Omniscope 'Best Practise' Multi-file Reference Architecture

Omniscope data flows typically encompass data import, integration, transformation and ultimately distribution of interactive, highly-visual reporting in various (some auto-refreshing) formats. Although it is possible to configure all these steps within a single Omniscope IOK file, this is rarely the 'best practise' architecture. Best practise usually involves spreading the data flow processing across multi-file 'chains' spanning 24/7 'always-on' servers, power user desktop machines and end-user desktop/mobile devices. Each IOK file in the chain should be configured to play a specific role. The Scheduler in Desktop/Mobile Servers can be used to ensure on-going automated refresh, repeated transformation and (personalised) report delivery on either a chronological or event-driven basis triggered by backend systems.

Configuring 'chains' of IOK files with specialised roles:

1. Server-side 'Timeslice' Source IOK file(s) should be used to periodically 'snapshot' or 'timeslice' each incoming data source (e.g. database views, feeds, API cloud sources or auto-synchronised folders of files) by periodically connecting to and refreshing from the source, performing data quality checks and generating alerts, then automatically saving a time-stamped copy of the data in an archive folder, which accumulates compact IOK files over time and acts as a batch append source 'data warehouse' driving other downstream files. There should be one Timeslice Source IOK file set per data source and these files should do all field/column re-naming and other transformations involving incoming data from each source, including checking for duplicates and unexpected values that might affect subsequent aggregation, merge and append operations;

2. Server-side Integration/Aggregation IOK file(s) should be used to manage row pre-aggregation, user-friendly field/row and value-renaming, repair of fragmented records etc. such that multiple incoming datasets (including entire folders of time-stamped IOK Timeslice files) can be integrated (merged/joined or appended) and with automated data quality checks and pre-calculation of formulae involving fields from multiple sources. Even unrelated data sets sharing no columns/fields can be appended (typical for a KPI 'dashboard'). Fields containing alert formulae and tab filters/queries settings not needed further downstream for actual inclusion in reports should be removed

3. Server-side Datamart IOK file(s) refreshes and integrates all (pre-processed/corrected/transformed) fields/columns that are used in various master Report IOK files. This file integrates across all incoming sources and acts as the 'single-source of truth' and 'semantic layer' from which all 'downstream' master Report IOK files are populated/refreshed. At this point in the data flow, all field names and values might be modified from those found in the original source systems.

4. Server-side Batch Configuration IOK file(s) are auto-generated, special-purpose IOK files with a pre-defined set of columns controlling the 'distribution list' i.e. the rules for personalisation and delivery of automatically-customised files in various formats using various file delivery/deployment options. The Batch Configuration IOK file specifies how and where copies of the incoming data set (above each Batch Output block in the DataManager workspace data flow) should be published and delivered. Initially blank, the values entered into each row of the Batch Configuration IOK file collectively contain all the instructions for performing one automated publish/delivery action. Each refresh and auto-publish enables any number of recipients to receive any number of personalised variants of each data set, each using potentially a different role-based IOK template file. For example, a version of each Report IOK file could use a special IOK template file with the window size layout set and optimised for small mobile tablet screen sizes, perhaps also omitting some columns from this particular copy of the data set to conserve screen space.

DataExplorer Tabs: The data visualisations configured on the tabs of typical server-side Timeslice, Integration/Aggregation and Datamart IOK files should be focussed only on enforcing data governance, i.e. ensuring validation, data typing consistency, timeliness, completeness, and auto-generation of alerts/flags, plus manual correction and write-back files to support centralised master data management. These files above are purely internal and never seen by end user recipients of the reports.

5. Server-side Master Report IOK(s) are fully configured, highly-visual multi-tab, multi-view interactive 'dashboards'/presentations targeted for specific groups of recipients and saved in network-accessible locations. These usually have the server-side DataMart IOK file as their source. Automated personalisation or formatting for different device screen sizes is achieved using pre-configured template files (empty or containing stale data) and the Batch Control file entries specifying the sub-set of the DataMart fields and the template file to be used for each recipient.

6. Distributed Report IOK(s) are copies of the central Master Report IOK file(s) that are automatically distributed and (licensing permitting) can live refreshed even while open directly from the web/network accessible server-side Master Report IOK file. The DataManager tabs of Report IOK files typically show only a single IOK Source block pointing to the Master Report IOK with static formula values and no formula details.

Multiple File Data Refresh cycles

A multi-file data flow will contain a series of sequential refresh cycles. Refresh cycle 1 (original sources refresh to Timeslice IOK files), 2 (Integration/Aggregation IOK files refresh from Timeslice files) and 3 (Datamart IOK refreshes for all upstream files) are straight-forward and can be manual, but are fully-automated in the Server Edition either chronological (every hour on the hour, etc.) or event-driven watchfolder triggered operations.

Refresh cycles 4 and 5 (Master Report from Datamart, and Distributed Report from Master Report) usually employ a trick that is also automated by the Scheduler.

For Users to be able to live-refresh their Distributed Report IOK files that have been either 1) e-mailed to them, 2) web-downloaded or 3) folder-synchronised, the central server-side Master Report IOK has to be saved with ITSELF as its own Source. This ensures that the distributed copies of Report IOK will all look back to the web/network-accessible Master Report IOK file for their live refresh updates.

But if you do this, how can the 4th refresh cycle, whereby the Master Report IOK refreshes itself from the previously-refreshed Datamart IOK be accomplished given that the source of the Master Report IOK file has been set to itself?

The answer is that as part of refresh cycle 4, the automated Scheduler Task List is used to change the source of the Master Report IOK to the Source Datamart IOK, refresh and save the updated Master Report IOK file, then change the source of the Master Report IOK BACK TO ITSELF, and save the file again in its web accessible distribution location ready for e-mailing or download/folder synchronisation as Distributed Report copies of the Master IOK file.

Populating Data Warehouses and Agile BI

Omniscope reporting data flows can also include 'by-product' write-back of output to SQL database tables, enabling population of new/existing databases/warehouses with scrubbed, transformed data from Omniscope. One reason to do this is to be able in future to write bespoke programs that access the accumulating (clean, transformed) 'warehoused' data sets, and/or to use other features of the BI stack availableon the database e.g. OLAP cubes. Omniscope can connect to MDX-compatible multi-dimensional OLAP 'cubes' as well.