

# EDWH Architecture for Global Data Loading Strategy



## Applies to:

System Architecture for EDWH having multiple regions across globe. SAP NetWeaver 2004s BI 7.0 version.

For more information, visit the [Business Intelligence homepage](#).

## Summary

This Whitepaper summarizes how to design an EDWH in SAP BW system for an organization where the operations involved are across globe.

Main intention of this article is to explain how to achieve 24 X 7 data availability for the business reporting and how to overcome the challenges associated with it by designing the EDWH as above.

**Authors:** Rahul Kulkarni / Maxim Sychevskiy

**Company:** Infosys Technologies Ltd.

**Created on:** 15 April 2009

## Author Bio

Rahul Kulkarni is certified BI consultant with over 7 years of industry experience including over 4 years of consulting in SAP BW / BI projects.

Maxim Sychevskiy is certified BI and Enterprise Architect with over 7 years of Business Consulting experience including over 6 years of experience in SAP systems implementations mainly focused on SAP BI and SCM.

Rahul and Maxim are presently working with Infosys Technologies Ltd. They are working on SAP BI 7.0 and are mainly responsible for execution of SAP BW/BI projects.

## Table of Contents

Business Scenario .....	3
EDWH Architecture.....	4
Modeling Data Acquisition Layer .....	5
Structuring DataMart Layer.....	7
Local DataMart layers .....	7
Regional DataMart layer .....	7
Global DataMart layer .....	7
Integrated DataMart Layer .....	8
Reporting / Analytical Layer Designing:.....	8
Transaction Data Loading Strategy .....	9
Master Data Loading Strategy .....	10
Achieving 24 X 7 Data Loading Requirements.....	11
Consideration for 24 X 7 Scenario .....	11
Steps in Data Loading.....	12
Time Zone Considerations During Loading .....	12
Concept to Handle Volume and Frequency.....	13
Summary.....	14
Disclaimer and Liability Notice.....	15

## Business Scenario

Hardware and software evolution with its low cost storage capacity and servers virtualization made it possible to consolidate enterprise data warehouse solutions in one single SAP BW system to cover the enterprise wide reporting structure. Many organizations, especially those who have a global footprint, realized the advantages of globally harmonized reporting solutions and already started BW consolidation projects or consider starting them in the nearest future.

Many questions arise immediately after this decision is made. For example these questions we have faced from our customers: How to deal with complexity of the new consolidated environment spread across the globe; how the system could be 24x7 available; How to deal with different volumes and frequency of data loads; how it can be efficient from performance point of view; how the workload could be balanced, etc.

The following document explains how to overcome these issues by defining a sound Data Load Strategy as part of EDWH architecture while ensuring consistency in current and future reporting solutions.

## EDWH Architecture

Global consolidated BW systems tend to be more complex when several regional systems taken altogether due to additional effort required for initial design and build and then to govern the maintenance and design changes. There are many studies available to explain how to deal with complexity emerged with consolidated architectures. Among them are “Simple Architectures for Complex Enterprises” by R. Sessions, TOGAF 9, etc. The main point of these studies is that consolidated architecture needs to be partitioned in order to reduce the complexity of overall architecture.

Partitioning of BW architecture as Common Data Layer of Enterprise Architecture could be done by applying several layers where each layer has its own optimized structure and services for the administration of data within an enterprise data warehouse. This to ensure less cross layer interactions and less interdependencies, which in consequence reduce the overall complexity.

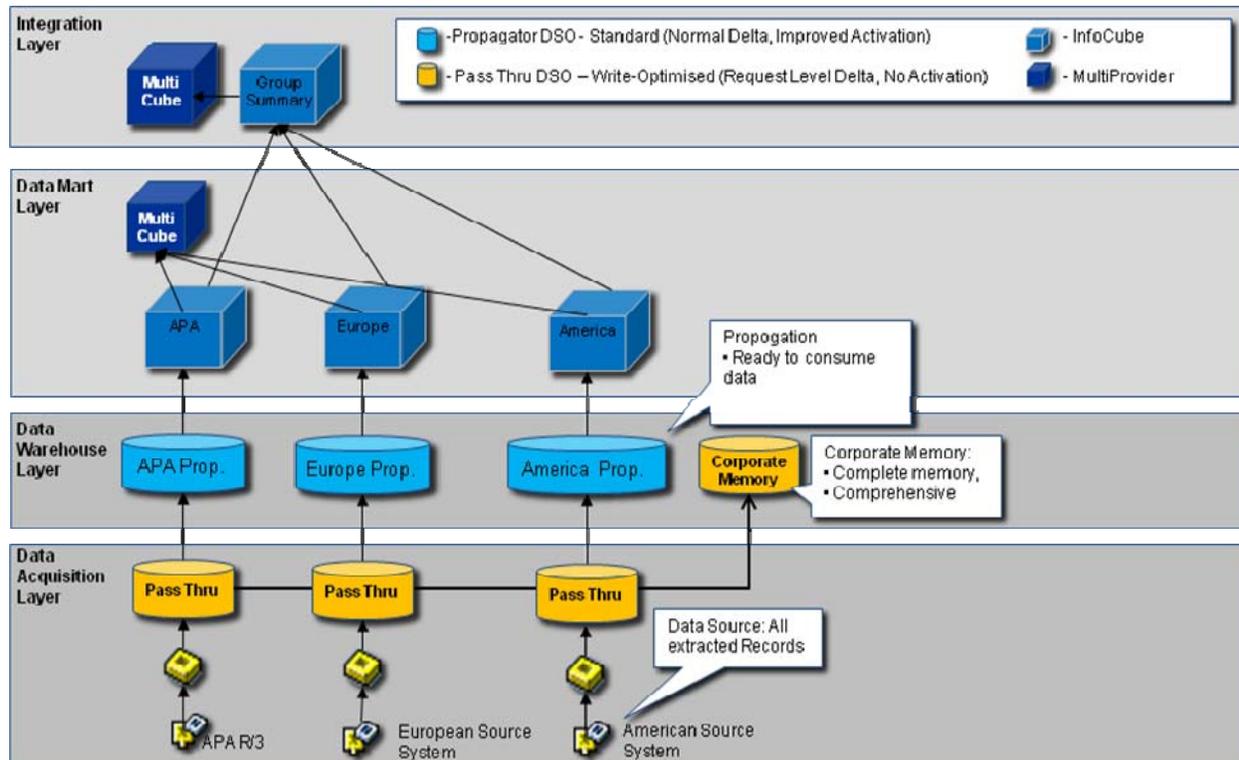


Fig. 1.1 Recommended EDWH Architecture

For SAP NetWeaver BW 7.0 the recommended Layered System Architecture is as explained in (fig. 1.1).

**Data Acquisition layer**

- Deals with collecting the data from various source systems and hold it persistently.
- Also this layer helps to decouple the source system from EDW system by loading data fast from source system.
- PSA works as an Acquisition layer.

**Data Warehouse layer**

- Modeled using DSO
- Corporate Information Repository of EDW.
- No aggregation of reporting-relevant data.
- Common master data definition to retain cross-system integration of system-dependant master data.
- Normally no reporting targets – Exception: operational reporting on line items.

**Data Mart Layer**

- Delivers benefits like Common Master data definition
- Data Manipulation with business logic, e.g. Calculation of process time KPI.
- Modeled using InfoCubes & DataStore objects.
- Analysis and reporting layer.

**Integration Layer**

- Modeled mainly for fulfilling global and cross function reporting requirements.
- All reports & KPIs having common definition across the regions in the organization are defined on the reporting objects in this layer.

**Modeling Data Acquisition Layer**

It is not an unusual situation when organizations have operations across several regions like Asia Pacific, Europe, Japan, North America, LatAm and each region consists of several sub-regions - countries. Each of those countries usually have separate source systems which could but not always represented as separate clients of regional SAP R/3 (ECC) system.

This results in complex loading schedule, with different loading windows for every country (SAP R/3-ECC client) need to be established.

Data Acquisition layer design for data retrieval from different source systems in the scenario above is most important while designing EDWH architecture. It should have characteristics like:

- Load data as it is without transformation
- Quick transfer of data from source system to allow multiple times data load.
- Should extract all available fields in the DataSource even though these fields are not required for reporting currently but may be required in future. This also facilitates to minimize modification in the DSO and DataSources when requirement arise in future.



## Structuring DataMart Layer

The DataMart layer can also be partitioned into different layers according to different data requirements, like those specified by TOGAF: Level of detail (data granularity), Viewpoint, Subject Matter etc. Many organizations decide that it can be local DataMart, regional DataMart and global DataMart.

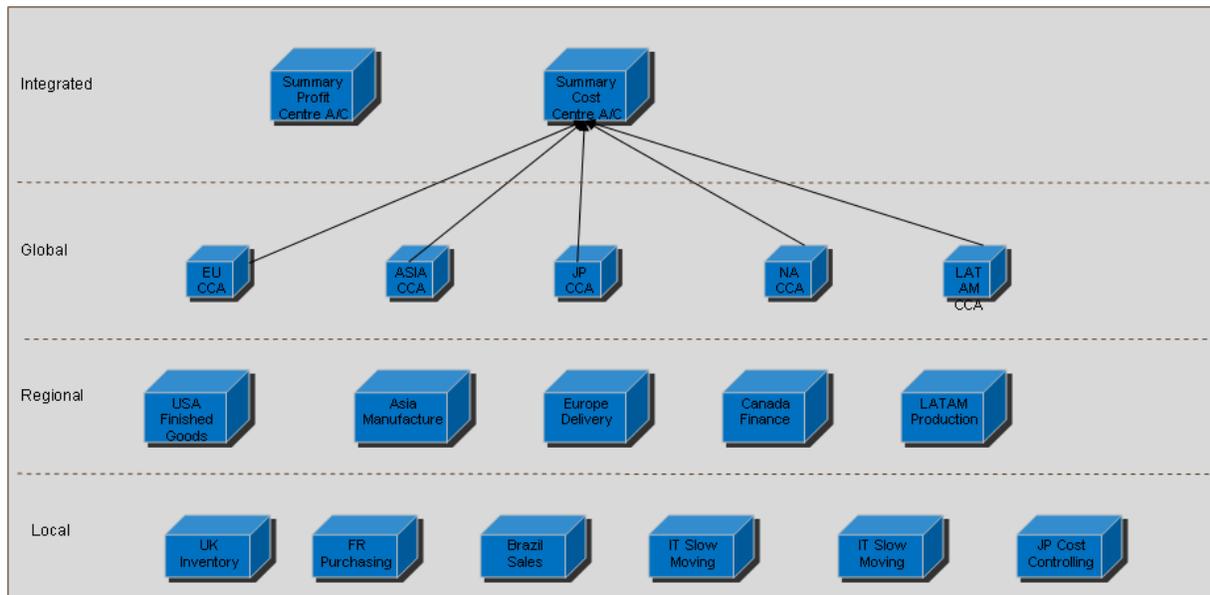


Fig. 3.1 Data Acquisition Layer design

### Local DataMart layers

1. It will be satisfying business reporting requirements at a state or country level.
2. Here only objects not having common data definition across countries will be located.
3. Isolating them from common reporting requirements at Regional or Global Level.

### Regional DataMart layer

1. This layer will have objects where there is a common data definition in that particular functional area across all countries in that region.
2. These are similar to the local layer but the data definitions are common in that region but two different regions have different definition.

### Global DataMart layer

1. This layer consists of objects which are having common data definition across all the regions.
2. In the above fig. 3.1 we are having one cube per region e.g. EU CCA, ASIA CCA or NA CCA. All these cubes are having same data model as the reporting requirement and data definition are same across the regions. The obvious question will come. "Then why there are different cube per region?" This is mainly to facilitate data loading at different time zones and ensure good reporting performance as it will be equal to applying logical partitioning to the cubes.
3. The design of the global layer can change according to volume of data and data load frequency, which will be explained later in this document.

## Integrated DataMart Layer

1. Integrated layer is mainly for Executive Information.
2. Here Cross functional data is combined together. e.g. in above case The PCA data and CCA data at global level is combined together. This layer is hypothetical layer, which can be present in each reporting layer where cross functional reporting is present.

## Reporting / Analytical Layer Designing:

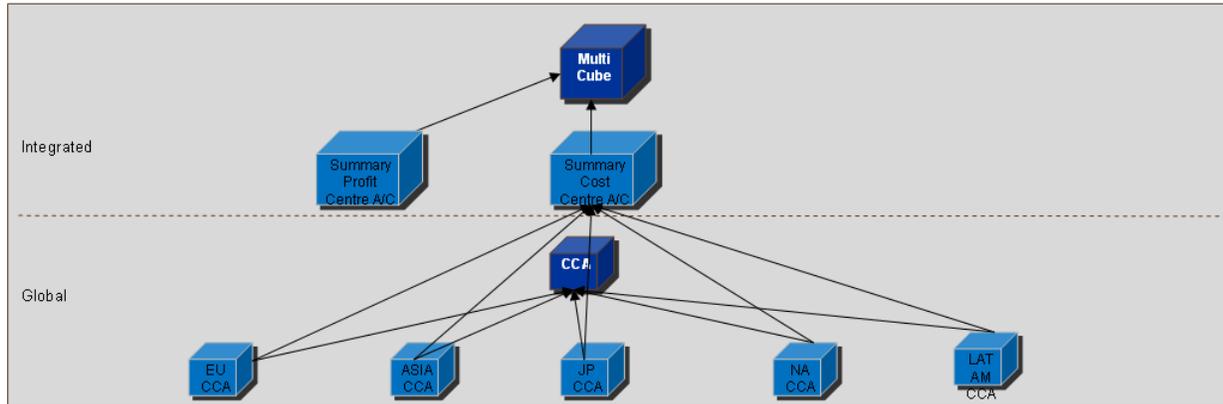


Fig. 4.1 Reporting / Analytical Layer design

1. As a standard practice all the reporting will be done on MultiProviders.
2. This makes the underlying structure scalable by providing a scope for addition of further reporting objects.
3. Also all benefits of MultiProvider reporting can be availed.
4. The reporting layer can be at different layers like local, regional, global and integrated. The integrated layer is used at enterprise level mainly for cross functional reporting.

## Transaction Data Loading Strategy

Here we will understand how global transactional data loading strategy can be defined. Following fig. explains the importance of the above data flow design, which is designed so as to enable various time zone specific loads without affecting reporting requirements at different times across the globes. In this fig. we are explaining data loading strategy for a delta enabled business content data source.

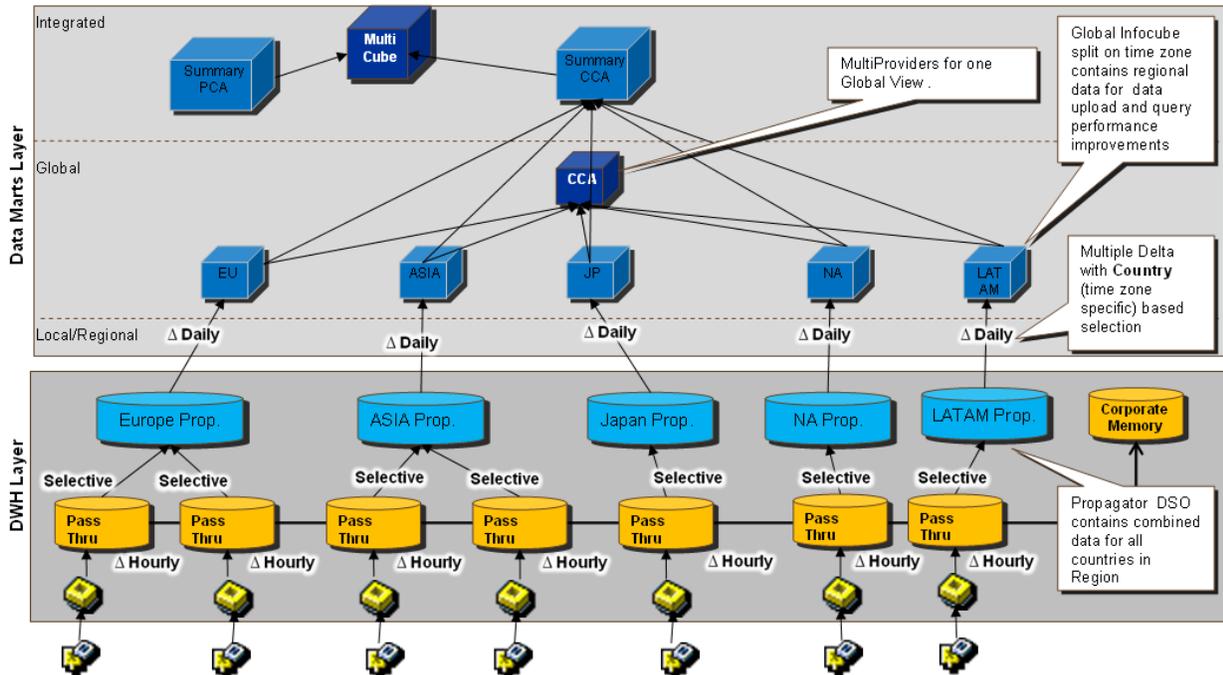


Fig. 4.1 High Level Data loading schema

Steps involved in data loading are as follows

1. As explained earlier, data will be pulled up from the source system on hourly basis to the Pass Thru DSO.
2. From the Pass Thru DSO's data will be pulled up into regional propagation DSO's.
3. Simultaneously it will also load to corporate memory as a single source of historical data in the EDWH.
4. From the propagation layer DSO's data will flow after required transformation to the regional DataMart objects (here InfoCubes like EU CCA. JP CCA etc.)
5. While loading into regional DataMarts we can set different Delta loads with respect to company codes to allow Time zone specific loadings. e.g. in Regions like Asia Pacific the countries involved are geographically situated far away from each other. Here countries like Australia need their data much earlier than countries like India because of time zone difference. So having delta process set according to country code or company code can be advantageous in such case.
6. The MultiProvider named CCA shown above will be used for global reporting.
7. For executive reporting requirements only summarized data is used and that can be loaded to the 'Summary CCA' cube. As a thumb rule MultiProvider are used for cross-functional reporting.

## Master Data Loading Strategy

Like the Transactional data the Master data objects can also be categorized like Local, Regional or Global InfoObjects. These three layers represent different viewpoints of the same entity, like Customer, and designed to store the layer specific data attributes about the entity.

The challenge which comes along with globally consolidated BW system is that master data from all layers need to be updated prior to provide a reference to transactional data loaded according to its own schedule. From overall system point of view it would be a constant cycle of loading transactional data to the system but in different regions which would require master data to be always up-to-date.

Hence these objects will be placed in different layers and will have different data loading strategy according to their position.

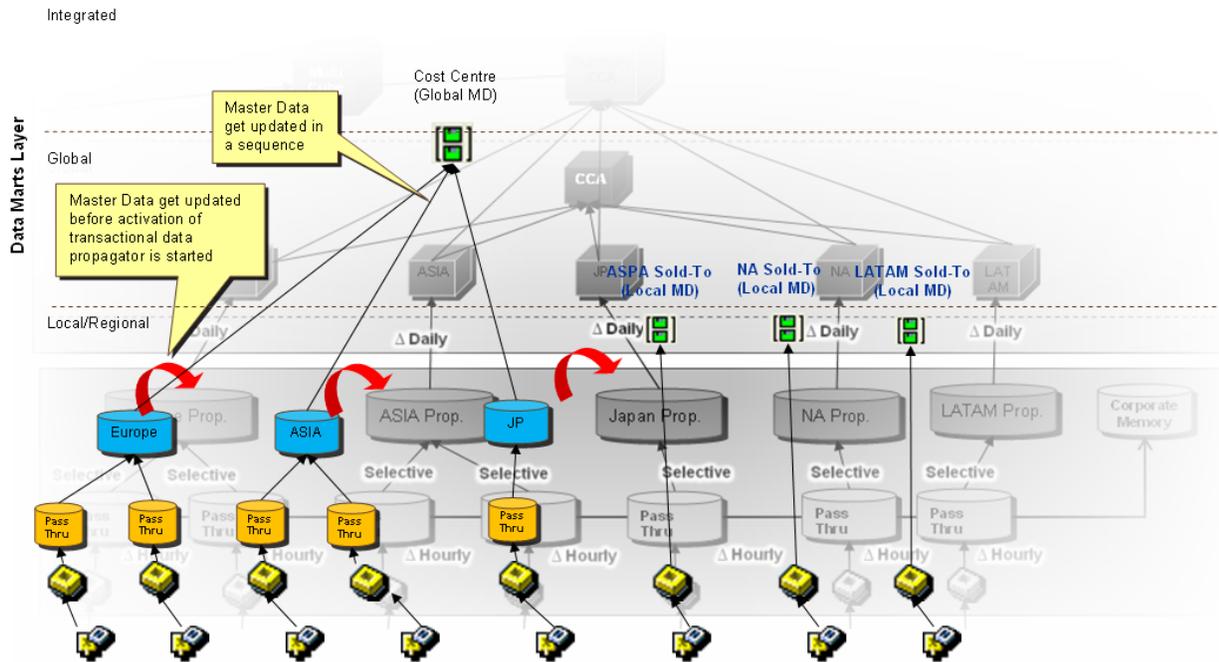


Fig. 5.1 Master Data (Attribute / Text) Loading strategy.

1. The Local InfoObjects e.g. country specific customer number, generally has a single source and does not require harmonization of Master data or Attributes. Hence for such InfoObjects attributes or texts are directly loaded from the DataSource.
2. However InfoObjects having common definition across globe like Global Material master data or Company code master data should not have any discrepancy and calls for harmonization of master data before loading transactional master data.
  - a. This can be achieved by loading attributes to the ODS object then loading it to InfoObject.
  - b. The master data will be loaded to the ODS and it is activated before loading the transactional data into the propagation ODS. This will facilitate lookup on InfoObject MasterData in transformation while loading transactional data.
  - c. This schema has one MasterData ODS per region to allow data loads in a single BW system at different times according to time zones.
  - d. Later master data ODS load data into the InfoObject.

## Achieving 24 X 7 Data Loading Requirements

Now let's understand the benefits of above data model while performing regular data loads to the system, and find out how to achieve 24 X 7 data availability for reporting and workload balancing.

For this we need to look at how the system works in different regions at the same time as well as at one region for different time of the day. In the diagram below we have depicted the system workload for one moment in time when the time is 10:00AM Europe time (CET).

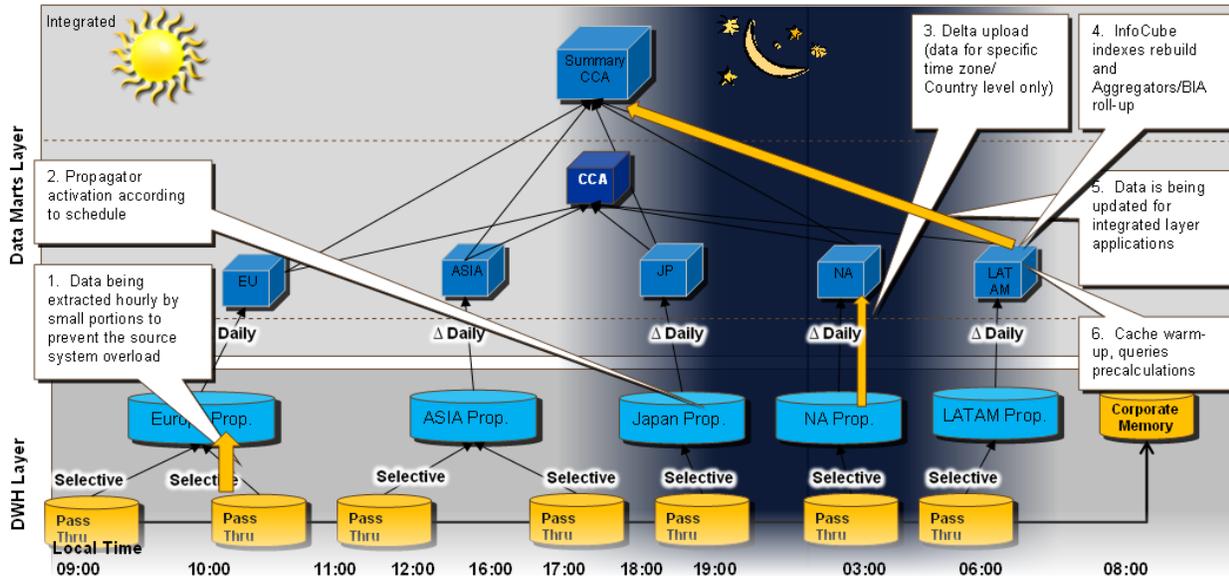


Fig 6.1 Dataloading schema in a single day

### Consideration for 24 X 7 Scenario

- In this particular region of Europe there is normal reporting activity is being performed during the day time starting 8 am till 7pm.
- Also during the same time transactions are being posted in the OLTP system.
- While at the same time in other region say Japan (CET+8) it is an evening time 06:00PM, so there soon will be no reporting in BW system or transaction postings in the OLTP system and data soon will become available for upload to BW. It means that data in Japan's propagator will be activated soon according to schedule.
- A little bit further round the globe, in North America (West Coast) it is still a night time 01:00AM, so data is being uploaded to InfoProviders from propagators.
- In some regions like Brasilia in Latin America the time is early morning 06:00AM when the data is already loaded and cache data is being prepared to speed up the reporting during the day.

We need to remember that the same scenario exists for every region and all of them are happening at the same time with shift in time according to their time zone. The constant cycle of reporting and loading activities changes provides the stable overall workload on the system without peaks associated with loading activities. This is to ensure that server's hardware utilized in best efficient way as possible.

## Steps in Data Loading

1. Data being extracted hourly by small portions to prevent the source system overload. It loads to PassThru DSO which does not require activation, also making fast data load feasible
2. Propagator DSO activation according to global time schedule.
3. Delta uploads to the regional DataMart object (data for specific time zone / Company code only). The advantage of having 1 DataMart Object per region is visible here. As the reporting is being done on the Multicube (CCA in this case) which executes different query on each underlying DataProvider hence when data is being loaded to the NA region InfoCube it not impacting reporting happening on other region.
4. InfoCube indexes rebuild and Aggregators/BIA roll-up.
5. Data is being updated for integrated layer applications, being summarized data the volume of data is very less and data load is pretty quick.
6. Cache warm-up, queries pre-calculations.

## Time Zone Considerations During Loading

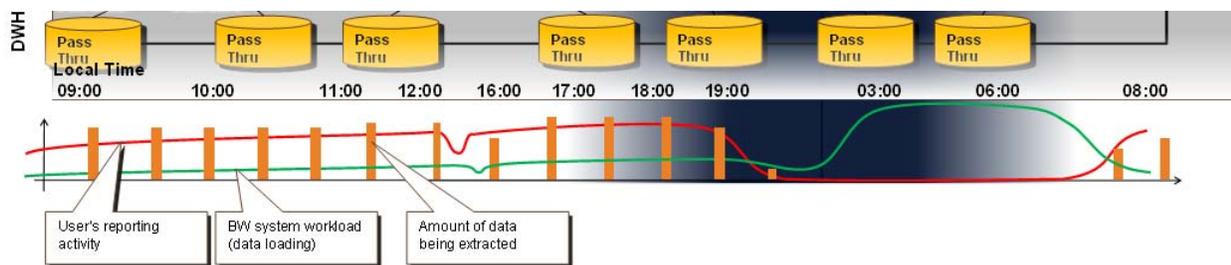


Fig 6.2 Breakdown of Reporting and Loading activities for one particular region

Figure 6.2 explains in detail the user activities both at Reporting in BW as well as OLTP transaction postings.

- In this figure the **Red** line shows the user activities.
- The **Green** line shows the BW system work load e.g. extracting data from the OLTP system or loading it to the other higher level DataMart and reporting objects in the system.
- **Bar graph** shows the amount of data extracted from source system as soon as the postings are done in OLTP and storing it in the PassThru objects on hourly basis.
- The amount of data being extracted every hour during the day corresponds to the user's activity on the hour before extraction.

Here are some advantages of using this approach:

- The data extraction activity during day time for that particular region is low as it is just loading data without transformation and storing in write optimized DSO. During night time this activity increases for objects related to that region as it is loading data to upper level DataMart Object having multiple levels and different transformations at each level.
- With this data loading strategy load on source system is maintained very low.
- In BW system propagation time to upper layers can be minimized just to 4 to 5 hours by proper planning. Hence giving each region related data loads even time in 24 hours and maintaining system resources dedicated for data load at constant as well as optimal level.
- Overall balancing of server's hardware utilization.

## Concept to Handle Volume and Frequency

While this approach has a lot of advantage, it also has a drawback. It comes with additional objects need to be created with similar design to handle this data load. Sooner or later, but usually by the time the design stage of the system did commence the question will arise about how many additional objects are necessary and how to justify the number of them to keep the maintenance costs of the system as low as possible when the it is stable and gone ahead in the steady state phase?

This can be handled by devising a mechanism to calculate the number of objects in the global layer based on the time for data extraction and volume of data being extracted for that particular application. Fig 7.1 explains this.

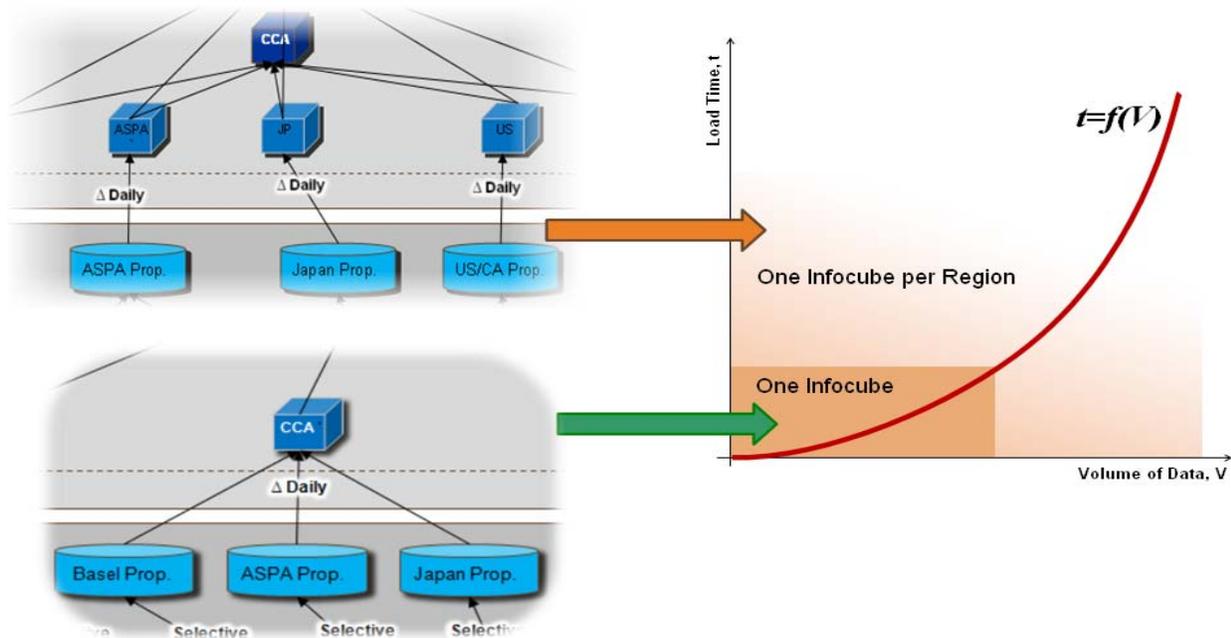


Fig. 7.1 Number of InfoCubes in Global layer

As shown in above figure the graph explains the function which is based on volume of data and time required for data extraction. As the Volume increases the time required to extract data also increases. But for certain amount of data volume the data extraction time is very less hence there is negligible impact on the reporting the end user experience due to data loading and activation.

Hence as long as the impact of the data volume on time for extraction is low we can have only one cube for all regions. This is indicated by the region in the graph pointed by Green arrow.

And whenever the time for extraction increases it impacts the reporting through the same InfoCube for global application. Or even volume of data increases which makes maintaining such a huge data in cube difficult as the age of data warehouse increases. So in both these cases, indicated by the Orange arrow, one should use one cube per region and place them in Global layer. Here the reporting will be done via MultiProvider.

## Summary

The architectural solution provided in this document is applicable to almost all standard business content data datasources delivered by SAP as most of them are supporting delta extraction of data. Those providing high volume of data per day are particularly fit for this solution as it will ensure the workload put on source system is spread during the day by extraction of small portions of data. Customer extraction solutions with delta support could also be adapted to be suitable for globally consolidated environment. We only need to ensure that delta discretization is based on timestamp rather than on calendar day.

However small amount of standard datasources delivered as business content as well as some customer specific datasources are only able to supply data in a full mode format. This would make this architectural solution unsuitable for them, especially for those who provide considerable amount of data every time as it will bring a very high pressure on source systems to supply all data several times a day.

## Disclaimer and Liability Notice

This document may discuss sample coding or other information that does not include SAP official interfaces and therefore is not supported by SAP. Changes made based on this information are not supported and can be overwritten during an upgrade.

SAP will not be held liable for any damages caused by using or misusing the information, code or methods suggested in this document, and anyone using these methods does so at his/her own risk.

SAP offers no guarantees and assumes no responsibility or liability of any type with respect to the content of this technical article or code sample, including any liability resulting from incompatibility between the content within this document and the materials and services offered by SAP. You agree that you will not hold, or seek to hold, SAP responsible or liable with respect to the content of this document.