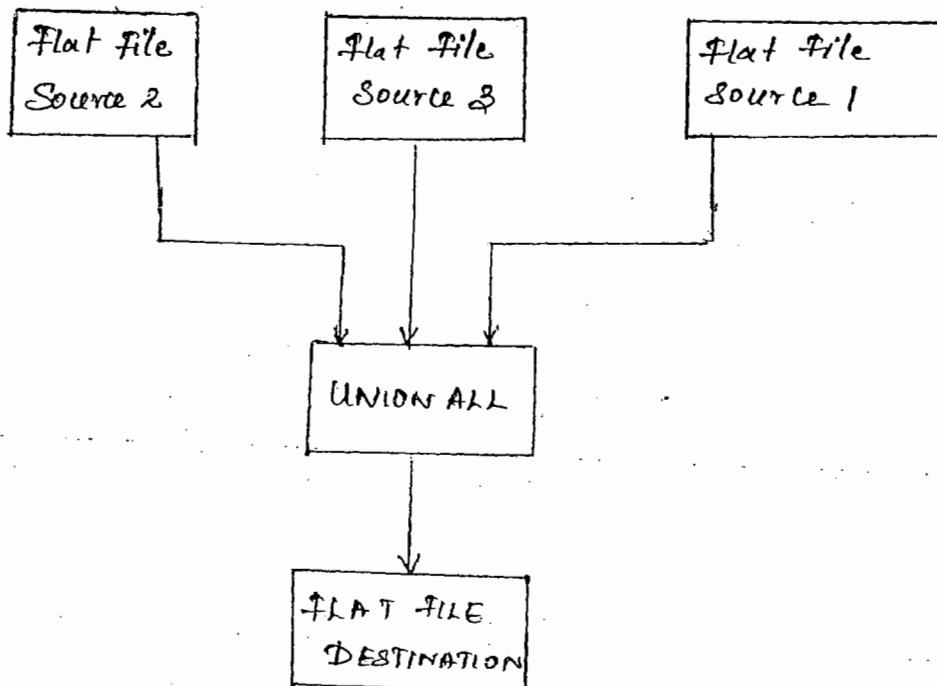
DEFAULT OUTPUT NAME

↪ UNMATCHED\_DATA



## UNION ALL TRANSFORMATION

- It merges multiple Input Sources (Two or More)
- No Need to take the Input in the Sorted Order, So that the output will also have Unsorted data.



### LIMITATIONS :-

1. Input Sources Structures Should be Same.  
[No. of columns, Order of Datatype of columns]

10911

## OLEDB Source Editor Properties:-

Data access mode:

- (i) Table or view → To retrieve the data from table (or) View
- (ii) tablename or view name variable → Table name (or) view name takes from the variable
- \*\*\* (iii) SQL command → we write a customized query to retrieve the data from the objects, so that required rows & columns retrieved and occupies less buffer every time.

Eg:- SELECT PartyID, partyName FROM Party where Party code in (20, 40, 60)

- \* (iv) SQL command from variable → we pass SQL command from a variable.

Note:- This variable generally recommended at the time of dynamic retrieval of data

## OLEDB destination additional Properties:-

Data access mode:-

- (i) Table or View

- \* (ii) Table or view - fast load → It loads the data very fast compared to normal view (or) Table loading.  
During the fast load there are couple of options we must select according to the situation.

- Keep Identity       Table Lock
- keep nulls       check constraints

Rows per batch 10000

Keep Identity uses identity column generated values.

Note:- This fast load option is useful when the table is having clustered index.



Q Table name (or) view name - fast load  
SQL command.

Q Question: In data accessing mode after selecting Table  
(or) view option and clicking new button to create a Table,  
then when does the table will be created.

Aq: After writing changes  
Immediately table created (but not at execution time).

### SSIS Expressions:-

Q Write expression when they are small because too many expressions and complex expressions decrease the performance

Q Generally we use expressions in various places.

① precedence constraints

② variables

③ for-loop

④ connection strings in the connection manager

⑤ Derived column Transformation.

⑥ conditional split



As part of expressions there are many functions, Typecasts, and operators available.

### Mathematical functions :- [MATHEMATICAL OPERATIONS ON NUMERICAL VALUES]

ABS, CEILING, EXP, FLOOR, LOG, LN, POWER, ROUND, SIGN, SQUARE, SQRT etc.

Eg:  $\text{ROUND}(4.82) \rightarrow 5$     $\text{ABS}(4.82) \rightarrow 5$     $\text{ceiling}(4.82) \rightarrow 5$   
 $\text{ROUND}(4.26) \rightarrow 4$     $\text{ABS}(-3.92) \rightarrow 4$     $\text{fLOOR}(4.82) \rightarrow 4$ .

### String functions:- [MANIPULATE'S STRING COLUMNS/ EXPRESSIONS.]

LENGTH, LOWER, LTRIM, REPLACE, CODEPOINT, FINDSTRING, HEX,  
REPLICATE, REVERSE, RIGHT, <sup>SUBSTRING</sup> etc.

Eg:- Lower ("ABC") → abc  
 LTRIM (" abc") → abc  
     ↳ it removes the space.  
 TRIM (" abc ") → abc  
 Reverse ("abc") → cba  
 find string ("VINNNAY","N",3) → 5 [ 3rd N position ]

Replicate ("a", 3) → aaa (29)  
 Replace ("VINAY", "Na", "NNa")  
     ↳ VINNAY.  
 Substring ("VINAY", 2, 2)  
     2 point string length 2 → IN.  
 find string ("VINAY", "N") → 3.

DATE/TIME functions :- [TO WORK WITH DAY, MONTH, YEAR IN DATE EXPRESSION]

DATE ADD, DATE DIFF, DATEPART, DAY, GETDATE, GETUTCDATE,  
MONTH, YEAR

UNIVERSAL TIME COORDINATE

DAY(DT-DATE) "2011-09-04" → 04

MONTH(DT-DATE) "2010-09-04" → 09.

DATEADD("MONTH", 4, (DT-DATE)) "2010-05-04" → 2010-09-04.

DATEADD("MONTH", 4, JDATE)

DATE DIFF ("MONTH", (DT-DATE)) "2011-07-09", (DT-DATE)) "2011-04-09"  
 "Day" → 03 monthly.  
 "Year" → 90 days.  
 → 0 years.

NULL FUNCTIONS :- [VALIDATE NULL ARGUMENTS.]

① ISNULL(Expression) → Result  
 → True/false.

② NULL(DT-DATE) → NULL [useful to display date value]

TYPECAST :- [CONVERTS ONE DATA TYPE TO OTHER DATA TYPE.]

→ DT\_I4 (column or Expression)

→ (DT\_STR, "length", "Codepage") (column or Expression)

→ (DT\_WSTR, "length") (col. or Exp.)  
 ↓ wide

→ (DT\_Numeric, precision, scale) ( )

Ex: (DT\_Numeric, 2, 6) (-123456) → 1234.56.



## Operators



○  $\&\&$  ----- Logical AND.

○  $\|$  ----- Logical OR

○ ?: ----- writing if condition.

○  $\langle \text{Expression} \rangle ? \langle \text{Success statement} \rangle : \langle \text{fail statements} \rangle$

Eg: ISNULL(PARTYNAME)? "Unknown": TRIM(PARTYNAME)

Display expression if date is null (or) the date length is zero (or) the date is having null date, display null otherwise display date.

ISNULL(JDATE) || Len(Trim((DT\_WSTR,10)JDATE)) == 0 ||

JDATE == "00-00-0000"? NULL(DT\_DATE): JDATE

List: ADD, concatenate<sup>(4)</sup>, Subtract, Negate, Multiply, Divide, modulo (%), parentheses ([]), equal (=), unequal (!=), Greater (>) ... etc.

Display an expression where the file name should be appeared with the current time<sup>stamp</sup> in this format.

FILENAME - YY YY MM - DD - HHMMSS.TXT

CDATE  $\rightarrow$  "2011-08-09 04:23:22:00000"

" FILENAME - "+

substring((DT\_WSTR,30) CDATE, 1,10) + " - " +

substring((DT\_WSTR,30) CDATE, 12,2) +

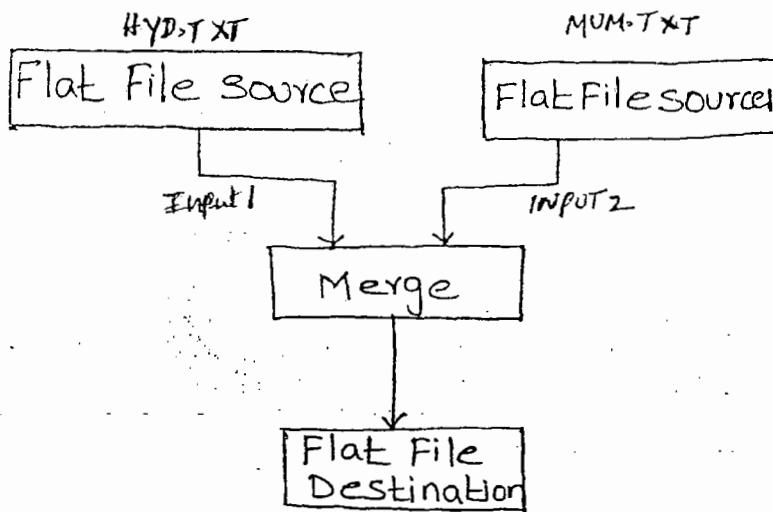
substring((DT\_WSTR,30) CDATE, 15,2) +

substring((DT\_WSTR,30) CDATE, 18,2) + ".TXT"



Note: Fast Load option is useful when the table is having clustered Index.

## (b) In case sorted input sources



0 → UNSORTED  
1 → SORTED

- Take 2 Flat Files and assign locations.
- Go to each flat file → RT click → show advanced Editor
- Input and output properties → Flat File Source output
  - IS SORTED: TRUE
  - Output columns → party ID → Datatypes: change DT\_I4
  - Sort key position: 1

Flat File Source Error output → IS SORTED: TRUE  
 Output columns → Flatfile source error output column:  
 Sort Key Position: 1 OK → OK

Note: Merge works with two sources at a time.

⇒ Differences between Merge & Union ALL

### Merge

only two sources

Input should be sorted

Sorted result

### Union all

No Limit

Not Applicable

Not Applicable



Merge JOIN - It performs merge operation along with joins. Generally it supports the below joins.

- (a) Inner Join
- (b) Left Join
- (c) Full Join

Emp Table

EID	ENAME	DID
1	A	10
2	B	20

Dept Table

DID	DNAM	NULL
10	IT	?
40	HR	?



Query:- Select cols/\* FROM <Table A> CrossJoin <Table B> No  
structure where condition

Inner Join <Table B> ON <condition>

Left outer Join <Table B> ON <condition>

Right outer Join <Table B> ON <condition>

Full outer Join <Table B> ON <condition>.

Ex:- SELECT E.EID, E.ENAME, E.DID, D.DID, D.DNAME FROM  
EMPE CROSS Join Dept D

Inner Join Dept D ON E.DID = D.DID

Left outer Join Dept D ON E.DID = D.DID

Right outer Join Dept D ON E.DID = D.DID

Full outer Join Dept D ON E.DID = D.DID.

Output:-



(34B)

E.EID; E.ENAME, E.DID, D.DID, D.DNAME

CROSS JOIN

[Cross product of two tables]

Table A → M rows  
n B → N rows  
CROSS JOIN: MXN rows

1	A	10	10	IT
1	A	10	40	HR
2	B	20	10	IT
2	B	20	40	HR

Inner Join

[Condition Matched data]

1	A	10	10	IT
2	B	20	10	IT

Left Join

1. CONDITION MATCHED  
2. UNMATCHED FROM LEFT

1	A	10	10	IT
2	B	20	?	?

RIGHT JOIN

1. CONDITION MATCHED  
2. UNMATCHED FROM RIGHT

1	A	10	10	IT
?	?	?	40	HR

FULL JOIN

1. CONDITION MATCHED  
2. UNMATCHED FROM LEFT and RIGHT

TABLES

1	A	10	10	IT
2	B	20	?	?
?	?	?	40	HR



O  
O  
C  
C  
C  
C  
C  
C  
C  
C  
C  
C  
C



## Cache creation:-

OLEDB SOURCE [DEPT TABLE]



Cache Transform → RC → EDIT → Cache Connection Manager New →  
Connection mgr name: Deptcache

use file Cache

[Browse]

OUTPUT FOLDER



DBP-Cache.CAW →

indexposition

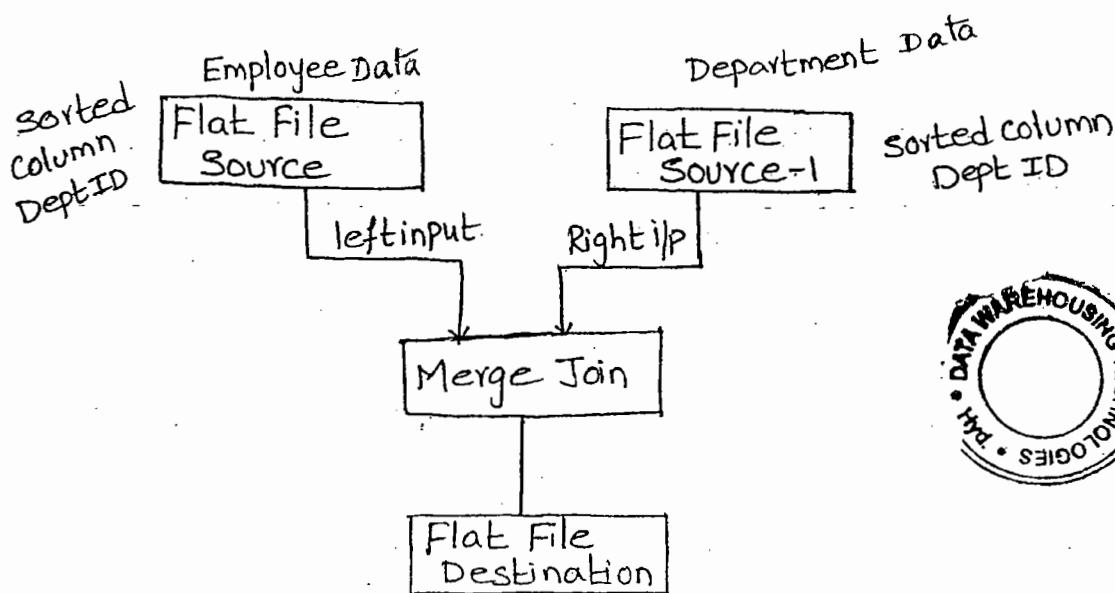
|

columns → DEPTID

OK → OK → Run

PACKAGE & See Cache  
generated





→ Merge Join → right click → Merge Join Transformation Editor →  
Join type: Left outer Join

Two tables are shown side-by-side. The left table has columns 'Name', 'Order', and 'Joinkey'. It contains three rows: (EID, 0, 0), (ENAM, 0, 0), and (DID, 1, 1). The right table also has columns 'Name', 'Order', and 'Joinkey'. It contains two rows: (DID, 1, 1) and (DNAM, 0, 0).

### Look Up Transformation

IT Looks up <sup>the</sup> required values on target and fetches relevant result. [EXACT RESULT]

#### Real Time Usage

- ① To fetch relevant value
- ② while working <sup>WITH</sup> ~~SCDs~~ [SLOWLY CHANGING DIMENSIONS] to
- ③ To have exact match with destination and improve query retrieval fast (IT uses caches)

## Types of caches:-

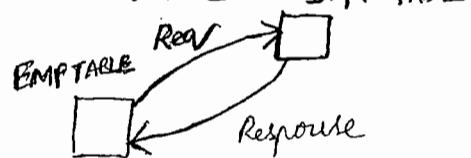
- (a) Full cache
- (b) No cache
- (c) Partial cache.



No Cache: Here, no cache to the target table. So every time source query hits the database and fetches the result.

### Advantage:

- If the source data changing frequently and less no of records are there it is recommended.



### Drawbacks:

- 1. Hits on the target increases and traffic also high.

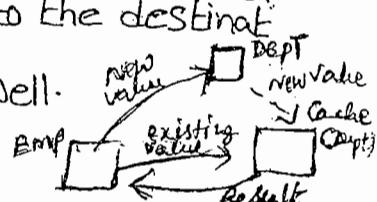
Full Cache: Here there is a cache for the target table, so every source request goes to the cache and fetches the data.



Adv: 1. If the target is not changed and having more records.

Partial Cache: Initially the cache is empty, for every new request source query hits the database and fetches the information to the cache. For existing record, source query hits the cache.

Adv: 1. More & more new records are added to the destination there is a huge usage in existing records as well.



## Look up result:-

If there is no match in the look up we can go for either of the below (ways) options.

- (a) Ignore failure
- (b) Redirect rows to error output
- (c) Fail component
- (d) Redirect rows to no match o/p.

If the source is having multiple matches in the destination it returns "First Match"

## Look up performance improvement:-

- (a) Increase (or) decrease the cache memory according to the target table size, because more rows with the big size cache gives bad performance.
- (b) Instead of taking table, take an SQL query to have required no of rows and columns in the cache and to perform look up operations.

Look up → right click → Edit → connection → mark use results of an SQL query option. and write this type of customized query.  
"SELECT DeptId, DeptName FROM Dept WHERE DEPTID IN (10, 20)"

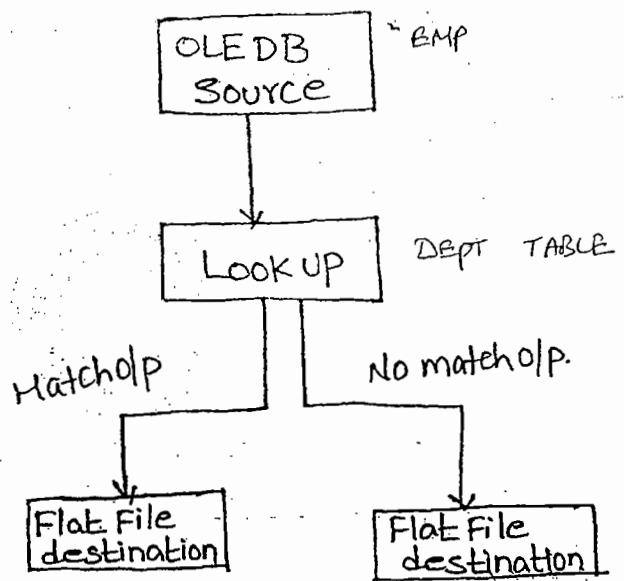
Eg:- Retrieving the deptid, name from Dept table based on the match from Emp table. (USING FULL CACHE)

i) Matched records in one destination & unmatched records to another destination.

Sol:- 1. OLEDB Source (Emp Table)

2. Look up → RC → Edit →  
General  
Select full Cache





specify how to handle rows with no matching entries  
i.e., Redirect rows to no match output

connection

OLEDB connection manager : DB-MSBI

USE Table or View : Dept.

columns

connect DID from EMP TO DEPT and Select DID,  
DName from Dept.

OK

OK.

- Take two destinations and connect matched result to one destination and unmatched result to another destination.

Note: Unmatched result destination structure is like Source table only and it contains source unmatched records.

- Lookup Operations we can perform only on tables (we can't perform on flatfiles) relational

## working with Partial CACHE :-

39

In the above steps do the below changes

### Navigation:-

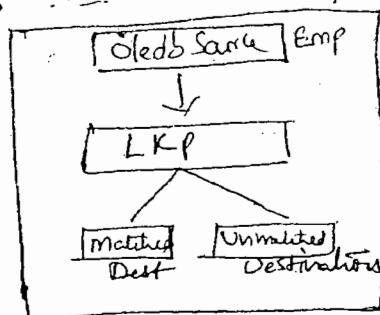
Lookup → RC → Edit →

General

Select partial cache.

Advanced

Take required amount of memory



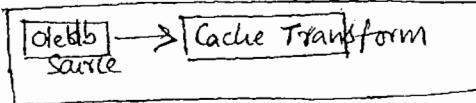
If required Enable the cache for no matching entries option.

Note: The cache created and drop automatically



## Creating the Named Cache:-

1. This cache is shareable across multiple packages.
2. This is recommended if the lookup table data is not changed such a long time.



### Navigation:-

start → programs → open BIDS → New → project → package → Control Flow Task → Data flow Task → Drag (or) select the OLEDB source → Cache transform → step (Refer page no 34)

## Using the precache (or) named cache for operations:-

To do the operations, some changes in the "LOOKUP" named cache take

i.e., Navigation:- Look up → rclick → Edit → Look up Editor

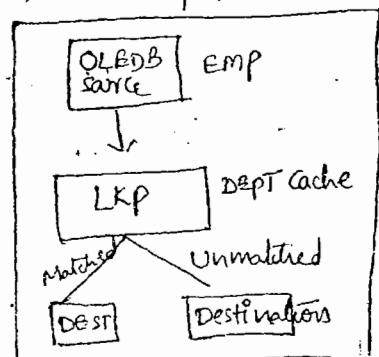
Properties → General

Cache mode:

Select Full cache

connection Type:

CACHE connection Manager



- ① Open up how to handle rows with no matching entries
- ② Redirect rows to no match output
- ③ Goto connection:
- ④ Cache connection manager → New → Name →
- ⑤ check use cache file name
  - OK
  - OK.
- ⑥ check columns, OK, OK..



### Fuzzy look up transformation

This transformation is designed to get the result from the destination based on the similarity but not exact match.

While doing the operations, three more columns for destination are added.

similarity: It displays how much similar source row with destination row.(column).

### Similarity\_Column name

It displays how much similar each source column with each destination column.

### Confidence →

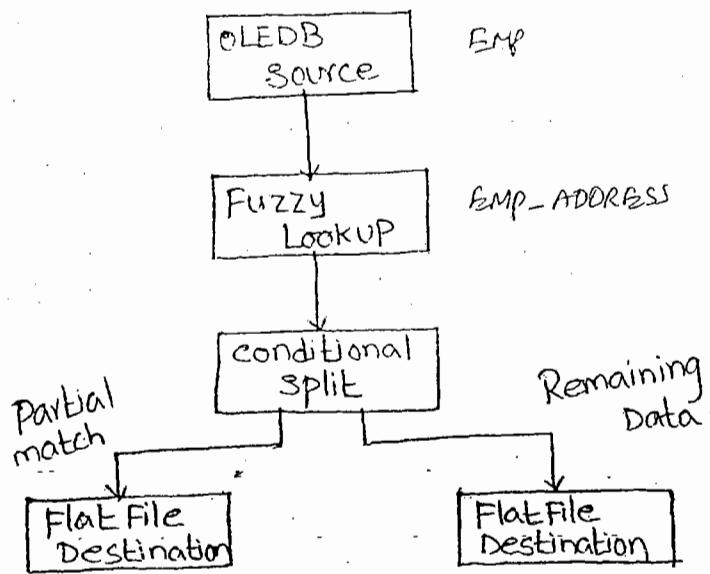
How much the system is having confidence to give the result.

Note: Generally we go for similarities for string values.

S.R.	EMP TABLE		
	EMPID	EMPNAME	EMP Address
1		Madhu	Chennai
2		Krishna	Hyderabad
3		Rajesh	Banglore
4		Hohan	Mumbai
5		Mohan	Kolkata

EMP-ADDRESS TABLE	
EID	ENAME
1	Madhu Mohan
2	Vinay Krishna
3	Rajesh Yadav
4	Ramana
5	Murali Mohan
6	

(41)



### Navigation:-

#### Fuzzy Lookup:

1. Take OLEDB Source as emp.
2. Fuzzy look up → Rclick → Edit →

#### Reference Table

OLEDB connection manager : DB-MSBI

Table or view : Emp\_address

columns:

Connect Ename From EMP to ENAME in EMP\_address  
and Select ENAME and E\_address columns From EMP\_address  
OK OK.

3. Take conditional split → Rclick → edit

<u>Order</u>	<u>Outputname</u>	<u>Condition</u>
1.	Partial-match	- similarity < 0.7
2.	Good-match	- similarity > 0.7

4. Take two destinations and connect each condition.

5. Run PACKAGE.

## PIVOT Transformation



It converts rows information into columns.

Creates less normalization of data.

Q: PIVOT i/p:

Party ID, SAL Component, Salamount

1	HRA	20,000
1	DA	2,00,000
1	TA	100,000
2	HRA	30,000
2	DA	300,000
2	TA	150,000
3	HRA	40,000
3	DA	400,000
3	TA	200,000

PIVOT O/P:

PRTYID HRA DA TA

1	20000	200000	100000
2	30000	300000	150000
3	40000	400000	200000



PIVOTED Column:

It doesn't participate in pivoting.

PIVOTED Column: column values which are converted from rows to columns.

PIVOTED Value: The values which are moved to pivoted columns.

column name      PIVOTED KEY USAGE

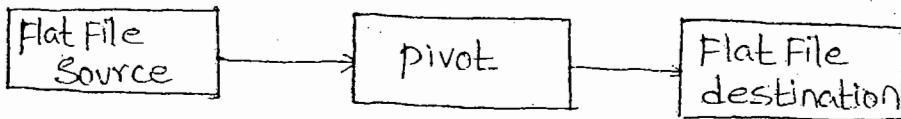
Party ID                  1 → Unpivot

SAL component            2 → Pivoted columns

SAL amount                3 → Pivoted values.

LINEAGE ID: It is a unique ID taken by a system for every column which is mapped.

(43)



### Navigation:

1. Take PIVOT-SRC as source

2. PIVOT → rclick → edit →

Input columns : select all columns

Input output properties:

PIVOT Default Input :

Do the below settings

Party ID → PIVOT usage → 1

SAL component → PIVOT usage → 2

SAL Amount → PIVOT usage → 3

Note: identify lineage ids of party ID, salcomponent, sal amount

Eg: column name lineageid  
Partyid 349  
Sal component 371  
Sal amount 375

PIVOT Default Opt

→ click add column → Rename column to Party ID → Goto properties and set PIVOT key value: Party ID, Source column: 349

→ click add column → Rename column to HRA → Goto properties and set PIVOT key value: HRA, Source column: 375

→ click add column → Rename column to TA → Goto properties and set PIVOT key value: TA, Source column: 375

→ click add column → Rename column to DA → Goto properties and set PIVOT key value: DA, Source column: 375

Audit: It displays audit information for every coming from source.

(o) It adds audit information to the source data.

Eg:- Audit Types- Execution instance GUID

Package ID

Package name

Version ID

Execution start time

m/c name

User name

Task name

Task ID



Character Map: It applies string operations such as Lower

To upper and viceversa etc.

Copy column: It creates multiple copies of the column.

Export column: It exports column value from rows in data set to a file.

Eg:- Exporting Images from column to a file.

into

Import column: It imports column values from a file.

Eg:- Loading images from file to table rows.

Fuzzy Grouping: It groups similar rows.

LEDB command: It executes an SQL command for each row in a data set.

Percentage Sampling: It takes sample percentage no of rows from source data set.



(45)

Eg:- Taking 20% of Sample rows From a dataset.

Row Sampling: It displays the specified no of Sampled rows. Eg: Display 10,000 sample rows from a table.

Row count: It counts the no of rows in a dataset.

Unpivot: It converts columns information into rows i.e, It creates more a normalized representation of dataset.

Script component: It executes a custom script (VB.NET/C# .Net)

(a) C# .Net

Term Extraction: It extracts terms from a data to column.

Term Look up: It counts the frequencies that terms in a reference table appears in a dataset.

Control Flow Items: It contains containers & Tasks.

Container: It contains other tasks

Eg: For each loop container

for loop container

Sequence container,

as well as.

Tasks: There are normal (and) maintenance tasks.

For loop container: It executes the underlying task to the specified no of times. [iterative Count we know here]

It has 3 sections.

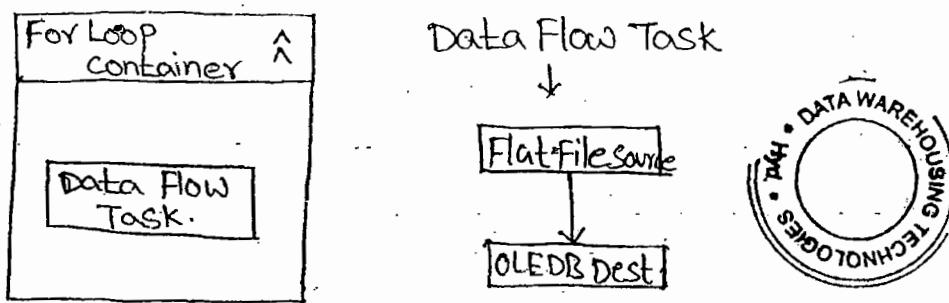
② Initial Section:

The variables are initialized in the container.

6 Assignment section:- The variable incremented or decremented.

7 Evaluating section:- Here the condition verified.

For loop requires a "variable" to do the operations.



Create Variable:

SSIS menu → variables → Add variable →

Name: Counter

Data Type: Integer

Take for loop container on control flow → RC → Edit.

INITIAL Expression: @Counter = 0

Eval Expression: @Counter < 5

Assign Expression: @Counter = @Counter + 1

Take dataflow

Take FlatFile source (To be copied) and OLEDB destination connect it.

Note! The above Eg is for loading a file data 3 times to table

For Each Loop container:

→ It is designed to load group of similar objects or working with similar objects whose count is "unknown".

For loop is having a condition, so that we know the count; whereas for each loop we don't know the count.

e.g. Loading set of records to dataset or similar dataset  
records one by one

- Loading similar files from a folder to a Table etc.
  - It uses enumerator for its operation.
  - The enumerator that supports are:

For each File enumerator

- " " item " "
  - " " ADO " "
  - " " ADO .Net Schema Rowset enumerator.
  - " " From variable enumerator.
  - " " Nodel ist " "
  - " " SMO " "

→ Enumerator values are "not changed" within a package (variable values are changed).

**Eg:** Loading the available files in the same structure from a folder to a table.

## Navigation:

1. Take Foreach loop container → right click → edit collection:

Enumerator: For each file enumerator.

Folder: C:\output\Group

File : ".TXT"

Retrieve file Name: Select fully qualified  
Select Transverse Folders

## Variable mappings:

Variable Drop Down → select New Variable –

Name: Group var → ok.

2. Take Data Flow Task in the for each loop container.



### 3. Data Flow Task:

- o Take Flat File Source → specify one file in the group.
- o Go to source connection manager → Rclick → properties.
- o → Expression → click Ellipse →
- o Property Drop down list: select connection string.
- o Expression: click Ellipse.
- o Variables: Drag & drop group var to the expression section.
- o OK → OK

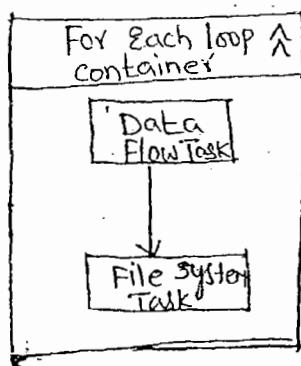
Take OLEDB destination and connect.



File System Task: It performs file and folder options.

such as copying, moving, deleting, creating ... etc.

- Ex: 1. Moving the files from folder 'x' to folder 'y'
- 2. In above for each loop container example load successive files into another folder [i.e., which are not loaded that can be easily tracked].



### Navigation:

File System Task → Rclick → Edit  
Destination:

IS destination path variable: False

Destination connection: Browse to Success Folder.

Overwrite Destination: False.

Operation: move file.

Source:

IS Source path variable: False.

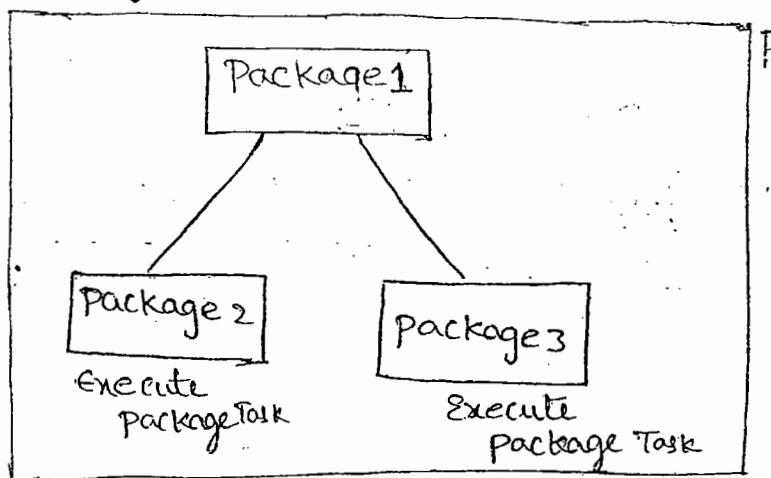
Source connection: Browse to Source connection manager (which was taken at DataFlow Task to source level)

## Execute Package Task:-

(49)

IT executes the packages which are available in "file system" and "SQL Server data base".

This is designed to execute another package with in main package we can control the flow b/w these packages.



1. Take Execute package Task on control flow

2. R-click → Edit →

Location: File system

connection: specify Desktop \ any folder package.

Password: If password is there, specify

Execute out of process: False.



**NOTE:-** Execute out of process true means, each package runs with a separate process.

⇒ **Execute SQL Task**: IT executes SQL of any database

(oracle; Teradata, Excel, SQL Server---etc).

To connect to the corresponding database we must specify the corresponding type.

IT executes queries, commands in the corresponding data base

**Note:-** the execute out of process option

\*If we turn into True the subpackages runs separately from

the main package process.

① Take Execute SQL Task on control flow.

② RT click → Edit

Connection Type : OLEDB

Connection : Local Host : DB - MSBI

SQL statement : Delete from party,

Bypass prepare : True



If select "False" instead of "True", in this situation SQL converts into the another Query and it runs every time on the target Database.

### Working with Procedure

```
Create procedure [dbo].[Samp] @PID Integer, @PName  
VARCHAR(30) AS BEGIN Insert into For each - Tab(PartyID,  
Party Name)  
values (@PID, @PName);
```

END

EXEC SAMP 10, KKK

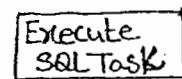
Select \* From foreach - Tab;

Drop procedure Samp;



### Executing the procedure from Executing SQL Task

→ SQL statement : EXEC SAMP 20, triven



Script Tasks It supports the scripting of VB .Net (or)

C# .Net

RT click → Edit

Script language : Microsoft Visual Basic 2008

click Edit Script → Add the below statement in the main()

MsgBox ("MSBI class")

Save → OK

- Real Time:**
1. To reuse the existing code of OLTP Systems.
  2. To write custom coding

Eg: Loading Multiple worksheets data in a single excel sheet to a table

### Bulk Insert Task:

- It loads bulk data with max speed into the tables.
- It cannot perform any intermediate operations.
- Before loading <sup>into</sup> the table, <sup>table should already be created.</sup>
- It loads <sup>... files only.</sup> [Direct file data to already created table]

### Bulk Insert Task Navigation:

Rt click → Edit →

Destination connection:

Connection: Local Host: DB-MSBI

Table: Party

Format:

Row Delimiter: {CR}-{LF}

column Delimiter: , (comma)

Source connection:

File: Browse the file.

### Backup Database Task:

It takes backup of SQL server databases.

Backup → Rt click → Edit → Backup Database Task →

Name: SRC

Backup: Full

Databases: DB-MSBI

Backup to: ODisc OTape

Create a backup file for every database → click OK!

## Send Mail Task

- This is designed to send e-mails to corresponding recipients.
- It requires an SMTP Server (SIMPLE MAIL TRANSFER PROTOCOL)

### Navigation:

Send mail Task → RC → Edit → Mail → SMTP connection →

New → Specify Name → SMTP server (IP address) → click OK

From (which user) → To (which user) → Subject (Job successfully )

Message Source Type: Direct Input → priority: High → click OK.

Send mail  
Task

→ RC → edit

## Active Script Task

It passes and executes active scripts.



Analysis Services Execute DDL Task – It executes DDL operations of analysis services.

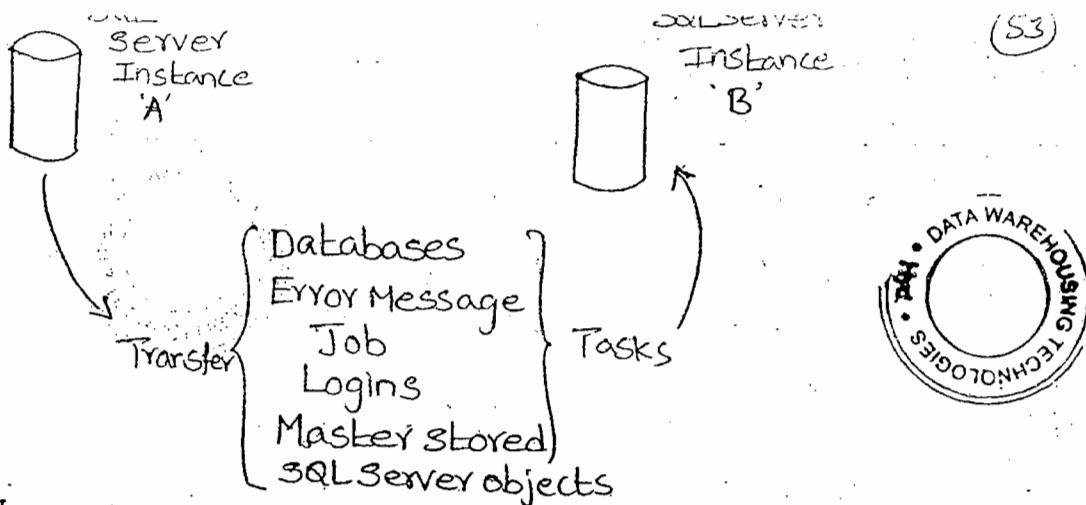
Analysis Services Processing Task – This is used to process the data of facts cube, dimensions etc. In analysis Services we use this task.

Execute DTS 2000 package Task – It executes SQL Server 2000 DTS package.

Execute Process Task – It executes Win 32 Executable tasks.

FTP Task – It performs file operations such as sending, receiving files.

TRANSFER DATABASES / ERROR MESSAGES / JOBS / LOGINS / MASTER STORED PROCEDURES
SQL SERVER OBJECT TASK.



→ The above Tasks transfers the specified objects from "one SQL Server instance" to "other SQL Server instance".

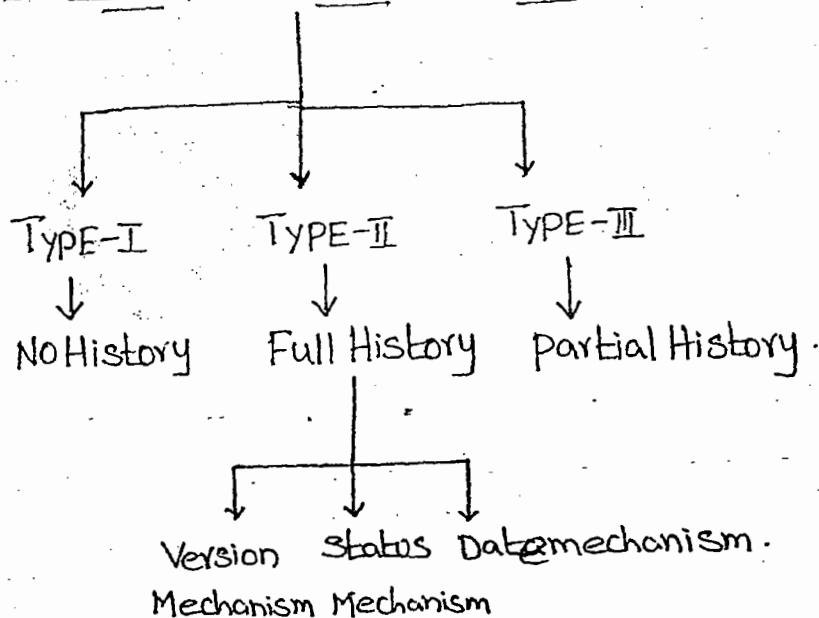
### Slowly Changing Dimensions : SCDs

→ To process the data from granularity <sup>Tables</sup> to main tables, we follow a mechanism called Slowly changing dimensions type.

Eg: There is a customer table where it holds customer details. If there is any change in SCD's there should be a manipulation in the process.



## Slowly Changing Dimensions [SCDs]:



Type-I → (i) New Record: Inserts  
(ii) Old Records: Update/Modifies

EMP-Daily			EMP-History			EMP-History		
EID	ENAME	ELOC	EID	ENAME	ELOC	EID	ENAME	ELOC
3	Krishna	USA				1	Vinay	HYD
1	Vinay	USA				2	Madhu	HYD
								USA

Type-I

Here in EMP-History after Type-I performance is done  
the location of Vinay is changed from the Location HYD  
→ USA (i.e., replaced).

One customer only one location.

## Type-2 (Version Mechanism):

- (i) New records inserted with Version '0' (zero)
- (ii) Old records are inserted with increment VERSION  
(i.e., either 1 or any).



EID	ENAME	ELOC
3	Kristna	USA
1	Vinay	USA

EMP\_Daily

Ver	ENAM	EID	ELOC
0	VINAY	1	HYD
0	MADHU	2	MUM

EMP\_History

Type-2

Ver	EID	ENAME	ELOC
0	1	Vinay	HYD
0	2	Madhu	MUM
0	3	Kristna	USA
1	1	Vinay	USA

EMP\_History

Note: Highest Version indicates the current location of the customer.

### Type-2 (status Mechanism):

(i) New record inserted with status = "Current"

(ii) Old record @ inserted with status = "Current"

(B) Modifies earlier record status to "expired"

EID	EN	ELoc
3	Kristna	USA
1	Vinay	USA

EMP\_Daily

EID	ENAM	ELOC	Status
1	VINAY	HYD	Current
2	MADHU	MUM	Current

EMP\_History

Type-2

EID	ENAM	ELOC	Status
1	VINAY	HYD	Current
2	MADHU	MUM	Current
3	Kristna	USA	Current
1	VINAY	USA	Current

expired

EMP\_History

Note: Status = 'Current' means the customer is in that location.

### Type-2 (Date Mechanism):

(i) New record inserted with start-date and End-date as '9999-12-31'

(ii) Old record (a) inserted with startdate and End-date as '9999-12-31'

(b) Modifies old record End-dt to new record Start-date.

Note: Current Location of customer identified with End-date = '9999-12-31'

EID	ENAM	ELOC	Start date
3	Krishna	USA	17/09/11
1	Vinay	USA	17/09/11

EID	EN	ELOC	Start Date	End Date
1	Vinay	HYD	16/09/11	9999-12-31
2	Madhu	MUM	16/09/11	9999-12-31

ID	N	EL	Start Date	End Date
1	Vinay	HYD	16/09/11	9999-12-31
2	Madhu	MUM	16/09/11	9999-12-31
3	Krishna	USA	17/09/11	9999-12-31
1	Vinay	USA	17/09/11	9999-12-31

EMP\_Daily

EMP\_History

EMP\_History

Type-3:- For every customer previous/current locations maintained.

EID	ENAM	ELOC
3	Krishna	USA
1	VINAY	USA

EID	ENAM	Current location	Precious location
1	Vinay	Hyd	
2	Madhu	MUM	

ID	EN	Current location
1	Vinay	USA
2	Madhu	MUM
3	Krishna	USA



Note:- If you want to increase the number of history rows those many columns should be added to the history tables, which is burden to the system.

In real time we use frequently Type-2 (date mechanism) is history maintenance mechanism. [See st page no for practicals]

Maintainance cleanup task:- It removes files left over from maintainance plan.

Notify operator task:- It sends an e-mail msg to any SQL Server Agent operator.

Built Index task:- It arranges (or) reorganises data in the data index pages by rebuilding the indexes. This improves the performance of index scans and seeks.

Organize Index task:- It defragments and compacts clustered and Non-clustered indexes on tables and views.

Shrink Database task:- It reduces the disk space consumed by database and logfiles by removing empty data & log pages.

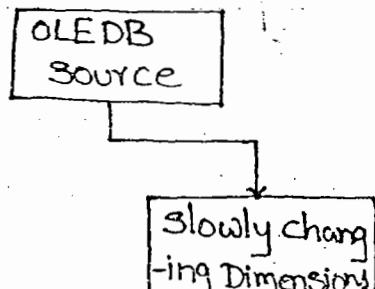
Update Statistics Task: It updates statistics of the object if there are already collected.

Execute T-SQL Statement Task: It executes T-SQL Server database, commands, Queries....etc.

Go to Execute T-SQL Task → rt click → Edit → Execute  
T-SQL statement task → Connection → SRC (take any connection)  
T-SQL statement: USE DB\_MSBI;  
Delete from EMP\_NEW1;

Implementation of Slowly changing Dimensions (SCD) (through wizards): In ssis along with Type-1 & Type-2 we implement fixed attribute.

Fixed attribute: In this situation the important business information field taken as "fixed". If there is any change for values in the column we can take appropriate actions like failing (or) ignoring the operation.



### Navigation:

1. Take OLEDB Source.

2. SCD → rt click → edit → Next →

Connection Manager: DB\_MSBI

Table or View: EMP\_HIST.

Specify the column EID as business key.  
only dimension columns

① Set change type: fixed attribute.

② Click Next.

Select Fail the transformation if changes are detected

in a fixed attribute → Next → Next → Finish.

### SCD Type-1 :- (Changing attribute)

Like above with only below two changes

(i) Set dimensions column: partyloc.

(ii) Set change type: changing attribute.

### SCD Type-2 :- (Status Mechanism) :-

Like above with only below changes

Set change type: Historical Attribute.

Select use a single column to show current and expired records.

Column to indicate current record: status.

Value when current: current

Expired value: Expired.

### SCD Type-2 :- (Date Mechanism) :-

Select use start and end dates to identify current and expired records.

Start Date Column: start\_date

End Date column: End\_date.

Variable to set Date values: System::Container

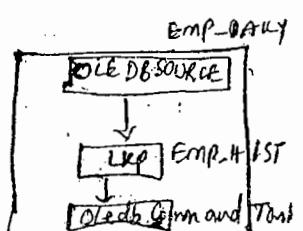
- Start time.

### SCD Type-2 :- (Date mechanism manually) :-

In Control Flow Task → DF\_UPDATE →

(FIG-(A))

DF\_Insert



→ (See next)

Create Ado-Net connection for the source table

1. SSIS Menu → New connection → Ado-Net → Add →

Data connections : MSBI-DB.

2. Data profile task on control Flow..

Rt click → Edit

General :

Destination Type: File connection.

Destination : New connection → createfile →

C:\ Party-Profile-Request.

click Quick Profile →

Ado-Net connection : MSBI-DB

Table or view: party

compute:

Select the required options

click OK → OK.

3. Execute data profile Task & observe the file.

4. Start Menu → programs → SS 2008 → Integration

Services → Data profile viewers → open → specify

File path (C:\ Party-profile-result) and monitor the analysis

by selecting each attribute.

Note: In other ETL tools for profiling we need to go for 3rd party tools.



## Various ways of Debugging

Q There are three ways of Debugging. They are:-

(i) By executing the package Partially

If multiple tasks are present in a package then we can execute a specific task.

Rightclick that task → click Execute Task.

(ii) By Break points.

To stop the execution of package at a particular event and to continue (or) stop the events, break points are used.

We perform break points only in control flow.

Consider the (Fig-(A)), taking breakpoint after DF\_update successful.

### Navigation:-

→ DF\_update → rightclick → Edit break points → Mark Break while the container receives the on post execute event option → click OK.

→ Execute the task so there is a break after DF\_update.

→ Monitor the statistics (or) observe the expected result, based on the result

① In case of continuation → goto debug menu

↳ click continue.

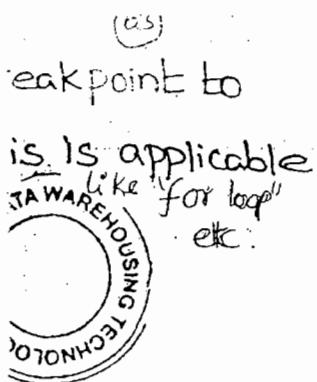
② In case of stopping → goto debug menu

↳ click stop debugging.



Hit count: We will take it along with the breakpoint to

have a break according to the condition. generally when we are having container and usable (except the type always).



Types:

1. Always: Execution is always suspended when the break point is hit. Eg. ALWAYS
2. Hit count Equal to: Execution is suspended when the no of times the breakpoint has occurred is equal to hit count.
3. Hit count greater than or equal to: Execution is suspended when the no of times the break point occurs is equal to or greater than the hit count.
4. Hit count multiple: Execution is suspended when multiple of the hit count occurs.

Eg: If we set this option to five, it takes a break for every fifth time.

### (iii) Data viewers:-

These are used only in the data source, destination and transformation view option only in the links between them.

Go to any link in the data flow task → right click → Data Viewers → click add → Select Grid → OK.

### Precidence Constraints:

These are useful to control the flow between various tasks in control flow.

flow task b/w  
take this data  
components.

click → Data

See below

FF

red

→ link → double click →

Lo

Data viewers →

OLE

click Add →

cmd

Select Grid → OK → OK

## Constraints options:-

There are two evaluation operations.

a) constraint :-

Success (Green link) :- If previous component is succeeded it executes the other.

Failure (Red link) :- If the previous component is failed it executes the other.

Completion (Blue link) :- If the previous component either success or failure it runs the other.

b) Expression → It supports an expression. When it satisfies (or) reaches, it executes the other task.

Eg:- @counter = 6

counter is a variable. (may be for loop (or) any).

c) Expression and constraint → If both are satisfied then only the other task executed.

d) Expression or constraint → If either of these are succeeded it executes the other task.

## Implementation of Expression

1. Declare a variable called "Counter".

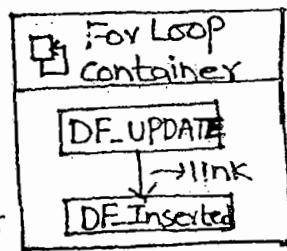
2. Take a For Loop and do the below settings

→ rclick → properties:

Initial Expression: @Counter = 6

Evaluation Exp.: @Counter < 9

Assign Expression: @Counter = @Counter + 1



(65)

→ Rclick the link → Edit

Evaluation operation: Expression and constraint

Value: Success

Expression: @counter==6

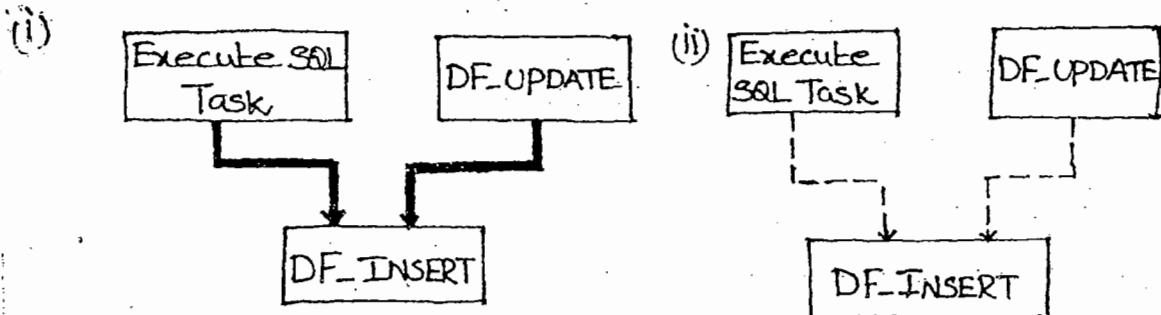


Multiple constraints:-

We take multiple constraints to interoperate and to control the execution of the constrained task.

Two ways:-

- (i) Logical AND. All constraints must evaluate to True.
- (ii) Logical OR; one constraint must evaluate to True.



Note:- Workflow can be controlled by precedence constraints.

Check point:- check point configuration helps us to resume from the last task in the package. i.e., If multiple tasks are there in a package, If there is any failure in task it stores the fail task point in the check point, once we restart the package the check point helps to start from last point specified in check point file. Once the package succeeded it - the check pt file.

- ① Take Two Execute SQL Tasks, one task with correct SQL command and second task with improper command (so that it fails).
- ② Go to each task → Properties → Fail package on failures: True.

(66)

control Flow → rtclick → properties →

checkpoint File Name: Desktop → check File.Txt

check point usage: If exists.

Save checkpoint: True.

Execute the package, as 2nd task is failed, checkpoint file is generated.

Rectify the SQL command in 2nd task and Rerun the

Package, then it starts from 2nd task instead of 1st task.

Logging: It uses various log providers to take the log information at particular event. The log providers are:

SSIS log provider for Windows Eventlog

Text files

XML files

SQL Server

Profiler

In "real time" this log information is used to perform the below tasks.

To eliminate bottlenecks

To troubleshoot the package



This log information is different from progress tab information becoz (i) it contains auditable information.

(ii) Start point, End point of the tasks available in the package. (iii) Machine name, operator name etc

Navigation: SSIS Menu → Logging → select package (or)

Task in the left hand side panel →

Right hand side

Provider Type: SQL Server provider for text files → click add →

## EVENT HANDLING:-

Implementing action at a

particular event is called event handling.

Eg of events :-

a) Onpost Executi-

b) Onpre Executi-

c) OnInformation

d) OnError ETC--



Event handling eg:-

1. Sending an Email after successful execution of package.

Navigation:-

GoTo Event handler tab



Select-

Executable

Package

Eventhandler

Onpost executi-



Drag and drop Send mail task, do the Configuration.

Run package and observe the result.

2. Delete the data before loading the data into GIRPLOAN table.

Navigation:-

GoTo event handler tab



Executable

Dataflowtask

Event handler

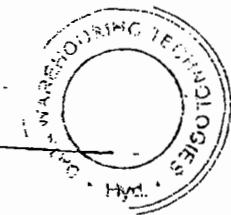
Onpreexecuti-

Take execute SQL task and do the below settings

Connection: DB\_MSBT

SQL Statement:- Delete FROM GRUPLOAD;

Execute package and See the expected Result.



Configuration → New connection → create file → specify  
file location → OK → OK.

Details Section → events → select on post execute →  
click OK → OK.

### Package Configuration:-



These are helpful while migrating (or) moving the packages from one environment to another environment.  
Development to Testing

Testing to production ...etc.

There are many ways we create configuration.

- ① XML configuration.
- ② Windows <sup>REGISTRY ENTRY</sup> Variable
- ③ Environmental variable
- ④ Parent package variable
- ⑤ SQL server database.

Note:- In realtime the frequently used configurations are "XML and SQL Server DataBase".

As XML is industrial standard with platform independent, most of the organisations are preferring it.

### XML Configuration:-

1. Take Data Flow Task → Flat File source (C:\Hyd.Txt) → FlatFile destination (C:\Hyd\_opt.Txt)

2. SSIS Menu → package configurations → check enable package configuration → click add → Next

configuration Type: XML Configuration File.

configuration FileName: Browse and specify filename (New)  
click Next → check the connection string property for source &

- ① Connection managers
- ② click Next → Finish → close.
- ③ Goto configuration file → open using any editor like Microsoft Visual studio Version selector (or notepad or wordpad...etc).
- ④ change Source filename to C:\Banglore.Txt
- ⑤ change destination filename to C:\Banglore.opt.Txt
- ⑥ Save
- ⑦ Run the package. Now the package runs with the configuration file settings.

### Deployment & Security :-

Providing a runnable solution at testing (or) Production generally we go for deployment (Moving the developed application from one environment to another environment)

In SSIS there are two deployments.

- 1) File system Deployment: In this case the packages deployed to a file system (i.e., to a specified drive and folder).
- 2) SQL Server Deployment: Here packages deployed in SQL Server Integration Services.
  - To deploy the packages we require manifest file.
  - Manifest File contains the information which is important at the time of deployment.

Note:- In real time we always perform the second type of deployment (i.e., SQL server deployment).

- It holds metadata information of package and its components (CONFIGURATIONS, SECURITY, PROTECTION ETC--)
- When we built the project (or) Solution manifest File updated file.

### Manifest Creation:

Solution Explorer

↳ project

↳ right click

↳ properties

Build : BIN

Deployment : Create Deployment utility:

Allow configuration changes: Type.

Create deployment utility: True.

Deployment output path: BIN\Deployment.

Build menu → Build SSIS project

GOTO SOLUTION → BIN → DEPLOYMENT → OBSERVE MANIFEST FILE.

### ① File System Deployment:

Go to Manifest File → right click → Deploy → Next →

Select File system deployment → specify Folder to deploy.

→ click Next → Next → Finish.

Go to deployed folder and observe the packages and configuration files deployed (or) not.

Execution: Package → right click → open with → SQL Server

2008 package Execution utility → run (Execute).

## ① SQL Server Deployment

70

Manifest File → click → Deploy → Next

→ Select SQL Server Deployment → specify

Server Name: Local Host

Package path: \ Maintenance plan.

click Next → Next → Finish.



### Observing:-

Go to SSMS → Integration Services → stored packages → MSDB → Maintenance plan.

Running packages: Maintenance plan → packages → click → Run package → execute.

Note: The property allow configurations true, allows configuration changes after the deployment.

### Applying security:

Two levels — (i) BIDS Level → password protection.  
(ii) SSMS → Role base security.

Password protection: It helps us to prevent from

- (i) Unauthorized deployment
- (ii) unauthorized manipulation to the packages.

At BIDS level for having better security, along with password we take "LEVEL" protection also.

There protection levels available are:

- ① Don't save sensitive
- ② Encrypt sensitive with user key.

- (71)
- ④ Encrypt sensitive with password.
  - ⑤ Encrypt all with password.
  - ⑥ Encrypt all with userkey.
  - ⑦ Server storage.



### Sensitive Information:-

Generally package connection strings, user defined variables, enumerators ... are considered as sensitive information.

⇒ Go to control flow → Rclick → properties → security →

In package password option assign password → In protection level option select "Encrypt all with password" option.

Build solution and test in either of the ways

(a) By opening the solution again.

(b) By deploying the Manifest file.

In the above two situations it asks password.

### SSMS Level Security:-

A user or group assigned to a role and every role will have responsibilities, so the users act according to the responsibilities.

### 1. Creating User/Group:-

My computer → Rclick → Manage →

Local users and Groups → users → Rclick → New User. User Name: VINAY → click OK.

Start Menu → SSMS → Database engine → security

→ logins → rclick → New login.

Login Name → Search → vinay → ok.

click ok.

System Databases → MSDB → security → users →

→ rclick → New user

userName: VINAYUSR

Login Name: Rawan\vinay.



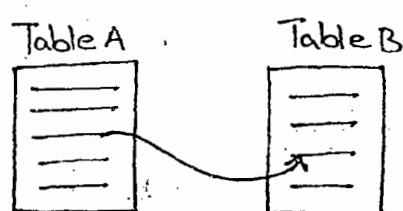
Check the required owned schemas & role members.

Eg: check DB-Data reader, DB-Backup operator...etc.

Sol: SSMS → Integration Services → connect →  
→ stored packages → MSDB → Data collector → rclick on  
→ my package → click Package roles → specify the roles in the  
reader and writer sections.

## Incremental Loading:

DailyData → Table  
(Granularity data)



1) Maintaining History [SCD]

2) Direct Loading

3) ETL Loading (Indirect load).

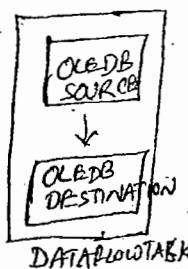
Ex: Loading the specified days data from one table to another table

Sol: Goto SSIS → variables → click add & Create two variables

Name	Scope	DataType	Value	(73)
EDATE	Package	Datetime	6/21/2011 12:36 AM	
SDATE	Package	Date	6/21/2011 12:36 AM.	

73

- ↳ OLEDB Source → rightclick → edit →



OLEDB Connection Manager: Local Host::DB\_MSBI

Access Mode: SQL command.

SQL command text:

Select \* From Emp\_Daily where ST\_DT between?  
And ?

click parameters:

Parameter 0 : USER:SDATE

parameter 1 : USER:EDATE

3. OLEDB Destination → EMP\_HIST table.

4. SSIS Menu → package configurations → check enable package configuration → click add → Next

Configuration Type: XML configuration file.

Configuration FileName: Browse and specify file name  
click Next → (New)

Goto SDATE & EDATE and check the 'value' sections.

Next → Finish → close

Open configuration file → change SDATE and EDATE & run the package.

## Working with Import Export Wizard

It performs operations between database to database, database to file, file to file.

## BIDS Level:-

Go to Solution explorer

↳ rt click SSIS package

↳ select SSIS Import and Export wizard

↳ Next

↳ choose a data source.

Data source:

Server Name: Local host

Database: DB-MSBI

↳ click Next →



Choose a destination

↳ Database

↳ DB New

↳ click Next

↳ specify Table copy (or) Query

↳ select the option copy the

data from the tables

↳ click Next

↳ select the options

Copy tables or views .

the data

Write a query to retrieve from dataset

click Next

↳ select the tables (trying to move)

↳ click Finish.

Now the system creates a package according to the settings given, execute the package and observe the result i.e, go for DBNew in Databases and observe the tables.

(B) SSMS Level:-

## Database Engine

↳ Goto any database

→ rightclick

## L $\rightarrow$ Tasks

↳ Select Import Data  
Export Data (or)

Working with Transaction and Isolation level:

Transaction :- It is the logical collection of statements (or) steps which can be succeeded (or) Failed.

The transaction isolation level determines the duration that locks are held.

(i) Read uncommitted: This is often referred to as "dirty read" becoz we can read modified data that hasn't been committed and it could get roll back after you read

(ii) Read committed: It acquires share locks and wait on any data modified by a transaction in process. This is a SQL "Server default".

(iii) Repeatable Read:- Same as read committed but in addition share locks are retained on rows, read for the duration of the transaction.

In other words any row that is read cannot be modified by any other connection until the transaction commits (or) Roll back.



## ④ Serializable:-

- Same as repeatable read but in addition no other connection can insert the rows,
- If new rows would appear in select statement already issued.
- In other words if we issue a select statement in transaction using the serializable isolation level we will get the same exact result set if we issue the select statement again within the same transaction.

## Transaction Options:-

- Required → If transaction exists join it, else start a new one.
- Supported → If transaction exists join it (this is the default).
- Not Supported → Don't join in Existing transaction.

## Navigation:-

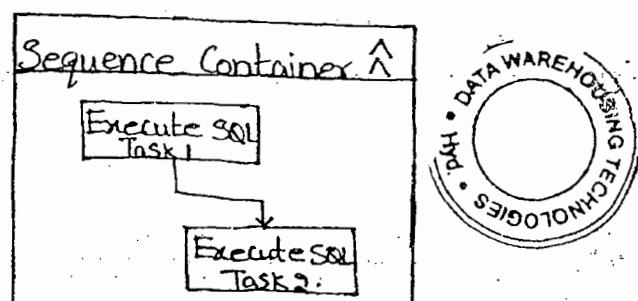
- Start the Service → DTC
- Take Sequence Container and Insert two Execute SQL Tasks.

Execute SQL Task 1

↳ Rclick → Edit

↳ SQL Statement: Insert into Emp  
values (9, 'VINAY', 50);

↳ Rclick → Edit Breakpoints → (check) break when  
the container receives on post executive event.



[DTC → DISTRIBUTED TRANSACTION COORDINATOR]

Execute SQL Task 2.

(7)

→ Rclick

→ Edit

→ SQL statement:

Insert into EMP(9, 'VINAY', 50); (wrong query).

3. Package → control flow → Rclick → Transaction option:

Required.

Sequence container → Rclick → properties →

Transaction option: supported.

4. Execute the package → After the first task Breakpt

Go to SSMS → Database Engine

→ DB-MSBI

→ Rclick

→ New Query:



Select \* from Party with (No Lock);

See the query result (It displays uncommitted data)

5. package → Debug menu → continue

[2<sup>nd</sup> Task failed, so Sequence container also failed]

6. SSMS → Database Engine → DB-MSBI → Rclick →

New Query: Select \* from Party;

See the query result (It displays old data, becoz newly added data is rolled back).

Note:- MSDTC (MICROSOFT DISTRIBUTED TRANSACTION CO-ORDINATOR).

## Creating and working with jobs:-

(78)

JOB: IT IS A PROCESS OF RUNNING A PARTICULAR TASK

[IS, AS (or) SET OF SQL QUERIES] AT A STIPULATED TIME.  
(SCHEDULE TIME).

- The jobs can be "one time" running jobs (or) "Iterative" jobs.
- To work with a job SQL Server agent should be in starting mode (it is in SSMS).

Ex: Running a file system package on every monday at morning 9:00 AM.

Job: Go to SSMS

- ↳ Database engine
- ↳ SQL server agent
- ↳ rclick → clickstart

Go to jobs

- ↳ rclick
- ↳ Select New job
- ↳ General

Name: SSIS-JOB → Steps →



New Job Step →

General:

Step Name: SSIS-STEP

Type: SSIS package

Package Source: file system

Package: specify the package location  
click OK.

Schedule → New

(79)

Name: Repeatable run.

Schedule type: Recurring

occurs: weekly

Recurse every: 1 week(s) on Monday

click OK → OK



Monitoring Job status: Go to the job → view history.

To see the log information regarding the job.

Job → right click

↳ View history

↳ Click export menu

↳ Specify file location and monitor

the job execution statistics <sup>in file</sup>.

Note:- THE BELOW SCHEDULERS ARE FREQUENTLY USED IN REAL TIME.

- Control-M
- Autosystem
- Trillium
- \$U (dollar U) ETC ...



### Performance Tuning:-

For more than 2 yrs of experience people this is mandatory concept and they must have good knowledge <sup>on</sup> this.

#### Situations to go:-

① To create a package with optimization.

② There is a package which is running such a long time.

In the above situations we need to identify the "bottle necks" and resolve it.



Q This Bottlenecks can be at many levels: (80)

- ① ④ package Level.
- ② ⑥ Source level
- ③ ⑤ Destination level.
- ④ ⑦ Transformation level
- ⑤ ⑧ Data Flow Task level
- ⑥ ⑨ System level.

### Identifying Bottlenecks:

By using "progress tab information (or) by using log providers" we identify the bottlenecks (Because they display step by step execution).

#### Package Level Tuning tips:-

① Implement check points to have better restartability of components in the package.

#### ② Disable event handlers:-

Event handlers decrease package performance so, unnecessary event handlers should be removed (or) disabled.

#### ③ Maximum Concurrent Executables:-

Increasing the no of executables will increase the parallelism of package and concurrently execute in less time.

#### ④ Maximum Error Count:-

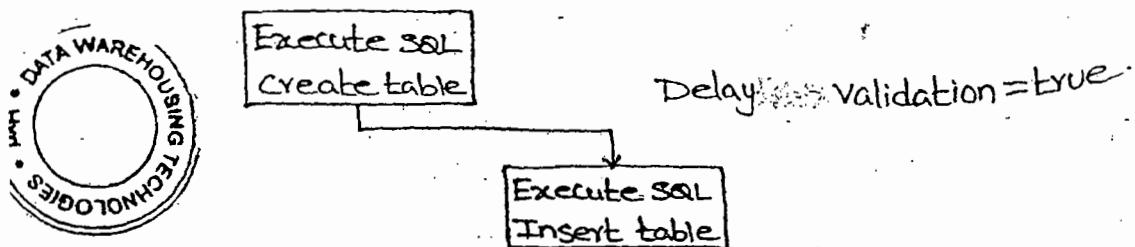
Default '1' means it fails for single error in the package. If you increase the error count it doesn't fail the package until it reaches the count.

## ② Dataflow task level tips:-

(81)

### i) Delay validation:- (Its True/False)

True means the validations of component is delayed until the execution of other component finished.



Description:- until "execute SQL Task 1" execution finishes.  
"Execute SQL Task 2" - validation not started.

### BLOB Temp storage path:-

specify this at the time of working with Binary Large Objects such as images, media files...etc.

### Default Buffer max rows and size:-

Increase or decrease according to the volume of data loading. i.e., for more volume rows and buffer size, for less volume, decrease rows and buffer size.

### Engine Threads:-

Default it takes '10', if we increase more threads it runs more parallelly and uses more processes to finish the data flow operations.

Run In optimized Mode:- If its true the dataflow avoids unnecessary transformations, conservations etc. operations usage.

## 9 Source Level Tuning Tips:

### (a) In case of Flat File

- (i) Try to take the flat file local to the system.
- (ii) use the property "Fast Parse = True". so that the column uses faster, local neutral parsing routines and avoids unnecessary conversions.

### (b) If the Source is Table or View

- (i) Create Indexes on the source table so that it retrieves the data faster.
- (ii) Instead of taking a table, take an SQL query
- (iii) SQL command as data access mode to get the required columns and rows of data.

## A Destination Level Tuning tips:



### (a) In case of Flat File

- (i) Try to take the file local to the system

### (b) In case of Relational Destination (Table or view)

- (i) use data access mode as SQL command to load any required rows and columns.

- (ii) use data access mode as fast load to load the data much faster.

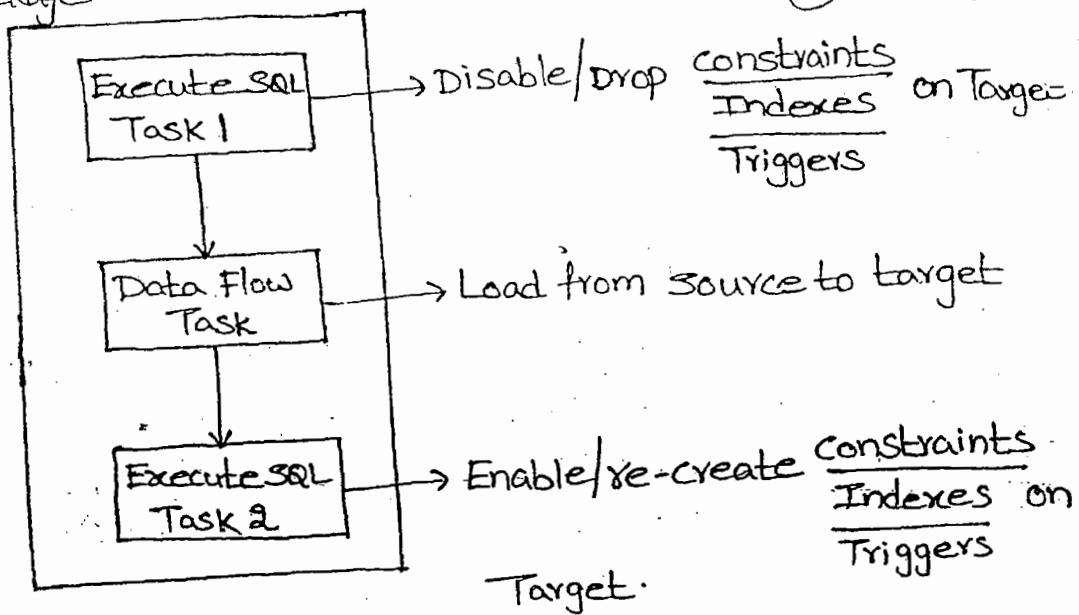
\*\* The table contains constraints, Indexes, triggers, then loading will be slow. so we need to disable (or) drop them, later once the loading is finished recreate (or) enable them.

To implement this there are many ways:

\* Eg:

Package

(83)



Another way to implement the above Package Functionality is by Event handling by

On pre execute of dataflow

↳ disable/drop

functionality.

On post execute of dataflow

↳ Enable/recreate

## ⑤ Transformation Level Tuning tips

We need to identify either the transformation is synchronous or Asynchronous.

Asynchronous transformation always takes time to run than Synchronous.

Synchronous Transformation:- Here the rows are processed as it is getting from the source

Eg:- Derived column, data conversion, etc.

This transformation occupies less memory & processes

## Asynchronous Transformation

- o In this case until the required rows, the system collects, it doesn't perform any operation.
- o It requires more memory and processes many rows.
- o These transformations either blocks partially (or) fully. so these transformation can also be called as "Partially blocking (or) Full blocking transformations"

Eg:- Sort, Aggregate, Merge ...etc.

In this transformation, until they receive all rows they will not perform any operation.

Use 'Multicast' transformation to read source data only one time and to perform different operations on the source in a single package.

Use 'Lookup' transformation to get the exact values.

Use 'SQL Query options' for Lookup to write customized query and to get required data.

→ Use 'Bulk insert task' to load the data from file to table. If the table already existed and there are no transformation required.



## SSIS PRACTICALS

### 1<sup>ST</sup> BUSINESS SCENARIO:

FILENAME	FILE COLUMNS	FILENAME	COLUMNNAME	TRANSFORMATION RULE	KEYVALUE
PARTY_TGT	PARTY_ID	PARTY_SRC	PARTYID	DIRECT MOVE	Y
PARTY_TGT	PARTY_NAME	PARTY_SRC	PARTYNAME	DIRECT MOVE	
PARTY_TGT	PARTY_LOC	PARTY_SRC	PARTYLOC	DIRECT MOVE	
PARTY_TGT	PARTY_INCOME	PARTY_SRC	PARTYINCOME	DIRECT MOVE	

**PARTY\_SRC.txt - Notepad**

PARTYID,PARTYNAME,PARTYLOC,PARTYINCOME  
 1,VINAY,HYD,800000  
 2,MADHU,MUM,400000  
 3,REJESH,CHENNAI,200000  
 4,KRISHNA,HYD,300000

**DATA\_VARTEXT.txt - Notepad**

ORGANIZATION,NUMBER,FIRSTNAME,LASTNAME,MIDDLENAME,PARTYLOC,PARTYCODE,JDATE,PARTYINCOME  
 IBM,100,MADHU,REDDY,SUDHAN,HYD,10,2011-01-19,80000  
 TCS,200,KISHORE,YADAV,KUMAR,MUM,20,2011-01-18,40000  
 WIPRO,300,RAJESH,MUKUNDAM,KUMAR,CHENNAI,30,2011-01-17,30000  
 CTS,400,VENKI,VELAMARTHI,MOHAN,HYD,20,2011-01-16,50000  
 INFOSYS,500,KRISHNA,REDDY,KUMAR,MUM,50,2011-01-18,60000  
 Cap Gemini,500,Deepesh,Agarwal,Agarwal,HYD,10,2011-01-18,30000  
 Visual Soft,500,Narne,yadav,Kishore,MUM,20,2011-01-18,40000  
 Accenture,500,Rupesh,Rudi,Medisetty,CHENNAI,30,2011-01-18,90000  
 CMC,50,Mohan,REDDY,Murali,CHENNAI,50,2011-01-18,80000  
 Satyam,101,KRISH,MOD,RAJE,HYD,40,2011-01-18,160000

### 2<sup>ND</sup> BUSINESS SCENARIO:

BDW TABLE NAME	BDW COLUMN NAME	FILENAME	COLUMNNAME	Mapping Remarks	Transformation Rule	Sort Order
PARTY	Party Id	DATA_VARTEXT	Organization Number	See mapping rule	Organization_Number concatenated with '_'	2
PARTY	Partyname	DATA_VARTEXT	Partyname	See mapping rule	Concatination of firstname,middlename and last name.	
PARTY	Email	DATA_VARTEXT		Hard Coded value	msbivinay@yahoo.co.in	
PARTY	Partycode	DATA_VARTEXT	Partycode	See mapping rule	DIRECT MOVE	1
PARTY	Partyincome	DATA_VARTEXT	partyincome	See mapping rule	DIRECT MOVE	
PARTY	Partyincome_inc	DATA_VARTEXT	partyincome_inc	PI value implementation	pi*12/100	
PARTY	Source_System_ID			Set mapping rule	Set the value as '41000'	

dropped by  
Satyam



### 3RD BUSINESS SCENARIO:-

TABLENAME	FILE COLUMNS	FILENAME	COLUMNNAME	TRANSFORMATION RULE	KEYVALUE
PARTY_HYD	USERNAME			USERNAME	
PARTY_HYD	PACKAGENAME			PACKAGENAME	
PARTY_HYD	ORGANIZATION	DATA_VARTEX_T	ORGANIZATION	TRIM OF ORGANIZATION	
PARTY_HYD	PARTYCODE	DATA_VARTEX_T	PARTYCODE	DIRECT MOVE	Aggregates
PARTY_HYD	JDATE	DATA_VARTEX_T	JDATE	DIRECT MOVE	
PARTY_HYD	PARTYINCOME	DATA_VARTEX_T	PARTYINCOME	DIRECT MOVE	
PARTY_MUM	USERNAME			USERNAME	
PARTY_MUM	PACKAGENAME			PACKAGENAME	
PARTY_MUM	ORGANIZATION	DATA_VARTEX_T	ORGANIZATION	DIRECT MOVE	
PARTY_MUM	PARTYCODE	DATA_VARTEX_T	PARTYCODE	DIRECT MOVE	
PARTY_MUM	JDATE	DATA_VARTEX_T	JDATE	DIRECT MOVE	
PARTY_MUM	PARTYINCOME	DATA_VARTEX_T	PARTYINCOME	DIRECT MOVE	
PARTY_CHEENAI	USERNAME			USERNAME	
PARTY_CHEENAI	PACKAGENAME			PACKAGENAME	
PARTY_CHEENAI	ORGANIZATION	DATA_VARTEX_T	ORGANIZATION	TRIM OF ORGANIZATION	
PARTY_CHEENAI	PARTYCODE	DATA_VARTEX_T	PARTYCODE	DIRECT MOVE	
PARTY_CHEENAI	JDATE	DATA_VARTEX_T	JDATE	DIRECT MOVE	
PARTY_CHEENAI	PARTYINCOME	DATA_VARTEX_T	PARTYINCOME	DIRECT MOVE	
LOCATIONS_SUMMARY	LOCATION	DATA_VARTEX_T	LOCATION	DIRECT MOVE	
LOCATIONS_SUMMARY	SUM_INCOME	DATA_VARTEX_T	See mapping Rule	Take sum of partyincome	
LOCATIONS_SUMMARY	AVG_INCOME	DATA_VARTEX_T	See mapping Rule	Take avg of partyincome	
LOCATIONS_SUMMARY	MIN_INCOME	DATA_VARTEX_T	See mapping Rule	Take min of partyincome	
LOCATIONS_SUMMARY	MAX_INCOME	DATA_VARTEX_T	See mapping Rule	Take max of partyincome	

Contd...  
20



#### 4<sup>th</sup> BUSINESS SCENARIO:

TABLENAME	FILE COLUMNS	FILENAME	COLUMNNAME	TRANSFORMATION RULE
PARTY_DETAILS	LAPARTYID	DATA_VARTEXT	PARTYID	DIRECT MOVE
PARTY_DETAILS	CITY	DATA_VARTEXT	PARTYNAME,PARTYLOC	PARTYNAME /First three characters of PARTYLOC
PARTY_DETAILS	INCOME	DATA_VARTEXT	See mapping rule	IF INCOME IS NULL DISPLAY 99999
PARTY_DETAILS	PARTYCODE DESC	DATA_VARTEXT	PARTYCODE	Use the below logic: PARTYCODEDESC='IT' if PARTYCODE=10 PARTYCODEDESC='HR' if PARTYCODE=20 FOR OTHER PARTYCOPES THE PARTYCODEDESC IS 'SALES & MARKETING'
PARTY_DETAILS	JDATE	DATA_VARTEXT	JDATE	Display only month and year in the format 'yyyy-mm-dd'
PARTY_DETAILS	ORGANIZATION	DATA_VARTEXT	ORGANIZATION	Uppercase of Organization

Deleted Column  
different

#### 5<sup>th</sup> BUSINESS SCENARIO:

TABLENAME	FILE COLUMNS	FILENAME	COLUMNNAME	TRANSFORMATION RULE	KEY/VALUE
LOCATIONS	USERNAME	PARTY_HYD	USERNAME	DIRECT MOVE	
LOCATIONS	PACKAGENAME	PARTY_HYD	PACKAGENAME	DIRECT MOVE	
LOCATIONS	ORGANIZATION	PARTY_HYD	ORGANIZATION	DIRECT MOVE	
LOCATIONS	PARTYCODE	PARTY_HYD	PARTYCODE	DIRECT MOVE	
LOCATIONS	JDATE	PARTY_HYD	JDATE	DIRECT MOVE	value
LOCATIONS	PARTYINCOME	PARTY_HYD	PARTYINCOME	DIRECT MOVE	
LOCATIONS	USERNAME	PARTY_MUM	USERNAME	DIRECT MOVE	
LOCATIONS	PACKAGENAME	PARTY_MUM	PACKAGENAME	DIRECT MOVE	
LOCATIONS	ORGANIZATION	PARTY_MUM	ORGANIZATION	DIRECT MOVE	
LOCATIONS	PARTYCODE	PARTY_MUM	PARTYCODE	DIRECT MOVE	
LOCATIONS	JDATE	PARTY_MUM	JDATE	DIRECT MOVE	
LOCATIONS	PARTYINCOME	PARTY_MUM	PARTYINCOME	DIRECT MOVE	
LOCATIONS	USERNAME	PARTY_CHENNAI	USERNAME	DIRECT MOVE	
LOCATIONS	PACKAGENAME	PARTY_CHENNAI	PACKAGENAME	DIRECT MOVE	
LOCATIONS	ORGANIZATION	PARTY_CHENNAI	ORGANIZATION	DIRECT MOVE	
LOCATIONS	PARTYCODE	PARTY_CHENNAI	PARTYCODE	DIRECT MOVE	
LOCATIONS	JDATE	PARTY_CHENNAI	JDATE	DIRECT MOVE	
LOCATIONS	PARTYINCOME	PARTY_CHENNAI	PARTYINCOME	DIRECT MOVE	

## Data Warehousing Technologies

#307, Annapurnablock, Adityaenclave, Ameerpet, Hyderabad-38. Ph:040-66469886



## 6<sup>TH</sup> BUSINESS SCENARIO:

FILENAME	FILE COLUMNS	TABLENAME	COLUMNNAME	TRANSFORMATION RULE	KEY VALUE
PARTY_REPORT	LAPARTYID	PARTY	PARTYID	DIRECT MOVE	Y
PARTY_REPORT	USERNAME	PARTY	See mapping rule	Get USERNAME where LOCATIONS. PARTYCODE=PARTY.PARTYCODE	
PARTY_REPORT	PACKAGENAME	PARTY	See mapping rule	Get PACKAGENAME where LOCATIONS. PARTYCODE=PARTY.PARTYCODE	Send later 2 days.
PARTY_REPORT	INCOME	PARTY	PARTYINCOME	DIRECT MOVE	
PARTY_REPORT	RATING	PARTY	PARTYINCOME	Use the below logic: Rating="Poor " if PINCOME <30000 Rating="Average " if PINCOME >30000 and PINCOME<50000 Rating="Good " if PINCOME >50000	
PARTY_REPORT	JDATE	PARTY	DATE	DATE FORMAT "DD-MM-YY"	
PARTY_REPORT	BUSINESS_DATE	PARTY	DATE	CURRENT DATE	
PARTY_REPORT	ORG_SHORT	PARTY	ORGANIZATION	FIRST 30 CHAR OF ORGANIZATION	

D:\DATA\_NORMALIZED.txt - Notepad

```

PARTYID,PARTYNAME,PARTYLOC,ACCOUNT TYPE,BALANCE
100,MUKESH KUMAR,VIZAG,SAVINGS,40000
100,MUKESH KUMAR,VIZAG,CURRENT,500000
100,MUKESH KUMAR,VIZAG,DMAT,20000
101,VINOD,CHENNAI,SAVINGS,30000
101,VINOD,CHENNAI,CURRENT,80000
101,VINOD,CHENNAI,DMAT,0
102,KUMARI,DELHI,SAVINGS,20000
102,KUMARI,DELHI,CURRENT,300000
103,RAJARAM,VIZAG,SAVINGS,30000
103,RAJARAM,VIZAG,CURRENT,0
103,RAJARAM,VIZAG,DMAT,12000

```

W : 040-6646983



Data Warehousing Technologies

#307, Annapurna block, Aditya enclaves, Ameerpet, Hyderabad-38. Ph:040-66469836



## 7<sup>TH</sup> BUSINESS SCENARIO:

TABLENAME	TABLESCHEMA	FILETYPE	FILENAME	COLUMNNAME	TRANSFORMATION RULE
COLUMNS					
PARTY_NORMALIZE_D	PARTYID	DATA_NORMALIZED		PARTYID	DIRECT MOVE
PARTY_NORMALIZE_D	PARTYNAME	DATA_NORMALIZED		PARTYNAME	PARTYNAME
PARTY_NORMALIZE_D	PARTYLOC	DATA_NORMALIZED		PARTYLOC	PARTYLOC
PARTY_NORMALIZE_D	SAVINGS	DATA_NORMALIZED		See mapping rule	Value of Savings for Partyid,Name,Loc combination
PARTY_NORMALIZE_D	CURRENT	DATA_NORMALIZED		See mapping rule	Value of Current for Partyid,Name,Loc combination
PARTY_NORMALIZE_D	DMAT	DATA_NORMALIZED		See mapping rule	Value of Dmat for Partyid,Name,Loc combination

PIVOT

## TABLES REQUIRED:

```
CREATE TABLE PARTY5(PARTYID INTEGER,PARTYNAME VARCHAR(30),PARTYLOC
VARCHAR(30),USERCREATED VARCHAR(30))
```

```
INSERT INTO PARTY5(1,'VINAY','INDIA')
```

```
CREATE TABLE PARTY6(PARTYID INTEGER,PARTYNAME VARCHAR(30),PARTYLOC VARCHAR(30),
STD_DT DATE,END_DT DATE)
```

```
INSERT INTO PARTY6(1,'VINAY','HYD',2011-01-19,2000-01-01)
```

## 8<sup>TH</sup> BUSINESS SCENARIO: FIXED WIDTH FORMAT (SCD TYPE)

BDW TABLENAME	BDW COLUMN NAME	FILENAME	COLUMNNAME	Mapping REMARKS	Transformation Rule
PARTY5	Party Id	DATA_FIXEDTEXT	PARTYID	PARTYID	DIRECT MOVE
PARTY5	Partyname	DATA_FIXEDTEXT	PARTYNAME	PARTYNAME	DIRECT MOVE
PARTY5	PARTYLOC	DATA_FIXEDTEXT	PARTYLOC	See mapping rule	if partyid exists location should be modified else insert
PARTY5	USERCREATED	DATA_FIXEDTEXT		See mapping rule	DEFAULT USER

old  
Lookup  
old

DATA_FIXEDTEXT	Notepad
File	Edit
Format	View
Help	
1VINAYUSA	
2MADHUHYD	
3KRISHNAMUR	



**9TH BUSINESS SCENARIO: VARIABLE TEXT FORMAT (SCD TYPE2-EFFECTIVE DATE RANGE)**

BDW TABLENAME	BDW COLUMN NAME	FILENAME	COLUMNNAME	Mapping Remarks	Transformation Rule
PARTY6	Party Id	DATA_VARTEXT	PARTYID	PARTYID	DIRECT MOVE
PARTY6	Partyname	DATA_VARTEXT	Partyname	PARTYNAME	DIRECT MOVE
PARTY6	PARTYLOC	DATA_VARTEXT	PARTYLOC	See mapping rule	DIRECT MOVE
PARTY6	ST_DT	DATA_VARTEXT	ST_DT	ST_DT	DIRECT MOVE
PARTY6	END_DT	DATA_VARTEXT		See mapping rule	1.If new record came inserted with start_dt and end date as '0001-01-01' 2. If partyid existed a should be inserted st_dt and end date as '0001-01-01' b.Earlier record end date should be new record start date

DATA\_VARTEXT\_2.txt - Notepad

1,VINAY,US,2011-01-20  
2,KRISHNA KANTH,hyd,2011-01-20

**TABLES REQUIRED:**

CREATE TABLE PARTY7(PARTYID INTEGER,PARTYNAME VARCHAR(30),PARTYCODE INTEGER)

INSERT INTO PARTY7(1,'KKKKK',20)

CREATE TABLE PARTY8(PARTYID INTEGER,PARTYNAME VARCHAR(30),PARTYCODE INTEGER,STATUS VARCHAR(30))

INSERT INTO PARTY8(1,'VINAY','HYD')



**Data Warehousing Technologies**

#307, Annapurnablock, Adityaenclave, Ameerpet, Hyderabad-38. Ph:040-66469886



## 10<sup>TH</sup> BUSINESS SCENARIO: DATA IN VARIABLE TEXT FORMAT

BDW TABLENAME	BDW COLUMN NAME	FILENAME	COLUMNNAME	Mapping Remarks	Transformation Rule
					1. Delete the record if transaction code is D 2. Do upsert operation if the transaction code is U
PARTY7	Party Id	DATA_VARTEXT	PARTYID	PARTYID	PARTYID
PARTY7	Partyname	DATA_VARTEXT	Partyname	See mapping rule	If partyid exists update it else insert it
PARTY7	Partycode	DATA_VARTEXT	PARTYCODE	See mapping rule	If partyid exists update it else insert it

DATA\_VARTEXT.txt - Notepad

D,1,KKKK,20  
U,3,KRISHNA,30  
U,2,VINAY KONNA,30

## 11TH BUSINESS SCENARIO: VARIABLE TEXT FORMAT (SQL TYPE 2 CURRENT EXPIRED MECHANISM)

BDW TABLENAME	BDW COLUMN NAME	FILENAME	COLUMNNAME	Mapping Remarks	Transformation Rule
PARTY8	Party Id	DATA_VARTEXT	PARTYID	PARTYID	DIRECT MOVE
PARTY8	Partyname	DATA_VARTEXT	Partyname	PARTYNAME	DIRECT MOVE
PARTY8	PARTYLOC	DATA_VARTEXT	PARTYLOC	See mapping rule	DIRECT MOVE
PARTY8	STATUS	DATA_VARTEXT	STATUS	See mapping rule	1. If new record came inserted with STATUS 'Current' 2. If partyid existed a. should be inserted with Status as 'Current' b. Earlier record Status should be 'Expired'

DATA\_VARTEXT\_3.txt - Notepad

1,VINAY,HYD  
3,KRISHNA,MUM  
2,MADHU KISHORE,MUM



CREATE TABLE PRODUCT(PRODUCTID INT,PRODUCTNAME VARCHAR(30),SUBPRODUCT  
 VARCHAR(30),PRODUCT\_DESC VARCHAR(30),SUBPROD\_DESC VARCHAR(30),PRODUCTTYPE  
 VARCHAR(30),BRAND VARCHAR(30))  
 INSERT INTO PRODUCT VALUES (0001,'SHIRT', 'HALF-SHIRT','Cotton shirt','Half coton  
 shirt','COTTON','PEPPY');  
 INSERT INTO PRODUCT VALUES(0002,'SHIRT','FULLHAND-SHIRT','Cotton shirt','FULL coton  
 shirt','COTTON','PANTALOONS');  
 INSERT INTO PRODUCT VALUES(0003,'TROUSURE','HALF-TROUSURE','JEAN TROUSURE','JEAN  
 TROUSURE','JEANS','PEPPY');  
 CREATE TABLE TIME(TIMEID INT,YEAR INT,QTR INT,MONTH INT);  
 INSERT INTO TIME VALUES(1,2009,1,8);  
 INSERT INTO TIME VALUES(2,2009,4,10);  
 INSERT INTO TIME VALUES(3,2009,3,9);  
 INSERT INTO TIME VALUES(4,2009,2,7);  
 INSERT INTO TIME VALUES(5,2010,5,12);  
 CREATE TABLE IKU(IKU VARCHAR(30) IKUFLAG VARCHAR(30));  
 INSERT INTO IKU VALUES('OK','STOCK AVAILABLE');  
 INSERT INTO IKU VALUES('NOK','REGRET STOCK');  
 CREATE TABLE LOCATION(LOCID VARCHAR(30) LOCNAME VARCHAR(30),LOCATIONTYPE  
 VARCHAR(30));  
 INSERT INTO LOCATION VALUES('HY','HYDERABAD','RENTED');  
 INSERT INTO LOCATION VALUES('CH','CHENAI','OWN');  
 INSERT INTO LOCATION VALUES('BL','BANGALORE','RENTED');  
 INSERT INTO LOCATION VALUES('MUM','MUMBAI','LEASED');  
 CREATE TABLE RAWMATERIAL(RAWMATERIALID INT,RAWMATERIAL\_FUNCTION VARCHAR(30),  
 UBFUNCTION VARCHAR(30),MEASUREMENT VARCHAR(30));  
 INSERT INTO RAWMATERIAL VALUES(1,'CHEMICAL','H2O2','LTRS');  
 INSERT INTO RAWMATERIAL VALUES(2,'OILS','KKK','LTRS');  
 INSERT INTO RAWMATERIAL VALUES(3,'CLOTH','COTTON','MTRS');  
 INSERT INTO RAWMATERIAL VALUES(4,'CLOTH','POEYSTER','MTRS');  
 CREATE TABLE TEXT\_FACT(PRODUCTID INT,RAWMATERIALID INT,TIMEID INT,  
 LOCATIONID VARCHAR(30),IKU VARCHAR(30),ACTUALCOST INT,ESTIMATEDCOST INT,  
 ACTUALRATE INT,ESTIMATEDRATE INT,VARIANCE INT,STANDARDDEV INT);  
 INSERT INTO TEXT\_FACT VALUES(1,1,1,'HY','OK',200,300,200,300,200,300);  
 INSERT INTO TEXT\_FACT VALUES(1,2,1,'CH','NOK',300,400,300,400,200,300);  
 INSERT INTO TEXT\_FACT VALUES(1,1,3,'HY','OK',200,300,200,300,200,300);  
 INSERT INTO TEXT\_FACT VALUES(1,2,2,'MUM','NOK',400,500,200,300,200,300);



## Data Warehousing Technologies

#307, Annapurnablock, Adityaenclave, Ameerpet, Hyderabad-38. Ph:040-66469886



## DWH & BI FUNDAMENTALS

DATA :- Data is composed of observable and recordable facts that are often found in operational or transactional systems.

OLTP :- OLTP is abbreviation of On-Line Transaction processing. This system is an applicable application that modifies data on the instance it receives and has a large number of concurrent users.

OLAP :- OLAP is abbreviation of On-Line Analytical processing. This system is an application that collects, manages, processes and presents multidimensional data for analysis and management purposes.

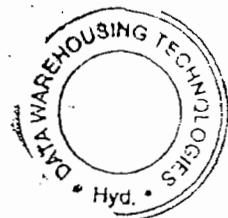
DATA MINING :- Data Mining is the process of analyzing data from different perspectives and summarizing it into useful information.

BI :- BI is the leveraging of BW to help make business decisions and recommendations. Information and data rules engines are leveraged here to help make these decisions along with statistical analysis tools and data mining tools.

### DWH DEFINITIONS :-

(1) Bill Inmon, considered to be the father of Data Warehousing provides the following definition:

\* Data Warehouse is a



Subject Oriented,  
Integrated,  
Nonvolatile,  
and time variant.

Collection of data in support of management's decisions.

Sean Kelly, another leading Data Warehousing practitioner defines the data warehouse in the following way.

The data in the data warehouse is,

1) Separate  
Available  
Integrated  
Time Stamped  
Subject Oriented  
Nonvolatile  
Accessible

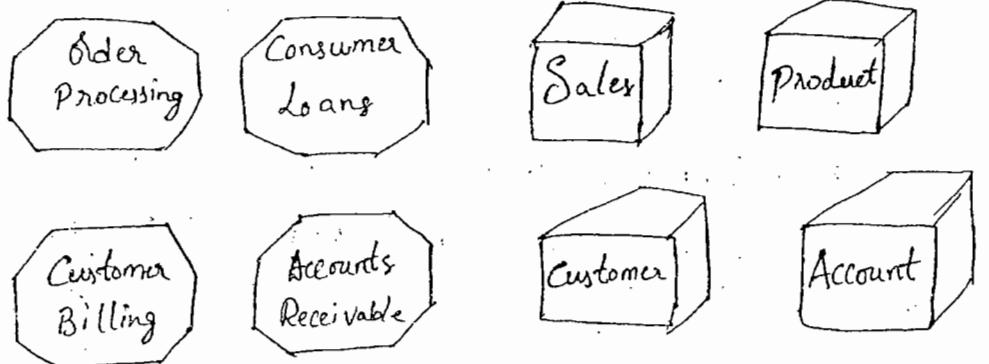


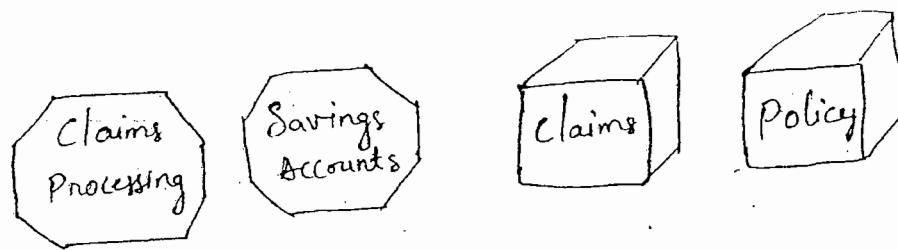
### Subject Oriented :-

In the data warehouse, data is not stored by operational applications, but by business subjects.

#### Operational Applications      Data Warehouse Subjects

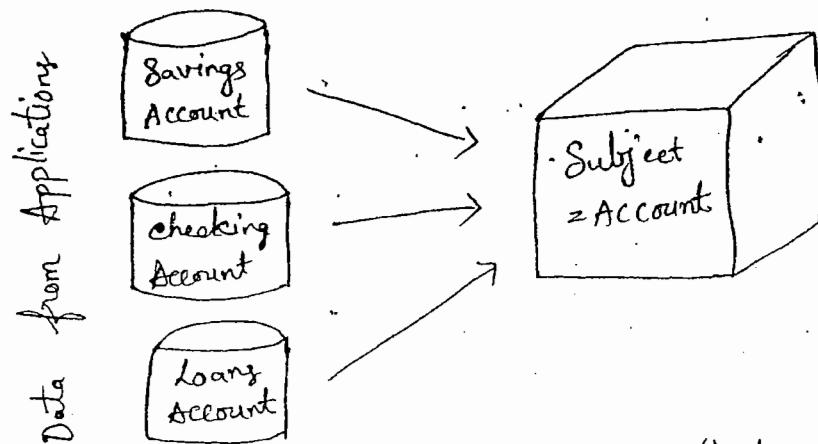
Eg:-





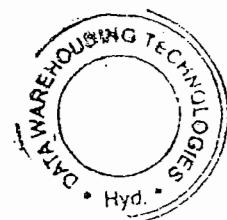
### Integrated :-

Data inconsistencies are removed; data from diversified operational applications is integrated.



Here are some of the items that would need standardization:

- \* Naming Conventions
- \* Codes
- \* Data attributes
- \* Measurements



### Time Variant Data :-

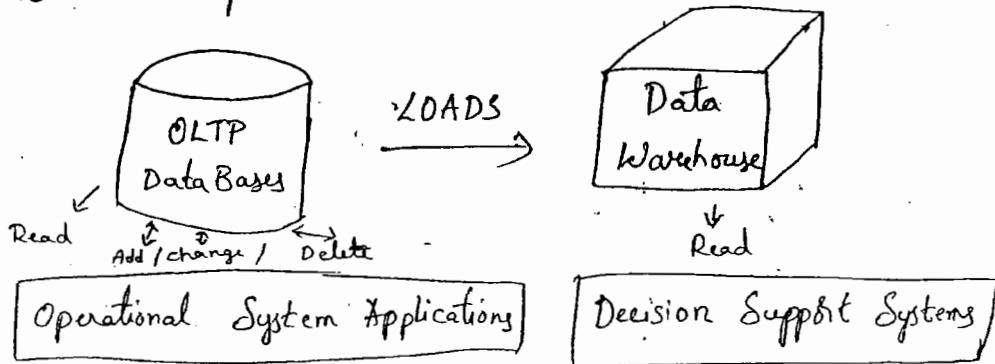
The time-variant nature of the data in a data warehouse

- Allows for analysis of the past
- Relates information to the present
- Enables forecasts for the future



### Non Volatile Data:

Usually the data in the data Warehouse is not updated & deleted.



### Data granularity:

Data granularity in a data warehouse refers to the level of detail. The lower the level of detail, the finer the data granularity.

- Depending on the requirements multiple levels of details may be present. Many data warehouses have at least three levels of granularity.

## Three data levels in a Banking Data Warehouse

<u>Daily Detail</u>	<u>Monthly Summary</u>	<u>Quarterly Summary</u>
Account	Account	Account
Activity Date	Month	Month
Amount	Number of transactions	Number of transactions
Deposit/ Withdrawal	Withdrawals Deposits Beginning Balance Ending Balance	Withdrawals Deposits Beginning Balance Ending Balance

## Data warehouse Approaches:-

RALPH KIMBAL

(Bottom up Approach)

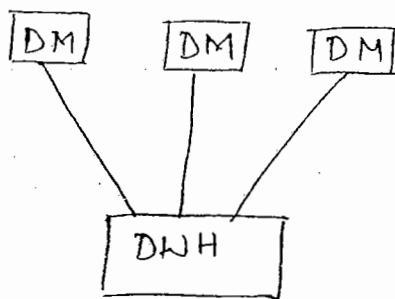
W.H DNMON

(Top down Approach)

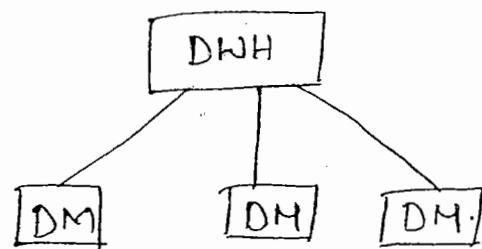
Here Datamarts designed first, later from Datamarts to warehouse

Enterprise DWH constructed first, Next Datamarts.



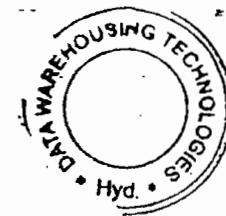


Data in "3nf"



Data is in dimensional model

### Top - Down Approach:-



The advantage of this approach are:

- A truly corporate effort, an enterprise view of data.
- Inherently architected - not a union of the ~~separate~~ data marts.
- Single, central storage of data about the content.
- Centralized rules and control.
- May see quick results if implemented with iterations.

The disadvantages are:

- Takes longer to build even with an iterative method.

- High exposure risk to failure.
- Needs high level of cross-functional skills.
- High outlay without proof of concept.

### Bottom - Up Approach:-



The advantages of this approach are:

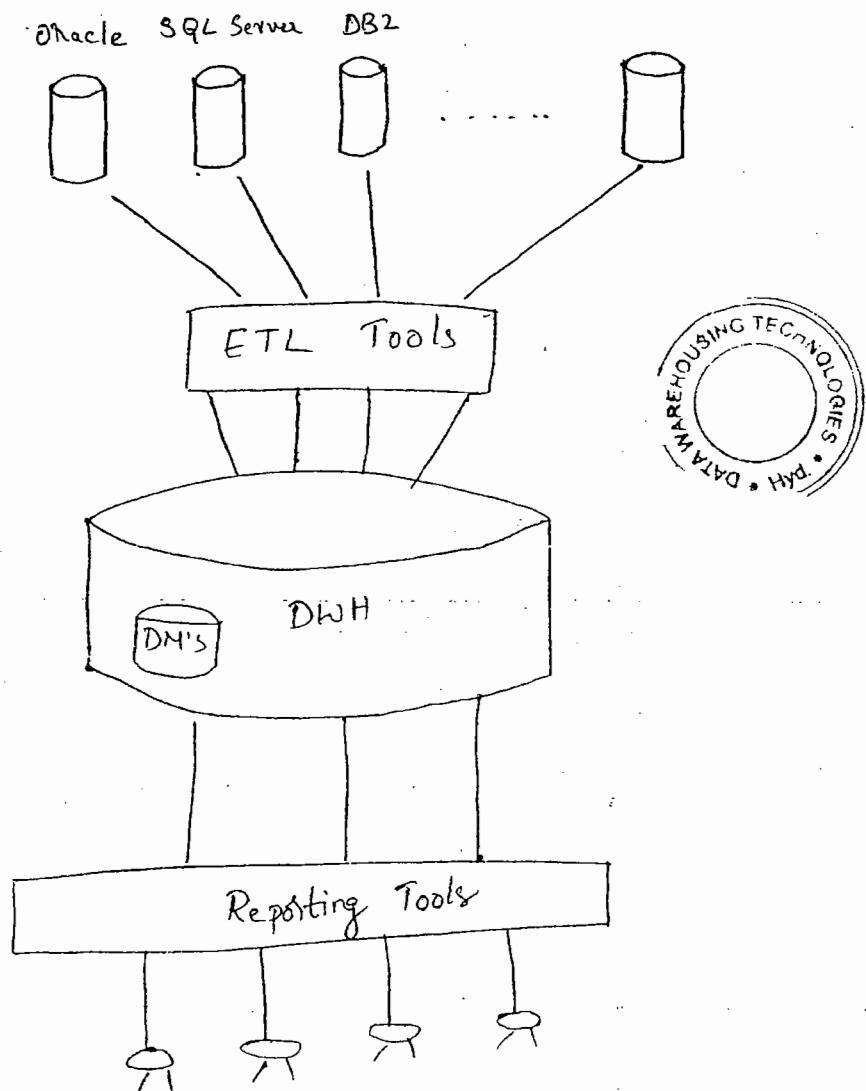
- Faster and easier implementation of manageable pieces.
- Favorable return on investment and proof of concept.
- Less risk of failure.
- Inherently incremental: Can schedule important data marts first
- Allows project team to learn and grow.

The disadvantages are:

- Each data mart has its own narrow view of data.
- ~~Permeates~~ <sup>permits</sup> redundant data in every data mart.

- Perpetuates inconsistent and irreconcilable data.
- Proliferates unmanageable interfaces.

### DWH LIFE CYCLE:-



## Difference between Data Warehouse & OLTP:-

### OLTP

### DWH

- Designed for transaction processing.
  - Volatile.
  - Store current data only.
  - Store detail data.
  - More joins.
  - Normalized (MORE TABLES)
  - Less indexes.
  - Low level managers access.
  - Read, update, insert, delete.
  - Access frequency high and in sec.
- Designed for decision support.
  - Store Non-Volatile.
  - Store historical data.
  - Store summarized data.
  - Less joins.
  - Denormalized (LESS TABLES).
  - More indexes.
  - High level managers access.
  - Read.
  - Low, Sec to minutes.



## Types of OLAP :-

### ROLAP:- (Relational OLAP)

Applied on relational sources, both data and aggregate information store under relational sources.

Eg:- BO, Cognos, Crystal reports, Micro strategy.

### MOLAP:- (Multi dimensional OLAP)

Here, Analysis will be done on multi dimensional applications.

Here, data and aggregate information store under multi dimensional sources.

Eg:- Cognos, SSAS, Hyperion, Micro strategy etc....

### HOLAP:- (Hybrid OLAP)

Here data store in relational sources and aggregated values store under multi-

dimensional Source "cubes"

Eg:- Cognos, BO, Micro strategic ....

DOLAP:- (Desktop OLAP)

Here, the analysis will be done on desktop application.

Eg:- MS-access, Excel, etc....

ETL:- a) CODE Based.

Development cost, testing, Maintenance.

Eg:- sas base, sas access, teradata utilities, SQL, PLSQL.

b) Gui Based ETL - informatica, datastage, Abinitio etc -

Data Extraction:-

a) Relational Sources:-

ORACLE, SQL SERVER, TD, DB2, INFORMIX, SYBASE, RED BRICK.

B) ERP Sources:-

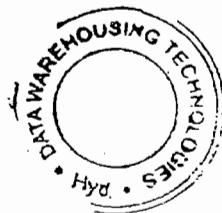
SAP R/3, PEOPLESOFT, J.D. EDWARDS,  
BANN, RAMCO MARSHALL.

c) Main frames Files:-

COBOL FILES, IMS FILES, JCL FILES,  
DB2, JCL

D) File Sources:-

FLAT FILES (TEXT FILES), XML FILES.



E) Other Sources:-

WEB LOGIC FILES, TIBCO M2 SERIES,  
EXCEL SHEET, PDF ETC....

## DATA MODELING

What is Data Model?

A data model is an abstraction of some aspect of the real world (system).

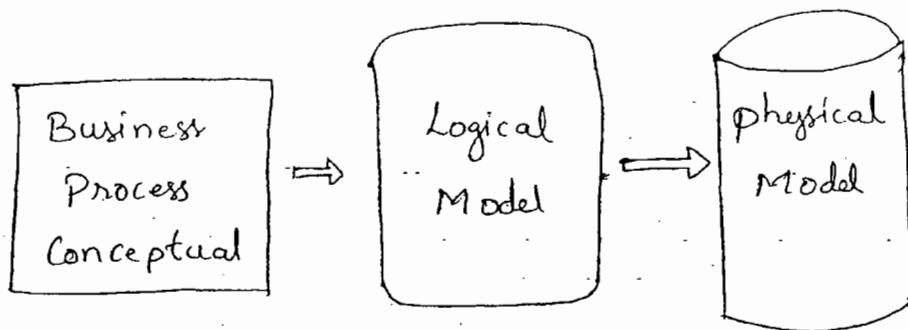
Why a Data Model?

- Helps to visualize the business.
- A model is a means of communication.
- Models help elicit and document requirements.
- Models reduce the cost of change.
- Model is the essence of DW architecture based on which DW will be implemented.





## Levels of Modeling:-



### Conceptual Modeling:-

- Describe data requirements from a business point of view without technical details.

### Logical Modeling:-

- Refine Conceptual models.
- Data structure oriented, platform independent.

### Physical Modeling:-

- Detailed specification of what is physically implemented using specific technology.

## Conceptual Model :-

- A conceptual model shows data through business eyes.
- All entities which have business meaning.
- Important relationships.
- Few significant attributes in the entities.
- Few identifiers & candidate (PK candidate) Keys.

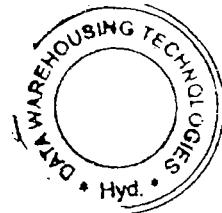
## Logical Model :-



- Replaces many-to-many relationships with associative entities.
- Defines a full population of entity attributes.
- May use non-physical entities for domains and sub-types.
- Establishes entity identifiers.
- Has no specific for any RDBMS or Configuration.

- ER Diagram > Key Based Modeling > Fully Attributed Model.

### Physical Model :-

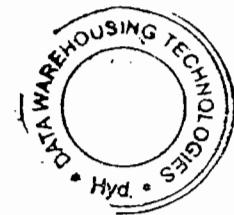


- A Physical data model may include:
  - Referential Integrity
  - Indexes
  - Views
  - Alternate keys and other constraints.
  - Tablespace and physical storage objects.

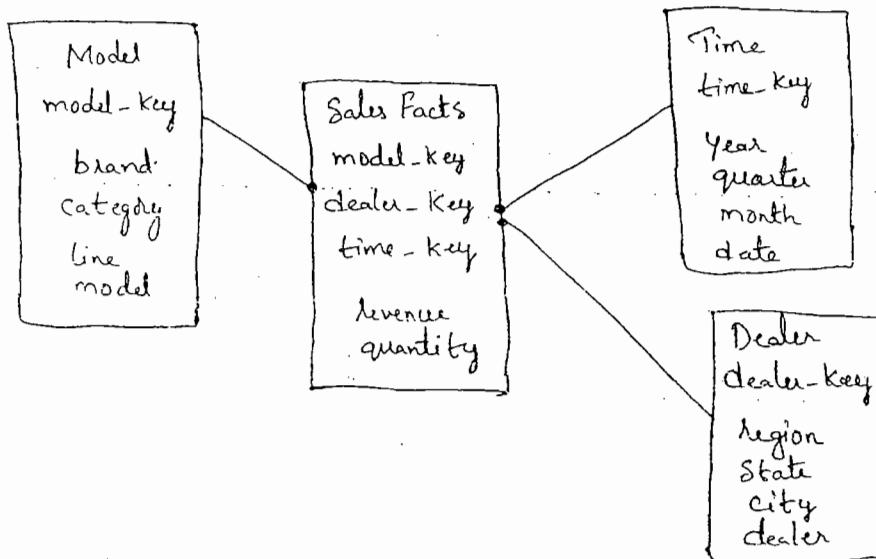
## Types of Facts :-

### Fully Additive Facts:-

- Can be Summed across any and all dimensions.
- Stored in fact table
- Eg:- revenue, quantity.



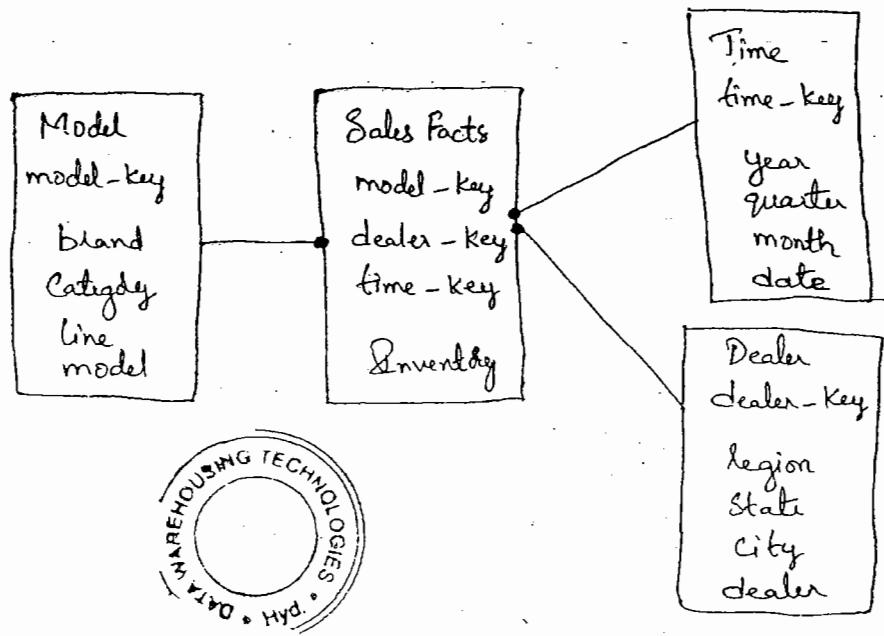
### Eg:-



### Semi Additive Facts:-

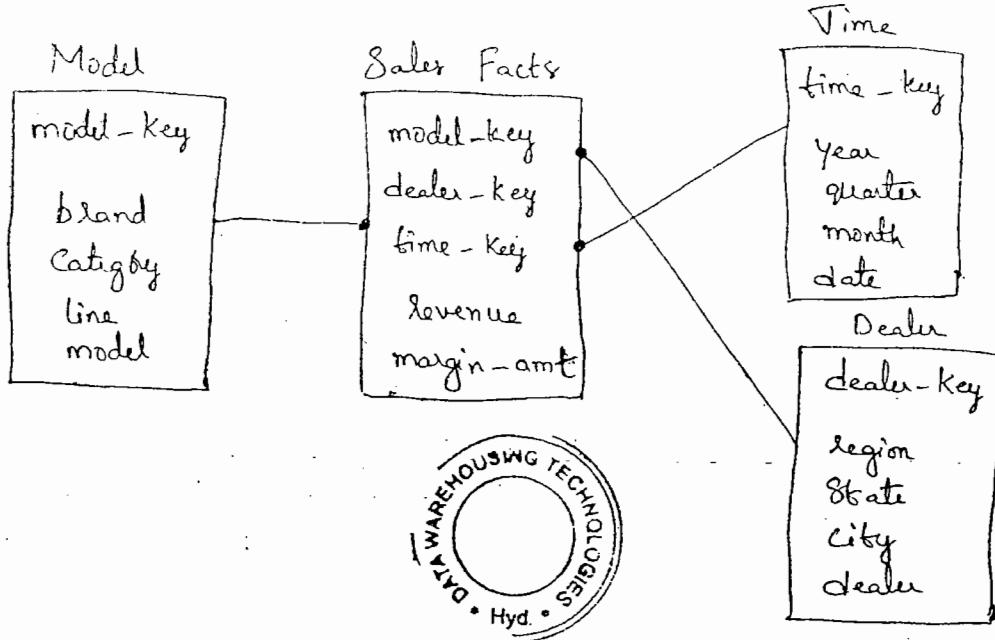
- Can be summed across most dimensions but not all.
- Eg:- Inventories quantities, account balances & personnel

- Anything that measures a "level".
- Must be careful with ad-hoc reporting.
- Often aggregated across the "forbidden dimension" by averaging



### Non Additive Facts :-

- Cannot be summed across any dimension.
- All ratios are non-additive.
- Break down to fully additive Components,  
store them in fact table.



Margin - rate is non additive

$$\text{Margin - rate} = \text{margin - amt} / \text{revenue}$$

### Fact Less Fact Table :-

- A fact table with no measure in it.
- Nothing to measure ...
- ... Except the convergence of dimensional attributes.
- Sometimes stde or "1" for convenience.
- Eg:- Attendance, Customer Assignments, Coverage.

## Surrogate Keys:-

The surrogate keys are simply system-generated sequence numbers.

Used generally in two situations

- 1) To replace the use of primary key (if it is Composite).
- 2) To generate unique values.

We use this surrogate key both at Dimension table and fact table.

## Difference Between Star and Snow Flake Schema

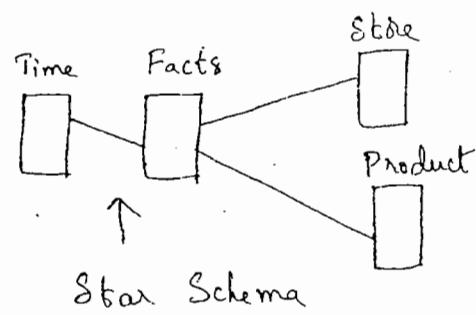
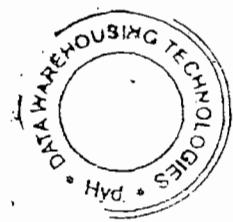
### Star Schema

- Here all dimensions directly connected to fact table.
- Dimensions doesn't contain any hierarchies
- Denormalized

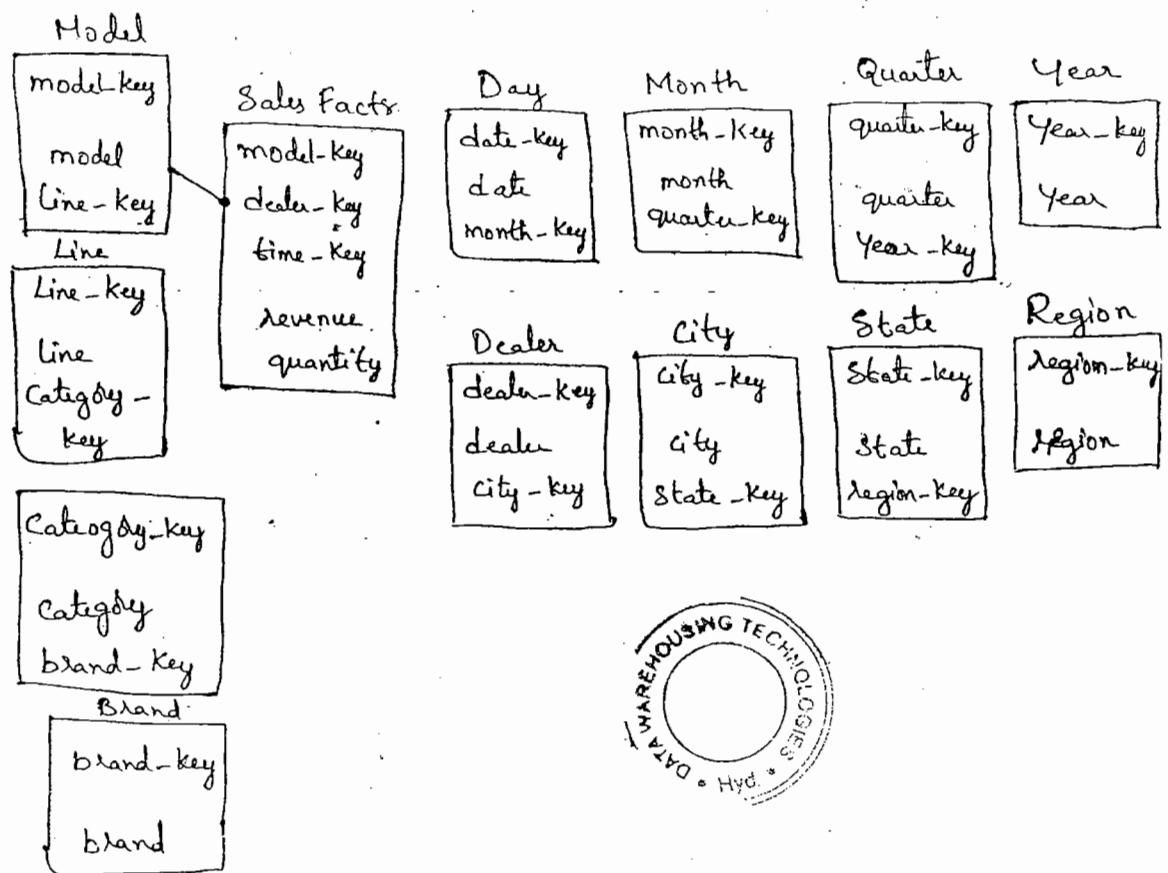
### Snow Flake Schema

- It is just like star schema but dimensions contain sub dimensions. So dimensional hierarchies are available
- Normalized

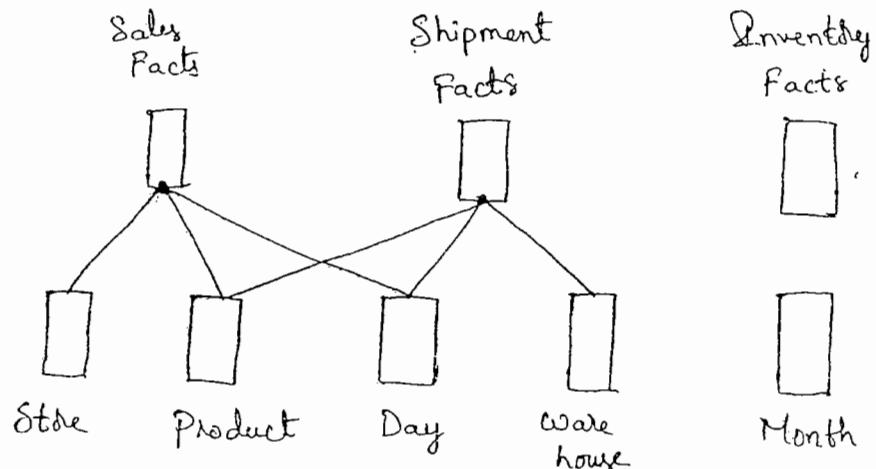
- Less joins, Query performance faster.
  - More memory occupies.
  - Business people can easy to understand.
  - Designed for DWH: OLAP applications.
  - Fewer Tables, Consolidated.
  - One to many relationship between fact and dimensions.
- More joins, Query performance slow.
  - Less memory occupies.
  - Complex to understand.
  - Designed for OLTP mode and OLAP less.
  - More Tables, Elaborated.



## Snow Flake Schema Egr:-



## Multiple Facts And Dimensions Connected :-



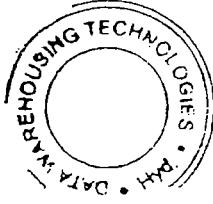
## Slowly Changing Dimensions:-

- Dimension source data may change over time -
- Relative to fact tables, dimension records change slowly.
- Allows dimensions to have multiple profiles over time to maintain history.
- Each profile is a separate record in a dimension table.

## Slowly Changing Dimension Types:-

Three types of slowly changing Dimension Types.

- Type 1
  - Updates existing record with modifications.
  - Does not maintain History.



### - Type 2

- Adds new record.
- Does not maintain history.
- Maintains old record.

### - Type 3

- Keep old and new values in the existing row.
- Requires a design change.

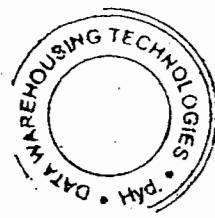
\_\_\_\_\_ X \_\_\_\_\_

## NORMALIZATION & T-SQL

The process of applying normal forms one by one at the time of data designing is called Normalization.

### Advantages:-

- ① Reducing unnecessary memory.
- ② Removing duplicates / Redundancy.
- ③ Having Proper Relationships.



Hint :- Helps the optimizer to provide an optimized execution plan.

### Hint applied at 3 levels :-

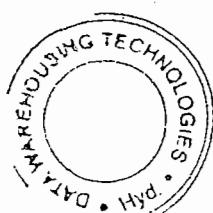
- ① At join operation :- It gives an efficient join plan.
- ② At table plan :- Locking mechanism on table control.
- ③ Query Level :- Select, Insert, Delete, Update operations will get an efficient execution plan.

SYNONYMS :- It is stored alias name for any object. We can use this instead of calling the main object name.

Ex:-

Create SYNONYM X for EMP

Select \* from X



### VARIABLES:-

- a) Global Variables → Defined by system  
→ Read only.

Ex:- Select @@ Servername, @@ Version, @@ Error.

## LOCAL VARIABLES:-

Created Manually. Defined for read, write ... etc operations.

Syntax:- Declare Variable Datatype.

Declare @ Var Varchar (30)

Set @ Var = 'MSBI CLASS'

Print @ Var.

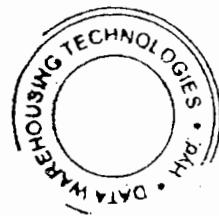
Select colnames / \* from <tablename>

Where <condition> / Sub Query

GroupBy <columns>

Having <Condition>

Orderby <columns>



NOTE:- GROUP BY, ORDER BY we can use numeric.

EMP

<u>Eid</u>	<u>Ename</u>
1	A
2	B
3	C
1	D
2	E
4	F

Select \* from EMP order by Eid (or)

order by 1

<u>Eid</u>	<u>Ename</u>
1	A
1	D
2	B
2	E
3	C
4	F

Identify duplicate records:-

Select Eid from emp Group by Eid having count (\*) > 1

<u>Eid</u>
1
2

Sub Query: Query inserted another query.

3 types:- ① Simple Subquery

② Nested Subquery (Subquery Inserted another subquery)

③ Correlated Subquery

The Innerquery preference by using scalar values machine operators.

like (<, <=, >, >=, is, <>, ----)

List of values multiple operators (IN, NOT IN, ----)

List of rows multiple operators (EXISTS, NOT EXISTS).

### Simple Subquery

① Inner query is executed first

② Outer query is executed next

③ Inner query is executed only one time

### Correlated Subquery

① Outer Query is Executed first.

② Inner query is executed Next.

③ Inner query is executed many times

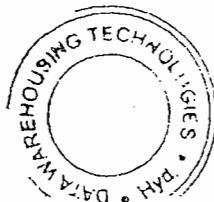


### Simple Subquery ex:-

Select max (partysal) from party

Where partysal not in

(Select max (partysal) from party).



### Correlated Subquery ex:-

Select P. partyincome from party P

Where P = (Select count (distinct (Q.partyincome)) from

Party Q Where P. Partyincome < Q. Partyincome).

surfaces from user (only 2008)

- Delete from trip where trip % % Physloc % %  
NOT IN (Select min.(%) from trip group by  
Eid having count(\*) = 1)
  - Physloc :- Physical location of a crew.
  - There are 3 more ways to delete the duplicates
    - ① By using Row Number.
    - ② By using Rank / Dense Rank.
    - ③ By using Intermediate tables.



Duplicate Records:- Select Eid, Ename into x from Emp Group By Eid, Ename having count (\*) > 1.

Delete Emp data for  
 deleting the records  
 available in X } → Delete from Emp, X where Emp.Eid.  
 = X.Eid and  
 $\text{Emp.Ename} = \text{X.Ename}$

Loading Unique records = Insert Into emp Select \* from X

## Working with multiple tables :-

To work with multiple tables the below objects required

- (3) Sub Queries      (6) Joins      (2) Procedures      (9) Triggers
  - (5) Set operations      (4) Views      (7) Functions      (8) Cursors
  - (1) Extended Procedures.

NOTE:- → for 5<sup>n</sup> max salary take  $n-1 = 5-1 = 4$ .  
→ for top 5 sal take  $5(5\geq)$

There are many ways to find out top sal & max salaries.

a) By using Rank() over(...)

Row-number over(...)

Dense\_Rank() over(...)

Top keyword

TOP:- It displays top values from the table.

Syntax:- Select top (Number) \* (or) <columns> from <tablename>.

Ex:- ① Display top 2 rows in the table.

Select top (2) \* from Emp.

② Display top 3 employees sal.

Select top (3) \* from Party ORDER BY partyincome DESC.

CTE (Common Table Expressions) :-

It is also called Temporary named result set, within the scope of an executing statement that can be used within a select / insert / update / delete / create view / Merge Statement.

Syntax:- WITH <CTENAME> <columns> As (Query)

Ex:- WITH xx (EID, FNAME)

As (Select EID, FNAME from Emp) Select \* from xx

NOTE:- Insert, select statements use many times CTEs.

### SET OPERATIONS :-

- It performs operations row wise b/w result set.
- It follows set theory and set operators.
- UNION:- It merges rows from result set excluding duplicates.
- UNION ALL:- It merges rows from 2 data sets including duplicates.
- EXCEPT:- It takes result set from 1st data set excluding duplicates.
- INTERSECT:- Takes common rows from both the datasets number of columns and order of columns should be same.





**ssas**



# SSAS INDEX

"Very for notes"

	<u>PAGE NO</u>
1. NEED OF ANALYTICAL APPLICATIONS	1-2
2. SSAS COMPONENTS & OPERATIONS.	2
3. SS S SAMPLE PROJECT	3-5
4. CUBE CREATION, DEPLOYMENT, BROWSING	5-13
5. WORKING WITH CUBE STRUCTURE	13
6. MEASURE GROUP, MEASURE	13-14
7. ADDING DIMENSION, HIERARCHIES, LINKED OBJECT	14-18
8. WORKING WITH DIMENSION USAGE WIZARD	18-22
9. MDX, EXPRESSIONS, QUERIES, EXAMPLES	23-35
10. CALCULATIONS, CALCULATED MEASURE, NON EMPTY	36-40
11. ACTIONS (Report, Drill through, URL)	41-45
12. PERSPECTIVES	46-47
13. TRANSLATIONS	47-48
14. KPI (KEY PERFORMANCE INDICATOR)	50 - 52
15. PARTITIONS (QUERY, TABLE), STORAGE SETTINGS PROACTIVE CACHING	51-57 65-71
16. FACT, DIMENSION PROCESSING, PROCESSING TYPES	58-64
17. NAMED QUERY, CALCULATION	73-74

18. IMPLEMENTING SCENARIO DIMENSION 74-75
19. APPLYING SECURITY 76-77
20. VARIOUS WAYS OF DEPLOYMENT 78-82
21. AGGREGATIONS DESIGN 97-100
22. PERFORMANCE TUNING 83-96
23. CREATING CUBE WITHOUT DATASOURCES - 104



## SSAS

(1)

### SQL SERVER ANALYSIS SERVICES

#### NEED OF ANALYTICAL APPLICATIONS :-

To create a Multidimensional objects and to provide multidimensional Analysis, we require Analytical Applications.

Ex:- COGNOS, BO, MICROSTRATEGY, SSAS, HYPERION, OBIEE etc..

→ SSAS, COGNOS, MICROSTRATEGY...etc tools are Pure 'MOLAP' tools.

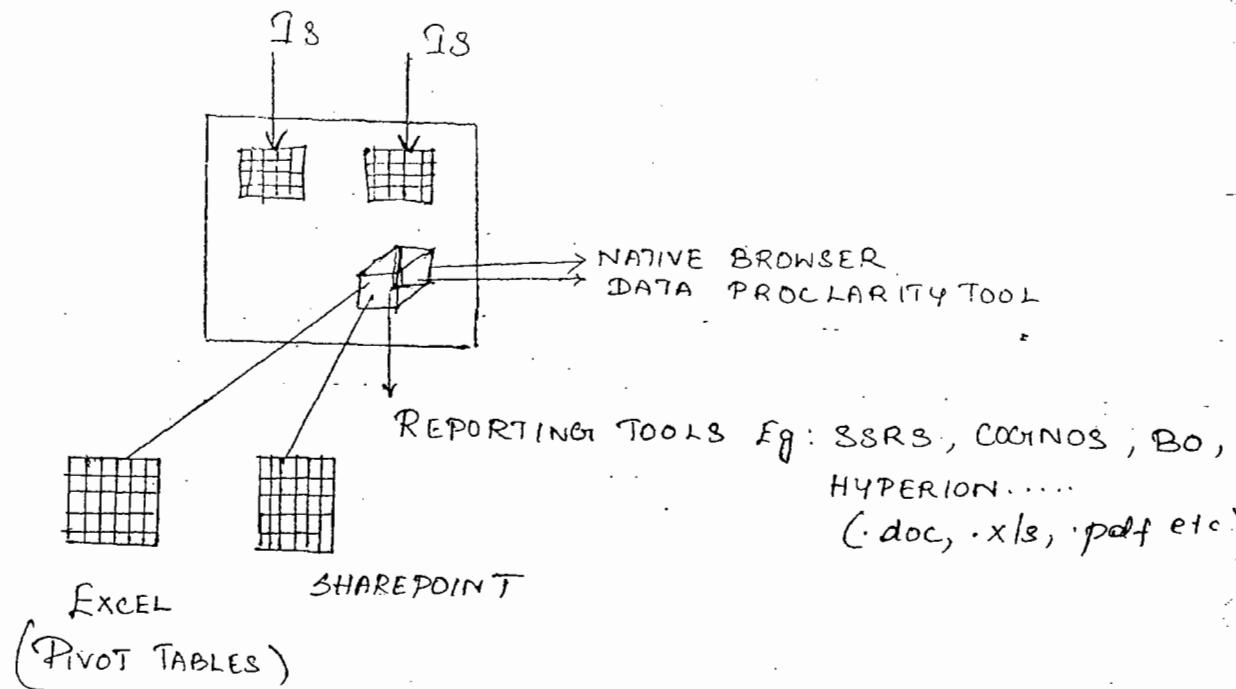
#### STRUCTURE OF MULTIDIMENSIONAL OBJECTS:-

Generally dimension Modeling structure (Star Schema, snowflake schema etc) used for multidimensional objects. Providing many-many relationship with E-R Modeling is difficult.

#### MULTIDIMENSIONAL OBJECTS USAGE (CUBE USAGE)

It Provides Support to End User (or) client by using various tools and Applications.





### ANALYSIS SERVICES COMPONENTS AND OPERATIONS !—

- a. Cube designing.
- b. Cube Creation
- c. Writing calculations
- d. Performing Actions
  
- e. Implementing KPI's (Key Performance Indicators)
- f. Creating and working with Partitions.
- g. Providing Multilanguage Support Using Trans.
- h. Designing the Aggregations.
- i. Creating Named calculation and Queries.



## SAMPLE PROJECT

### Description      TEXTILE MANUFACTURING INDUSTRY

A textile manufacturing industry manufactures various types of shirts, Trousers, etc...

Shirts also are different varieties (Half hand, Full hand sleeves) etc. Trousers also with different varieties (full, half, 3/4 etc...)

- These product sale is different location in different time zones (time period). <sup>happens</sup>
- Various Raw material requires are (chemical, oils, clothes...) to manufacture this product.
- There are different storage areas (coldown, Item keeping units) to maintain the product storage.
- For this project client require an Analytical solution where they can take appropriate decisions on the business.

### STEPS REQUIRED FROM END TO END TO IMPLEMENT THE ABOVE PROJECT

1. Need to collect the decisions from major to minor.
2. According to the decisions Analytical team collects or gathers the required Business information from the project.



(4)

3. Analysing the Business Requirements.
4. Identifying dimensions and facts for the above Business and creating ... LDN's (Logical Data Modeling).
5. Converting LDN's to PDM's (Physical data Modeling).
6. Based on the PDM's we go for designing of the dimensional model. (Star schema, Snowflake Schema-- etc)
7. Writing code to the design document in Analysis services for Analytical application (BO, cubes, HYPERION ..etc).

NOTE: Except the last step Remaining all steps are platform independent.

#### CONCLUSION :-

For the above business, technical team identifies (datamodelers, Data designers, Data Architects, domain Experts, SME's (subject Modern Experts). the below dimension and fact Tables.

#### Dimension and Fact Tables

- a. Time Dimension
- b. Product Dimension
- c. Location Dimension



d. Raw material dimension

e. Item Storage Unit dimension (true dimension)

One fact table for the above all dimensions.

## CUBE

It is an multidimensional object constructed with dimensions and facts in a particular design for taking multidimensional decisions.

## CUBE CREATION.

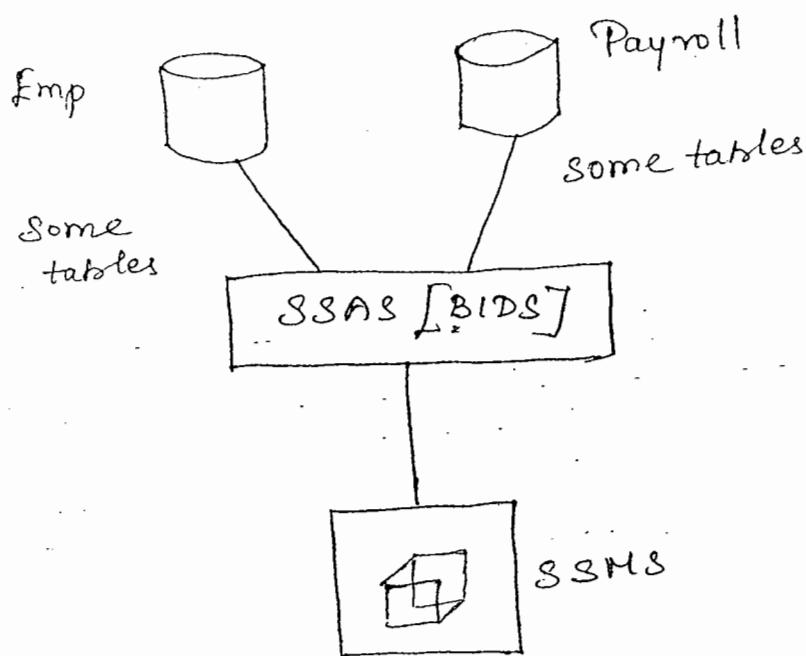
### Steps.

1. Open BIDS
2. Create Data Source
3. Create Data Source View
4. Provide relationship between dimensions and facts.
5. Create a cube
6. Manipulate the Components (Action, KPI.. etc)
7. Deploy the cube
8. Browser cube (or) perform reconciliation (or) unit testing.

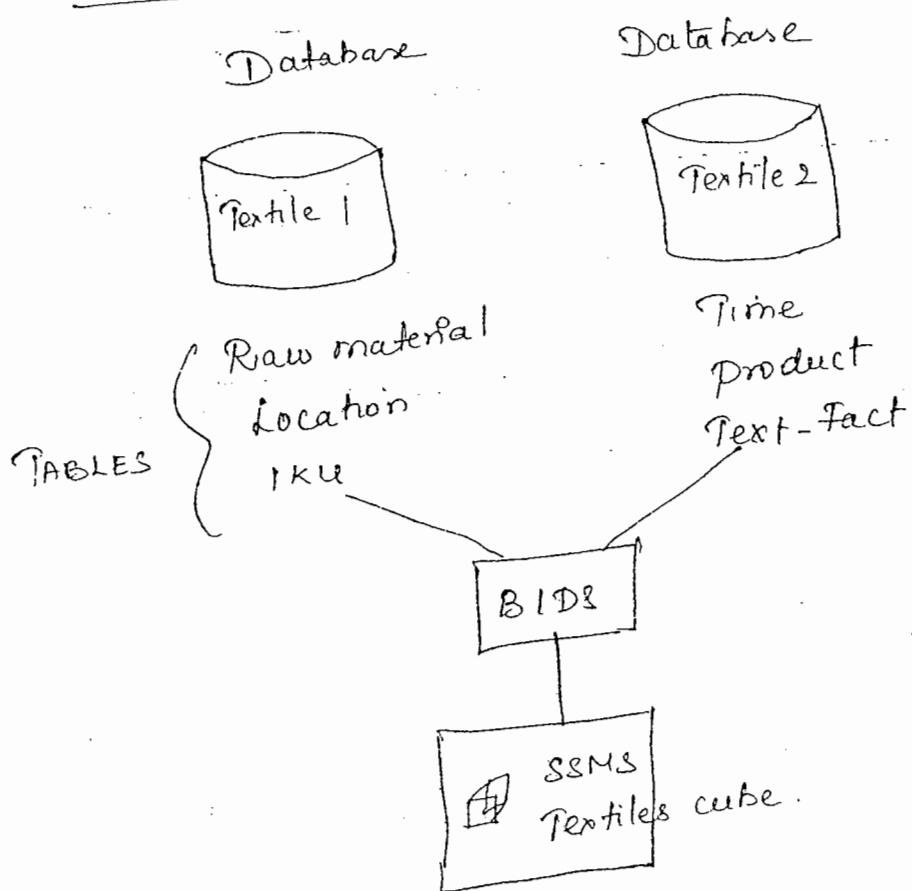


⑥

## Real time Example.



## CLASS ROOM EXAMPLE



## PRACTICAL IMPLEMENTATION OF CUBE

1. Open BIDS
2. File → New → Project → Template → Analysis Services Project → Project Name —

Name : TEXTILES\_CUBE

Location : C:\Documents and settings\Vinayaka

Solution Name : TEXTILES\_CUBE

3. View, Solutions Explorer

4. Create two Data Sources DS\_Textile1,  
DS\_Textile2 with the below procedure

Data Sources → RC → New Data Source

→ Next → New → SERVER NAME →  
LOCAL HOST

① Select or enter Database name.

LOCAL HOST ; TEXTILEST → OK → NEXT →

② Inherit → Data Source name : DS\_Textile1



Like this create another data source DS\_Textile2.

Finish

- (8)
5. Data Source views → RC → new Data Source view → Next →
- Relational data Sources.
- DS-textiles1 → Select → Next →
- Create logical relationships by matching Columns
- ↓
- Next → Select
- Available objects
- |             |                                     |
|-------------|-------------------------------------|
| RAWMATERIAL | <input checked="" type="checkbox"/> |
| LOCATION    | <input type="checkbox"/>            |
| IKU         | <input type="checkbox"/>            |
- RAWMATERIAL  
LOCATION  
IKU
- ↓
- NEXT →
- Name: DSV-TEXT → FINISH.

6. GOTO DSV-CUBE\_DB, for taking remaining (TIME, PRODUCT, TEXT\_FACT) tables into it follow this process.

DSV-CUBE\_DB Design → RC → ADD/REMOVE tables

↓

Data Source : DS-textiles 2

Available objects: Included Objects

TIME	<input checked="" type="checkbox"/>	TIME
PRODUCT	<input checked="" type="checkbox"/>	PRODUCT
TEXT_FACT	<input type="checkbox"/>	TEXT_FACT

↓

CLICK OK

7. PROVIDE RELATIONSHIP BETWEEN FACT TABLE TO REMAINING DIMENSION TABLES. By Dragging and drop Column mappings from fact table columns to dimension

... .

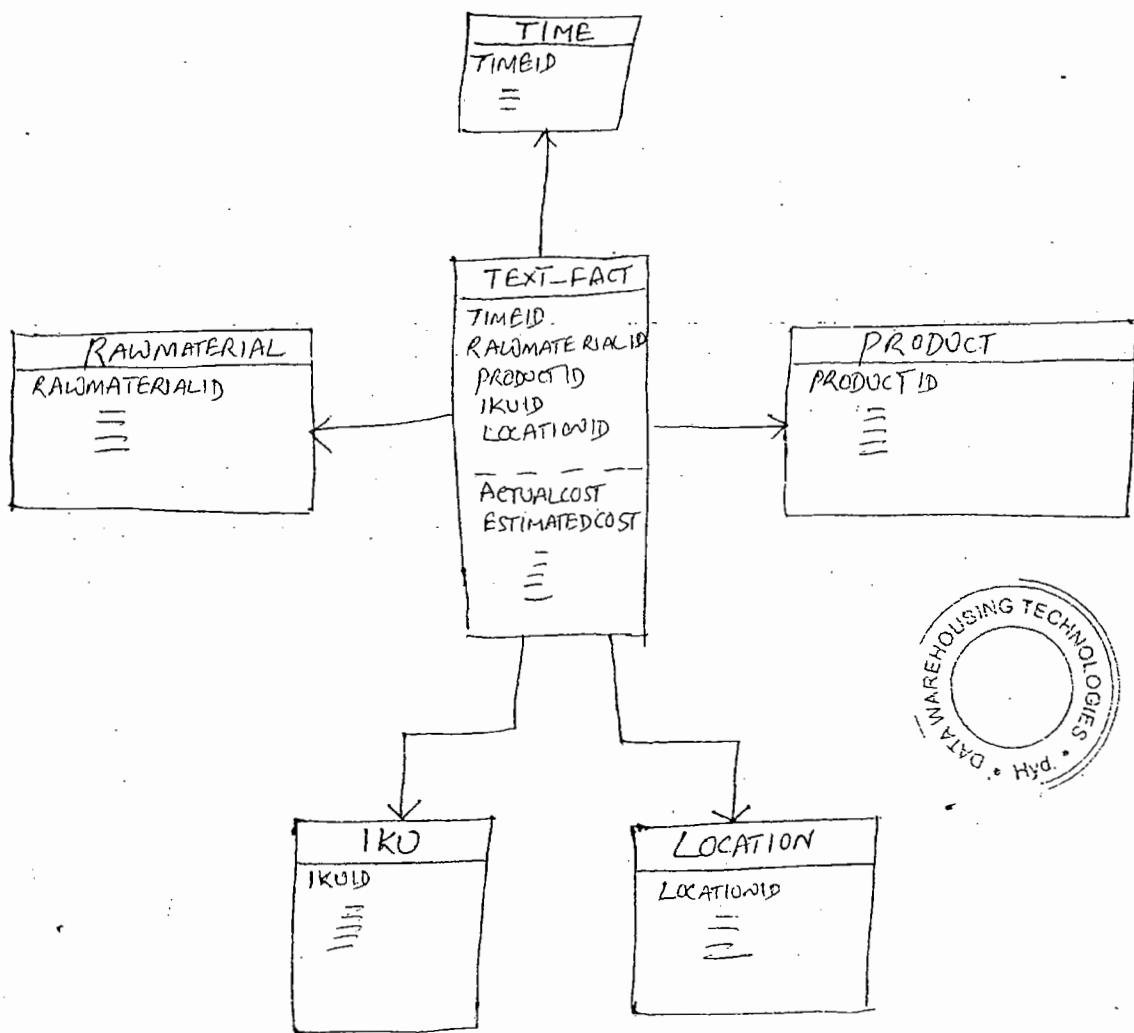
while connecting from fact column to dimension column  
it displays a message, click OK.

The destination table of the newly created relationship  
has no primary key defined. Would you like to define  
a logical primary key based on the columns used  
in this relationship?

YES

After all dimension Columns Connections,

DS - cube - DS → RC → Arrange tables, then it  
looks like this



8. CUBES → RC → New Cube → Next → (10)

① Using Existing tables → Next →  
Measure group tables

Select  TEXT\_FACT



NEXT



NEXT



NEXT



Cube name: Textile\_Cube



Finish

Now various tabs opened and we can see the  
Cube structure as well. [FACTS IN YELLOW, DIMENSIONS IN  
Build → Deploy → TEST The cube. BLUE COLOR] \*\*

#### Note:- Important options.

a) Build → Deploy :- If Cube Structure changed in BIDS  
to have the same in Cube database, this option useful

b) Build → process :- If data sources data and structure  
changes to have the same in Cube database, this option  
useful

c) Build → Build Solution:

It takes the required set-up files in the Solution folder

## CUBE DEPLOYMENT

(4)

- We need to ensure the cube is deployed successfully.
- To do this follow the below two general approaches.
  - a) In BIDS, goto Cube browser try to analyse and see the data.
  - b) Source any table data should match the cube database table data.

Eg:- No of rows in source (Text-Fact) data (40 rows)

$$= \text{SELECT [Measures].[Text-Fact Count] on columns}$$

$$\text{from [Textiles-Cube]}$$

$$(40rows)$$

Fire the above query in the below navigation.

SSMS → Analysis Services → Textile\_Cube → RC → MDX →  
Query.

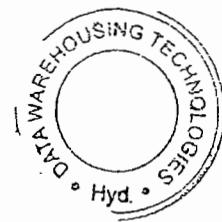
## GENERAL ERRORS IN THE LAB

1. If we are using other than Dimension key column values in fact table to the corresponding key, you may get error because of foreign key violation.

Eg:- Assume there is a LOCATION table with the below locations

"HYD  
MUM  
USA"

These



If we are using other than locations in the fact table, then we get errors.

(12)

## USING THE CUBE DATABASE

There are many ways

- a) Analysing the cube database data in BIDS browser
- b) Using "pivot table" in Excel application to connect and work with cube database.
- c) Using Reporting Tools (cognos, BA, SSRS - - -) to generate reports.
- d) By writing MDX queries in the Cube database.
- e) Using "data proclarity" tool to analyse the data.

## ANALYSING IN THE BIDS BROWSER

→ Take dimensions (or) facts either rowwise (or) column wise and analyse.

→ Go to menu bar on the top for filtering the data in the browser.

This bar can also be called as "FILTER BAR"

Eg:- Take actualcost, Estimatedcost on columnwise and locationID, productID, RawmaterialID on rowwise and Analyse.

RAWMATERIALID	LOCATIONID	PRODUCTID	ACTUALCOST	ESTMATE
1	HYD	SHIRT	900	800

## TO SEE THE DATA IN THE DIMENSIONS/FACTS

There are two different ways

a) GOTO Data Source View → Select table → RC



Explore Data

b) GOTO Cube Structure → Select table → RC



Explore Data.

## WORKING WITH CUBE STRUCTURE

→ It displays Cube Design, Measure groups, Measures, dimensions etc.

→ We preview the data here for dimensions & facts.

→

**MEASUREGROUP** It Contains Collection of measures.

→ Default measure group table is 'fact table' of the cube.

→ We can add new measure group tables.

Adding new measure group:-

1. Take measure group table in Data Source view.

2. Measures

CUBE DATABASE NAME



RC



New Measure Group



Select the measure group table



OK



Note:- Now 2 measure group tables are available in the cube.

### MEASURE

- It is the numerical presentation value in fact table.
- It describes a business information.
- May be Simple Value OR Aggregated Value (SUM, AVG, MIN, MAX ETC - - - )

Eg:- Taking SUM(ACTUALCOST) as a measure to the measure group table.

Measure Group table → RC

↓  
New Measure

↓

Usage: SUM

Source Table: Measure Group Table

Source Column: ACTUALCOST

↓  
OK

### ADDING CUBE DIMENSION

1. Add the table(s) in the datasource view.

2. Solution Explorer → Dimensions → RC →

New Dimension → Next → Use an existing table

Next → Data Sourceview: DSV\_Textfiles

Main table: Select the table (xx)

Key Columns

XXID

- Ensure Attribute type : Regular
- ↓
- Next
- ↓
- Finish.

3. Go to Cube Structure → Dimensions → RC →  
 Add Cube dimension → Select the table(s) ↓  
 OK.

### EDIT DIMENSION

i. we edit dimension to manage attributes and to create hierarchies.

(a) Taking all attributes to display in browser & Analysis

Select Dimension (Eg: TIME) → RC → Edit Dimension

↓  
 Select the required Columns (Eg: year, att-  
 in Data Source view, drag and drop in  
 Attributes section

↓  
 SAVE → DEPLOY

GO TO Browser → Reconnect & See all  
 attributes of the tabl

(b) Creating Hierarchies:

→ It is designed to provide topdown and bottom up analysis.



→ While analysing we can drill down for deep dive, we can drill up

for high level information.

→ Hierarchies contain multiple levels and members

→ Each hierarchy should have '2' levels.

Eg:-

Country hierarchy

COUNTRY

  └ STATE

    └ DISTRICT

      └ MAN. DDL

Time hierarchy

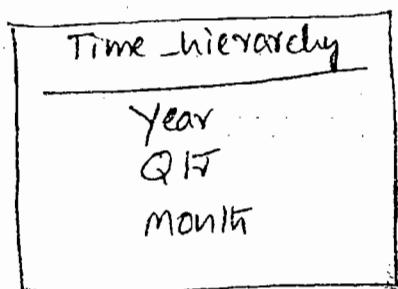
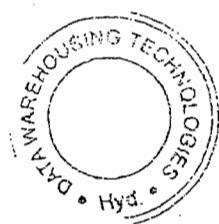
YEAR(2009)

  └ mon(JUN)

    └ 25(04)

Creating time hierarchy:-

TIME → Edit dimension → Drag year, QTR, month one by one to hierarchy section & Rename the hierarchy



SAVE → Deploy.

To See the hierarchy usage:- Go to cube browser →

Take Time hierarchy in browser pane, take locations, product dimension attributes, Actual Cost, estimated measures and see the time hierarchy drilldown

## NEW LINKED OBJECT

- ⇒ This wizard is useful to link measure groups and dimensions in another Analysis Services database or cube to the current database or cube.
- ⇒ Linked objects appear the same to users as other measure groups and dimensions in the cube.
- ⇒ We can also use this wizard to Import KPIs, Calculations and Actions.

Eg:- Importing a calculation from another cube  
(eg:- SumCost) (TextCube2)

New linked object → next →  
Analysis Services data sources → :  
New data sources → next → new →  
Server (or) filename : LOCALHOST  
Initial Catalog : Textcube2

↓

OK

↓

Next :

↓

Inherit

↓

Next

↓

Finish

↓

Next

↓

Select Calculations

↓

Primary



Next



Finish

(18)

Now the Calculations are imported, Deploy and Use.

### WORKING DIMENSION USAGE WIZARD

- This wizard is useful to add, remove dimensions and their relationships.
- We can add dimensions to another dimension (or) measure group table.

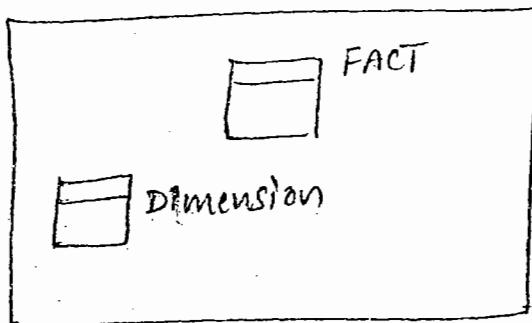
Relationship Types:

1. No Relationship
2. Regular
3. Fact
4. Referenced
5. Many -to- Many
6. Data mining

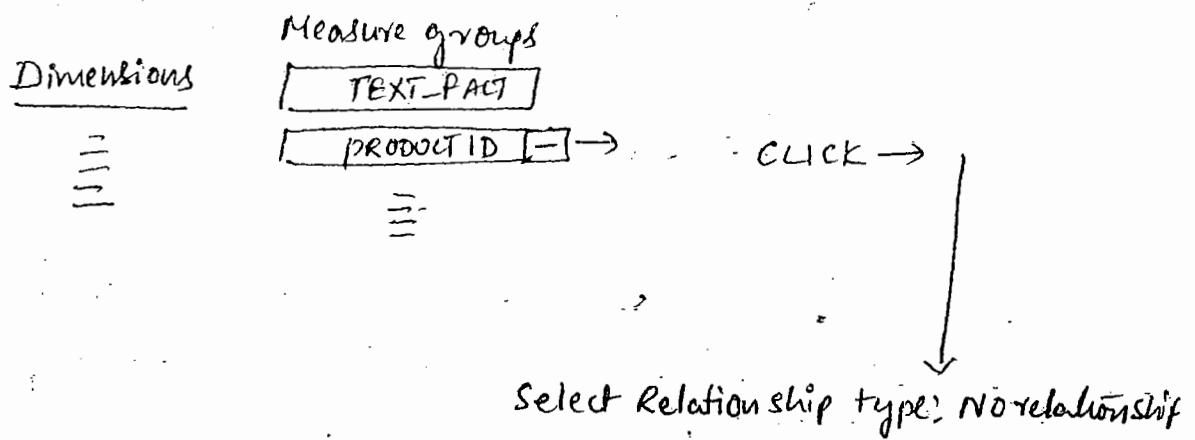
### NO RELATIONSHIP

- The dimension and measure group are not related
- The dimension available in the cube but at the time of analysis it doesn't participate with its values.

pic



Eg:- Removing the relationship between product dimension and measure group table.

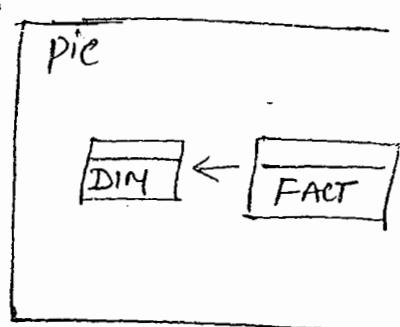


Deploy, GOTO Browser and analyse OK  
Now the product dimension doesn't participate in analysis.

### REGULAR RELATIONSHIP

- Here dimension adds to a measure group table directly.  
(Generally represents Star Schema structure)
- When we are going for this, we should have proper key relationships between Dimension & FACT.

Eg:- Adding product dimension to the measure group table



## Measure Group

(20)

[productID] → click →

Select RelationshipType : Regular

Granularity attribute : PRODUCTID

Dimension table: PRODUCT

Measure Group table: TEXTFACT



Dimension Columns	Measure Group Columns
productID	productID

OK.

Deploy → Go to Browser & analyse, now product table participated in the analysis.

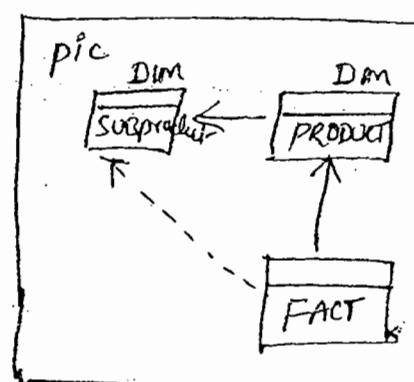
## REFERENCED RELATIONSHIP

→ The dimension table is joined to an intermediate dimension table, which in turn, is joined to the fact table.

→ It provides 'Snowflake'

schema type of structure.

→ We require appropriate column references between subtable to main table.



Eg: - Connecting subproduct dimension to a product dimension

Sub productID → CLICK →

Selected relationship type : Referenced

Reference Dimension : product-

Intermediate :: : RAW/ SUBPRODUCT

Reference Dimension attribute : productID

Intermediate :: : SubproductID

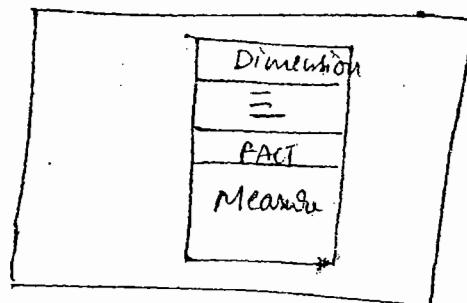
Deploy

OK.

### FACT RELATIONSHIP

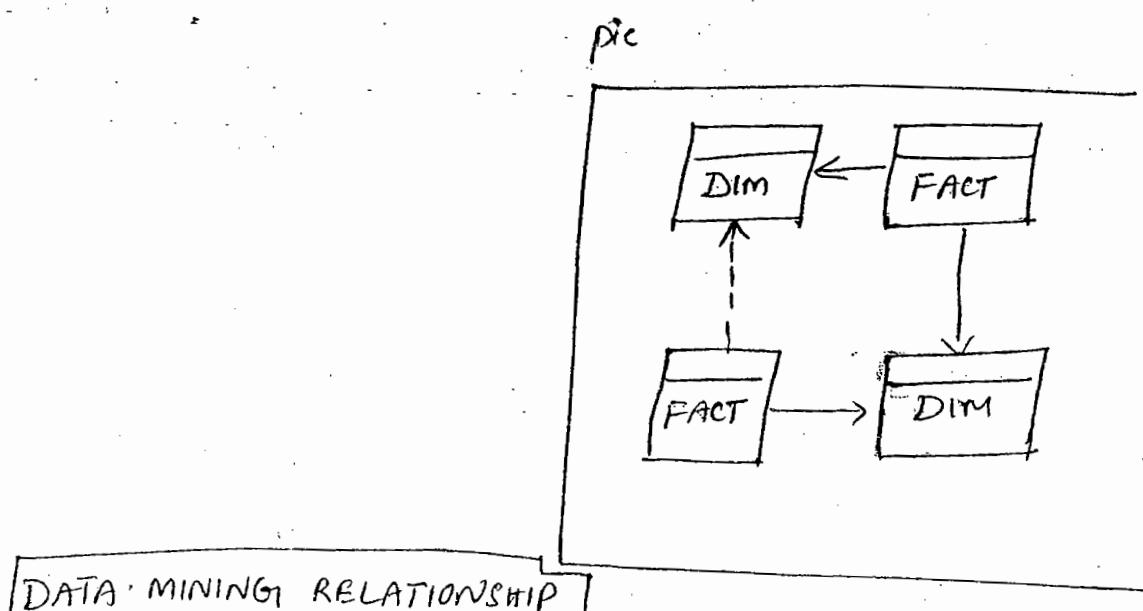
- The dimension is the fact table here.
- Generally textual information represents dimension information and numeric " " fact information in this table
- These types of cubes are called "Standalone Cubes".

pic



## MANY - TO - MANY RELATIONSHIP

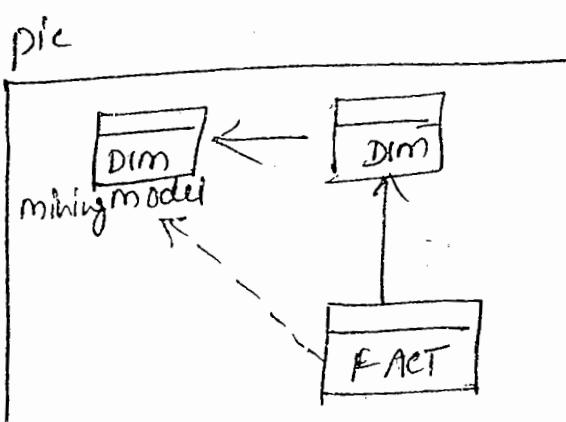
⇒ The dimension table is joined to an intermediate fact table. The intermediate fact table is joined, in turn, to an intermediate dimension table to which the fact table is joined.



## DATA MINING RELATIONSHIP

⇒ The target dimension is based on a mining model built from the source dimension.

The source dimension must also be included in the cube.



## MDX [MULTIDIMENSIONAL EXPRESSIONS]



→ To work with Normal two-dimensional applications, two dimensional Programming languages are enough (C, C++, .Net ... etc).

→ To work with Two-dimensional databases, two-dimensional Query language SQL is enough (Oracle SQL, T-SQL, Teradata SQL - etc)

→ To work with multidimensional applications and multidimensional databases the above specified or not enough. So we go for a separate expression and Query Language MDX

### Important terminology in MDX

a. Member: Dimension attribute is called Member.

→

Syntax:-

[DimensionTable Name]. [Attribute Name]

Ex:- [Product]. [Product Name]

[Location]. [Location Name]

(24)

MEASURE: Fact attribute is called Measure.

Syntax: [Measures]. [Measure Name]

Ex : [Measures]. [Actual cost]

[Measures]. [Estimated cost]

TUPLE:

Collection of Measures or Members

is called Tuple.

a. Starts with (

b. Ends with )

Example: ([Measures]. [Actual cost], [Measures.  
Estimated cost])

SET:

Group of Tuples are called as SET:

a. Starts with {

b. Ends with }

Ex:

{

( [Measures]. [Actual cost], [Measures. Estimated  
cost] ),

( [Measures]. [Actual cost], [Measures. Estimated  
cost] )

}

## MDX USAGE IN SSAS

1. For creating cubes, KPI's, Actions, Calculations, partitions etc ... Objects.
2. MDX Usage in other applications, such as
  - a. Hyperion Fesbase
  - b. SAP Netviewer

## MDX QUERY

1. Generally MDX Queries we write in Analysis Services cube database
2. For Retrieving data from Cube database we use Select statements.

### Syntax:

Select {measures/members} ON columns,  
{measures/members} on Rows  
From <cube name> where < condition >;

## SOME FUNCTIONS IN MDX AND THEIR MEANINGS

There are two types of functions

1. SOME FUNCTIONS TAKES PARAMETERS

Ex: TOP COUNT, BOTTOM COUNT, ISEMPTY etc  
(TOPCOUNT), (BOTTOMCOUNT)

## 8. FUNCTIONS WITHOUT PARAMETERS

Ex:-

- MEMBERS, • ALLMEMBERS, • CHILDREN,
- PREVMEMBER, • CURRENTMEMBER ETC.

### • MEMBERS :-

It displays the child members without including Total.

### • ALLMEMBERS :-

Display Allmembers and their Total.

### • PREVMEMBER :-

Display previous members to the current member.

### • CURRENT MEMBER :-

Display the current cell member value.



## FUNCTIONS WITH PARAMETERS

### ISEMPTY :-

It verifies whether the Member is empty or not.

Syntax :-  $\text{IsEmpty}(\text{set})$

### TOPCOUNT :-

Display Top values.

Syntax :-  $\text{Popcount}(\text{set}, \text{value})$



### BOTTOM COUNT :-

Display Bottom Count of Values.

Syntax :-  $\text{Bottomcount}(\text{set}, \text{value})$

### FILTER :-

It filters the given set based on the condition.

Syntax :-  $\text{Filter}(\text{set}, \text{condition})$

### ORDER :-

It displays the set by keeping a descending or ascending on the given column.

DISTINCT () :- It displays the set values

$\text{DISTINCT}(\{\text{set}\})$

## WORKING WITH HIERARCHIES:

We refer hierarchies member value in two ways.

- [Dimension]. [Hierarchy]. [Members]
- [Dimension]. [Hierarchy]. [Level]. [Members]

NOTE:-

If we do not specify the level, it displays all member values.

## CROSS JOIN :-

CROSS JOIN (Eset<sup>3</sup>, Eset<sup>3</sup>) (OR)

Eset<sup>3</sup> \* Eset<sup>3</sup>

## IMPORTANT MDX QUERIES:-

### NAVIGATION :-

SSMS → Analysis Services → TEXTILES\_CUBE

→ RC → New Query → MDX.



## QUERIES

1. DISPLAY FIRST MEASURE SUM

Syntax :- select from [TEXTILES-CUBE]  
                          ↳ cebename

2. DISPLAY NO. OF ROWS IN THE CUBE

Syntax :- select [Measures]. [TEXTFACTCOUNT]  
on columns from [TEXTILES-CUBE]

3. DISPLAY ALL BRANCHES ACTUAL COST

Syntax :- Select [Measures]. [Actualcost] on  
columns, [Product]. [Brand] on Rows  
from [DSV-textiles-cube]

Actual cost

9800.

4. DISPLAY ALL BRANCHES ACTUAL COST AND  
SUM OF ALL THE BRANCHES

4. DISPLAY BRANCHES AND THEIR ACTUAL COST

Syntax :- Select [Measures]. [Actualcost] on  
columns, [Product]. [Brand]. children  
on rows from [Textile-cube]



5. DISPLAY ALL BRANCH ACTUAL COST AND SUM OF ALL ACTUAL COST.

Syntax

Select [Measures].[Actual cost] on columns,  
[Product].[Brand].ALLMEMBERS on Rows  
from [TEXTILES\_CUBE]

6. DISPLAYING EVERY RAW MATERIAL AND LOCATION, THEIR ACTUAL COSTS

Syntax :-

Select ([Measures].[Actual cost], [Measures].[Estimated cost]) on columns,  
([Rawmaterial].[Rawmaterial ID].children,  
[Location].[LocName].children) on Rows  
from [TEXTILES\_CUBE]

7. DISPLAY THE FIRST RAW MATERIAL ACTUAL COST

Syntax

Select [Measures].[ACTUAL COST] on  
columns, [Rawmaterial].[Rawmaterial ID]  
Firstchild on rows from [TEXTILE\_CUBE]

## 8. DISPLAYING TOP TWO VALUES OF THE LOCATION

Syntax :-

Select [Measures]. [Actual Cost] on columns,  
TopCount ([Location]. [Locname].  
Children, 2) on rows from [TEXTILE-CUBE]

## 9. DISPLAYING BOTTOM TWO LOCATION VALUES

Syntax :-

Select [Measures]. [ActualCost] on columns,  
BottomCount ([Location]. [Locname]. Children, 2)  
On ~~columns~~ rows from [TEXTILE-CUBE]

## 10. DISPLAY THE LOCATION TOOLS ACTUAL COST

is > 1000

Syntax

Select [Measures]. [ActualCost] on columns,  
filter ([Location]. [Locname]. Children,  
[Measures]. [ActualCost] > 1000) on rows  
from [TEXTILE-CUBE]



## 11. DISPLAY THE DATA IN ACTUAL COST SORTED ORDER IN ASC

Syntax :- Select [Measures]. [ActualCost] on columns,  
order ([Location]. [Locname]. Children,  
[Measures]. [ActualCost]), Asc on rows  
from [TEXTILE-CUBE]

(3)

12. DISPLAY THE CROSS PRODUCT OF LOCATION

PRODUCT AS WELL AS THEIR ACTUAL, ESTIMATED  
COST.

Syntax

Select { [measures]. [Actual cost],  
[measures]. [Estimated cost] } ON  
columns, crossjoin ([location]. [Locname].  
children, [product]. [product name].  
children) on rows from [TEXTILE-CUBE]  
children )

(OR)

Select { [measures]. [Actual cost]. [measure].  
[Estimated cost] } on columns,  
([location]. [Locname]. children) \*  
([product]. [product name]. children)  
on rows from [TEXTILE-CUBE]

13. DISPLAY DISTINCT FROM DISTINCTWISE  
VALUES AND LOCATION WISE.

Syntax

Select { [measures]. [Actual cost], [measures].  
[Estimated cost] } on columns, DISTINCT  
([LOCATION]. [LOCATIONNAME]. children, [product].  
[product name]. children) on rows from [TEXTILE-CUBE]

14. DISPLAY 2009 YEAR ACTUAL COST SCEN

Syntax:-

Select [Measures].[Actual cost] on columns  
from [Textile\_cube] where [Time].[Year].  
& [2009]

(OR)

Select [Measures].[Actual cost] on columns,  
[Location].[Locname].children on  
rows from [DSV\_Textiles\_cube] where  
& [Time].[Year].&[2009]



## CONDITIONAL EXPRESSIONS

If:-

IF (< condition >, success stmt, failed stmt )

Eg : IF ([Measures].[Actual cost] - [Measures].[Estimated cost] > 0, 1, 0)

CASE:-

Evaluates against Multiple conditions.

CASE

When < condition 1 > then < statement 1 >

" < 2 > " < " 2 >

" < 3 > " < " 3 >

" < 4 > " < " 4 >

Else < Statement 5 >

END

Ex:-

case when [Measures].[Actual cost] - [Measures]

[Estimated cost] > 0 Then -1 / where [Measures]

[Actual cost] - [Measures].[Estimated cost] < 0

Then 1

When [Measures].[Actual cost] - [Measures]

[Estimated cost] = 0 Then 0

END



## CALCULATIONS :-

These are the intermediate Operations performed between data source to cube database.

→ These are Created at BIDS Level.

→ It uses MDX syntax commands to perform operations.

→ It supports different operators, set functions, methods, members ... etc in its calculations.

We write calculations in 'two' ways

1. FORM VIEW

2. SCRIPT VIEW.



### FORM VIEW :-

Here one by one calculation.. Prepared manually.

Ex:- Creating a calculated member, finding the difference between Actual & Estimated costs.

### Navigation

Calculation tab → New calculation Member →

Name: DiffCost

parent hierarchy: Measures

Expression : [Measures].[Actual cost] - [Measures]  
 [Estimated cost]

FORMAT string : standard

Visible : True

Associated Measure Group : (undefined)

Color Expression :

For color : 10711080 /\* Blue \*/

Back color : 12682256 /\* Silver \*/

Click OK → Save.

Build → Deploy, Auto Browser, select  
 Actual cost, estimated cost, DIFFcost and  
 See the result.

### SCRIPT VIEW :-

We write Scripting for creating multiple calculations one by one, if we are having hands on experience in writing coding.

In script view it is easy to move the calculations from one environment to other environment

(Dev → Test, Test → production)



Syntax :-

CALCULATE ;

CREATE MEMBER < MEMBERNAME > AS

< EXPRESSION >;

< SETTINGS >

Eg:-

CALCULATE ;

CREATE MEMBER CURRENT\_CUBE.[MEASURES]

SUM CUST AS

[MEASURES].[FACTUAL COST] + [MEASURES].

[Estimated cost],

FORMAT-STRING = "STANDARD ,

BACK-COLOR = 123456 /\* SILVER \*/ ,

FORE-COLOR = 432345 /\* BLUE \*/ ,

VISIBLE = 1 ;

CALCULATED MEASURE :-

If the Parent Hierarchy

is measures, then the calculated members

is called "calculated Measure"



## \*\* NON-EMPTY BEHAVIOR :-

It is used to resolve the NonEmpty Queries in MDX while calculations.

- If the Non-Empty behavior Property is blank the calculation must be evaluated repeatedly to determine a member is empty.
- The Non-Empty behavior Property contains the name of the measure the calculation is empty if measure is empty.
- Generally we go for Ratio calculations.

Eg :-

There is a calculation ACIEC, If 'EC' is empty it throws an error. In this situation if we set up non-Empty behavior Property to 'EC' then system avoids calculations and displays the calculated results are empty if the EC is empty.



## Additional Expressions :-

1. Display Actual/cost as '99999', if its Empty.

Syntax : If ([Measures]. [ActualCost]),  
99999, [Measures]. [ActualCost])

2. Display in Time Hierarchy Each member and its participation in the entire actual cost sales.

Syntax

( [Measures]. [ActualCost], [Time]. [Time-Hierar].  
& [ ]) / ([MEASURES]. [ACTUALCOST],  
[TIME]. [TIME-Hierarchy]. & [ ])

3. Display Actual cost of the Member to its Parent.

Syntax

( [Measures]. [ActualCost],  
[TIME]. [TIME - HIER]. & [ ]) /  
([MEASURES]. [ACTUAL COST]  
[TIME]. [TIME - HIER]. PARENT)



## ACTIONS

(41)

These are performed at the time of Events.

→ Generally, the Event is "clicking" the cell.

→ There are '3' Types of Actions

- a. Statement (or) URL (or) Record set Action
- b. Drill through Actions
- c. Report Action.

a. URL ACTION :-

Calling the URL, while analysing the data is called URL action.

Eg:- Going to the locations website, when we click location attribute value.

Navigation :-

Action → New Action

Name : URL - Action

Action Target

Target Type : Attribute numbers

Target Object : [location]. [Locname]

Action Context

Type : URL

Action Expression : "http://www.alllocation.com"



→ Event is clicking 'Location', Action is Opening the URL.

Build → Deploy → Go to cube browser, highlight Location → RC → click Action (URL-Action)

b. REPORT ACTION :-

→ Calling a report while Analyzing the data is called Report Action.

→ It requires report Servername, Report name.  
(servername and their report names)  
Eg:- COINOS, BO, HYPERION...etc.

Navigation:-

Action → New Action

Name: Report-Action

Action Target

Target type: Attribute members

Target object: Location . Locname

Report Server

Server Name: RAWAN : 8080

Report path:

Report-  
Server-DW / ---

Report format: HTML 5

(44)

Build → Deploy → Go to cube browser →

RC → Location → Click Report Action.

Note:- It takes parameters to display the Parameterized report content.

### c. DRILL THROUGH ACTION :-

- Drilling to other columns (or) for opening a separate Analytical window for the required columns, we must go for drill through action.
- Simply when we are moving from one Analytical to another analytical window for required details, we go for drill through Action.

Eg:- Drilling through locations, Time, Raw materials, Dimensions while Browsing the Actual cost.



Navigation :-

2. (70)

Action → New drill through Action

Name : Drill through - Action

Action Target

Measure group members ; Text - fact

### Drill through columns

Dimensions

Return columns

Measures

Actual cost

Location

Loc Name

Time

Year

Raw material

Raw material Function

Build → Deploy → go to Browser → Reconnect →

RC → Actual cost → click Drillthrough  
Action.



## PERSPECTIVE

→ It is helpful to provide limited visibility to cube objects.

→ In this case, what are all the dimensions, fact, KPI's, calculations.. etc.

You want to provide visibility to the user take in perspectives.

→ Now, the user see only the selected dimension, facts, KPI's, calculations.



### Navigation:

PERSPECTIVE → NEW PERSPECTIVE →

CHOOSE PERSPECTIVE NAME, CHECK (OR).

UNCHECK the required objects and items.

E.g.  TEXTILE FACT  TIME  
 PRODUCT  LOCATIONS

Build → Deploy → go to cube Browser →  
change the perspective and see the  
Limited items (or) objects.

Note:

To provide 8 fact tables and only 5 dimensions visibility to a particular user in the cube of the five fact tables and 50 dimensions perspectives are useful.



TRANSLATION &

This is helpful to provide multilanguage support.

→ In 2000 the translation and multilanguage support provided at web page level.

But from 2005 onwards this support is at BIDS Level

(48)

### Navigation:

TRANSLATION → NEW TRANSLATION →

SELECT LANGUAGE TELUGU (INDIA) → CLICK OK

Eg:-

OBJECT	DESCRIPTION
TEXT FACT	VILUVALA PATTIKA

A CTUAL COST	A SALU DHARA
--------------	--------------

ESTIMATED COST	COCHIN CHINA DHARA ETC.
----------------	----------------------------

Build → Deploy → Go to Browser → And

Change the language to select Telugu;

See the result.



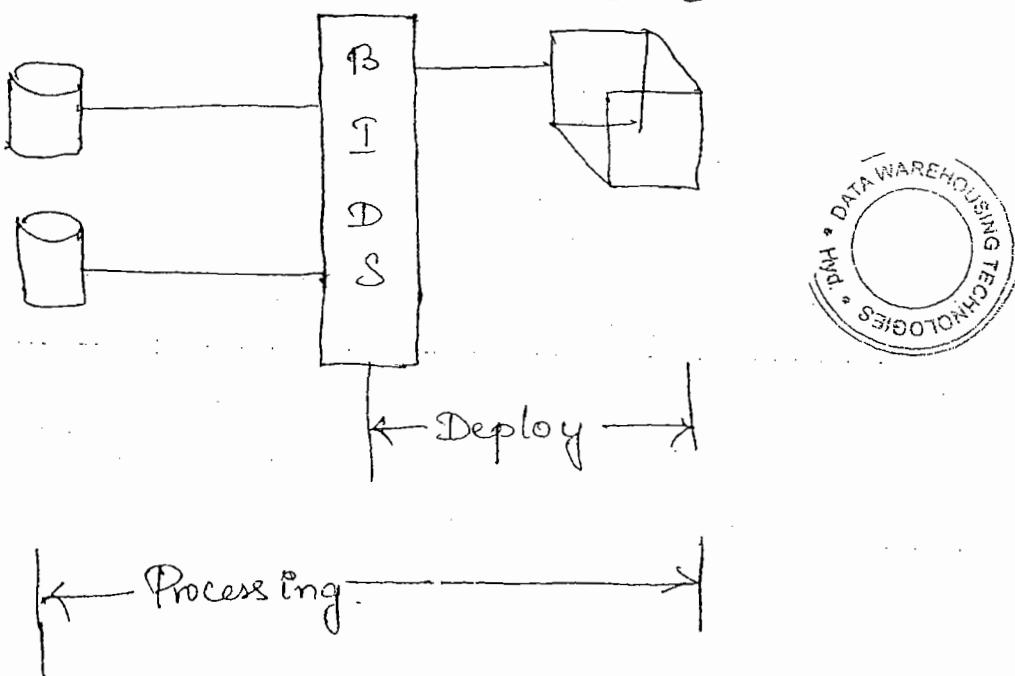
Build → Deploy :- If the cube structure changes in BIDS level, to effect the same in cube database level.

Build → Process (processing data) :-

The data (or) structure in the data sources changed, to effect the same in cube database.

Data source.

Cube DB.



(50)

## KPI (KEY PERFORMANCE INDICATOR)

- These are the important business information representors.
- Generally we go for KPI, when ever we want to represent the high-level and important business information in graphical items.

Eg: of graphical items :-

GAUGE

REVERSED GAUGE

THERMOMETER

CYLINDER

FRACES

VARIENCE ARROWS etc .....



Standard Arrows.

↓ DOWN

→ LINEAR

↑ UP

(51)

## TRAFFIC SIGNAL NUMERIC

RED → -1

YELLOW → 0

GREEN → 1

Eg :- Display in Red color, If the Actual cost  
is greater than ( $>$ ) Estimated cost.

Display in Green color, If Actual cost is  
less than ( $<$ ) Estimated cost,

Display in Yellow color, if both are Equal  
in a graphical item traffic signal.

### Navigation:

KPI → New KPI →

NAME: Sales-KPI

ASSOCIATED MEASURE GROUP: TextFACT

VALUE EXPRESSION: [MEASURES].[ACTUAL COST]

GEOAL EXPRESSION: [MEASURES].[ESTIMATED  
COST]



(5L)

STATUS

STATUS INDICATOR : TRAFFIC LIGHT

STATUS EXPRESSION :

CASE

WHEN [MEASURES].[ACTUAL COST] - [MEASURES].

[Estimated cost] > 0 then -1

WHEN [MEASURES].[ACTUAL COST] - [MEASURES].

[Estimated cost] < 0 then 1

WHEN [MEASURES].[ACTUAL COST] - [MEASURES].

[Estimated cost] = 0 then 0

END.



TREND INDICATOR : Standard Arrow

TREND EXPRESSION : < same as status

expression >

Build → Deploy

Go to 'KPI BROWSER VIEW' and then

see the result.

## PARTITIONS

### Realtime Usage :-

Multiple partitions will process the data more parallelly.

→ We can run the specified partition to process the required data (so, that limited system resources will be utilized).

While creating Partitions we go for 'data binding' between partitions to the table data.

→ There are '2' types of bindings available

#### a. Table Binding.

Here the fact table measure group directly bound to the partition.

#### b. Query Binding.

Here based on the query the Partition is created.



Note :-

In Realtime we use Query binding than table binding.

### Storage Settings :

While Creating the Partitions, we must specify either of the below storage settings.



#### a) ROLAP (Relational OLAP)

In this case data and Aggregations store under relational sources.

#### b) MOLAP (Multidimensional OLAP)

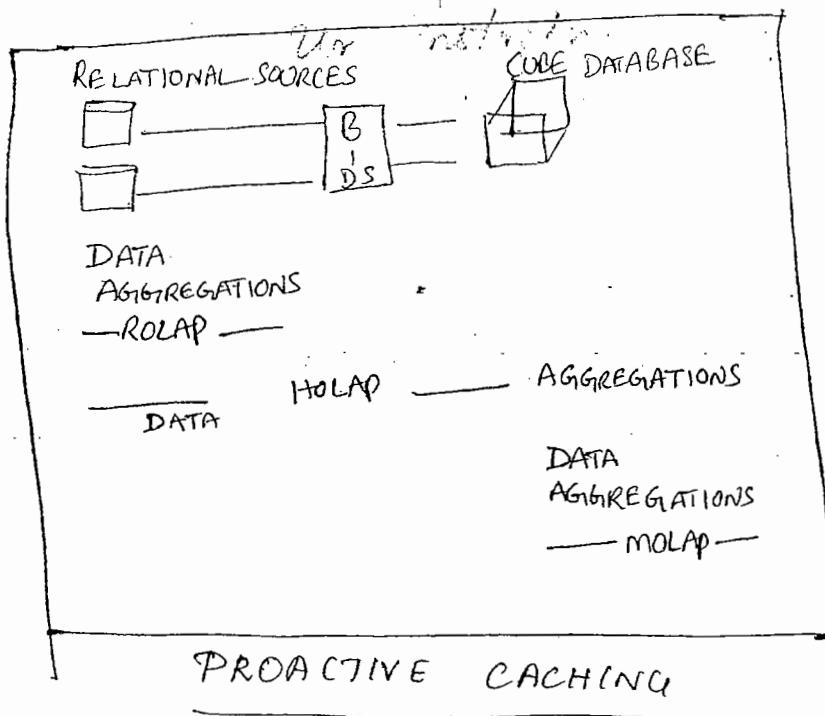
Here data and aggregations store under multidimensional sources.

#### c) HOLAP (Hybrid OLAP)

Here data stores under Relational sources and aggregations store under Multidimensional sources.

(55)

Star diagram of the flow plan



This feature helps the cube in sink with the Relational data sources

→ It takes the latency time, schedule time and event table to capture the changes from Relational data sources to cube databases.

(3)

## CREATING TABLE BINDING PARTITION:

By default a table is binded to the partition. In this case, the table is FACT TABLE. Along with this FACT TABLE Binding if you want to add more FACT TABLE bindings (the cube is having many fact tables). We go for the below process.

### Navigations:

Partition → New partition → click Next.

Measure group : TEXT FACT

Available Tables :  TEXT\_FACT →

CLICK NEXT → Specify the Query to restrict rows and remove where condition →

Next → processing location

CURRENT SERVER INSTANCE

Storage Location → click next → Name: Table-text\_fact\_partition → click Next finish.

Build → deploy.



## WORKING WITH QUERY BINDING:-

(1)

- This mechanism we use frequently in real time.
  - Generally we create partitions based on its frequency of data processing and its columns.
- Assume we are processing data into fact table based on 'OK' and 'NOK' flags, then create partitions on those columns.

### ① Creating 'OK' partition

1. Delete the existing table partition

2. partitions → new partition click →

Measure group : TextFACT

Available tables

TextFact

Specify a query to restrict rows

SELECT

FROM [dbo].[TextFACT]

Where [dbo].[TextFACT].[IKU] = 'OK'

Next →

② Current Server instance → Next

Name : OK\_FACT

Aggregation option

③ Design aggregations table

Finish



② Like above process create NOK partition with below change:

SELECT

From [dbo].[Text-Fact]

Where [dbo].[Text-fact].[IKO] = 'NOK'

### Processing partitions

There are two ways

a) Build Menu → process

b) partitions → select partition / partitions → RC

### Processing FACT TABLE

↓  
process

If the data (or) structure in fact table changed to effect the same in cube database level we go for fact processing.

Navigation:-

partitions tab → Select partition → RC → process

### Fact processing Options:-

- a) process default
- b) " full
- c) " Data
- d) " Incremental
- e) " Index
- f) Unprocess.



## DIMENSION PROCESSING

If dimension table structure or data changes in data sources to effect the same in cube database we go for dimension processing.

Navigation:-

VIEW → SOLUTION EXPLORER → DIMENSIONS

↓  
Select Dimension (Ex-  
TIME)

↓  
RC  
↓

PROCESS.

processing options:-

- a) process default
- b) " full
- c) " Data
- d) " Index
- e) " Update
- f) Unprocess.



## PROCESSING Options for OLAP Objects:-

The objects that you can process in SSAS are database, cube, measuregroup, partition, dimension, mining structure, and mining model.

Among these objects, only dimensions, partitions, and mining structures store data.

⇒ When you process an object, the server creates a processing plan.

PROCESSING OPTION	DESCRIPTION	APPLIES TO
process Default	performs the minimum number of tasks required to fully initialize the object. The server converts this option to one of the other options based on the object state.	ALL OBJECTS
process Full	Drops the object stores and rebuilds the object's metadata changes, such as adding a new attribute to a dimension, require process full.	ALL OBJECTS
process Update	Applies member inserts, deletes and updates without invalidating the affected cubes	Dimension
process Add	Adds only new data	Dimension partition
process Data	Loads the object with data without building indexes and aggregations	Dimension, Cube, Measure Group / partition
process Index	Retains data and builds only indexes and aggregations	"

PROCESS OPTION	DESCRIPTION	APPLIES TO
Unprocess	Deletes the object data or the data in the Containing objects.	ALL Objects
Process Structure	Deletes the partition data and applies process Default to the Cube Dimensions.	Cube
Process Incremental	Incremental data it adds	partition, Cube

Note- 'process data Add' is not available in dimension and (fact-) partition processing options.

### PROACTIVE CACHE PRACTICAL IMPLEMENTATION

1. Create a table binding partition
2. Go to storage settings  
Specify
  - ① Standard setting
    - Automatic MOLAP



## ① Custom Setting

(64)

Options



CLIK



Storage mode: MOLAP

Enable procedure caching

General

Cache Settings

Update the Cache when data changes

Silence interval 20 seconds

Silence Override interval 20 seconds

Notifications

Specify tracking table

[dbo].[Text\_fact]

Build → Deploy  
click OK



Observations:-

- 
- Go to browser → take some fields and see grand total.
  - add some rows in source "Text\_fact" and after 20 seconds if you see cube automatically processed and grand total changes.

Note:- No manual intervention.

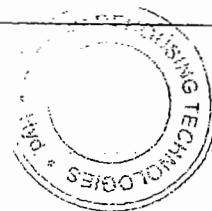
## UNDERSTANDING STORAGE MODES

MODE	QUERYTIME	LATENCY	PROCESSINGTIME	STORAGESIZE
MOLAP	Fast	High	Fast	Medium
ROLAP	Slow	Low	Slow	Large
HOLAP	Medium	Medium	Fast	Small

## STANDARD PARTITION STORAGE SETTINGS

Mode	QUERY TIME
REAL-time ROLAP	As with standard ROLAP, partition data and aggregations are stored in the relational database. The server maintains an internal cache to improve query performance. When the change notification is received, the server drops the ROLAP cache to ensure that data is not out of sync with the data in the relational database.
Realtime HOLAP	As with standard HOLAP, partition data is stored in the relational database, and aggregations are stored in the cube. Aggregations are rebuilt as soon as a data change notification is received.
Low-latency MOLAP	The MOLAP Cache expires in 30 minutes.
medium-latency MOLAP	Similar to Automatic MOLAP except that the MOLAP Cache expires in 4 hours.

Mode	Querytime
Automatic MOLAP	The default silence interval is set to 10 seconds. As a result the server will not read if the data change batches are fewer than 10 seconds apart. If there is not a period of silence, the server will start processing the cache in 10 minutes.
Schedule MOLAP	Same as MOLAP - except that the server will process the partition on a daily schedule.
MOLAP	The partition storage mode is standard MOLAP. proactive caching is disabled. You need to process the partition to refresh the data



## PROACTIVE CACHING

(65)

### UNDERSTANDING PROACTIVE CACHING

As noted, with MOLAP and HOLAP storage modes, SSAS caches data (MOLAP storage mode only) and aggregations (both MOLAP and HOLAP) on the server.

When you take a data "snapshot" by Processing a cube, the data becomes outdated until you process the cube again. The amount of OLAP data latency that is acceptable will depend on your business requirements.

In some cases, your end users might require up-to-date or even real time information.

A New Feature of SSAS 2005, PROACTIVE CACHING, can help you solve data latency problems.



(6)

## TIPS FOR EXAMS

PROACTIVE CACHING IS ESPECIALLY USEFUL WHEN THE RELATIONAL DATABASE IS TRANSACTION ORIENTED AND DATA CHANGES AT RANDOM.

WHEN DATA CHANGES ARE PREDICTABLE —

Step 1. SUCH AS WHEN YOU USE AN EXTRACT, TRANSFORM, AND LOAD (ETL) PROCESS TO LOAD DATA

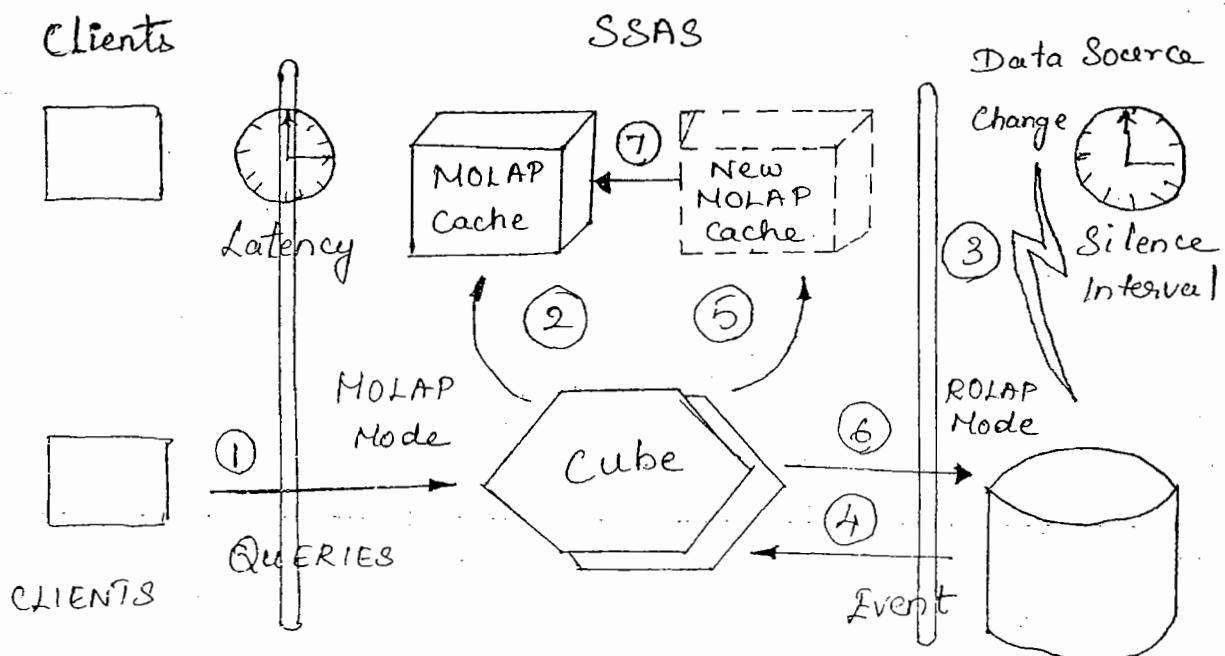
2. CONSIDER PROCESSING THE CUBE EXPLICITLY.

WHEN THE DATA SOURCE IS TRANSACTION ORIENTED AND YOU WANT MINIMUM LATENCY CONSIDER CONFIGURING THE CUBE TO PROCESS AUTOMATICALLY BY USING PROACTIVE CACHING.



## HOW PROACTIVE CACHING WORKS

When you enable PROACTIVE CACHING, the server can listen for Data change notifications and can update dimensions and measures dynamically in an "AUTOPilot" mode.



## STEADY STATE

In steady mode, no changes are happening to the Relational data.

Step 1 :-

Client Applications submit Multidimensional Expressions (MDX) queries to the cube, pls check in diagram before page.

Step 2 :-

The cube satisfies the Queries from a MOLAP Cache. The Server listens for a data change notification event, which could be one of the following three types.

■ SQL SERVER : This option uses the likes of Server trace events that the relational Engine raises when data is changed (SQL Server 2000 and later).

■ CLIENT INITIATED : In this case, a client application notifies SSAS when it changes data by sending a NotifyTableChange XML for Analysis (XMLA) command.

- I SCHEDULED POLLING : With this Option, the Server periodically Polls the required tables for changes.

## UNSTEADY STATE

### STEP 3 :-

At some point, a data change occurs in the data source as shown in figure.

### STEP 4 :-

This change triggers a notification Event to SSAS server.



### STEP 5 :-

The Server Starts two stop watches.

The Silence Interval stopwatch measures the time elapsed between two consecutive data change events. This will reduce the number of false starts for building the new cache Until the database is quiet again.

(70)

### FOR EXAMPLE

If data changes are occurring in batches, you do not want to start rebuilding the cache with each data change event.

Instead, you can optimise proactive caching by defining a silence interval that allows a predefined amount of time for the batch changes to complete.

After data in the relational database is changed, the server knows that the MOLAP cache is out of date and starts building a new version of the cache.

### STEP 6:

The latency stopwatch specifies the maximum latency period of the MOLAP cache, the administrator can also predefine the maximum latency period.

During the latency period, queries are still answered by the old MOLAP cache.



When the latency period is exceeded, the server discards the old cache.

While the new version of the MOLAP cache is being built, the server satisfies client queries from the ROLAP database.

#### STEP 7:-

When the New MOLAP Cache is ready, the server activates it and redirects client queries to it.

PROACTIVE CACHING enters a steady state again until the next data change event takes place.



## NAMED QUERY

(75)

→ These queries are helpful to construct a table with customized query. Here advantages are

a) restriction of rows and columns

b) multiple columns from multiple tables

→ The name of the query result acts like a table.

→ This option available in data source view.

Navigation:-

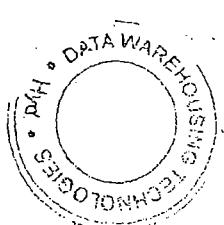
GOTO DATASOURCEVIEW

↓  
RC  
New  
named query

name:- EMP-Dept

Data Source: DS-textiles (primary)

Query : SELECT E.EID, E.ENAME, E.DID,  
           D.DID, D.DNAME  
         FROM EMP E  
       INNER JOIN DEPT D ON E.DID=D.DID



OK

Now Connect this resulted table to the cube.

## NAMED CALCULATION

→ Like named queries, named calculations will also improve performance [because both run at source level]

→ It is recommended in real time that if we are using a calculated value as permanent ... as few named calculations

→ Instead of doing calculations at cube level, if the calculations are at source level, always better performance the query gives.

Eg:- Creating ActualCost increment field, by increasing 12% of actualCost

Navigation:- Go to dataSourceview

↓  
Select a table (Eg:- Text-Fact)

↓  
RC → new named Calculation

Column name: ACTUALCOST\_INC

Expression: ActualCost + ActualCost \* 12 / 100

OK

### IMPLEMENTING SCENARIO DIMENSION

- This is designed to implement "WriteBack" option.
- Here the data written back to the databases.
- Scenario dimension performs many operations, one among the operation is "WRITEBACK".

#### Steps:-

1. Create a table in data source (Eg:- PARTY)

2. Add the table in dataSourceview

3. Dimensions → New dimension →  
    ○ Use an existing table

Next

Data Source View : Textiles1\_DB

MainTable: PARTY

Key Columns: PARTYID, PARTNAME

...PARTLOC → NEXT →



Attribute Type

└ Other

└ Scenario

└ Scenario

Name: PARTY-SC → **[FINISH]**

Dimension Structure

Attributes

└ PARTY-SC → RE → properties → writeEnabled : true

SAVE, BUILD → DEPLOY

GoTo Dimension browser

↓  
ReConnect

↓  
click writeback  option

↓  
GoTo DATA ROW

↓  
RE → CREATE SIBLING

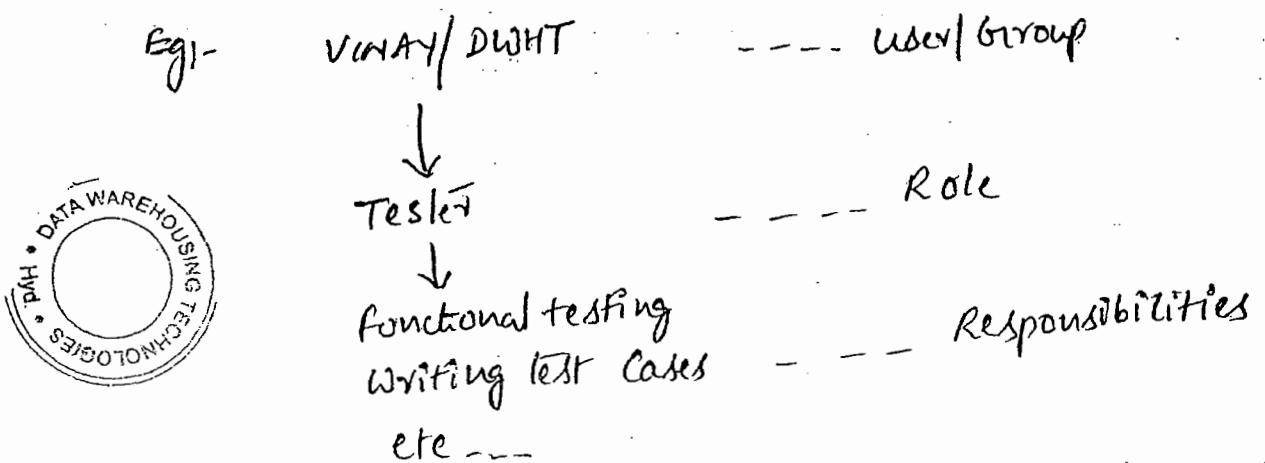
Add row data like below  
PARTYID PARTYNAMES PARTYLLOC  
4 MADHU HYD

Once you change your mouse location to another row, the row saves in the party table.



## APPLYING SECURITY

- SSAS BIDS level it supports 'role based' security.
- User (or) group assigns to a Role, each role contains set of responsibilities. Now the user/group satisfies the responsibilities (objects privileges - Access, Read, Write)



→ Eg:- Creating two roles (Developer-SSAS, Tester-SSAS) with different objects and different privileges and observing the result.

1. Solution Explorer → Roles → RC → New Role →

General

Role name: Developer-Role\_SSAS

Role Description:- Developers have full privileges

Full Control (Administrator)

Membership

[Add] → Click →

Enter the object names to select-

VINAY → Check Names → OK

Note:- IN SS2005, even we Select full Control, we can edit objects and its security. That means the tabs (data sources, cubes — ) enabled for operations.

IN 2008, full Control option disables remaining tabs.

- 2. Role → New Role →

### General

Role name: Tester - Role - SSAS

Role Description: No privileges  
to tester (or)

process database limited privileges

### membership

Add → madhu → checknames →

### DataSources



DS\_Textiles1 Read

DS\_Textiles2 None

### Cubes

Textiles\_Cube Read Drillthrough

### ETC --

3. DEPLOY, GOTO CUBE BROWSER, RECOMMEND

4. CLICK CHANGE USER  icon → ROLES → SELECT

DEVELOPER - ROLE - SSAS & SEE THE RESULT

5. ROLES → SELECT → TESTER - ROLE - SSAS and observe

O  
O  
C  
C  
A  
C  
C  
B  
G  
C

## VARIOUS WAYS OF DEPLOYMENT

There are many to deploy cube objects or cubes.

- a) BIDS (BUSINESS INTELLIGENCE DEVELOPMENT STUDIO)
- b) Deployment wizard
- c) XMLA Script
- d) Synchronize Database wizard
- e) Backup and Restore
- f) Analysis Management Objects (AMO)



OPTION	Recommended Use
BIDS	Deploying the latest changes to your local server for testing
Deployment Wizard	Deploying to a Test (or) production environment when you need more granular control
XMLA Script	Scheduling a deployment task
Synchronize Database Wizard	Synchronizing two cubes, such as a Staging Cube and a production cube.
Backup and Restore	Moving a cube from one server to another
Analysis Management Objects (AMO)	Handling deployment programmatically.

a) BIDS:

SOLUTION EXPLORER → Project (Text-File) →

RC → properties →

Build

Output path :- bin\

Deployment

Options

Processing Options: Default

Transactional Deployment: False

Batch and multi-dimensional analysis



Target

Server: <Servername to be deployed>

Database: Text-Fact-Cube-DB

OK

## ⑥ Deployment Wizard:-

Start → programs → Microsoft SQL Server 2008 →

Analysis Services → Deployment Wizard

↓

Next

↓

Database files C:\ VINITAYAKA\Textfiles\bin

Textfiles.asdatabase

↓

Next

↓

Server: <Servername to be deployed>

Database: Textfiles

↓

Next

↓

Partitions

① Deploy partitions.

Roles and members

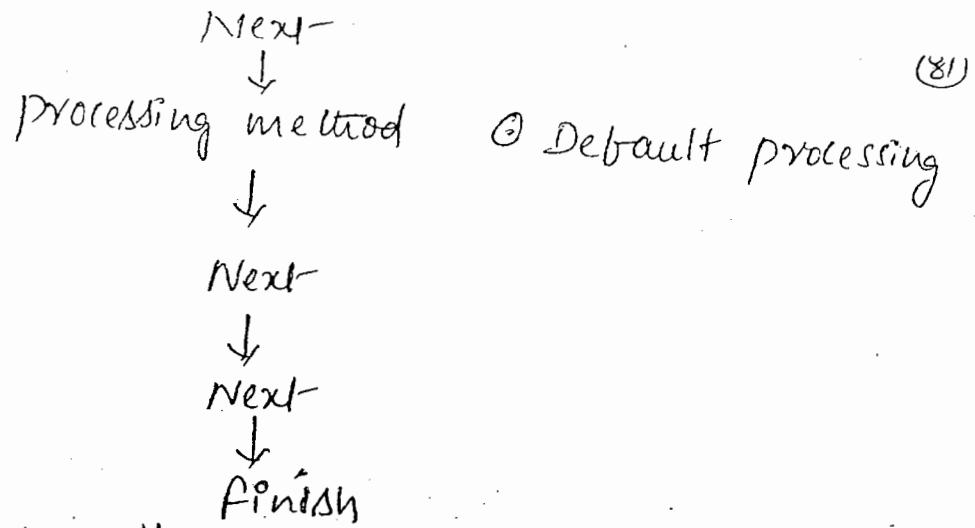
② Deploy roles and retain members

↓

Next

Retain Configuration Settings for existing objects





Now the "analysis Service database" deploys in the specified in the Server.

Note:- This wizard deploys (asdatabase) analysis Services database file, as created in the Output-directory by project build.

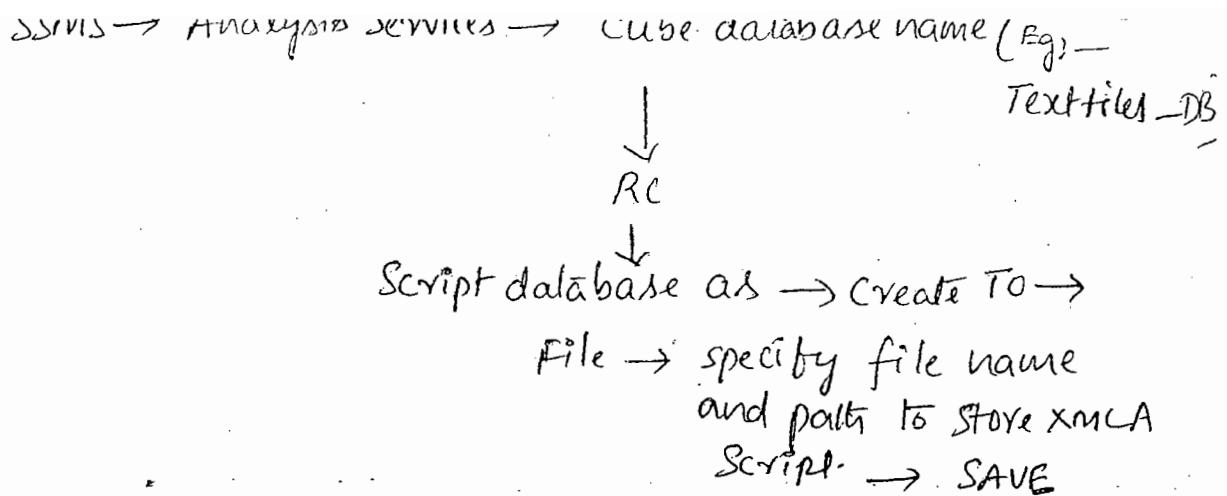
### (c) USING XMLA SCRIPT:-

- XMLA refers to "XML for Analysis Services".
- This is the frequently used mechanism in real time.
- It Contains Cubename, Servername, dimension, fact tables and their attribute names etc-- objects information.

This information is useful to modify/add current/ new settings.

- ① NAVIGATION :- (Creating an XMLA Script file.)





GOTO THE FILE & MONITOR THE DETAILS.

## ② Using XMLA Script:-

- a. GOTO XMLA Script file
- b. Do the changes whatever required, save file.
- c. SSMS → Open → File → specify XMLA Script file path → Open
- d. Execute the Script
- e. Observe the cube database on the specified Server according to the settings.

## ③ BACKUP and RESTORE:-

- a) BACKUP:- SSMS → ANALYSIS Services → cube database → RC →  
 Backup → Backup file: 9AM-news.abf [Browse]  
 Password: VINAY  
 Confirm #: VINAY → OK

- b) RESTORE:- Analysis Services → Databases → RC → Restore →  
 Backup file: Browse to 9AM\_news.abf  
 Restore Database: VINAY\_9AM\_NEWS  
 PASSWORD: VINAY



## SSAS PERFORMANCE TUNING

### 1. OPTIMIZE CUBE AND MEASURE GROUP DESIGN

- Define cascading attribute relationships  
for example, Day > Month > Quarter > Year  
and define user hierarchies of related attributes  
(called natural hierarchies) within each dimension as appropriate for your data.
- Remove redundant relationships between attributes to assist the query execution engine in generating the appropriate query plan. Attributes need to have either a direct or an indirect relationship to key attribute, not both.
- Keep cube space as small as possible by only including measure groups that are needed.
- Place measures that are queried together in the same measure group. A query that retrieves measures from multiple measure groups requires multiple storage engine operations.



- e. Minimize the use of large parent-child hierarchies.

In parent child hierarchies, aggregations are created only for the key attribute and the top attribute.

(for example, the All attribute)

Unless it is disabled.

- f. Optimize many-to-many dimension performance, if used. When you query the data measure group by the many dimension, a run-time join is performed between the data measure group and the intermediate measure group using the granularity attributes of each dimension that the measure groups have in common.

## 2. DEFINE EFFECTIVE AGGREGATIONS

- a. Define aggregations to reduce the number of records that the storage engine needs to scan from disk to satisfy a query.



(85)

- b. Avoid designing an excessive number of aggregations.

Excessive aggregations reduce processing performance and may reduce query performance.

- c. Enable the Analysis Services query log to capture user query patterns and use this query log when designing aggregations.

## 2. USE PARTITIONS

- a. Define partitions to enable Analysis Services to query less data to resolve a query when it cannot be resolved from the data cache or from aggregations. Also define the partitions to increase parallelism when resolving queries.
- b. For optimum performance, partition data in a manner that matches common queries. A very common choice for partitions is to select an element of time such as day, month, quarter, year or some combination of time elements.



(86)

- c. In most cases, partitions should contain fewer than 20 million record size and each measure group should contain fewer than 2,000 total partitions. Also, avoid defining partitions containing fewer than two million records.  
Too many partitions causes a slowdown in metadata operations, and too few partitions can result in missed opportunities for parallelism.
- d. Define a separate ROLAP partition for Real-time ROLAP partition in its own measure group.

#### 4. WRITE EFFICIENT MDX



- a. Remove empty tuples from your result set to reduce the time spent by the query execution engine serializing the result set.

b. Avoid Run-time checks in an MDX calculation that result in a slow execution path.

If you use the Case and If functions to perform condition checks which must be resolved many times during query resolution, you will have a slow execution path.

Rewrite these queries using the SCOPE function to quickly reduce the calculation space to which an MDX calculation refers.

c. Use Non-Empty-Behavior where possible to enable the query execution engine to use bulk evaluation mode. However, if you use Non-Empty-Behavior incorrectly, you will return incorrect results.

d. Use EXISTS rather than filtering on member properties to avoid a slow execution path.  
Use the NonEmpty and Exists functions to enable the query execution engine to use bulk evaluation mode.



(88)

e) Perform string manipulations within Analysis Services stored procedures using server-side ADO.NET rather than with string manipulation functions such as StrToMember and StrToSet.

f) Rather than using the LookupCube function, use multiple measure groups in the same cube wherever possible.

g) Rewrite MDX queries containing arbitrary shapes to reduce excessive subqueries where possible.

for example

the set  $\{(\text{Gender}.\text{Male}, \text{Customer}.\text{USA}), (\text{Gender}.\text{Female}, \text{Customer}.\text{Canada})\}$  is an arbitrary set.

You can frequently use the Descendants function to resolve arbitrary shapes by using a smaller number of subqueries than queries than ~~are~~ return the same result that are written using other functions.



- a) Rewrite MDX queries that result in excessive prefetching where possible. Prefetching is a term used to describe cases where the query execution engine requests more information from the storage engine than is required to resolve the query at hand for reasons of perceived efficiency.

#### 5. USE THE QUERY ENGINE CACHE EFFICIENTLY

- Ensure that the Analysis Services Computer has sufficient memory to store query results in memory for re-use in resolving subsequent queries.
- Define calculations in the MDX script. Calculations in the MDX script have a global scope that enables the cache related to these queries to be shared across sessions for the same set of security permissions.



(70)

- c. Rewrite MDX queries containing arbitrary shapes to optimize caching.

For Example

In some cases you can rewrite queries that require non-cached disk access such that they can be resolved entirely from cache by using a subselect in the FROM clause rather than a WHERE clause.

In other cases, a WHERE clause may be a better choice.

## 6. ENSURE FLEXIBLE AGGREGATIONS ARE AVAILABLE TO ANSWER QUERIES.

- a. Note that incrementally updating a dimension using ProcessUpdate on a dimension drops all flexible aggregations affected by updates and deletes and, by default, does not re-create them until the next full process.



- b. Ensure that aggregations are re-created by processing affected objects, configuring lazy processing, performing Process Indexes on affected partitions, or performing full processing on affected partitions.

## 7. TUNE MEMORY USAGE

- a. Increase the size of the paging files on the Analysis Services Server or add additional memory to prevent out-of-memory errors when the amount of virtual memory allocated exceeds the amount of physical memory on the Analysis Services Server.
- b. Use Microsoft Windows Advanced Server® or Datacenter Server with SQL Server 2005 Enterprise Edition (or SQL Server 2005 Developer Edition) when you are using SQL Server 2005 (32-bit) to enable Analysis Services to address up to 3 GB of memory.



- c. Reduce the value for the Memory/LowMemoryLimit property below 75 percent when running multiple instances of Analysis Services or when running other applications on the same computer.
- d. Reduce the value for the Memory/TotalMemoryLimit property below 80 percent when running multiple instances of Analysis Services or when running other applications on the same computer.
- e. Keep a gap between the Memory/LowMemoryLimit property and the Memory/TotalMemoryLimit property - 20 percent is frequently used.
- f. When query thrashing is detected in a multi user environment, contact Microsoft support for assistance in modifying the MemoryHeapType.



(73)

g. When running on non-uniform memory access (NUMA) architecture and VirtualAlloc takes a very long time to return or appears to stop responding, Upgrade to SQL Server 2005 SP2 and contact Microsoft Support for assistance with appropriate settings for pre-allocating NUMA memory.

## 8. TUNE PROCESSOR USAGE

- To increase parallelism during querying for servers with multiple processors, consider modifying the Threadpool\Query\MaxThreads and Threadpool\Process\MaxThreads options to be a number that depends on the number of server processors.
- A general recommendation is to set the Threadpool\Query\MaxThreads to a value of less than or equal to two times the number of processors on the server.



For Example.

If you have an eight-processor server,  
the general guideline is to set this  
value to no more than 16.

- C. A general recommendation is to set the Threadpool\Process\MaxThreads options to a value or less than or equal to 10 times the number of processors on the server. This property controls the number of threads used by the storage engine during querying operations as well as during processing operations.

For example:

If you have an eight-processor server, the general guideline is setting this value to no more than 80.



7. (95)

## 9. SCALE UP WHERE POSSIBLE

- a. Use a 64-bit architecture for all large systems.
- b. Add memory and processor resources and upgrade the disk I/O subsystem, to alleviate query performance bottlenecks on a single system.
- c. Avoid linking dimensions or measure groups across servers and avoid remote partitions whenever possible because these solutions do not perform optimally.

## 10. SCALE OUT WHEN YOU CAN NO LONGER SCALE UP.

- a. If your performance bottleneck is processor utilization on a single system as a result of a multi-user query workload, you can increase query performance by using a cluster of Analysis Services servers to service query requests.



- ① Requests can be load balanced across two Analysis Services Servers, or across a larger number of Analysis Services Servers to support a large number of concurrent users (this is called a server farm). Load balancing clusters generally scale linearly.
- b. When using a cluster of Analysis Services Servers to increase query performance, perform processing on a single processing server and then synchronize the processing and the query servers using the XMLA Synchronize statement, copy the database directory using Robocopy or some other file copy utility, or use the high speed copy facility of SAN storage solutions.



## Aggregations Design

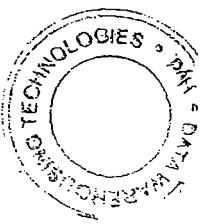
(97)

→ Aggregations are precalculated summaries of data from cells.

→ Aggregations improve query response time by preparing answers before questions are asked.



Eg:- When a datawarehouse fact table contains millions of rows, a query requesting the weekly sales totals for a particular product line can take a long time to answer if all the rows in the fact table have to be scanned and summed at query time to compute the answer.



However, response can be almost immediate if the summarization data to answer this query has been precalculated.

→ Aggregations are stored in the multidimensional structure in cells at coordinates specified by the dimensions.

→ Depends on storage mode [ MOLAP,ROLAP, HOLAP - ] and percentage of calculations.

→ If 0% Aggregations, storage less but query <sup>does</sup> <sub>full</sub> scanning and aggregations performed at Cube.

Need of Aggregations:- fact table with more rows, to

benefit from long running queries and instead of doing scanning at leaf levels, aggregations create some level of data, so query processing time will be reduced.

What is an Aggregation:-

(78)

- It's a copy of data in your fact-table, preaggregated to a certain level
  - Created when the cube is processed
  - Stored on disk
- [This result is similar to Group by statement in SQL Query]
- 2. It makes queries fast because it means SSAS doesn't have to aggregate as much data at everytime.

Drawbacks:-

1. Aggregations created at processing time, Building more aggregations takes more processing time.
2. Also increases disk space used by the cube.

Eg:-

Database Explosion

10	20
30	40

DATA WAREHOUSING TECHNOLOGIES

10 cell	20 cell	30 cell
3 cell	4 cell	7 cell
40 cell	60 cell	100 cell

5 more aggregations added,  
even though we have 4 values

When aggregations are used:-

- ① Aggregations are useful when the storage engine has to fetch the data from the disk.

- ② Aggregations will not be used if the data is in the storage engine cache. (99)

### Aggregated Design:-

- ① Each measure group can have 0 or more aggregation design objects.
- ② Each partition in group can have 0 or 1 aggregation design objects.

### Aggregation Design methodology:-

- ① Run aggregation design wizard to build some aggregations.
- ② perform usage based optimization for at least a few weeks, regularly.

### Usage based Optimization:-

Usage based optimization lets you adjust the aggregation design for a measure group by analyzing the queries that have been submitted by client applications.



### Navigation:-

Aggregations tab → Design Aggregations → Next →  
Exclude (or) Included required Objects →  
Next → click Count (or) manually  
enter the object counts →

..... aggregations now

(100)

② Estimated storage reaches 120 MB

Name:- XX - design

Click Start → Next → ③ Deploy and process now →  
Finish.

To See aggregations designed:-

Aggregations → Advanced view →

Measure group: Text-Fact Aggregation Design: XX-Desig

Starting: <specify if required> Range: <specify if required>

<Observe the aggregations created for each table>

Note:- If there are no aggregations, default aggregation is "GROUPING"



## CREATING CUBE WITHOUT DATA SOURCES

This is possible in ways.

(a) Without Template (b) With template

### 1. Without Template:-

- On the fly we need to create measures, measure groups, attributes and dimensions.
- The on the fly created objects we can store in a database, so that we can process data later.

### Navigation:-

Solution Explorer → Cube → New Cube → Next →

① Generate tables in the data source

Template  
(None)

Next →

Add new measures

Measurename	Measuregroup	Datatype	Aggregation
ActualCost	Text-FACT-NEW	Single	SUM
EstimatedCost	"	"	Avg

====

Next →

Select dimensions from template

Time --

Name	SCD	Attributed
------	-----	------------

Product1	<input checked="" type="checkbox"/>	- - -
----------	-------------------------------------	-------

Location1	<input checked="" type="checkbox"/>	- - -
-----------	-------------------------------------	-------

Next → Next →  Regular Calendar

Dimension Usage:

Dimension	Measure Group
-----------	---------------

Time1	<input checked="" type="checkbox"/>
-------	-------------------------------------

Product1	<input checked="" type="checkbox"/>
----------	-------------------------------------

Location1	<input checked="" type="checkbox"/>
-----------	-------------------------------------

Next → Cube Name: CB-Test

Generate Schema now → Finish → Next →

① Create a new data source view

Data Source View name: DSV-CB-TEST

Data Source: CB-TEST-DS

[if required click New & Specify - - -]

Next → Next → Next → Finish → Close

Now ② The Time1, Product1, Location1, Text\_fact\_new tables created in a datasource (CB-TEST-DS)

③ Cube opens in BIDS.

Deploy Cube, process Cube by adding data in source tables (Time1, Product1---).

## 2. WITH Template

(10.2)

→ Here there "Templates" and we select / create dimensions, facts On the fly (dynamically).

→ Navigation:-

Cubes → RC New cube → Next →

(i) Generate tables in the data source

Template

Select template (Adventure Works  
Enterprise Edition)

Next → Select measures, measure group

(iv) Create      "      "      "

Next → Select dimensions

(v) Create      "

Next - - - < Rest of the process like  
above >

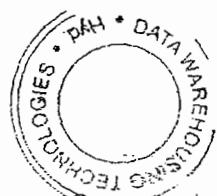


## Developer Responsibilities

1. Understanding analysis services documents and business needs.
2. Creating Calculations, Actions, KPIs
3. Adding, Manipulating measures
4. Adding, removing dimensions and relationships
5. processing dimensions and facts
6. Working <sup>on</sup> named queries, Calculations
7. Writing MDX queries, expressions
8. Calculated Measures, members etc --
9. Creating partitions (if required)

## DBA Responsibilities

1. Deployment
2. Security
3. creating data sources/ data source views
4. Manipulating structures
5. Managing partitions
6. Creating Cube (optional)



## Modellers/ Designers Response

1. Designing a model (Star Schema, Snow Flake etc --) for the project
2. Creating Cube, partitions etc -
3. Defining relationships etc --



**ssrs**



# SSRS INDEX

" Vinay Sir notes " 11

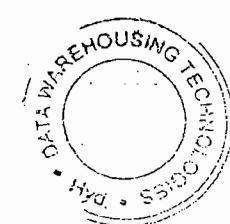
## CONTENT



## PAGENO

1.	SSRS BASICS, NEED, REPORT TYPES	1-2
2.	SSRS 2008 ADDITIONAL FEATURES	3-5
3.	ARCHITECTURE & EXPLANATION	6-9
4.	REPORTING SERVICES CONFIGURATION MANAGER	10-13
5.	CREATING REPORTS, STEPS, DATA SOURCES	13-15
6.	CREATING REPORT THROUGH WIZARD	15-16
7.	STANDARD REPORT CREATION	16-18
8.	REPORT ITEMS, DATA REGIONS	18
9.	TABLE REPORT, OPTIONS, PROPERTIES	19-22
10.	TEXTBOX PROPERTIES	23
11.	EXPRESSION ITEMS, CUSTOM CODE, CUSTOM ASSEMBLY	24-27 28 ,B,C,D
12.	ACTIONS (REPORT, URL, BOOKMARK)	28

13. LOOKUP ENHANCEMENTS IN 2008R2 28C,D  
 14. VARIABLES ( REPORT, GROUP ) 29-30  
 15. QUERY FIELD, CALCULATED FIELD 30  
 16. DATASET PROPERTIES 30  
 17. PASSING PARAMETERS 30-36  
 ( QUERIED, NON QUERIED, CASCADED,  
 MULTIVALUED )  
 18. DRILL DOWN, DRILL THROUGH REPORTS 37-38  
 19. SUBREPORTS, CHART, MATRIX, LIST,  
 GUAGE REPORTS 39-46  
 20. REPORT MODEL, BUILDER 47-50  
 21. REPORT MANAGER OPERATIONS 51-65  
 ( FOLDER, CACHE, SNAPSHOT, SUBSCRIPTION,  
 SECURITY, SITE SETTINGS -- )  
 22. CONFIGURATION FILES, COMMAND LINE 66-68  
 UTILITIES  
 23. PUBLISHING REPORTS 68-70  
 24. PERFORMANCE TUNING 71-85



## SSRS [SQL SERVER REPORTING SERVICES]

Reports are required in general but situations

a. For Internal Reports

b. For External Reports



### Internal Reports

Generally, Company Internal Operations such as Pay slip, Salary slip, Relieving letter, Internal Auditing etc.

### External Reports

This Reports generally submitted to 3rd party authorities such as IT department, STPI etc....

### CONCLUSIONS :-

To Create this type of reports we go for different Reporting Applications like COINOS, BO, SSRS, CRYSTAL REPORTS, MICROSTRATEGY ETC

### REPORT USAGE TYPES

#### a. Standard Reporting

Here there will be a Centralized database. Multiple users connect to database and they generate their own Reports.

### b. ADHOC REPORTING:-

This Reports also can be called as dynamic Reports and the content and layout changed Frey time.

### c. EMBEDDED REPORTING:-

Here the Reports are Embedded with the 3rd party applications like Java, .Net etc.

## ENTERPRISE REPORTING LIFE CYCLE :-

There are 3 phases

### a. COLLECTING

Here design, identify, content, placed, Layout specification

### b. MANAGING PHASE

Once the Report is generated Security and the manageable Operations for the Report applied here

### c. ACCESS AND DELIVERY PHASE

In this phase Report Publishing, accessing ... etc taken place.

## SRS 2008 ADDITIONAL FEATURES AND ARCHITECTURAL CHANGES THAN 2005

### 1. IT IS REMOVED IN 2008

As there is a conflict to other by other Applications they removed the dependency of ITs in 2008 by Replacing with the Components.

- a. SQL 08
- b. SQL CLR
- c. SQL NETWORK INTERFACE
- d. COMMUNICATIONS THROUGH http, SIS protocol.

### 2. RICH MEMORY MANAGEMENT ADDED.

- a. Server Infrastructure for Process memory monitoring.
- i. Dynamic, self-managing with Memory Pressure
- ii. Reduces throughput in memory Pressure situations.
- b. Reporting Processing uses a File System Caches to adapt to memory Pressure
- c. Receives memory Events from Server
- c. Administrator is able to set Targets (Min, Max)
  - i. Minimum threshold defines the amount of memory the server thinks 'belongs' to it.



(4)

- ii). The memory is only used if a requests need it.
- iii) Maximum threshold defines that not to exceed value.
- d. Adapts to other Processes Consuming Memory



### 3. RS 2008 REPORT ENGINE CHANGES.

- a. Report Processing
  - i. On-demand Processing
  - ii. Hierarchical Cursor-Based Object Model.
- b. Rendering
  - i. New Rendering Architecture
  - ii. Renderer Rewrites.

### 4. SCALABILITY

- a. Reports In SQL 2005 are memory bound
  - i. Memory Usages is Proportional to data size.
  - ii. Large datasets can cause Out of Memory exceptions.
  - iii. Memory Usage in Problem Renderers (pdf, Excel, csv)



- b. Very large Reports can starve (or) fail many smaller reports.

#### 5. DUNDAS ACQUISITION:-

- a. The SQL SERVER Reporting Services team has acquired dunda's software data visualisation products.
- b. chart, Gauge, Map, Barcode and calendar for Reporting
- c. chart, Gauge for sharepoint
- d. chart Program/Enterprise, Gauge, Map, OLAP chart and Tab controls for Visual studio (windows and web).

#### 6. TABLIX - NEW DATA REGION

- a. Tablix provides a combination of the best features of Tables and Matrix data regions.
- b. Build Versatile Reports.
- c. Allows a flexible layout with Multiple Row and column.



# SSRS 2008 Architecture



### HTTP Listener

(7)

It monitors the incoming request directed to HTTP.SIS on a specific code on the local computer. The Host name and port are specified on a URL request while you configure the server.

When the HTTP listener process the request it forwards the application layer to verify the user Identity.

### Authentication Layer

It verifies the userid, password (or) the identity of user (or) application that makes the request. The following authentication that supported are

- (a) Windows Integrated Security.
- (b) NTLM Authentication.
- (c) Forms Authentication.
- (d) Basic Authentication.
- (e) Anonymous Access.



### Report Server

It is the heart of Reporting Services which is implemented as windows service.

It consists of

- (a) windows service
- (b) Report manager
- (c) Web Service
- (d) Back ground processing.

(e) Windows Service: (Provides report scheduling & Delivery Services): Both the services are used in designing, saving, executing, managing and publishing the reports.

Reporting Services hosts the report manager, the report server, web service & background features running in their own.

Report Manager - It provides client frontend access to the report server items and their management.

④ Web Service - It provides access to report server via report builder.

⑤ Background Processing - There are many processing happens in this background processing.

- ① Report processing
- ② Data processing
- ③ Model processing
- ④ Data rendering
- ⑤ Data authentication extensions.
- ⑥ Scheduling
- ⑦ Subscription
- ⑧ Data base maintenance.



### Report processing

Report server has 2 cool processors

- ① Report processor
- ② Scheduling and delivering processor

### Report Server Backend

Report Server stores folders and files just like file system.

- The Report you create exists as a file in the file system with extension ".RDL" (Report Definition Language).
- When the report is published it will be stored in the Report Server database.
- The deployment uses a SQL Server relational database for internal storage.
- ④ Report Server database (ReportServer, ReportServerTempDB).

- ④ Report Server Tempdb → It stores temporary data, session information and caching information.

### Data Processing:-

It is designed to retrieve a specific type of data source and provide extended functionality during report design and processing.

### Data Rendering Extensions:-

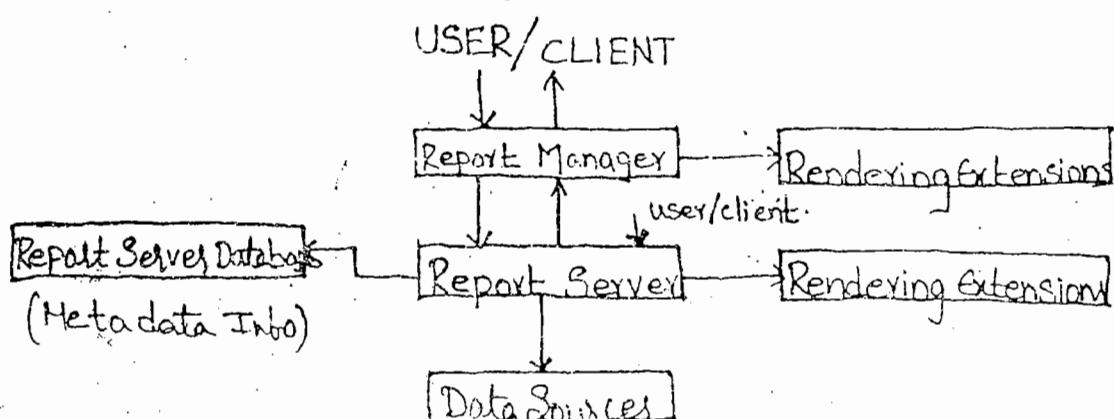
There are three rendering formats available

- ① Data Render:- Data only Display.  
Eg: CST & XML.
- ② Soft page break Render:- Maintain format & Layout  
Eg: MSWord, Excel, MHTML, Report viewer controls etc
- ③ Hard page Break Render:- It supports gif & pdf formats.

### Scheduling Delivery Extensions:-

- ① Report Server Email.
- ② Report Server file share.
- ③ Custom extensions
- ④ Subscriptions.

### Simple SSIS Architecture:-



## WORKING WITH REPORTING SERVER CONFIGURATION

Generally this tool is available for administrators to configure the Reporting Services Components and items.

### Navigation:

START → PROGRAMS → CONFIGURATION TOOLS →

REPORTING SERVICES CONFIGURATION MANAGER.

#### 1. SERVICE ACCOUNT:

It specifies or takes Built-in Account (or) windows domain User account to run the Report Server Service.

#### 2. WEB SERVICE URL:

This is Report Server URL, if we want to change we can change it.  
→ Virtual Directory : ReportServer  
→ Click Apply.

Default URL :  $\text{Http://} <\text{servername}>: <\text{portno}>/$   
 $<\text{Reportservername}>$

#### 3. REPORT MANAGER URL:

Default URL :  $\text{Http://} <\text{servername}>: <\text{portno}>/$   
 $<\text{REPORTMANAGER}>$

Eg :  $\text{http://} <\text{RAWAN}>: <8080>/\text{Report\_DW H}$

(11)

#### 4. DATA BASES :-

It takes and displays Report Server databases.

Default : ReportServer DB.

Report Server Temp DB.

#### Navigation :-

Start → programs → SqlServer 2008 → Management Studio



↓  
Database Engine

↓  
< Observe Database >

#### New database creation :-

Change Database → Next → Servername: Localhost →  
DatabaseName: ReportServer → DWH → Next →  
Next → Finish.

#### 5. EMAIL SETTINGS :-

These settings are useful, required

at the time of subscription and delivery.

It requires below settings.

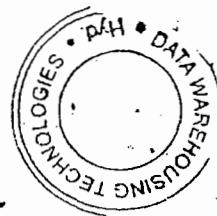
Sender address: MSBI\_Viray@gmail.com

Current SMTP delivery method: Use SMTP server.

8 MTP Server : 127.0.0.1 click Apply

### Note :-

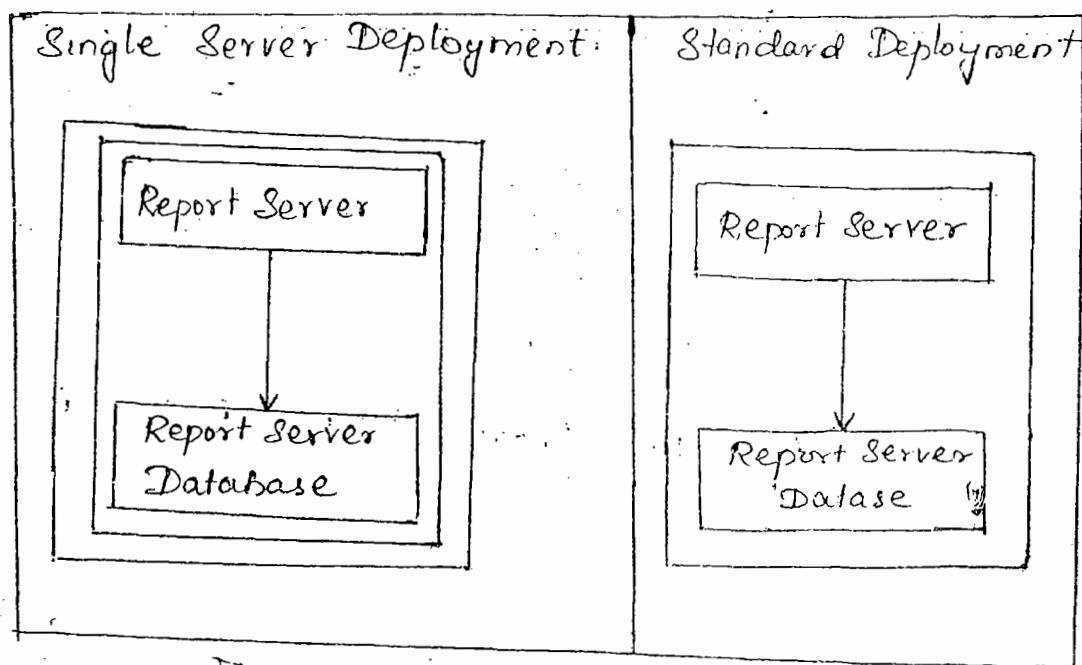
from the specified mail ID on the specified Server the mail sent to the corresponding Recipients.



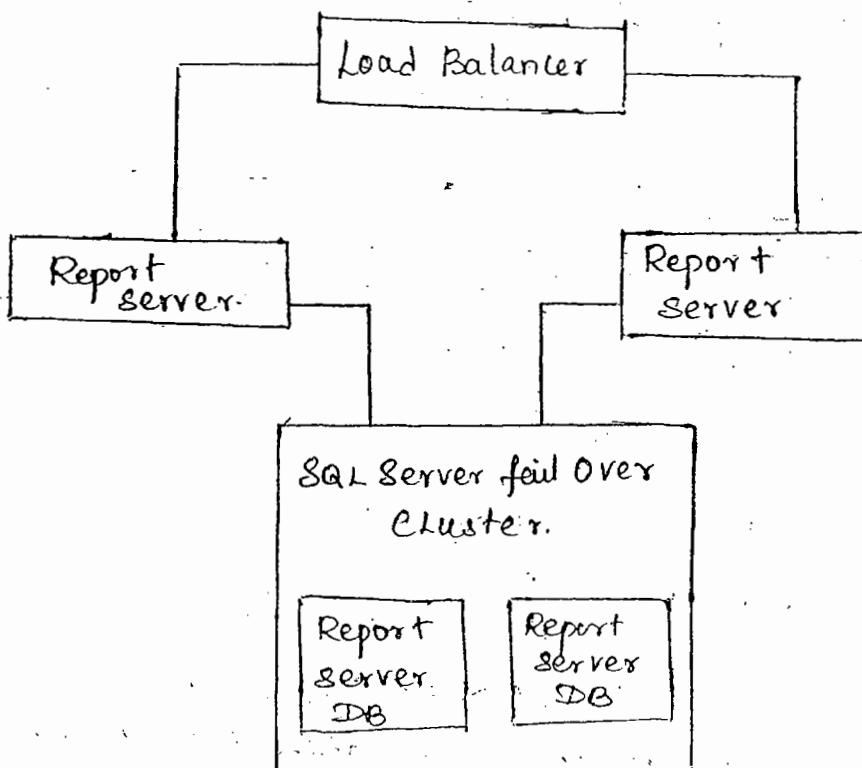
## 6. SCALE OUT DEPLOYMENT

In Realtime there are two Types.

- a. Single Server deployment → Report Server and database on same machine.
- b. Scale Out Deployment → Report Server and database on different machines.



## SCALE OUT DEPLOYMENT



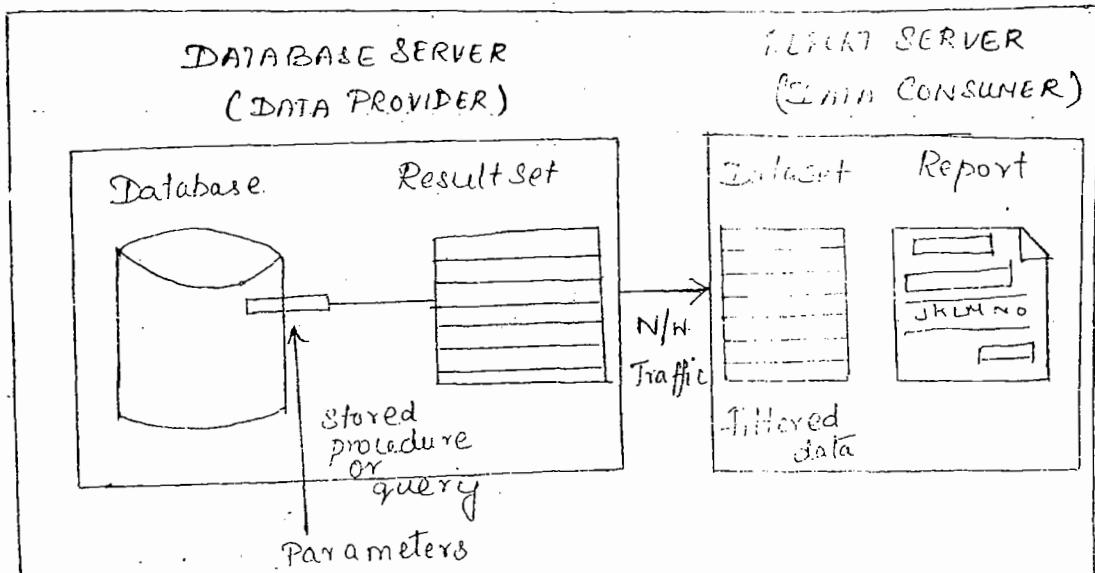
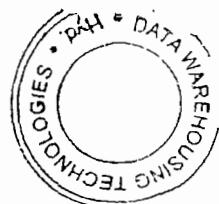
## CREATING REPORTS

There are 3 ways to Create Reports

- a. By Using Report Server Project Wizard
- b. By Using Report Server Project
- c. By Using Report Model & Builder.

## STEPS TO CREATE A REPORT

1. Create a data source
2. Create a Data Set
3. Take the Report Layout.
4. Design the layout.
5. Use the Data Set fields in the layout
6. Do parametrisation (or) Implement Actions (Optional)
7. Publish the Report
8. View the Report from the Report Manager (or)  
Report Server (or) Control from any 3rd party  
Applications (or) Report Viewer etc.
9. Render the Report (optional)



(15)

### Data Source

It is a connection to the Database.



### Data Set

It is a logical Representation of tables (or) tables.

### Shared data Source

It is a Data source which can be used across any report in the same solution.

### 1. Create a Report by Using Report Server Project Wizard:

Open BIDS → FILE → NEW PROJECT →  
SELECT TEMPLATE AS PROJECT & SERVER PROJECT  
WIZARD → ASSIGN NAME → CLICK OK →

NEW DATA SOURCE → TYPE → CLICK EDIT

NEXT → CLICK QUERY BUILDER →

QUERY DESIGNER → ADD TABLE →

SELECT REQUIRED FIELDS → CLICK OK.

NEXT → SELECT THE REPORT TYPE (Tabular, Matrix),

→ NEXT → DESIGN THE TABLE →

CHOOSE

CHOOSE the Available Fields into Page, Group and Details.

Remove → click Next → choose the table layout → stepped (or) blocked → Next → specify colour → Next → choose the deployment location → Report Server.

http://RAWAN:8080/Report Server → DWH → Next → Report Name → Finish.

In Wizard → Press Build → click Deploy Option.



### CREATING STANDARD REPORTS (Real Time Creating Reports)

1. Open BIDS → New Project → Select Template as Report server project → Click OK.
2. View → Solution Explorer → Shared Data Sources → R+ click → Add new Data Source and create a Data Source for DB-MSBI.
3. Reports → RT click → Add → New Item → Report → Specify Name (Report-party.rdl)

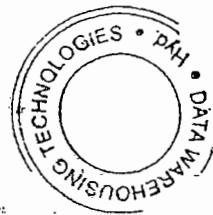
4. Add → View → Toolbox → Drag the Table Region on to the layout.
5. View → Report data → New → Data Source  
→ Specify any Name (DST-DBMSBI) → Click the option "Use shared data source Reference" → click OK.
6. New → Data Set → Specify Name (DST-DBMSBI)  
→ Data Source (DS-DB-MSBII) → Query Type  
→ Click Text Option → Query designer (If we do not know how to write a query).
7. Drag and Drop Data set fields on the layout (or) Select the fields in the layout
8. Preview
9. View → Solution Explorer → Solution  
Rt click → Properties → In "Target Server URL" option we should Enter the URL Manually  
i.e. `http://Rawan:8080/reportServer_DWH`  
→ click OK.



10. Build → Deploy
11. Go to the Manager and Observe the Report.

### Data Regions in the Report :-

- a. Table
- b. Matrix
- c. List
- d. chart
- e. Sub Report etc
- f. Image



### Report Items :-

- a. Image
- b. Pointer
- c. Line
- d. Textbox
- f. Rectangle

(17)

## WORKING WITH TABLE REPORTS :-

It displays data in Rows and Columns format

It contains 3 sections

a. Header Section :-

Here column headings (or) Aggregate Information displayed.

b. Footer Section (Optional) :-

This section displays Aggregate Information generally.

c. Data Section :-

Displays Actual data Presented to be in the Report.

Note:- We can take one more section to split the data in the data section.

i.e. Group section.

## ADDING COLUMNS LEFT AND RIGHT :-

SELECT → TABLE → HIGHLIGHT THE COLUMN →

RIGHTCICK → INSERT COLUMN → LEFT, RIGHT.

## PROPERTIES

SELECT THE OBJECT → PRESS F4 (or)

CLICK VIEW OPTION AND SELECT PROPERTIES WINDOW.

## Properties for changing colors, fronts, Borders, alignment

e.g. Eg:

Border colour → purple

(20)

Border style → solid.

Border width → 2pt

Background color → paleTurquoise.

Front → century

Front size → 11 pt

→ Taking Border to rows → highlight border section →

Press F4 → Select Border colour → change to blue.

→ format menu

- Foreground color
- Back ground color
- Front
- Justify
- Align
- Make same size
- Horizontal spacing
- Vertical space
- Centre in form
- Order.



→ Table Properties → These are the properties common for table and matrix.

↳ Table → click at the Top corner → Select

Table properties.

↳ General

Name: party\_table

Tool tip: party and address table info

Datasetname: DST\_DBMSBI

Pagebreak options

Row Headers & column headers

→ Visibility - change displays options.

↳ Show - displays the object

↳ Hide - hides the object

↳ Show or hide based on an expression.

[ ] → [fx] ← press the fx

### Expression window:

set expression for: Hidden

=IIF(sum(fields!PARTYINCOME.Value,"DST-DBMSBI")<5000000,  
False, True)

Eg: Display the report to the entire sum less than 5,00,000/-

Sol: view the sol. at top.

↳ Display can be toggled by this report item:

This property can be useful while during drill down analysis

Eg: Display the report when

Row section.			
Party ID	PARTYNAME	PARTYLOC	PARTYSALARY
[Party ID]	(Party name)	[Party loc]	[Party salary]

↳ FILTERS: These are used to display the report based on filter condition.

Eg: Display the report where the party location is HYD.

Filters → 'Add' → Expression [Party loc] → = "HYD".  
operator =

↳ Sorting = Displays the corresponding column data in the specified sort order (ascending or descending).

Sorting → click Add → sort by: [party loc] A to Z

22

→ Taking full total of the INCOME

Rt click → Row section → Insert row → select outside Group-below

Note: By default the aggregation is sum. If we want to change the aggregation → goto textbox expression and change it.

Taking Groups → generally we group the data if we are having duplications or to keep similar item records in a page  
Eg: HYD data in one page, Bangalore data in one page etc.

→ Select Row section → click Add group → Row group → Parent group.

Table group

Group by: [Party loc] → Mark the options → click OK



→ Taking subtotal (Group by total) → Goto group footer section and select Income

Working with group properties:

→ Goto group section → click → group properties.

↳ General:

Name: Loc-Group

↳ Page breaks:

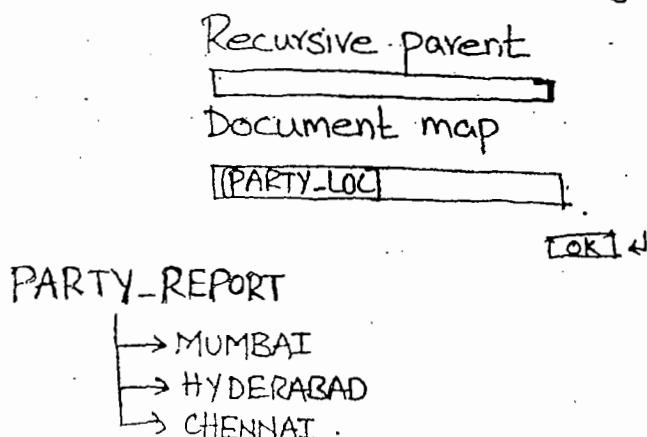
Page break options:

Between each instance of a group

Also at the start of a group

Also at the end of a group

- ↳ Sorting
- ↳ visibility
- ↳ filters
- ↳ Variables — Group variables which are used in certain groups.
- ↳ Advanced — set additional group properties.



### Taking sub groups:-

→ select group section → Add Group → Rowgroup → child group.

Tablet group

Group Name: [PARTY-LOC]

OK

### Working with text box properties:-

→ Each and every cell in a Textbox.

→ Textbox is having below properties.

④ Expression → Here we write expressions to display the data according to the expression in the column.

Eg:- party income increment by 12%.

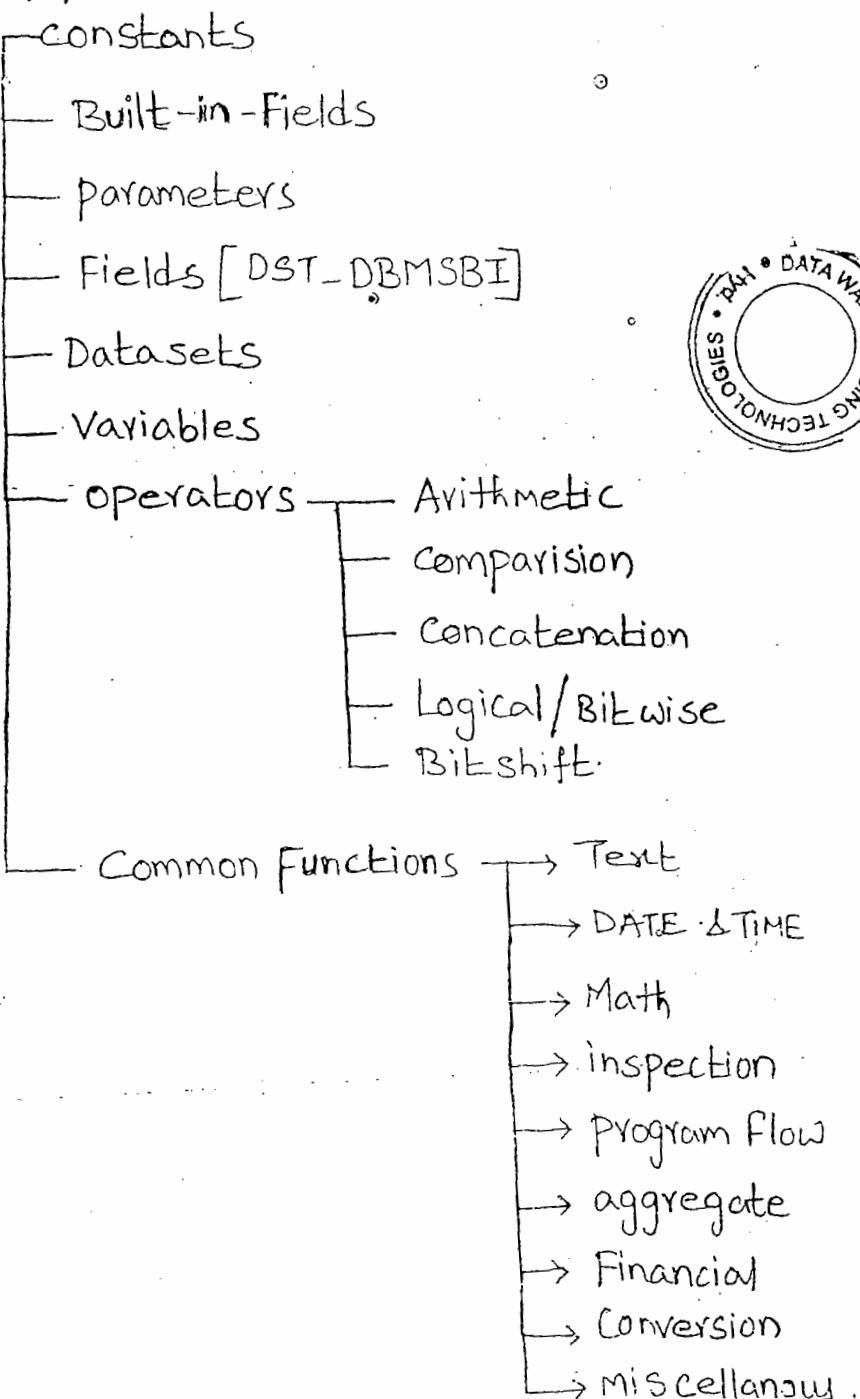
Sol:- R click on selected column → In expression window → Set expressions for: Value → = Fields! PartyIncome.Value \* 12/100.

The below options properties we use an expression,

i.e.,

## EXPRESSION ITEMS:-

(24)



Statement :- Many Reports

(25)

Date Functions: = Today() → displays current date

Adding months: = DateAdd("D", 6, Today())

= DateAdd("M", 6, Today())

= DateAdd("y", 6, Today())

Display year of the date: = Year(Today())

String functions:

= Fields! PartyName.Value & VBCRLF & Fields! PartyLoc.Value  
→ Enter char. here

Length of a string:

= LEN(Fields! PartyName.Value)

= Right(Fields! PartyName.Value, Len(Fields! PartyName.Value) - (IIf(Fields! PartyName.Value, "\\"))).

Eg: VINAY/HYD → O/p: 9-6=3 of Right i.e., HYD.

= CSTR(Fields! PartyName.Value, "\\")

→ It converts into string (i.e., it changes the numeric

value from one position to another position (i.e., left to right & vice versa).

Decision Expressions:

IIF: IIF(condition, <success statement>, <fail statement>)

SWITCH: Switch(condition, stmt1, cond.2, stmt 2, cond.3, stmt 3)

NOTHING: It is NULL value in Visual basic.

RowNumber: Displays the row number of a row in the report.

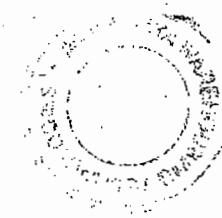
## Report Functions:-

SUM- It totals the values in Grouping and data regions. This function can be useful in the header (or) footer in the table.

= SUM (fields).lines Total .value, "partyIncome")

## Page Header & Footer Functions:-

- (a) Built in functions.
- (b) First/ Last/ Sum ...etc.



## Page Break Expression:-

Eg:- Required a page break for a group for every 25 rows.

= INT (RowNumber (Nothing)-1)/25

Custom code- we use this code in report by embedding the report (or) store in a custom assembly which is used in the report.

Custom Functions- There are many functions available in reporting services but still if we want to write a custom function we can write.

Eg:- Adding a function call get coloured and using it.

Sol- Report Menu → Report properties → CODE →

Public Function GetColor (By VAL Pcode as string) as string;

IF Pcode = "10" then

Return "YELLOW"

END IF

IF Pcode = "20" then

Return "RED"

END IF

IF prode = "40" then

Return "GREEN"

END IF

END Function.

Goto color section ie, Expression in Party code box  
= code. Getcolor(Fields! party code . Value)

### Text Box properties:

#### ① General

Name : partycode

Value : partycode

Tool tip → options

② Number → Here it sets Number and Date formatting option

③ Alignment → Horizontal & Vertical alignment with left and right side padding options

④ Font

⑤ Border

⑥ Fill → It changes the background colour, images etc

Eg: Displaying different background colours for different party code values.

Set IIF(Fields! Party code . Value <= 20, "red", "Green")

⑦ Visibility

⑧ Interactive Sort → It does interactive sorting on the column based on the user action.

27



(28)

- click the option → Enable interactive sort on this text box → detail row.
- Sort by: party name.
- click ok.



### ① Action:

ⓐ Goto report → Refer to the drill through report concept for this action.

ⓑ Goto URL: → Calling a URL while doing report actions.

Eg:- After clicking respective party location value it displays respective URL Location.

Sol:- = Switch(Fields! partyLoc.value = "Hyd",  
                      "Http://www.Hyderabad.com",

Fields! partyLoc.value = "CHE",

"Http://www.Chennai.com",

Fields! partyLoc.value = "MUM",

"Http://www.Mumbai.com",

IS NOTHING(Fields! partyLoc.value), "Http://www.other  
location.com").

ⓒ Goto Book Mark: → Refer to the chart report concept where there is an example for book mark option.

## More examples for Custom Coding

28A

- Take this code in Report properties

```
public function VinayMsg (ByVal opp as String) as String
```

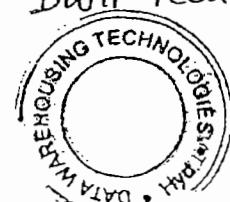
Return "MSBI opportunities at" & opp

End Function

- In textbox → expression write the below

```
= Code.VinayMsg ("DWH Technologies")
```

## Adding Custom Assemblies



- Open an instance of VS.NET and create a new class library template. Here I am using C# (you can use any language). Give it a name "VinayLib" and specify its location.

- Rename Class1.cs to WelcomeClasses, open it and add the following methods to the class:

```
public string VinayMsg (String name)
{
    return "MSBI Opportunities at " + name;
}

public static string SvinayMsg (String name)
{
    return "MSBI Opportunities at " + name;
}
```

- In the above case two methods (a) static (b) normal.
- 3. Build project, take "vinaylib.dll" in Bin\Debug folder to "C:\IDE\privateAssemblies" in your machine.
- 4. Go to report, report → Report properties → References tab  
Browse: VinayLib.dll ← Add ←

Add (or) remove classes

click →

ClassName	Instance Name
VinayLib.WelcomeClass.cs	MyWelcomeClass

- 5. Take textbox and call the functions with the below syntaxes.

- 
- a) = Code. classInstanceName. InstanceMethodName (List of params)
  - b) = AssemblyName. className. StaticMethodName (List of params)

Eg - a) = Code. MyWelcomeClass. VinayMsg ("DWHTechnologies")  
 b) = VinayLib. WelcomeClass. VinayMsg ("DWHTechnologies")

- 6. Copy "vinaylib.dll" in "ReportServer\bin" folder in your machine.

- 7. Deploy and see the results.



## Look up enhancement in 2008 R2

28C

The Look up enhancement functions available in 2008 R2 are listed below:

Go to Text box → Right click → Expression → Category →

(Table)

Common Functions → Miscellaneous → Item

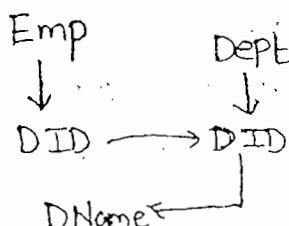


- In scope
- Lookup
- Lookupset
- Multi Lookup

### Syntax:

- **Lookup**(Source column, Destination column, Result set, dataset)
- **Lookupset**(Source column, Destination column, Result set, dataset)
- **MultiLookup**(Source column, Destination column, Result set, dataset)

## Look up function usage (single value relations)



① Create 2 datasets for Emp, Dept i.e., EDS, DDS.

② Expression properties:

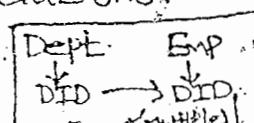
= **Lookup**(Fields! Emp.DID, Fields! Dept.DID, Fields! Dept.DName)

O/p: 

EMP Table			Dept Table		
EID	EN	DID	DName		"DDS")
1	X	10	IT		
2	Y	11	HR		

→ Lookup set: Set of values (or) Result set relations

Eg: For DID in Dept, multi emp ID's in emp



(28D)

= JOIN (LookUpSet (Fields! Dept. DID,  
Fields! Emp. DID, Fields! Emp. EID, "EDS"), ",")

O/p:- DID DNAME EID

10	IT	1 2 3 4 5 6 7 8 9
----	----	---

### MULTI LOOK UP :-

Generally MultiLookup use at the time  
of Multivalued the parameters because  
Users can send multiple values.

This Lookup Returns based on every value,  
Passes from the source In multivalue list.

JOIN (multilookup (Fields! Dept. DID, Fields!  
Emp. DID, Fields! Emp. EID, "EDS"), ("\"))

----- X -----

## Displaying Different colors for Text box Value 29

→ party code text box → press F4 → color properties →

In Expression Window → = code.Getcolor(Fields! Partycode  
· value

### Variables:-

The Variables are of two types. They are

- ① Report variable
- ② Group variable.

③ Report Variable: This Variable is used across anywhere in the report.

### Navigation:-

Report Menu → Report → properties → variables →

Add → Name : PINC

→ Expression window:

Set the Expression

\* 12/100

click OK.

Goto textbox → rclick expression → Expression window

= Fields! party income.value \* Variables! PINC.value

click OK.

- ① Note: we cannot use in calculated or query fields. ③
- ② Group Variables: These are used within the same group only.
- ③ Group → Group properties → Variables → Add → NAME (pic)  
→ Expression (12/100)

### Creating a query field:

- ④ → These query fields added at dataset level
- ⑤ → No calculations allowed for these fields.
- ⑥ → Data set (DST\_DBMSBT) → rt click → Dataset properties  
→ Fields → Add → Select query field → OK.



Adding calculated field: Here the fields are created based on calculation:

Navigation: dataset properties → Fields  
→ Dataset → rt click → Add → Calculated field → party INC\_DWII  
→ Expression → = Fields! partyincome.value \* 12/100 → OK

Note: This calculated field performs the operation while it retrieves from source (according to situation).

2: This field is usable across anywhere without rewriting calculation.

### Data Set properties:

- ① Query: ① Name, Data source, Query Type → Text → query:  
\* Refresh fields → If the data in the sources change, to get the fresh data, this option is useful
- ② Parameters
- ③ Fields
- ④ Options → Here options are case sensitivity, collation etc..

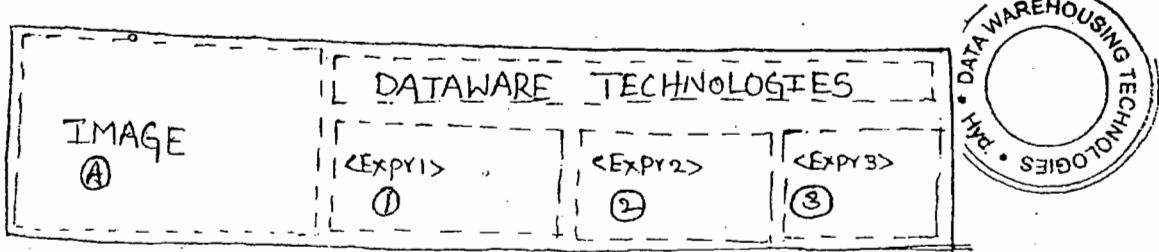
⑥ Filter: It Filters the data of the dataset. (31)

### Working with page Header and Footer:

In Page Header & Footer titles, logo, Page numbers, Page names, Report names etc... we can take & display.

Report menu → **Add page header**      ① Toolbox → Image option  
                        **Add page Footer**      → Take image →

Embedded → Import image → Image properties → General → Name:  
→ Size → In Display → select the option Fit to size.



For Expression → Right click → Textbox properties → Expression

① Globals! Reportname & VBCRLF & Globals! ExecutionTime

② Globals! Pagenumber & "-To-" & Globals! TotalPages

③ = "Total INCOME::" & Sum(ReportItems("partyincome").Value)

### Passing parameters:

Parameters provide user interactions to the report.

There are 4 ways we pass the parameters:

① Queried Parameters

② Non queried parameters

③ multivalued parameters

④ Cascaded Parameters

② → In this case Parameter values are not getting from the database based on the query. That means the parameter values are written manually.

- ① Provide flexibility to the user to go to the corresponding location based on the location selection.

## Non Queried (Default values)

(32)

View → Report data

1. parameters → R click → Add parameter

General

Name : VARLOC

Prompt : Please select the location::

Data type : Text

Set parameter visibility

Select visible option

Available values

Specify values

Add the below

Label	Value
Hyderabad	HYD
Mumbai	MUM
Chennai	CHE



Default values → No default value

2. Dataset → R click → Dataset properties

Query → ADD The where condition to the existing query  
where partyLoc=@Varloc.

OK.

Browse and see the result.

Note: Label = It is a displayable field.

Value = The value which is passed to the dataset.

### ③ Multivalued parameters-

We can create multivalued parameters for both queried & Non-queried reports.

It allows multiple values to select in the parameter list.

Navigation- In the above steps do the below changes:

#### 1. General

check allow multiple value option.

2. Dataset → Where clause should be "Where PartyLoc  
in [@VarLoc]"

### queried parameters-

Here the parameter values are taken from query.

This is a recommended mechanism in real time.

① Create a new dataset to take only distinct location values.

View → Report data → ADD Dataset →

Name: DSTLoc

Query: Select DISTINCT (PartyLoc) from Location.

② Create a parameter based on the above dataset result  
parameters → right click → Add parameter

#### General

Name: VARLOC

Prompt: Please select the Location::

Data Type: Text

Set parameters visibility:

Select visible.

## ① Available values.

(34)

Get values from Query.

Dataset : DSTLOC

Value field : party Loc.

Label field : party Loc.

Default values → No Default value.

- ③ Use the above parameters in the actual report dataset (which is displaying values in the report).

Dataset → right click → Dataset properties.

Query: ADD the Where condition Where partyLoc = @VARLOC  
click OK.

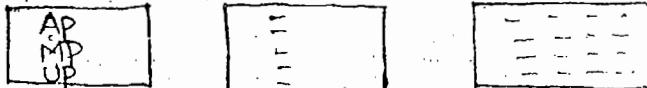


## ④ Cascaded Parameters:

Based on one parameter value if we display data in other parameter values then it is cascaded.

Eg: Based on the location Selection, corresponding party codes should be selected and based on the party code selection the report should be displayed.

States → Districts → District Reports



1. Create a new Dataset to take only DISTINCT Location Values.

View → Report Data → Add Dataset →

Name: DST LOC

Query: Select DISTINCT (PartyLoc) From Location

2. Create a parameter based on the above dataset result.

Parameters → click → Add parameter.

General

Name : VARLOC

Prompt : please select the Location ::

Data Type: Text

Set parameter visibility:

Select visible.

Available values

Get values from Query.

Dataset : DST LOC

Value field: PartyLoc

Label field: Party Loc

Default values → No Default value.

3. Create a new dataset to take only DISTINCT CODES based on Location.

View → Report Data → Add Dataset

Name : DST code

Query: Select DISTINCT (Partycode) from party where  
PartyLoc = @ VARLOC.

4. Create parameter based on the above dataset result

parameters → rclick → Add parameter

General:

Name: Varcode

Prompt: please select the party code::

datatype: Text

Set parameter visibility.

Select visible.



Available Values:

Get values from Query.

Dataset: DST code

Value field: party code

Label field: party code

Default values:

No Default value.

5. Use the above parameter in the actual report dataset

(which is displaying values in the report)

Dataset → rclick → Dataset properties.

Query: Add the where condition where

PartyLoc = @VARLOC and

Partycode = @VARCODE

click OK and

Execute the report.

## Drill down and Drill Through Reports

(37)

### Drill Down Report

With in the same report we want to do navigations like expansions, collapsing...etc then we go for drill down report.

Group 1	Group 2	Party ID	Party Name	Party Loc	Party Income	Party Add.
		PID	PNM	PLOC	PIINC	PA

Select all

↓  
press F4

Hidden: True

Toggle Location: Group 1.

Hidden: True  
Toggle Location: Group 2.



NOTE: This report is useful to perform Top down (or) "Bottom up" Analysis.

## through Reports

(38)

Moving from one report to another report based on report action.



### Navigation:

1. Create a New report → report 2 (Name specify).

2. Create a parameter like the below in the report.

View → Reportdata → parameters → rclick  
→ add parameter →

Name: Varloc

click OK

3. The report 2 → dataset → query:

Add the below condition where partyLoc = @VARLOC

4. Goto Main report.

report 1 → Goto Textbox → partyLoc → rclick → Text  
box properties → actions → Goto report (option) → select  
the report from the list: **Report2**

**ADD**

Name	Value
VARLOC	PARTYLOC

→ **OK**

5. Browse the report 1 and click the location.

so that it goes to the other report (i.e., report2)  
and displays the corresponding content.

## Taking Sub-Reports

Sub Report displays the content in the main report.

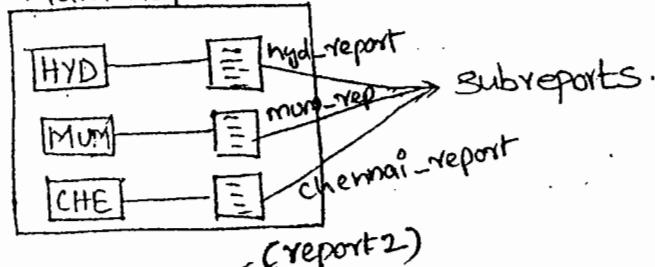
(or) It is the report in the main report.

It can take parameters.

Generally sub reports are taken in real time to display particular information for every row (or) set of rows in the main report.

Eg:- Display every branch and its report in the main report.

Main Report



(report 2)

1. Create a report and take a parameter to the report
2. Take a New column in the Main Report and Drag and Drop subreport from toolbox to the column and specify the below subreport properties:

General :

Name: Location - SUB\_Report

Use this report as sub report : party2

Visibility: Select SHOW

Parameters: Click ADD

Name	Value
VARLOC	[PARTYLOC]

Border: Style: Solid

Width: 2pt

## Working with Chart Report:-

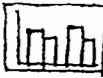
To display the data in more analytical way chart reports are useful.

These display data in required chart type.

The chart report contains 3 sections.

- (a) Data Section → Here actual data presented.
- (b) Category section → Here group wise information displayed.
- (c) Series section → It displays the information related to series of data presented in the report.

## TYPES OF CHARTS:-

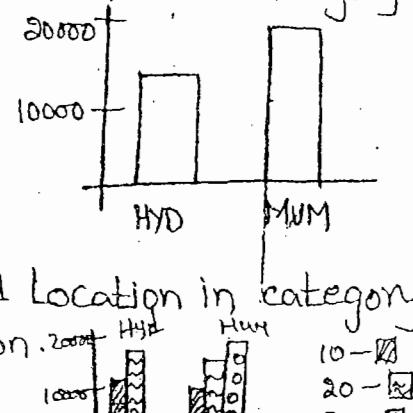
- (a) COLUMN  
- (b) LINE
- (c) SHAPE
- (d) BAR
- (e) AREA   
- (f) RANGE
- (g) SCATTER
- (h) POLAR

- (1) Display locations and their total incomes.

Sol: Take income <sup>on</sup> data section and Location in the category section.

- (2) Display party codewise sum of income in every location

Sol: Take Income on data section and Location in category section & code in series section.



(41)

## CHANGING CHART TYPE

→ chart → rt click → shape → select shape.

## Taking Tables on chart

ON the chart → rt click → click show data tables.

### A) chart properties

    ↳ General

    ↳ visibility

    ↳ Filter

    ↳ Fill

    ↳ Border.

### B) chart Area properties

    ↳ 3D options, visibility, Fill, Border, shadow.

### C) Series properties

    ↳ series data:

        Value: field:

        [sum [Party Income]]

    ↳ visibility, Axes & chart Area, Fill ... etc.

### D) Legend properties

    ↳ General:

        Name: Location - code.

        Layout: Auto table

    ↳ visibility, Font, Fill, Border, shadow ... etc.

Note: chart report takes parameters and it supports action implementation.

(42)

- ① → For changing the default dataset values goto series properties and change the value field.

Matrix Report: This is useful to display cross tab

reports. Generally it displays data in the form of matrix like rows against columns. It contains 3 sections

① Row section → Row wise information displays.

② Column section → It displays column wise information.

③ Data section → It may have summarized (or) aggregate.

(or) Normal data against row and column section.

It also supports parameters, actions etc.

→ Take (Drag & Drop) Matrix on the layout

→ Take location in the row section, code in column section,  
Income in data section.

Loc* code	party code
Party Loc.	[sum[PRTYINC]]

→ To change colour, Font etc press F4 (or) generic method  
& change it.

Ex:

STATES* STATISTICS	SALES			Marketing		HR
	10	20	30	40	50	100
<u>AP</u>						
VSKP	8000	7000	6000	5000	70000	60000
RR	4000	30000	40000	40000	6000	70000
Medak	7000	20000	10000	30000	50000	80000
<u>UP</u>						
Alahabad	6000	71000	60000	50000	70000	50000
---	4000	60000	80000	70000	60000	70000
<u>MP</u>						
---	---	---	---	---	---	---



State → Row Group

District → Sub Group of Main row.

Sales → Column Group.

Department → Column sub Group.

Sum of rupees → Data section.

→ We can add groups in the row wise (or) column wise.

Report → Rclick → Insert row

→ Matrix report supports tablex properties.

→ Taking subtotal within the matrix report

a) Goto row section → Rclick → click insert row → OUTSIDE Group below

b) GOTO column SECTION → " → " → OUTSIDE Group right.

List Report - It displays list of contents.

→ It uses tablex properties.

→ This is used to display columnar reports (or) table reports

List box properties are rectangle properties:

→ Drag and drop Text boxes and design like below

[Party ID]	[Party ID]
[Party Name]	[Party Name]
[Party Loc]	[Party Loc]
[Party INCOME]	[Party INCOME]

To change the font, colour..etc goto properties (or) press F4.

→ Take line from Tool box of report items and take before (or) after (or) where ever you require in the list report.

- Displaying each row content in separate (44) page.
- Goto rectangle properties → General → Name: **Rectangle1**
- Page break options → Select 1<sup>st</sup> & 2<sup>nd</sup> options.
- Rectangle properties:-** change name and page break options.

↳ General → Name:

**Rectangle1**

Page break options:

- Add a page break before
- Add a page break after
- Omit border on the page break

↳ Visibility

↳ Fill

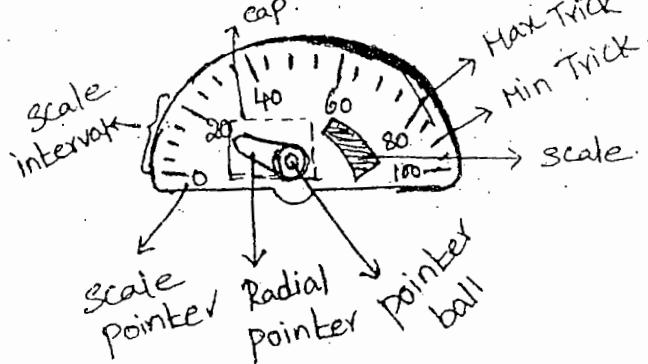
↳ Border



**Gauge Report:-** Newly added in 2008.

- It displays information like industrial meters, bike speedometer.
- 2 types of Guages.

(a) Radial Guage



(b) Linear Guage

- Take a guage in one of the column in the existing report.
- specify a radial pointer (or) Linear pointer as party code.
- To change the pointer values (or) work with pointers go to properties.

## → Pointer Options

Value: = Fields! partycode.value

Expression: = Fields! partycode.value

Pointer Type:  Needle

Needle

Marker

Bar

Needle style:

Tapered

Tapered with style

Placement and size options:

Placement (relative to scale) Distance

Inside

0

Width

15

→ Action

→ pointer fill

→ pointer border

→ Cap options

→ Cap Fill

→ Shadow → click ok

## Gauge panel properties

① General →

Name:

Tool Tip: MSBI Gauge

Dataset Name:

DB-MSBI

page break options:

⑥ Visibility

⑦ Filter

⑧ Fill

⑨ Borders

## range properties

### @ General

Start range at scale value:  End range at scale value:

Placement relative to scale:

Inside

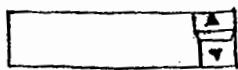
(other options outside, cross)

Distance from scale:

Size options:

Start width

End width



⑥ Visibility ⑦ ACTION ⑧ FILL ⑨ BORDER ⑩ SHADOW

## scale properties

### @ General

Reverse scale direction

Value and interval options

Minimum:

Maximum:

Interval:

Interval offset:

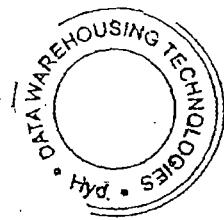
⑪ Action ⑫ Layout ⑬ Labels ⑭ Label Font ⑮ Number ⑯ Major Tick

Marks ⑰ Minor Tick Marks ⑱ Fill ⑲ Border ⑳ Shadow.

↓  
Width, Length ↑

↓  
10

width: 3 length: 20



### Working with report models:-

These are temp tables, contain dataset and data sources. It stores attributes which are supposed to display in various reports. Before taking attributes from various tables the tables <sup>should</sup> contain primary (or) foreign keys. These model statistics are stored under SMDL (semantic Model Definition Language).

Real Time Usage: It acts like a temp table to set of tables and columns. So no need to create data source, datasets on the fly <sup>for</sup> reports. (dynamic reports).

#### Steps:-

1. Ensure tables are having valid keys in the source database.
2. Goto start menu → programs → open BIDS → New project → select → Report model project → specify the name: Temp\_Models → View → Solution explorer →

- ① Create a data source → RT click → datasource view →
- ② Add new datasource view → Next → Next → Next → Select the option "use current model statistics stored in the data source view" → Next → RUN → Finish.
- ③ Goto project → RT click → properties → Deployment → Target Server URL → specify URL → click OK.
- ④ Build menu → Deploy Temp\_Models
- ⑤ Goto report manager and observe the report.



### Working with Report Builder:-

This is designed to create and work with on the fly (or) dynamic (or) situational reports.

There are two versions of report builders are available

- (a) Report builder 1.0 and
- (b) Report builder 2.0

2008 supports both these versions.

IE uses report model to generate the reports. So there is no need to create datasource and dataset.

Goto report manager → click report builder → Run → Select report model and report layout (Table (columnar), Matrix (cross Tab), chart) → click OK.

Take Design Report from menu → create a file and click run report option to run the report.

### Filtering the report:-

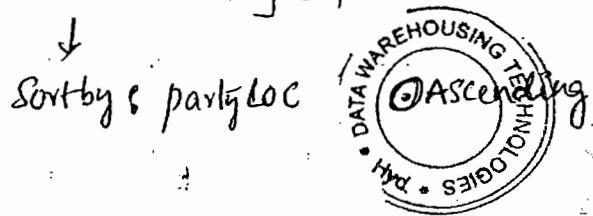
Goto report menu → click Filter.

(or) (49)  
Goto Toolbar → click Filter → click party with: →  
select from fields → partyLoc equals

### Sorting and Grouping:-

Goto report menu → click sort and group  
(or)

Goto Toolbar → click sort and group.



pagebreaks between groups  
<Select required option>

"OK"

### View Menu :-

- Design Report
- Run Report
- TOOL Bars
- Explorer
- Task pane ETC - - -

### Insert Menu :-

- Textbox
- picture
- Filter Description
- Total Rowcount

### Format Menu:- Number, Alignment, font, border, Fill

### Report Menu:-

- filter
- New Field
- Sort and Group
- Report properties

- Report builder supports below type of data regions
- @ Table
- Ⓛ Matrix
- Ⓜ chart



Report Builder → Report properties

choose the properties that you want to apply to this report.

User sorting

Allow users to sort the report data when they view it

Drill through

Allow users to drill to this report from other reports.

Fixed Headers

Lock down the report headings so that they are always visible when users scroll through the reports.

Drill down

→ We can use below report items in the layout.

Insert menu → Textbox

↓  
→ Picture

↓  
→ Total Row Count

Creating New fields

Report Menu → New field → specify field name →

Specify Formula for each party (Party income \* 12/100) → OK.

Report Saving: once we save the report it saves directly

in the Report Server (or) Report manager.

File menu → Save as → specify Server location → specify name → Save

(51)

## WORKING WITH REPORT MANAGER

1. It is a client tool to work with Client Reports and Manager Reports.
2. It does various operations such as
  - a. Subscription
  - b. creating and Manipulating Folders.
  - c. creating Snapshots, Caches, Maintaining Histories
  - d. creating and working with Linked Reports etc.

Navigation:-

<http://servername:IP/REPORTS>



### SHOW DETAILS :-

It displays Various Report Manager Items.  
They are creation date, Modified date, Created Users and Modified User Names etc....

### NEW FOLDER :-

It creates a new folder in the Report Manager.

Click New Folder → Name: → OK.

NEW Data Source → Data Source Credentials

Should be stored permanently in the

Report Server in the below situation.

a. Creating Caches, Snapshots, History

b. While creating Subscriptions



### CREATING A REPORT FROM EXCEL SHEET

#### CREATE A DRIVER FOR EXCEL SHEET

START → SETTINGS → CONTROL PANEL

→ ADMINISTRATIVE TOOLS → DATABASES

→ RT CLICK → ADD → ODBC MS EXCEL SETUP

→ DATA SOURCE NAME → DSE EXCEL →

SELECT WORK BOOK → DESKTOP → FOLDER

→ OK → OK → OK.

→ VIEW → REPORT DATA → NEW DATA SOURCE

→ GENERAL → NAME → DST EXCEL →

EMBEDDED CONNECTION → TYPE (ODBC) →

EDIT → CLICK OPTION STRING CONNECTION

(53)

→ BUILD → MACHINE DATA SOURCE

DRIVER NAME: DST\_EXCEL → CLICK OK →  
OK → OK.

→ NEW DATA SET → NAME → DST\_EXCEL

→ CLICK QUERY BUILDER → QUERY DESIGNER  
→ OK → OK

→ VIEW MENU → REPORT DATA  
TOOL BOX PROPERTIES → TABLE →  
DRAW ON TO THE LAYOUT.



⑤ While working with site settings... etc. New Database click 54

→ Name: DS-DB-MSBI

Description:

Data Source Type: Microsoft SQL Server

connection string: Data Source = .; Initial catalog = DB-MSBI

Connect using:

⑥ Credentials stored securely in the Report server.

User Name: Rawan\administrator

Password:

use the windows authentication.

click → apply.

**UPLOAD FILE:-** It uploads files to report manager.

Upload File menu → click → File to upload:  [Browse]

→ click OK

### Working with individual properties of a Report

View → It displays the report.

Properties →

⇒ @ General → It displays

Modified Date:

Modified By:

Created Date:

Created By:

Size:

#### Properties

Name: Party\_Report

Description:

#### Report Definition

Edit update

## Updating Report Definition:-

This option is helpful to directly update the definition of report without deploying from its position.

Report definition → update → File to upload:  Browse  
 → specify the rdl file → click OK → Apply.

Move: Moving the report to specific folder.

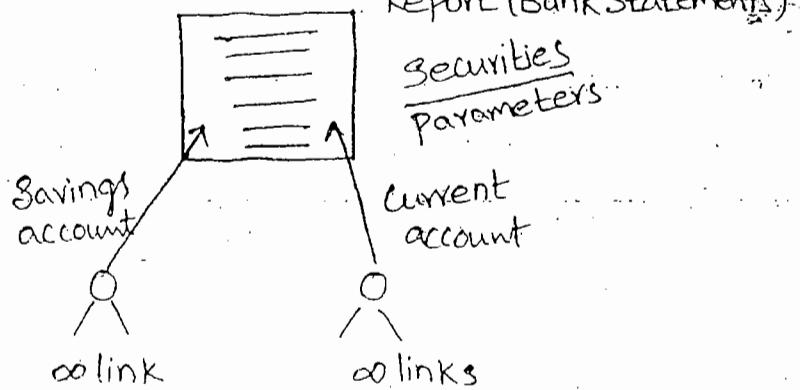
Delete: Deleting the report.



## Creating and Working with Linked Report:-

Create a linked report when you want to use different securities and parameters <sup>with</sup> the report (then it is recommendable).

Eg:-



From the above example Savings & Current accounts having their links to main report. Now connect and see the required parameter content from the report.

click create linked report → Name → change location → ok

Go to the corresponding folders & observe the Links.

(56) Creation.

- ⑥ Data Sources - Same as above like Datasource, in addition it displays the existing data source also (like page no 54)

### Working with Site Settings

It's having Shareable options across reports.

#### @ General

##### Properties

Name:

Select the default settings for report history:

① Keep an unlimited no of snapshots in report history

② Limit the copies of report history:

##### Report Execution Timeout:

③ Do not timeout report execution.

④ Limit report execution to the following no of seconds:

##### Custom Reports Builder launch URL:

URL:

Apply

#### ⑥ Security

#### ⑦ Schedules → <sup>click</sup> New schedule

Schedule Name: SC-SHARED

##### Schedule details

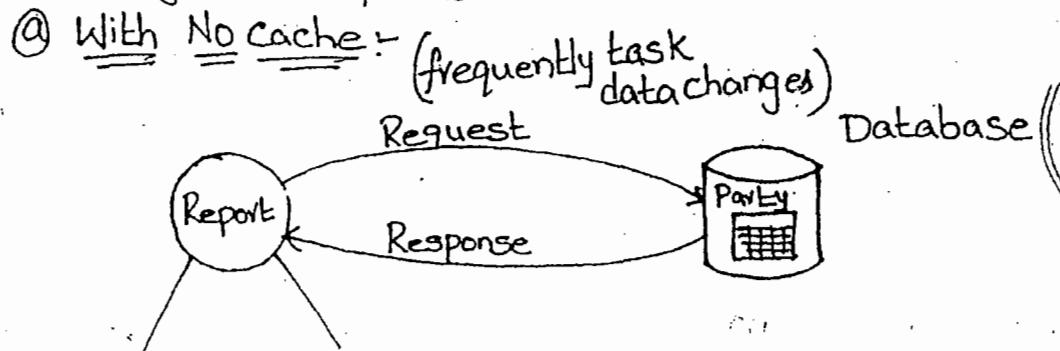
Define a schedule that runs on hourly, daily, weekly, monthly or one time basis.

Note: The above Security, Schedule, General options should be across all reports.

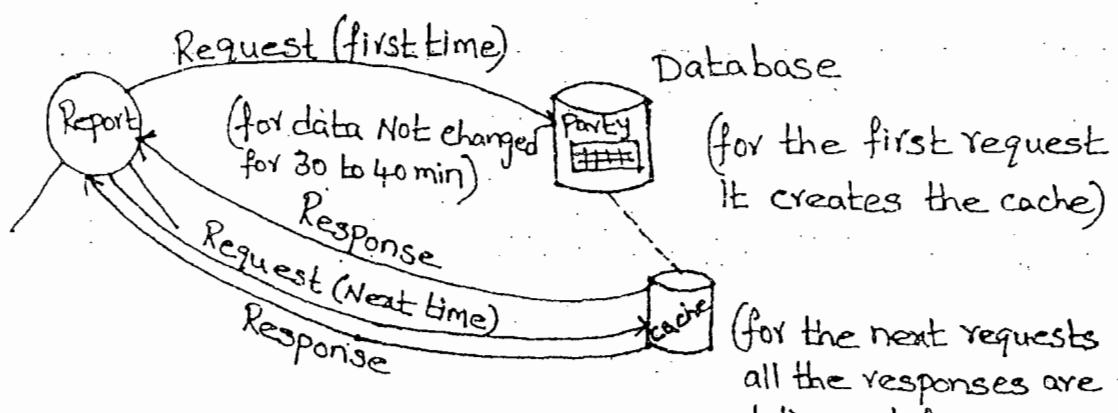
→ Upload File: It is used to upload any files to report manager.

① Execution: Every report executes either of below ways:

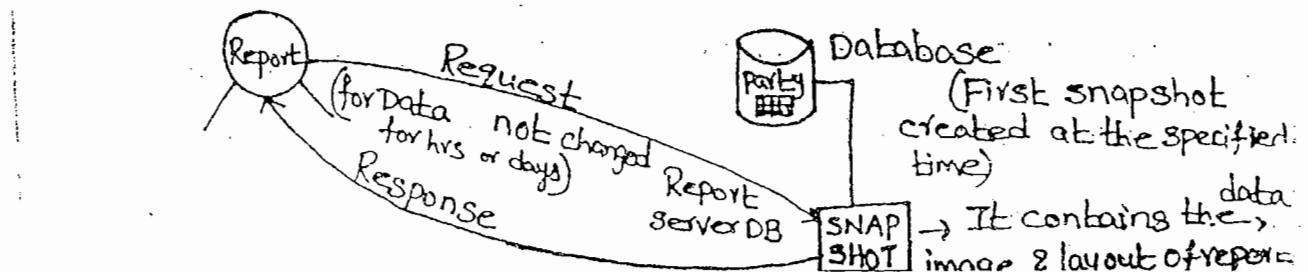
- ④ With No cache → If data is changing frequently.
- ⑤ With cache → If data changing for 30 to 40 minutes.
- ⑥ With snapshot → If data changing after hrs (or) day, we go for snapshots.



⑤ With Cache:



⑥ With Snapshot: It Maintains History.



## CACHES

1. No start time but end time is available
2. Created for the first request of the user.
3. No history is maintained
4. For small data reporting and frequently changed data (every 30 min, 40 min) it is recommendable.
5. Eg: changing mobile no in internet it takes 30min to update during the time we fire any no of queries it gets old number.

## SNAPSHOTS

(58)

1. There is start time but no end time.
2. Created at the specified time.
3. History is maintained.
4. For long running queries, for huge data we go for snapshots.
5. Eg: Sensex data is not changing evening 3:50 PM to next day morning 9:00 am.

No cache: If the table data is changing frequently we can not choose cache or snapshot because we need to get latest data everytime.

Properties → Execution →

- ① Always run this report with the most recent data
- ② Do not cache temporary copies of this report

Cache Implementation: For Cache there is start time but no end time.

- (59)
- ② Cache a temporary copy of the report. Expire copy of report after a number of minutes:
  - ③ Cache a temporary copy of the report. Expire copy of report on the following schedule.
    - ① Report-specific schedule
    - ② Shared schedule

To see the cache creation and separation:-

Goto → SSMS → Database Engine → Report server temp\_DB  
 → Execution cache

### Snapshot Implementation:-

Execution →

- ① Render this report from a report execution snapshot
  - use the following schedule to create report execution snapshots:

- ② Report - specific schedule

click Apply

- ⇒ ③ History → Here the settings related to history available like manual History maintenance, Schedule History maintenance etc.
- In times of SHM we need to specify start time to move snapshot to history
  - We can limit the No of snapshots in a history.

- Allow the report history to be created manually.
- Store all report execution snapshots in history
- use the following schedule to add snapshots to report history...

- ④ Report - specific schedule

At 9:02PM every MON of every week, starting 9/21/2011



Select the no of snapshots to keep: (60)

① Use default settings.

② Limit the copies of report history: [100]

[Apply] ←

Navigation:

Database engine

↳ report server database

    ↳ snapshot data

        ↳ monitor

            ↳ snapshot information.



Database engine → report server database → History.

⇒ ② Security: RS supports Role based history.

The default roles available are:

(i) Browser → May view folders, reports and subscribe to reports.

(ii) Content Manager

(iii) My Reports

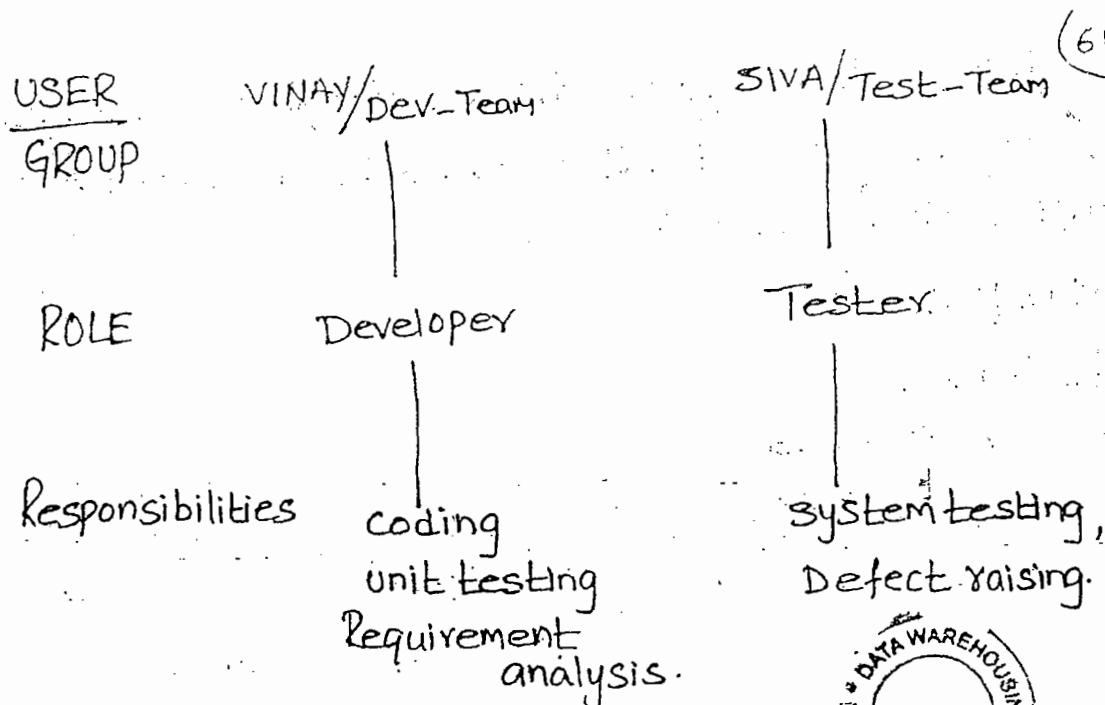
(iv) Publisher → May publish reports & linked reports to the report server.

(v) Report Builder → May view report definitions.

→ Every role will have set of responsibilities.

→ Users (or) groups assigned to those default roles, so that they perform the underlined responsibilities of the role.

Note: In 2005 we can create our own role and we can assign required responsibilities. But 2008 doesn't support own role and responsibilities creation.



### Navigation:

#### Security

↳ New Role Assignment

↳ specify Group or User Name:

(local computer n.).

↳ click the option Report builder

↳ click ok.

### Subscriptions:

Delivering the report to the corresponding persons in the corresponding format at a stipulated time is called Subscription.

These are of two ways to deliver the reports

They are:- ① File Share

② E-mail.

Note: We can deliver the report through customized extensions also.

2 types of Subscriptions

→ Standard Type

→ Data driven type.

(62)

To deliver the reports through E-mail we must perform  
(or) we must set E-mail settings in the report server configuration manager.

### Navigation:-

Start menu

    └→ programs

        └→ SQL Server 2008

            └→ configuration tools

                └→ Reporting Services

                    configuration Manager



connect:

    └→ E-Mail settings

        SMTP settings

To edit, change the fields & click the apply button.

### Standard Subscription:-

Here to the specified user, the report is delivered and it is event driven (schedule).

Here parameter values and rendering format can be set along with the report.

Generally small scale reports delivered through this.

Goto the individual reports → click Subscription → click New subscription.

#### Report Delivery options:

Specify options for report delivery.

Delivered by: Windows File share.

File Name: Party\_Report - Today

(63)

path: \\rawan\reports.

Render format: XML File with report data.

Credentials used      user name  
to access the file      password  
Share:

Overwrite options:

- ① Overwrite an existing file with a newer version.

### E-Mail Subscription:

Goto Report → Select New Subscription

Delivered by: E-mail

To: Administrator @ dwht.com

Cc: Vinay @ dwht.com

Bcc:

Reply To:

Subject: @ Report Name was executed at @ Execution Time.

Include Report Render Format:

MHTML(Archive)

Include Link

Priority: High

Comment:  

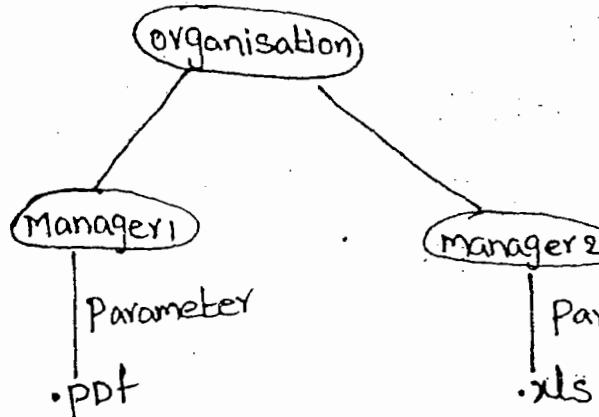
When the schedule run is complete

Select schedule.

### Data Driven subscription:

Here we deliver to any no of users with their required parameters (according to the reports with generally large parameters we go for this).

(64)



1. Create a table in the database with the below fields

FileName	Path	Render-Format	WriteMode	File	USR	PWD
				Extn.	Name	
Party_Report	\\rawan\\report	Excel	y(yes)	.DOC	Admin	VINAY
Party_Madhu	"	"	"	.xls	VINAY	MADHU

- 2) Go to the reports → Subscriptions → New data driven subscription →

- Step-1 :- create a data-driven Subscription: Party\_Report

Description:

specify how recipients are notified: Windows file share

- Specify for subscription only → click Next.

Step-2:-

Data Source Type: MS SQL Server

connection String: Dataset = Rawan; Initial catalog =

credentials used to access the fileshare      Username: "MSBI-DB";  
    Password:

use as windows credentials

(65)

click Next

Step-3: Enter a query : select \* From muluser → Next

Step-4:-

File Name

① Get the value from the database:

Path

② Get the value from the database:

Render Format

③ Get the value from the database: Render-Format

Write Mode

File Extension

User Name

Password

click Next

Step-5: click Next

Step-6: Notify recipients:

④ on a shared schedule

click Next

Null delivery provider:-

Here recipients are notified without any report.

→ Null delivery doesn't have any settings.

→ IF report cache is enable then the result of subscription will end-up here.

(60)

## Configuration Files:

Reporting Services stores configuration settings in various configuration files.

If we change the configuration files information automatically reporting services work with these settings.

This configuration files are useful when migrating from one environment to another environment.

### RS Web Application .config :-

Here settings of a report related to report manager available.

### RS Report Designer .config :-

Settings related to custom data processing and rendering available.

### RS Report Server .config :-

Here settings related to report manager, report server, subscriptions, security, caching .. etc are available

### Navigation:- Start menu

    ↳ Search option

        ↳ All files (or) folders

            ↳ rsreportserver.config

                ↳ click search.

To check out the config. results goto SQL Server Visual Studio → open and see the XML settings, for manager, Server, Subscriptions.. etc.

Note :- (i) To change the settings admin privileges required. (67)

(ii) These files are created at the time of installation.

(iii) Changes in these settings are not recommendable.

### Command Line utilities and their usage:-

RS config → It manages and encrypts connection settings in the server.

RS key management → It processes the server encrypted key set and writes to the file.

RS Active → It creates the encryption key to activate the server.

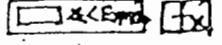
RS → It automates common server management tasks like @ publishing the report.  
⑥ configuring the report...etc.

Syntax: RS -i <Net script file> <Report server>

Eg: RS -i party.css

Note: RS, RS key mgmt ..etc are web services.

Display: alternative color to alternative rows

Goto the fields (select all the required rows) → select text box properties → Fill → Fill color:   
→ Set expression for: Backgroundcolor:  
= IIF(RowNumber(Nothing) mod 2, "light green", "light blue,

(68)

→ Displaying the specified no of rows in every page

Sol: write the below expression in the group Expression  
So that for every 25 records it take a break.

$$= \text{INT}((\text{RowNumber}(\text{Nothing})-1)/25)$$

## Publishing the reports:

CREATE a Report



→ Preview the report

→ Build the report [The report files and all contents move to a folder]

→ publish [Deploy]

Single Report Deploy

Multi report Deploy

Goto Solution Explorer

Report Name

Click Deploy

Note: Before Publishing ensure the following settings.

① Solution Explorer → Project → Properties →

Deployment

overwrite Data Sources: False

Target Data Source Folder: Data Sources

Target Report Folder: MRG - REPORTS

Target server URL: http://rawan-8080/Reports\_EWIT

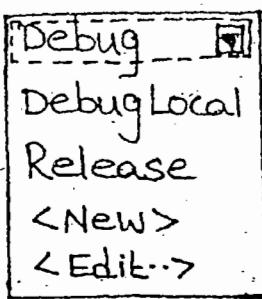
(69)

## ⑥ Configuration Manager →

Project contexts (check the project configuration to build or deploy)

Project configuration platform Build Deploy

MRG-Reports



Debug → Publishing to test environment

Debug Local → Previewing the report in the Local.

Release → Publishing to production server

⑦ The user who is publishing the report should belongs to Publisher Role.

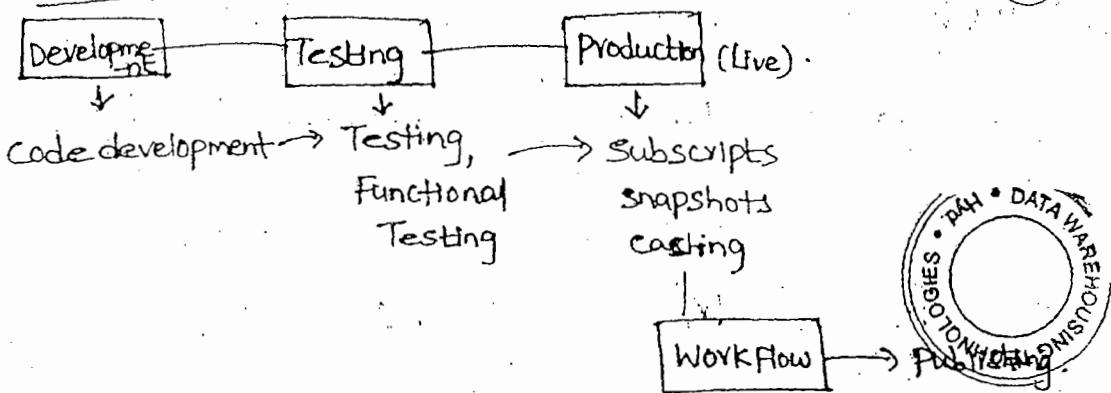
### Developer Responsibilities:-

1. Creating & Working with standard, complex, Adhoc reports by using Report Server project , Report builder.
2. Parametrization, Actions
3. Report Manipulations, Various Data regions, report items
4. Expressions, Calculations writing
5. Data Sources, Datasets, connectivity .
6. Linked reports . . . etc .

### DBA (or) Superuser (or) Managers Responsibilities:-

- |                                   |                               |
|-----------------------------------|-------------------------------|
| 1. publishing reports             | 5. Report definition updation |
| 2. Subscriptions                  | 6. Security . . . etc .       |
| 3. Snapshots, Caches              |                               |
| 4. folders, data sources creation |                               |

(70)



### Complex creation:-

1. complex single report with queried, nonqueried, cascaded multivalue, drill through and drilldown processes.
2. Create partial dashboard by using cross tab (or) Matrix reports.

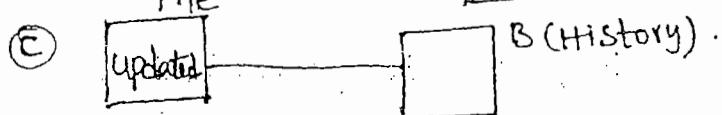
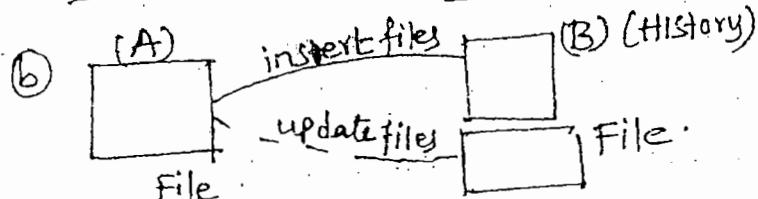
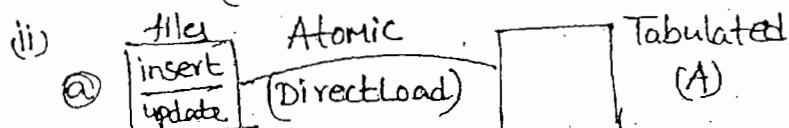
[Various actions implemented]

### SSIS in complex implementation (or) creation:-

1. SCD manual implementation.

(i) class rooms

(or)

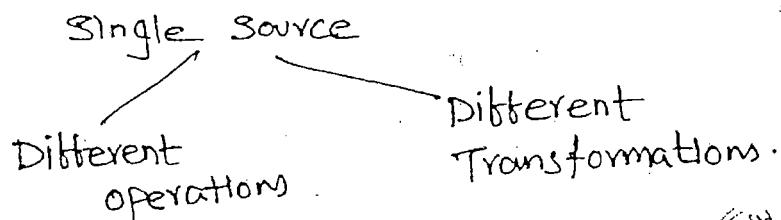


2. Multi files loading using forced load

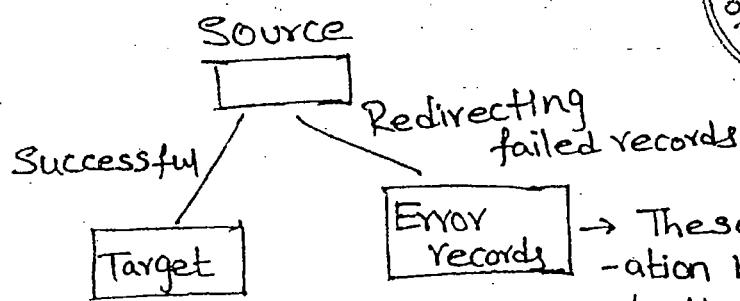
3. Multiple worksheets loading through Script task

4. Incremental loading

5. optimized mapping using various performance tuning methods



6. Error loading records splitting.



These error records information is sent to the client to their rectification.

(or)

We ourselves rectify these error records and reload to the source and send to the required target.

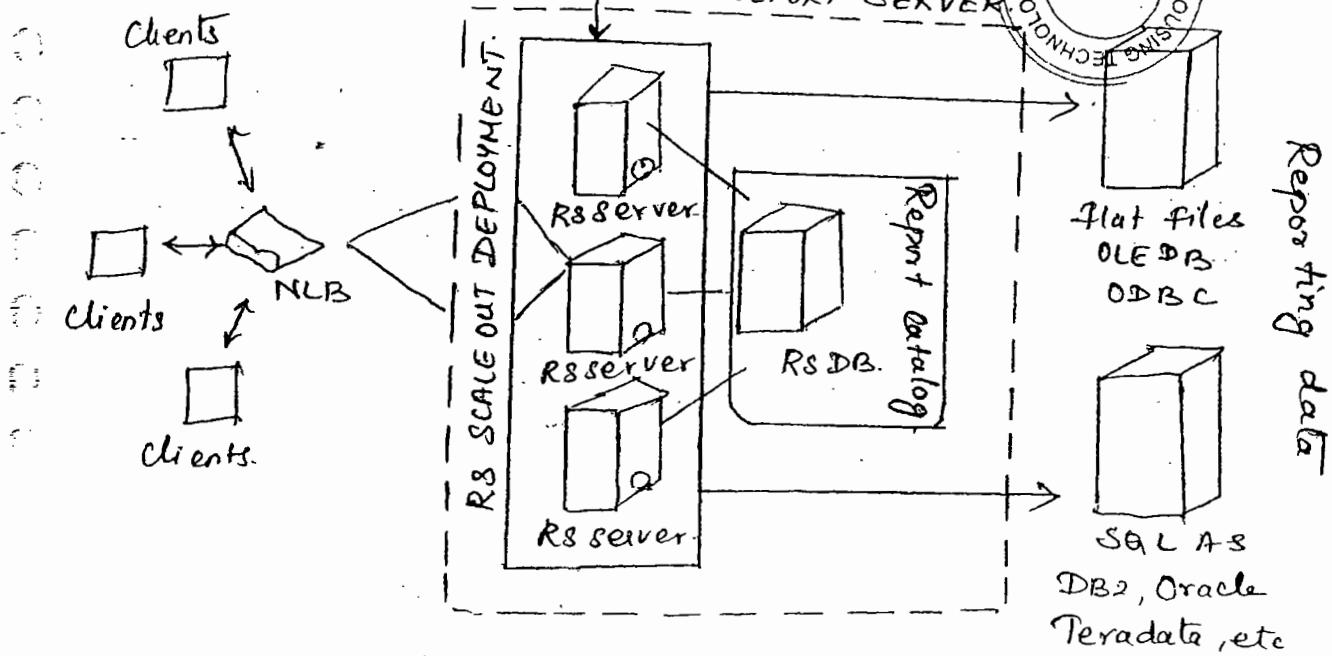
## SSRS PERFORMANCE Tuning

11

### ARCHITECTURE

TYPICAL SCALE OUT REPORTING SERVICES ENVIRONMENT

MARKE IN BOX



Reporting data

### How 64-bit Helps Report Server Service.

1. Most of the Reporting Services reports are memory-intensive, and the additional addressable memory made available by 64 bit will provide with more Scale.
2. For some loads 32 bit can execute faster, especially in scenarios where we have many small reports.

(72)

2. If you have more memory made available by 64-bit we can handle more concurrent Report users. Because there is less memory contention, report throughput will be higher.
4. We can allow more users to view and Export larger reports.
5. In SQL Server 2005 Reporting Services, report's data set is placed into memory; the more concurrent reports; the more memory used.
6. If we use 32-bit we can easily hit the 3GB ceiling, which can also result in Internet Information Services (IIS) process recycles, leading to report failure.

But as noted in the Reporting Services Scale-out Deployment Best Practices, technical note

SQL Server 2008 Reporting Services is not memory bound.

It is able to effectively use the file system in order to move data structure in and out of memory and file system if the Report Server experiences memory pressure.

These memory limits are explicitly configurable in SQL Server 2008 Reporting Services via R8ReportServer.config noted in the Memory Configuration for SQL Server 2008 Reporting Services.

When Reporting Services uses the file system the reports run more slowly, because it is much more efficient to request data from the memory than from disk.

The file system is used only if Reporting Services memory usage gets close to the configured memory limits.

If we overload the report server in SQL Server 2008 Reporting services with a large number of concurrent users and very large reports, all of our reports can still complete processing albeit more slowly because Reporting services can hold all of this data without running out of memory space.

In Enterprise environments, you will eventually run into situations where your servers will need to be able to handle many concurrent users and large load - the optimization in SQL Server Reporting Services (in comparison to SQL Server 2005 Reporting Service) is that while reports may run slower at times, they will complete.

### Exceptions.

Keep in mind that certain data providers are not available for 64-bit (for example the Microsoft JET provider or certain third party providers.) In these cases, customers will need to continue using 32-bit for their Reporting Services Environment.



### HANDLING A LARGE WORKLOAD

Two main issues concerning Enterprise Reporting environments are

- \* 1. The ability to handle concurrent user load and
- \* 2. Ability to handle a large workload (that is large report)

To mitigate the concurrency issue solution is to scale out to multiple report servers to handle user query load in Reporting Services scale out deployment.

Q  
O  
O  
D  
O  
D  
O  
F

To get the HIGHEST PERFORMANCE when handling large workloads that include user request for large reports. (15)

### CONTROL THE SIZE OF YOUR REPORTS

1. First want to determine the Purpose of these Reports and whether a large multipage report is even necessary.
2. If a large report is necessary, draft how frequently used. If provide users with smaller summary reports, can you reduce the frequency with which users attempt to access this large multipage report?
3. Large reports have a significant processing load on the report server, the report server catalog, and report data, so it is necessary to evaluate each report on a case-by-case basis.

Some common problems with those large reports are they contain data fields that are not used in the report or they contain duplicate datasets.

Often users retrieve more data than they really need. To significantly reduce the load placed on your Reporting Services environment,

(76)

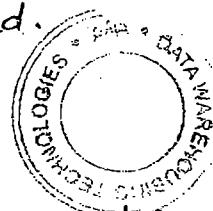
Create summary reports that use aggregates created at the data source and include only the necessary columns.

If you want to provide data feeds you can do this asynchronously using more appropriate tools such as SQL Server Integration Services, to provide the field data feed.

### USE CACHE EXECUTION

If you have Reports that do not need to have live execution, enable the Cache execution setting for each of your appropriate reports.

This settings causes the report Server to Cache a temporary copy of those reports in memory.



## CONFIGURE AND SCHEDULE YOUR REPORTS

1. For your large reports, use the Report Execution Timeouts setting to control how long a report can execute before it times out.
2. Some Reports simply need a long time to Run, so timeouts will not help you. There, but if reports are based on bad or runaway queries, execution timeouts ensure that resources are not being inappropriately utilized.
3. If you have Large reports that create data processing bottlenecks, you can mitigate resource contention issues by using Scheduled Snapshots. Instead of the Report data itself, a regularly scheduled report execution snapshot is used to render the report.
4. The Scheduled Snapshot can be executed during off-peak hours, leaving more resources available for live report users during peak hours.

## DELIVER RENDERED REPORTS FOR NONBROWSER

### FORMATS

1. The rendering performance of non browser formats such as PDF and XLS has improved in SQL Server 2008 Reporting Services.
2. To Reduce the load on your SQL Server Reporting Services Environment, you can place nonbrowser format reports on to a file share and/or SharePoint team services, so users can access the file directly instead of continually regenerating the report.

## POPULATE THE REPORT CACHE BY USING DATA DRIVEN SUBSCRIPTION FOR PARAMETERISED REPORTS

1. For Large Parameterised Reports you can Improve performance by prepopulating the report cache using data-driven subscriptions.
2. Data -driven subscriptions enable easier population of the cache for set Combinations of parameter values that are frequently used when the Parameterized report is executed.

Q

Q 3. Note

(79)

If you choose a set of parameters that are not used, you take on the cost of running the cache with little value in return.

Q 4. To Identify the more frequent parameter value combinations, analyze the ExecutionLog2 view.

Q 5. Ultimately when a user opens a report, the Report Server can now use a cached copy of the report instead of creating the report on demand.

6. You can schedule and populate the report cache by using Data Driven Subscriptions.

### BACK TO THE REPORT CATALOGS

You can also increase the sizes of your Report Server Catalog which allows the database to store more of the snapshot data.

### RUNNING THE WEB SERVICE

IIS and Http.sys tuning helps get the last incremental performance out of the Report Server computer. The low level options allow you to change the length of the HTTP request queue, the duration that

(80)

Connections are kept alive, and so on.

For large concurrent Reporting loads, it may be necessary to change these settings to allow your Server Computer to accept enough requests to fully utilize the Server Resources.

You should consider this only if your servers are at maximum load and you do not see full Resource Utilization or If you experience connection failures to the Reporting Services Process.

- For SQL SERVER 2005 Reporting Services

↳ tune IIS

- For SQL SERVER 2008 Reporting Services

↳ tune Http.sys within the Operating System : Windows @ 2003 or windows 2008 :

## MONITORING BY USING EXECUTION LOG2

The Reporting Services Execution Log2 View is a good starting point from which to analyze your current workload and understand its dataset size, performance and complexity characteristics.

In particular, this view contains a new AdditionalInfo column.

ExecutionLog2.AdditionalInfo contains information related to the size of memory-pressure responding data structures.

One way this information can be useful is to check whether you have reports with high values (10s or 100s of MBs) - these reports might be candidates for further review, focusing on the design of those reports and the dataset query sizes.

\* \* \* Some tips on how to view the Execution Log2 view to quickly understand potential performance bottlenecks are given below [P.70]

(82)

Linked is the Review Execution Logs Reporting

Services project which create summary and detail Reporting Services Reports on the last 1,000 Entries into the Execution Log View.

### Execution Log Summary

Summary report of Last 1000 Execution Log Events

Format	Request type	Report Action	Source	Status Log count
	Interactive	Drill through	Session	Success 28
	Interactive	Sort	Session	Success 5
	Interactive	Toggle	Session	Success 139
RPL	Interactive	Render	Live	Rs Processing 1 Aborted
RPL	Interactive	Render	Live	Success 67

Fig: Review Execution Logs (Execution Log)  
Summary Report.

## EXECUTION LOG DETAILS

103'

Query the most recent 1000 execution log details to better understand what is happening with RS server

1	2	3	4	5	6
Instance Name	ReportPath	User Name	Request Type	Format	Parameters →
MyServer	/ReviewExe logs/Exe Log	UserA	Interactive	RPL	
MyServer	/Review Exe Logs/Exe Log	UserA	Interactive	RPL	
MyServer	/SQL Auditing Reports/3- Overview-DDL Actions	UserB	Interactive	RPL	pMin Date = 01/09/2005 0000 00 & Max Date = 01/13/2009 00 00 00 & Report Key = 102

7	8	9
Report Action	Source	Status
Render	Live	RS Success
Render	Live	RS Success
Render	Live	RS Success

(84)

## LONG-RUNNING (LONG-RUNNING)

Sorting by ElapsedSec or RowCount helps you identify long-running reports.

If the value for TimeDataRetrieval is high, the data source is your bottleneck and you may want to optimize.

If there is a high value for RowCount, a lot of data is being retrieved and aggregated by Reporting Services - Perhaps have your data source do this to reduce the load on your Report Server.

## SUBSCRIPTIONS OR INTERACTIVE

Sorting by the RequestType field allows you to determine whether you have a lot of subscriptions;

You can then determine the bottlenecks and stagger-schedule the reports (that is schedule the subscription execution times of the reports at different times).



## LIVE DATA OR SNAPSHOTS

Sorting by the Source field Allows you to determine whether your reports are Typically Live data or Snapshots.

If the Reports can be Snapshots (for example, Yesterday's report), Create snapshots so you can avoid query execution, Report processing, and Report Rendering.



## LOAD BALANCED

Sorting by the Instance field can help you see whether your network load balancer is handling Report Request In a balanced fashion. You can also see if some nodes are down or not processing Request.



