



ADOBE ANALYTICS DATA WAREHOUSING

AUTOMATING ADOBE
ANALYTICS DATAWAREHOUSE
INTEGRATIONS



ACME MARKETING SOLUTIONS
AUGUST 2016

USING ADOBE ANALYTICS DATAWAREHOUSE API CUSTOMERS CAN BUILD POWERFUL AUTOMATIONS FOR EXTRACTING ANALYTICS DATA FOR OFFLINE REPORTING AND ANALYSIS

Adobe Analytics(formerly SiteCatalyst) is an industry leading Web Analytics tool that forms the reporting mainstay for a large number of high traffic websites each drawing hundreds of thousands of website visitors per month. The click stream data generated by these visits is typically collected by Adobe in its data centres and can be leveraged using primarily 3 tools-Adobe Report Builder (Excel plug-in), Reporting& Analytics (online GUI based reporting) and the Adobe Data Warehouse (offline data export). All these tools enable basic reporting capabilities that are usually adequate for ad-hoc, activity based reporting but for scenarios that require advanced business intelligence, these methods have limited utility.

Our position is that truly unlocking the potential of Adobe Analytics data requires a purpose built business intelligence solution that clearly decouples the data extraction, processing and storage of click stream data from the reporting and visualization layers. A well designed custom database that is purpose built for specific analysis needs and which automatically refreshes from the Adobe analytics data can then be used as the backend for a specialized reporting tool for advanced reporting and data analysis.

In this whitepaper, we focus on the automated data extraction and storage layer. In particular, we provide an overview of how the Adobe Data Warehouse product can be used for creating advanced digital business intelligence solutions that can provide analysts with unparalleled insights into business performance.

Before digging deeper into the solution though, here is a quick outline of the **top 3 use cases** wherein deploying such sophisticated data integration capability makes commercial sense.



1 TRACKING ADVANCED CALCULATED METRICS

Calculated metrics are either built using complex calculations and/or use multiple data points. Running totals, trends, lifetime values, ROI figures, voice of customer measures are all examples of advanced metrics constructed using complex calculations on more than one data point. Implementing these in the in-built Adobe Excel plugin is severely inefficient if at all possible.

2 NEED FOR ADVANCED DATA SLICING AND DICING CAPABILITIES

Meaningful analysis requires the ability to quickly slice and dice this data on the fly without extensive query/data model redesign. Differential reporting, co-relation analysis, trend forecasting, statistical analysis, etc. are all examples of advanced data analysis patterns that require specialized analysis tools which can work off bespoke databases designed for specific analysis needs.

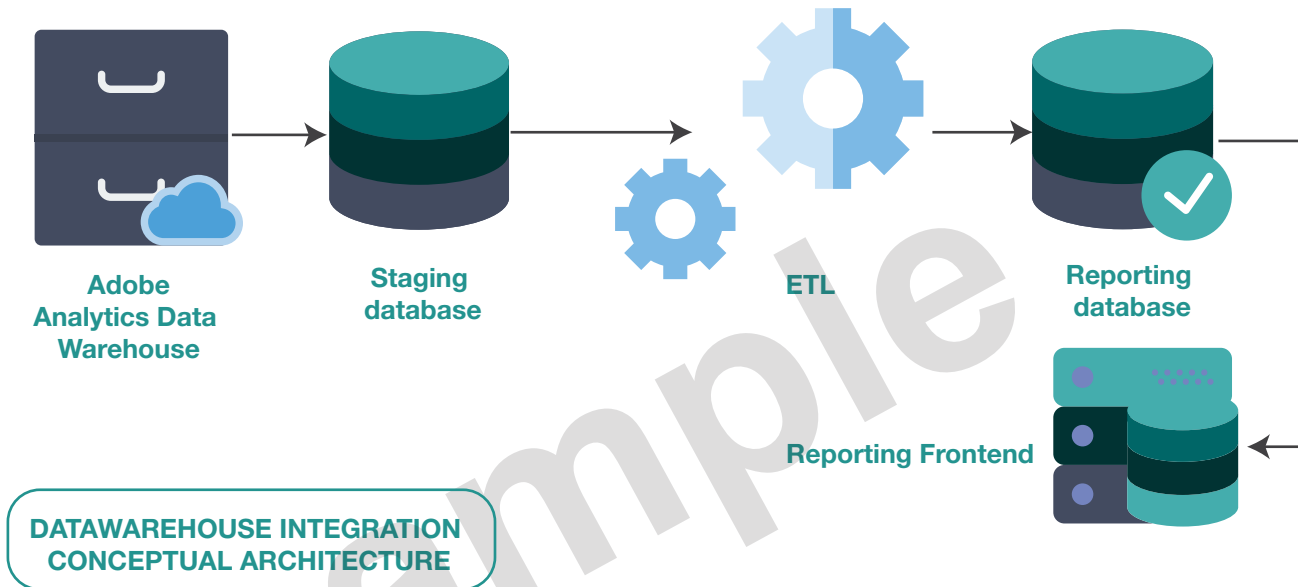
3 HANDLING LARGE VOLUMES OF DATA

Performing advanced, pattern based data analysis on click stream data is complex as is, but the complexity increases multi-fold when such analysis needs to be performed frequently & on large historical datasets. Neither Report builder, nor the online GUI based reporting tool is architected for meeting such advancednon functional requirements.

These are just some prominent challenges that hopefully amplify the need for exploring alternative architectures for analyzing analytics data. In the next section, we outline a conceptual technology architecture that can provide a generic solution to all the scenarios outlined above.

DATA WAREHOUSE INTEGRATION SOLUTIONS BLUEPRINT

We position that building bespoke databases for storing click stream data is the first step to extracting meaningful information out of user's online activity. Using dedicated logical databases for staging and reporting data lies at the heart of our conceptual architecture.



In this architecture, data is first extracted periodically (using FTP) from Adobe Analytics data warehouse into an intermediate data storage (staging area). ETL techniques are then applied to merge and transform the various FTP extracts into a reporting database (reporting area) that is used for final reporting.

1 RAW DATA EXTRACTION INTO STAGING DATABASE

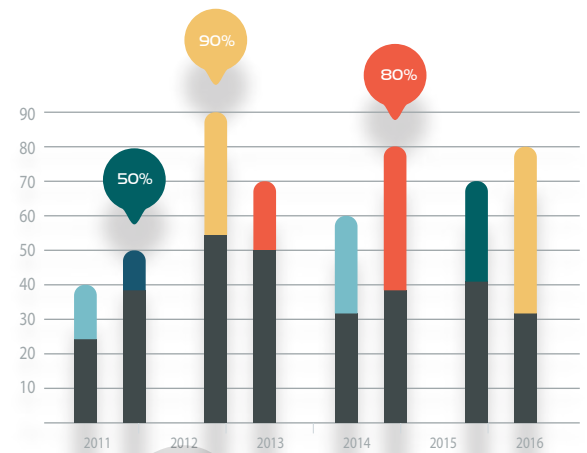
Using the Datawarehouse API, customers can setup data extracts that Adobe delivers via email or to FTP locations once the data is ready. Typically, data is requested for a particular date range and breakdown combination and Adobe provides a snapshot extract for each. The raw data extraction step includes logic to automatically combine the individual extracts into an RDBMS storage in staging area. ETL principles come into play here while automating the data extraction from zip files and then applying data transformation rules to consolidate this data into relational tables.

Using a dedicated staging area allows customers to a) host a copy of raw data in RDBMS tables, and b) implement data aggregation which might be useful when data feeds from multiple sources do not arrive at the same time or follow some kind of a sequential pattern.



2 STAGING TO REPORTING DATABASE

This step involves applying specific ETL rules for transforming raw data into a format that can be used for specific reporting or analysis requirements. The staging area provides temporary storage for assembling raw data and is typically designed for providing persistence and flexibility in being able to support multiple reporting requirements. The reporting database on the other hand is designed for specific reporting requirements in mind. With this architecture, new reporting requirements can be easily met by transforming raw data from staging rather than going back to Adobe for historical data fetch.



AUTOMATING THE DATA EXTRACTION WORKFLOW WITH DATA X

A common, less tech savvy approach to deploying such integrations is to create discrete scripts that perform each individual step in isolation. Command line utilities are then written to provide an overview of the execution for each run. An easy to use, visual interface to monitor the current and historical executions is almost always absent. Further, creating settings for database (staging/reporting), FTP access details are all handled manually. By far the biggest drawback of this approach is the need to involve IT which in many cases induces significant latency into the overall insights delivery process.



THIS IS WHY WE BUILT DATA X

DataX is our hosted marketing integration platform that addresses many of the issues outlined above. The platform uses Amazon RDS instances for staging and reporting databases (it can be configured to use other cloud platforms) and automates the extracts from Adobe Datawarehouse provided FTP dumps.

Key Features

- ⦿ Quickly configure integration steps using web based GUI
 - ⦿ Schedule the automated extract into staging area
 - ⦿ Setup data extraction into reporting database using event driven triggering (e.g. start reporting database upload as soon as extract into staging area completed)
 - ⦿ Monitor detailed execution logs for each run
 - ⦿ using DataX monitoring console
 - ⦿ Manage staging/reporting database
 - ⦿ connection details and access
- Easily increase size of databases without any physical database reconfiguration

We currently offer the Adobe Analytics datawarehouse integration as a custom built workflow app that can be quickly deployed within DataX platform. With end-to-end turnaround times of under a week in most cases (including extraction setup, database design and hosting and workflow configuration), Marketers can quickly get access to a rich data repository that can be easily plugged into any major reporting/analysis tool for deeply insightful click stream data.

DATA X IN A NUTSHELL



source data sink

Hosted FTP server with virtualized FTP Dump storage or Client FTP server (in-premis)



supported databases

Amazon RDS (Mysql,oracle, postgres), Azure SQL Server. Client database (in-premise)



access

Web-based GUI for full integration configuration, including job scheduling configuration and monitoring



pricing

Pay monthly from **\$250/month**+initial consulting fee

FIND OUT MORE

For more details on this integration and general information about DataX, please get in touch today at info@acme.org.



INF@ACME.ORG



+1 -408-223-1256

+1-800-166-9898(Toll free)



206, Market Street
San Jose 94159



ACME



WWW.ACME.ORG