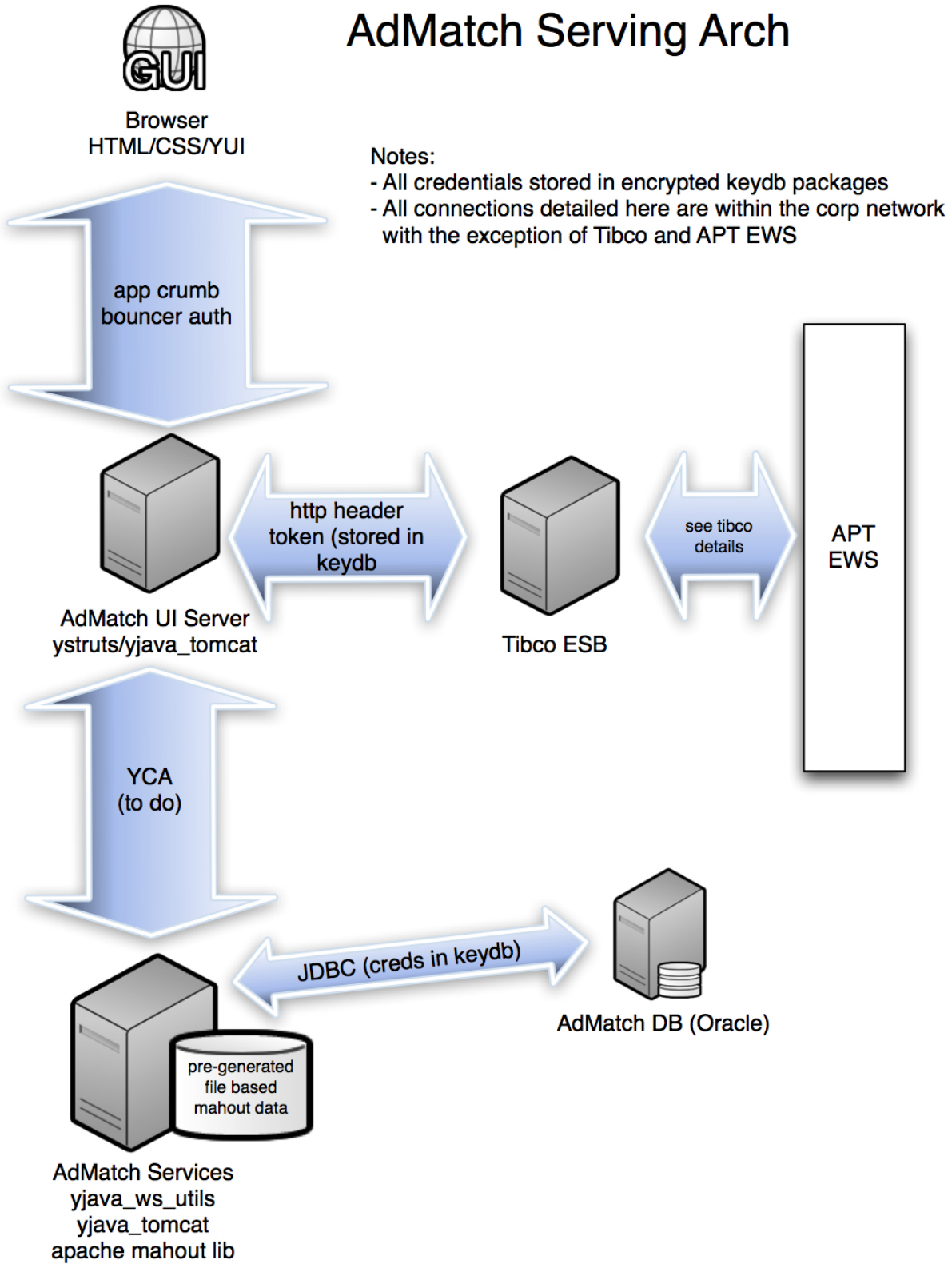
Table of Contents

Ad Match Architecture

Architecture

AdMatch Serving Architecture



Tibco AdMatch APT Integration Architecture

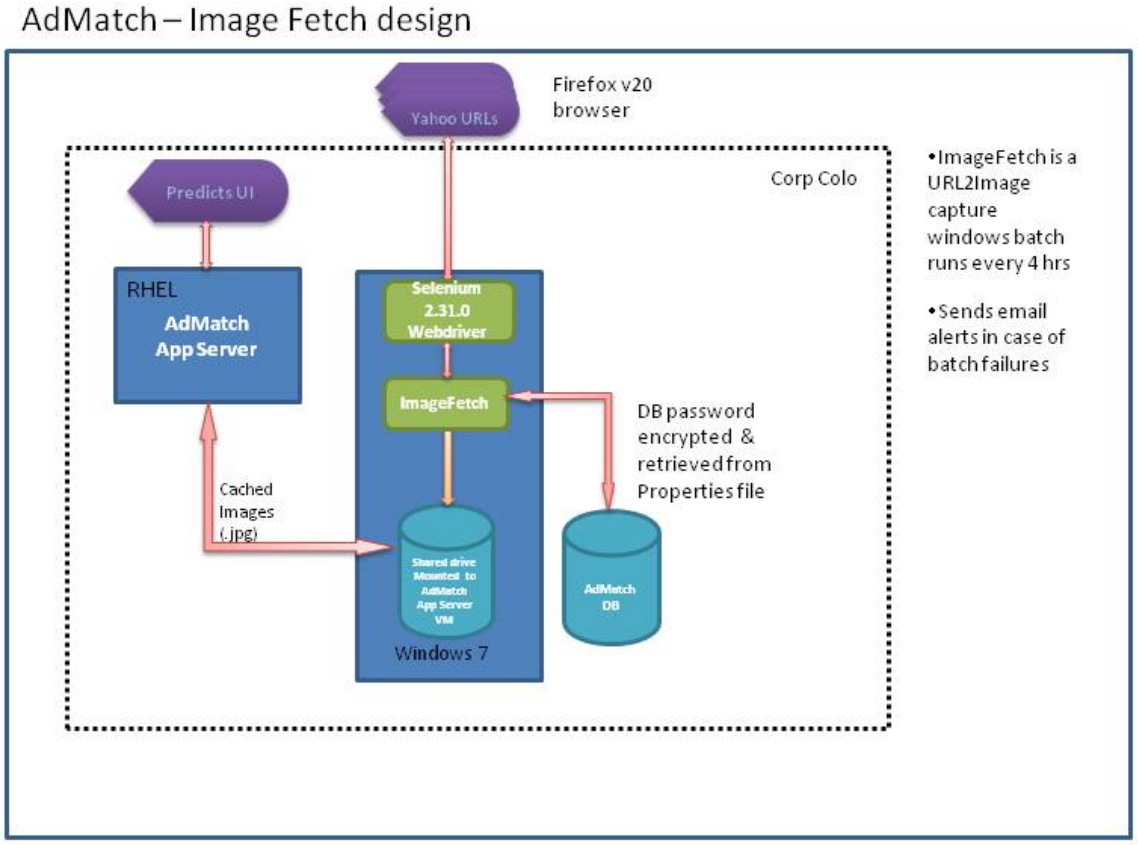
Architecture Diagram

Image Fetch@import "http://twiki.corp.yahoo.com/pub/System/TreeBrowserPlugin/dtree.css";

- Image fetch is a Java standalone application to take screenshot of various Yahoo property URLs. It connects to the AdMatch database to get the list of urls for which the screenshot needs to be captured, it uses Selenium webdriver 2.31.0 to load these URLs and take screenshot of them, in case the URL is not valid, redirected to search page or it shows a blank screen such URLs are marked as invalid and the database table is updated back to eliminate them from the next fetch cycle. This application runs in multiple threads and the number of threads can be configured in the app.properties file based on the memory size available on the Windows box. The database Password is encrypted and stored in the properties file. The application read the password from the properties file, decrypt and connects to the AdMatch database.

- SVN Path http://svn.corp.yahoo.com/view/yahoo/corpapps/salestools/predicts/ImageFetch/

- Image fetch initial design for Release 1.0:



Known issues

Issues
Selenium webdriver throws UnreachableBrowserException for some of the URLs

AdMatch Recommendation Engine

The AdMatch recommendation engine leverages Apache Mahout at its core

- Mahout is an open source scalable machine learning library from Apache.
- Mahout is a well established library for 'Collaborative Filtering' (Recommendations) and Clustering.
- The reason for using Mahout is that it works well with Hadoop.Mahout is already ported on Hadoop.
- http://mahout.apache.org/

AdMatch DataProcessors

Data processors are the processes which bring data from the source system to predicts tables and then create data files that are used by the algorithm. There are also some data processor programs which collect images, digits data, AIC attributes, calculate average CPM, get SiteStructureId, APT data and MDM Id from sources.

Historically we used YCRM as a main source of lines. We are in the process of switching to using UAD Staging as source. Kathiresan populates UAD Staging from UAD production data.

List of data processes

- 1.1 Fetch lines from source to predicts table
    - 1.1.1 Fetch lines from YCRM products to Staging table
    - 1.1.2 Fetch lines from YCRM Staging table to Predicts tables
    - 1.1.3 Fetch lines from UAD Production to UAD Staging (Kathiresan)
    - 1.1.4 Fetch lines from UAD Staging into Predicts tables
- 1.2 Generate predicts files
- 1.3 Fetch APT info (APT Id, Seat Id etc.)
- 1.4 Fetch MDM Id from YCRM
- 1.5 Fetch Site Structure If from YCRM
- 1.6 Fetch digits data
- 1.7 Fetch images
- 1.8 Fetch AIC Attributes
- 1.9 Calculate CPM

Future direction for 1.1 and 1.2 is that we may abandon the use of files and go directly to the database.

Data Sources

- The data is supplied to Ad Match from UAD, and in turn UAD already sources those specific data from APT & MME booking systems, YCRM, and Data Systems.
- UAD provides advertisers, orders, placements and CTR data to Ad Match. The required data are generated in UAD in AdMatch format, and pushed on demand at this time since both databases are not in same environment. The frequency of data push will be automated once a new AdMatch database server is setup in the same environment.
- Data for each property is fetched from Digits. The data is delivered via a db link from Digits DB to AdMatch DB

Staging Tables

- YCRM Staging Table Schema - This is a flat table containing customer and property information. This is based on order line from YCRM. In the long run this will be replaced by UAD Staging Table which will have richer set of information such as ROI etc.

Column Name	Source Table	Source Column	Data Type	Comment

- UAD Staging Table Schema - This schema is currently being built by Kathiresan. We will update this as soon as the schema is available.

Predicts Tables

Process Details

-- VadodkarYahoo -- 29 Jun 2012

Attachments

Attachments

Topic attachments

I	Attachment	Action	Size	Date	Who	Comment
jpg	imagefetch.jpg	manage	56.7 K	09 Apr 2013 - 22:16 UTC	HarrisjeYahoo	
jpg	AdMatchServingArch.jpg	manage	619.7 K	10 Apr 2013 - 00:20 UTC	HarrisjeYahoo	
pptx	AdMatchAPTIntegrationArchitecture.pptx	manage	136.5 K	10 Apr 2013 - 19:15 UTC	HarrisjeYahoo	

This topic: APAG > Predicts > PredictsArchitecture
Topic revision: r4 - 10 Apr 2013 - 19:15:05 - HarrisjeYahoo