**Data Aggregation Process**

**Specification**

Author: Justin Machata

David Muench

Date: December 16, 2011

# Overview

A requirement of the Greenfield POC is to demonstrate that two data sources can be used together to produce a single report (web page). This ability will be taken into consideration when making the final decision on where research data will be stored. The choice is to either store all research data in the Dimension system or to build a new system at AshmoreEMM to store the research data

This document describes how Research data from the SQL Server data source provided by AshmoreEMM will be combined with Holdings, Performance and Reference data provided by Ashmore (London). The AshmoreEMM SQL Server contains a set of eighteen (18) tables that can be accessed either directly or through stored procedures. The Ashmore data exists in an Oracle database where a large number of views are available to different types of data. The data is accessed through a WCF services that allows the requesting application to receive data from one view.

# Details

## SQL Server

In order to ease the complexity of data access stored procedures have been written to provide exactly the data required for each specific purpose. For example the Detailed Estimates web page uses a stored procedure named DetailedEstimates that takes three parameters company, Period Type (ANNUAL, First Quarter, etc.) and Estimate Type (NTP, ETS, etc.)

The Research data for the aggregation process will have a new stored procedure for the specific purpose of gathering the necessary data. The XREF column contains a unique ID for all the companies in the SQL Server database. The ISIN is a common field that can be used to match multiple data sources. The Company information coming from the SQL Server data source will be net Income (NTP).

From tblActual, we’ll pull the reported Net Income for the last two years (2009 and 2010) using: EstimateType = ‘NTP’, PeriodType = ‘A’, Value is in ActualValue field

From tblConsensusEstimate, we’ll pull the current Consensus Estimate for Net Income for the next three years (2011-2013). The row with the oldest expiration date where: EstimateType = ‘NTP’, PeriodType = ‘A’, Value is in Median field.

## Oracle

Ashmore has recommended that these views be used for the aggregation:

Holdings use ‘U\_POS\_EXP\_BASEVIEW’

Reference use ‘IRP2\_SEC\_MASTER\_BASEVIEW’

Performance any view beginning with ‘IRP\_PERF\_%’ they are performance tables aggregated up to different hierarchies

After reviewing the data in each view, the aggregation will use one view: Holdings (U\_POS\_EXP\_BASEVIEW). Later we may add data from the Performance view, once we have a better understanding of the data it contains.

## Aggregation Process

The user selects a Portfolio. Portfolio “AREF” will be used for the demonstration.

The web site requests the data from the U\_POS\_EXP\_BASEVIEW Oracle view.

The web site accepts all the data sent from the Oracle view.

The web site stores data where the column PORTFOLIOCODE = the selected value

The web site stores the following columns:

PFCD\_FROMDATE

PORTFOLIOTHEMEGROUPCODE

PORTFOLIOCODE

PORTFOLIONAME

PFCH\_BALBOOKVALPC

PFKR\_FXRATEQP

COUNTRYZONENAME

COUNTRYCODE

COUNTRYNAME

CURR\_QUORISK\_BEST

ISIN

SEC\_SECSHORT

SEC\_SECNAME

SEC\_INSTYPE\_NAME

The following table shows the column matching to be used between the two data sources.

SEC\_SECSHORT is the unique field in the Oracle data source for a company.

XREF is the Unique field in the SQL Server data source for the company.

ISIN is the identifier common to both data sources, but not populated well.

Therefore in order to join data from the two data sources the following table will need to be implemented. This is not an ideal situation, but will be used for the POC Aggregation process.

|  |  |  |  |
| --- | --- | --- | --- |
| **SEC\_SECSHORT** | **ISIN** | **NAME** | **XREF** |
| RUEVRAZE | US30050A2024 | EVRAZ GROUP SA - GDR REG S | 200056374 |
| RUEVRLNE | GB00B71N6K86 | EVRAZ PLC | 200160851 |
| RURUSHYDRODE | US4662941057 | FEDERAL HYDROGENERATING-ADR |  |
| RUGAZORE | US3682872078 | GAZPROM OAO-SPON ADR | 100069288 |
| RULSRGRPGDRE | US50218G2066 | LSR GROUP OJSC-GDR REGS | 200078912 |
| RULUKOYE | US6778621044 | LUKOIL OAO-SPON ADR | 100040772 |
| RURSXUSFC | US57060U5065 | MARKET VECTORS RUSSIA ETF | 100072790 |
| RUMECHSPE | US5838405091 | MECHEL-PREF SPON ADR | 200055386 |
| RUNOVATE | US6698881090 | NOVATEK OAO-SPONS GDR REG S | 200061914 |
| RUOKEYGDE | US6708662019 | O'KEY GROUP SA-GDR REGS | 200137901 |
| RUPHOSAGGDR | US71922G2093 | PHOSAGRO OAO-GDR REG S | 200153580 |
| RUROSGDE | US67812M2070 | ROSNEFT OJSC-REG S GDR | 200063865 |
| RUSBERPFDE | RU0009029557 | SBERBANK-PFD | 200037905 |
| RUSBERADRE | US80585Y3080 | SBERBANK-SPONSORED ADR | 200037905 |
| RUSEVESE | US8181503025 | SEVERSTAL - GDR REG S | 200078330 |
| RUSURGTE | RU0009029524 | SURGUTNEFTEGAS-PFD | 200027722 |
| RUSURNEFTADR | US8688611057 | SURGUTNEFTEGAZ-SP ADR PREF | 200027722 |
| RUTNKBPE | RU000A0HGPM9 | TNK-BP HOLDING-CLS | 200062690 |
| RUTNKBPREFE | RU000A0HGPN7 | TNK-BP HOLDING-PFD-CLS | 200062690 |
| RUVIMPADRE | US92719A1060 | VIMPELCOM LTD-SPON ADR | 100119292 |

Select the data from the POCAggregation stored procedure. Provide a list of unique SEC\_SHORT values from the Oracle data previously captured. The stored procedure will implement the above table to find the correct data to be returned.

The web site will capture the resulting data from the SQL Server database and store it.

The web site will connect the two data sets using the SEC\_SHORT field in both so that a new data set is created containing the following fields.

PFCD\_FROMDATE

PORTFOLIOTHEMEGROUPCODE

PORTFOLIOCODE

PORTFOLIONAME

PFCH\_BALBOOKVALPC

PFKR\_FXRATEQP

COUNTRYZONENAME

COUNTRYCODE

COUNTRYNAME

CURR\_QUORISK\_BEST

ISIN

SEC\_SECSHORT

SEC\_SECNAME

SEC\_INSTYPE\_NAME

COMPANY\_NAME

NET\_INCOME\_ACT\_2009

NET\_INCOME\_ACT\_2010

NET\_INCOME\_EST\_2011

NET\_INCOME\_EST\_2012

NET\_INCOME\_EST\_2013

The web site will display the columns in the new data set in a grid with the following headers.

|  |  |  |
| --- | --- | --- |
| **Data Field** | **Column Heading** | **Format** |
| PFCD\_FROMDATE | From Date | Date |
| PORTFOLIOTHEMEGROUPCODE | Portfolio Group | String |
| PORTFOLIOCODE | Portfolio Code | String |
| PORTFOLIONAME | Portfolio Name | String |
| PFCH\_BALBOOKVALPC | Book Value | Number (19.2) |
| PFKR\_FXRATEQP | FX Rate | Number(6.6) |
| COUNTRYZONENAME | Country Zone | String |
| COUNTRYCODE | Country Code | String |
| COUNTRYNAME | Country Name | String |
| CURR\_QUORISK\_BEST | Currency | String |
| ISIN | ISIN | String |
| SEC\_SECSHORT | Security ID | String |
| SEC\_SECNAME | Security Name | String |
| SEC\_INSTYPE\_NAME | Instrument Type | String |
|  |  |  |
| COMPANY\_NAME | Company Name | String |
| NET\_INCOME\_ACT\_2009 | Net income Act. 2009 | Number (19.2) |
| NET\_INCOME\_ACT\_2010 | Net income Act. 2010 | Number (19.2) |
| NET\_INCOME\_EST\_2011 | Net income Est. 2011 | Number (19.2) |
| NET\_INCOME\_EST\_2012 | Net income Est. 2012 | Number (19.2) |
| NET\_INCOME\_EST\_2013 | Net income Est. 2013 | Number (19.2) |