

## Parallel Data Warehouse PDW for Enterprise Kuwait Finance House

Suhail Ahmad Amanullah, Senior Manager – Data Warehouse , Data Warehouse Management team, I.T.,  
Kuwait Finance House (KFH), Kuwait.

### ABSTRACT

Kuwait Finance House (KFH), pioneering in the Islamic Finance industry in the Middle East region, has Enterprise Data Warehousing (EDW or DW) solution for its banking, commercial and Real Estate sectors. This paper outlines appropriate steps taken technically during its implementation on MS APS ® platform Windows environment. Essential concepts and guidelines covered are Naming conventions, synchronization, history data specifics, Error handling, Load process automation, and its control process, etc. Subsidiaries data from Malaysia, Turkey and Bahrain is also being loaded for some of its downstream applications. Informatica® v10.2 is used as a tool for extracting and loading data to the EDW. MS APS used as a server on windows platform, CA Erwin v12 used as a data modeling tool.

### INTRODUCTION

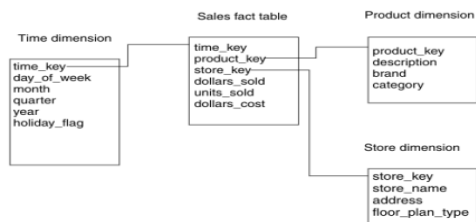
KFH is not merely a bank, but also deals in other businesses like Real estate, investment, commercial markets. In order to have a “Single Version of Truth” across its entire enterprise, KFH realizes a need to build a DW to host almost its entire source systems for loading data to Data Warehouse, its data marts and related downstream applications, like Basel-III ACM, ALM Ambit Focus, IFRS9 Moody’s , GTRP – Group Treasury Reporting, Facility Risk Rating - FRR Moody’s CL systems, CRM Customer Dynamics, AML OMNI, The scope will ultimately broaden to incorporate additional businesses, hence, more sources residing in various platforms and hence, more output reports.

### OBJECTIVE

Our objective is to gather data from various sources and store it on a single platform in an integrated way to be used by various users for their relevant business reporting needs, daily, quarterly or monthly.

### DATA MODEL

The Model design is snow-flaked and follows Ralph Kimball’s approach. A typical star Dimensional model is shown as under:



A Typical Dimensional Model

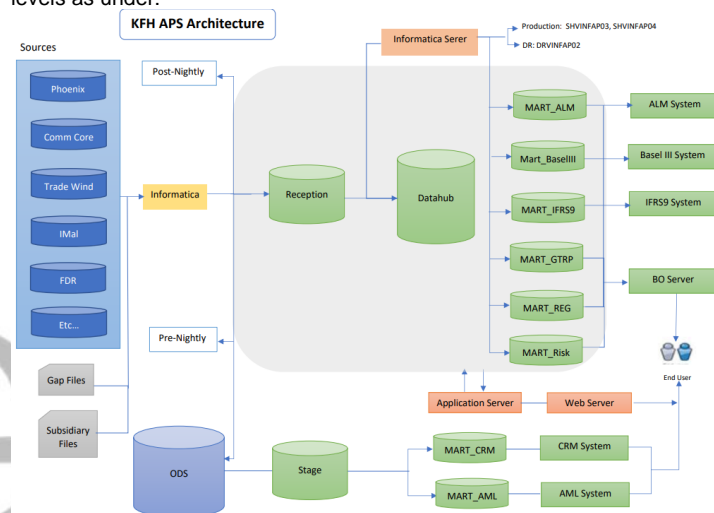
### USAGE

The uses of the data within the Data Warehouse are to:

- Generate consolidated reports to support Business Analysts, Operational and the Executive Managers.
- Develop bank’s Financial control, Risk and Operational reporting needs in terms of developing data marts pertaining to Basel-III, Regulatory, IFRS9, Risk, etc.

### DW TECHNICAL ARCHITECTURE DESIGN

The architecture design of MS APS server is broadly classified into 4 levels as under:



The KFH DW Technical Architecture Design

- PhoenixRep sources: Core Banking source tables are loaded as post nightly dump under this area.
- ODS: sources are read on 1-to-1 basis and loaded to ODS, without any transformations.
- Reception: ODS data is read and loaded to reception, whilst carrying out Data Transformation maps and scrubbing procedures as business requirement. Data is compared with the previous days to capture the changes (Change Date Capture - CDC).
- Hub: These are DW tables loaded from Reception, following SCD-2 increment loading methodology for Dimensions and data dump load to Fact.
- Mart: data from hub is read and business maps applied before loading to the mart.
- Xlerator®: data loaded from marts to Fintellix® reporting engine.
- BO: reports being generated reading from the data hub

### SOURCE DATA

Operations team extract and restore post nightly dump file from the Core Banking system for the T-1 day (yesterday).. The remaining number of (peripheral) sources originates from the other identified sources, like text files and MS Excel.

At present, building the DW is done in stage libraries. After appropriate data Transformation, it is loaded to target DW tables. The frequency of load is daily and follows SCD-2 methodology for loading. Informatica® ETL tool is used to extract the data from source system, cleanse, manipulate and deliver it to MS APS, the final destination. Business Objects and Fintellix Xlerator tool acts as a front-end tool to display the premeditated reports as per the business requirements.

**NAMING CONVENTION**

PascalCase naming conventions is followed as part of the implementation to maintain consistency across the EDW.

Objects	Standard example
Tables	ComRep_TTachsub
Views	E_SwhCtx_DebitCard
Dimension	Dim
Fact	Fct
Procedures	Usp_

Tables are designed as REPLICATE, DISTRIBUTION, HEAP, ROUND ROBIN as storage design option. For transactions we use distribution as HASH, and for dimensions or lookups we use REPLICATE or ROUND ROBIN, as the number of records is less as compared to the Facts. REPLICATE is used when a dimension is frequently joined with huge fact tables.

**THE LOAD DESIGN**

This section discusses the processes involved within the daily DW load. The daily execution of DW Load happens by a scheduled event on the DW server.

The DW data loads MS APS through the following 4 stages:

1. **Preparing ODS sources** : Core banking source data files are read from post nightly restores on phenix rep source, which is without shadow transactions. When the nightly process is in progress the shadow is ON, where the transactions are loaded to different DB. Shadow transactions for today are recovered in next days restores and loaded to the DW.
2. **Reception**: data loaded from ODS
3. **Data-hub**: The last, but the most essential phase of the load is the load to the data hub from reception area. This is in essence, where the complete transformation is applied and the DW tables updated with the latest restored data.
4. **Data Mart**: the data loaded to the data marts based on the business requirements.

- ODS load:

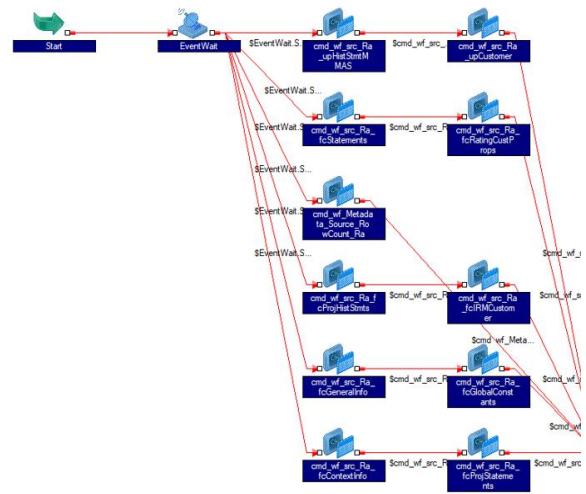
First workflow starts at around midnight daily which makes Business Date entry in *Load\_Execution\_Detail* table and generates ODS and Reception parameter files. *Load\_Execution\_Detail\_trigger.txt* file is created at the end to start or trigger of the successful load for the following processes. *Load\_Execution\_Detail* table contains *Load\_Execution\_ID* for each business Date for the respective database stages (ODS, RECP, DataHub, Marts, etc.) with Start and End Data dates. The same table is available in all DB layers to track the load process

Load_Execution_ID	Business_Date	Status	Start_Time	End_Time	First_Start_Time
1	1937	SUCCESS	2023-09-10 00:01:06.820	2023-09-10 04:32:58.993	2023-09-10 00:01:06.820
2	1936	SUCCESS	2023-09-08 00:00:00.000	2023-09-09 05:15:18.527	2023-09-09 00:01:06.130
3	1935	SUCCESS	2023-09-07 00:00:00.000	2023-09-08 11:16:36.733	2023-09-08 00:01:07.960
4	1934	SUCCESS	2023-09-06 00:00:00.000	2023-09-07 00:01:07.507	2023-09-07 00:01:07.507
5	1933	SUCCESS	2023-09-05 00:00:00.000	2023-09-06 05:12:06.610	2023-09-06 00:01:06.133
6	1932	SUCCESS	2023-09-04 00:00:00.000	2023-09-05 05:21:47.897	2023-09-05 00:01:07.970
7	1931	SUCCESS	2023-09-03 00:00:00.000	2023-09-04 05:34:19.577	2023-09-04 00:01:09.480
8	1930	SUCCESS	2023-09-02 00:00:00.000	2023-09-03 05:20:10.563	2023-09-03 00:01:05.663
9	1929	SUCCESS	2023-09-01 00:00:00.000	2023-09-02 06:27:27.977	2023-09-02 02:28:15.750
10	1928	SUCCESS	2023-08-31 00:00:00.000	2023-09-01 04:45:31.790	2023-09-01 00:01:06.133
11	1927	SUCCESS	2023-08-30 00:00:00.000	2023-08-31 05:35:30.330	2023-08-30 00:01:09.110
12	1926	SUCCESS	2023-08-29 00:00:00.000	2023-08-30 05:40:24.053	2023-08-29 00:01:06.850
13	1925	SUCCESS	2023-08-28 00:00:00.000	2023-08-29 06:34:57.940	2023-08-28 00:01:09.207
14	1924	SUCCESS	2023-08-27 00:00:00.000	2023-08-28 05:51:53.977	2023-08-27 00:01:07.153
15	1923	SUCCESS	2023-08-26 00:00:00.000	2023-08-27 05:35:26.207	2023-08-26 00:01:07.410
16	1922	SUCCESS	2023-08-25 00:00:00.000	2023-08-26 06:13:49.783	2023-08-25 00:01:06.400
17	1921	SUCCESS	2023-08-24 00:00:00.000	2023-08-25 06:55:16.640	2023-08-24 00:01:07.737
18	1920	SUCCESS	2023-08-23 00:00:00.000	2023-08-24 05:35:01.987	2023-08-23 00:01:05.920
19	1919	SUCCESS	2023-08-22 00:00:00.000	2023-08-23 06:36:58.703	2023-08-22 00:01:05.630
20	1918	SUCCESS	2023-08-21 00:00:00.000	2023-08-22 06:47:47.110	2023-08-21 00:01:07.493

Load Execution Detail table

Workflow scheduled at 12:10am looks for *Load\_Execution\_Detail\_trigger.txt* file and starts the loading from RiskAnalyst data source. Likewise, workflows are executed for other sources IBS, imal investment, FDR, Phoenix, Tradewind LC , LG, commercial, etc. The load is an iterative process and runs in loops until it is successfully

executed. Workflow updates the Status as Success from InProgress.



Workflows design of loading source data to ODS.

- Reception load:

Workflow starts at 2:30 am daily after checking all ODS data sources successfully loaded for reception load, If ODS trigger file exists, then it starts the respective Reception integrated ETLs. Business mapping of around 80% is implemented during reception data load. Metadata workflow table status is logged for start and end times of each individual workflow with its status (success/in progress)

Wf_Name	Start_Time	End_Time	Status	Load_Execution_Id
1 wf_src_XtmKtm_ResponseCode	2023-09-10 03:38:08.940	2023-09-10 03:39:11.360	SUCCESS	1937
2 wf_src_Tw_ReferenceAmount	2023-09-10 03:20:22.393	2023-09-10 03:21:32.467	SUCCESS	1937
3 wf_src_Phoenix_Ad_Gb_Crnccy	2023-09-10 00:07:55.233	2023-09-10 00:08:38.903	SUCCESS	1937
4 wf_src_Krcs_WrittenOffDeals	2023-09-10 05:41:45.467	2023-09-10 05:42:27.017	SUCCESS	1937
5 wf_src_Krcs_GofCategory	2023-09-10 05:39:19.027	2023-09-10 05:40:05.387	SUCCESS	1937
6 wf_src_Imal_TrnsGflimitDet	2023-09-10 00:30:16.383	2023-09-10 00:30:58.733	SUCCESS	1937
7 wf_src_Imal_TrnsCategory	2023-09-10 00:40:50.460	2023-09-10 00:41:39.560	SUCCESS	1937
8 wf_src_Imal_PmsSecurity	2023-09-10 00:36:08.747	2023-09-10 00:36:51.133	SUCCESS	1937
9 wf_src_Fdr_Request_Parameters	2023-09-10 01:07:12.403	2023-09-10 01:09:24.467	SUCCESS	1937
10 wf_src_ComTrd1_V_Gur_All	2023-09-10 02:31:12.223	2023-09-10 02:32:05.707	SUCCESS	1937
11 wf_src_ComTrd1_TTActivity	2023-09-10 02:31:11.187	2023-09-10 02:31:56.113	SUCCESS	1937
12 wf_src_ComTrd1_TT_GoodActivity	2023-09-10 02:30:12.947	2023-09-10 02:30:56.710	SUCCESS	1937
13 wf_Metadata_Source_RowCount_Krcs	2023-09-10 05:39:16.527	2023-09-10 05:39:31.133	SUCCESS	1937
14 wf_Metadata_Source_RowCount_ComTrd1	2023-09-10 02:30:25.572	2023-09-10 02:30:46.147	SUCCESS	1937
15 wf_src_RaArch_fcglobaConstants	2023-09-10 00:05:34.780	2023-09-10 00:06:30.380	SUCCESS	1937
16 wf_src_Ra_fcgContextInfo	2023-09-10 00:10:12.567	2023-09-10 00:10:55.377	SUCCESS	1937
17 wf_src_Krcs_TmpProvCaldNew	2023-09-10 05:39:52.230	2023-09-10 05:41:26.587	SUCCESS	1937
18 wf_src_Krcs_CustomerProfile	2023-09-10 05:40:13.143	2023-09-10 05:41:49.767	SUCCESS	1937
19 wf_src_Imal_PmsTradeDet	2023-09-10 00:31:22.750	2023-09-10 00:33:09.783	SUCCESS	1937
20 wf_src_IbsCr_Vw_Reports_Trans_Sum	2023-09-10 00:22:45.150	2023-09-10 00:25:16.897	SUCCESS	1937

Meta data table to log load status.

- Hub load

When reception completes then data hub starts to load

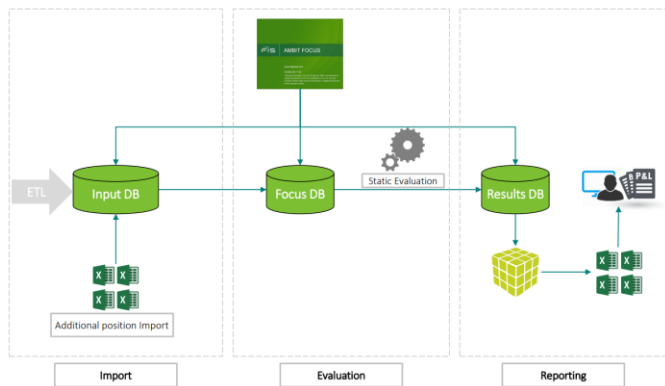
- Mart load:

Data marts run afterwards to load ALM, CBK data. It reads from ALM marts and then loads to downstream DB

- Downstream ALM architecture:

Data Flow diagram for ALM from Mart to till ALM MIS reporting is pasted below:

2 Architecture and data flow

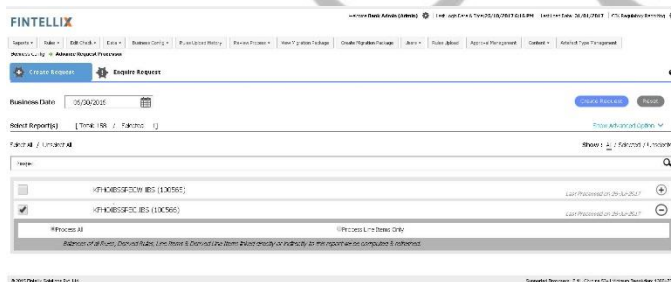


Data is then loaded to ALM reporting downstream Ambit Focus application

This procedure in turn calls other suite of programs/ETL projects in a sequential and controlled manner. Should an error or failure occur within the process, the complete load process halts, the problem is identified, rectified, and then the load continues by submitting the same (above mentioned) procedure to complete the remaining load.

In this case, the procedure detects the load incompleteness and continues loading from where it last stopped. The entire load process is hence, automated from the operational viewpoint. The load itself takes around of 12 hours elapsed time daily load and, 24hours for monthly load to complete. The frequency of load is daily.

Although our load completes loading till downstream data marts, further 4-5hrs are taken to read data from downstream for generating reports for Moody's and Ambit Focus applications



Xlerator load: business user requests to run any reports on demand for generating reports in Fintellix application – Xlerator

The DW load runs as following 2 types: *Full and Increment Load*: After the successful Full load process, the DW load following it is the Increment Load, also abbreviated as INCR. By executing this process, the tables in the DW are updated subsequently with the delta ( $\delta$ ) change data.

*Day-ID* - Every load submitted is tracked by a unique combination of the Day ID (*DayID*). Furthermore, the metadata workflow run status tracks the load process, the status is IN PROGRSS and if any step fails , it's updated to FAIL. After fixing the issue, and re running the same process, the load resumes from the point of failure and updates to SUCCESS when it completes successfully.

The status in the table can range from Processing, Successful or Failed based on the status of the DW load. Every successful load is preceded by any failure(s) that might have previously happened on the load day. In this case the Day ID remains the same, while the Batch ID increases by one (2,3,4, etc.) depending on the number of failures that have occurred from the last run.

**Task Alias**

The log tables on Different DB schemas can be accessed to track the load progress

Database	Load Start Time	Load End Time
EDW_ODS	12:05 AM	7:30 AM
EDW_RECP	12:05 AM	8:00 AM
EDW_DATAHUB	8:00 AM	12:00 PM
MART_ALM	12:00 PM	2:00 PM
Mart_REG (Daily)	8:30 AM	11:30 AM
Mart_REG (Monthly)	11:15 AM	11:45 PM
Mart_BasellIII	4:15 PM	11:15 PM

*Source Restore table*

The flow of the DW load levels is controlled (starting from the highest Control level-0 to the lowest level-3) with its relevant conditions mentioned

### SUBSIDIARY DATA

Data is received from subsidiaries, namely Malaysia, Turkey and Bahrain in csv files through MFTP process to a shared folder scheduled once monthly on the 1<sup>st</sup> day of the month at 1pm. This data is loaded to the EDW for downstream data marts, like IFRS9, EWS etc

### HISTORY DATA

The historical data in DW is preserved from the first data load, and no efforts were made to load any past data from the restored source dumps.

### DATA RELOAD AFTER SOURCE FAILURE

After the DW load is not run with some sources (trigger files) missing.

To retrieve data from dimension loaded with SCD#2 methodology for any particular previous day :

Where `dayid >= @day` and `update_date < @date`

### EVENT LOAD LOG

The log of every successful or failed event taking place during the load is stored in *metadata workflow run status* table.

### ERROR HANDLING

Every program has the following exception thrown to report an error on its occurrence and stop abends at point of failure. The related error code and description is written to the log.

```

END TRY
BEGIN CATCH

DECLARE @ErrorMessage NVARCHAR(4000)
        ,@ErrorSeverity INT
        ,@ErrorState INT;

SET @ErrorMessage = ERROR_MESSAGE();
SET @ErrorSeverity = ERROR_SEVERITY();
SET @ErrorState = ERROR_STATE();

EXEC dbo.usp_Add_Execution_Log (@Session_Id ,@Proc_name ,'' ,@Step_Code ,@Step_Desc ,@ErrorState ,@ErrorMessage ,@i_Load_Execution_ID ,@i_WF_Run);

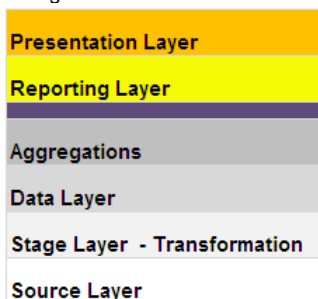
RAISERROR (@ErrorMessage,@ErrorSeverity,@ErrorState);
END CATCH

```

Exception Handler for each program

### AGGREGATIONS

After the data is loaded to the IWS, aggregations are done on separate fact tables based on business requirements. These are stored in their own fact tables to avoid contention of single large tables and for easier management



### CONSOLIDATED MAP METADATA

The (Business) metadata is created and maintained for user access. The table has the following columns:

- Source Location
- Source table name
- Source fieldname
- Source data type
- Transformation applied
- Target table name
- Target Field Name
- Target Data Type

*Metadata (Business) – created originally in MS Excel and uploaded to the IQ server as a table*

### DW TEAM

Three Business Analysts/ BRM – business Relationship Managers (one having knowledge of backend systems), responsible for:

- Coordinating with the Business users to gather requirements.
- Translates Business requirements into Technical, Analytical and/or Reporting specifications.
- Responsible for documenting/editing/updating Mapping Transformation Maps.

Two DW programmers, with one junior staff responsible for:

- Implementing business requirements and Transformation Maps gathered from the business analysts.
- Managing the overall development, implementation and maintenance of the Data Warehouse. This Includes: Logical/Physical Data Modeling, and Implementation of the ETL process.

One BO staff responsible for reporting

The Project Manager oversees/Checks the BI development and maintenance progress.

### CONCLUSION

Understanding Business requirements and translating it to technical specifications as Transformation Map document is a crucial step towards procuring correct results and meeting underlying deadlines.

This coupled with the management influential skills, careful planning, right level of resource skills and tools and rigorous testing significantly improves the successful implementation of the DW project.

### KEYWORDS

- Abbreviations used in this paper:
- ALM – Assets and Liability Management
  - AML – Anti Money Laundering
  - APS – Analytic processing System
  - BRM – Business Relationship Manager
  - CRM- Customer Relationship Management
  - EDW , DW – Enterprise Data Warehouse
  - EWS – Early Warning Signal
  - ETL- Extract – Transform- Load
  - MS - Microsoft
  - IFRS9 – International Financial Reporting System
  - INCR \_ increment load
  - SCD-2 - Slowly changing Dimensions Type2 – expire and insert
  - ODS- Operational Data store

### REFERENCES

[1] Ralph Kimball, Margy Ross, Warren Thorthwaite, Joy Mundy, Bob Becker, *The Data Warehouse Lifecycle Toolkit , Second Edition, Kimball Group*, (ISBN: 978-0-470-14977-5)

[2] MS APS is a registered trademark(s) of Microsoft® indicates registration in the United States of America. All other company and product names mentioned may be trademarks of the respective companies with which they are associated.

### ACKNOWLEDGEMENTS

This paper has been edited and written by Suhail A. Aman-Ullah, Senior Manager - Data-Warehouse, and is currently residing in Kuwait.

I acknowledge the efforts put in by all members of the KFH EDW team in bringing this project a success, for their professionalism, dedication to the project, and assistance in every phase of the project.

My special thanks to KFH Group CIO and Deputy CIO for their cooperation, inspiration, and encouragement throughout the project phase

## BIOGRAPHY OF THE AUTHOR



Suhail Ahmad Aman Ullah, is currently working at Kuwait Finance House in Kuwait as Senior Data Warehouse Manager.

He received B. Eng. (hons) in Electrical and Electronics Engineering from University of Engineering and Technology, Lahore, Pakistan, and MSc-IT from the West of Scotland University in Paisley, Scotland, UK.

He has worked in National Bank of Kuwait in IT for around 9 years as Data Warehouse Developer/System Analyst.

He is also certified as CDMP –Certified Data Management Professional DAMA, CBIP- Certified Business Intelligence Professional , CDS – Expert – Certified Data Expert. CAPM – Certified Associate Project Management from US.

His current work concentrates on building and maintaining Data Warehouses in dimensional or 3NF depending on business requirements.

His interests are in Big Data, Data processing, Data Management , Data Analytics etc.

© GSJ