

AWS Online Tech Talk

The Secret Sauce Behind GroupM's Data and Analytics Platform



Shai Ziv

Director of Operations,
Data and Analytics Platform

group*m*

Host: Ayush Jain, Sr. Manager – Product Marketing, AWS

Moderator: Teresa Hambelton – Sr. Partner Marketing Manager, David Lipowitz – Sr. Solution Architect

Oct 15, 2018

Agenda

- Introduction
- The problem: What the world used to look like
- The solution: What the world looks like today
- Steps we took to get there
- Our biggest hurdles and how we overcame them
- Summary

The Legacy: What the world used to look like at GroupM DAP



GroupM Data and Analytics Platform

BACKGROUND

- Who are we
- DAP was developed by GroupM on behalf of the agencies to manage advertiser data more consistently, efficiently and at scale.
- Enable us to be a better partner to our advertisers and cross-sell added value services.

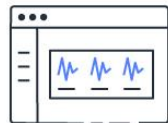
USE CASES



Automation of periodic
client deliverables



Cross-channel repository
for insights, simulation
and optimization

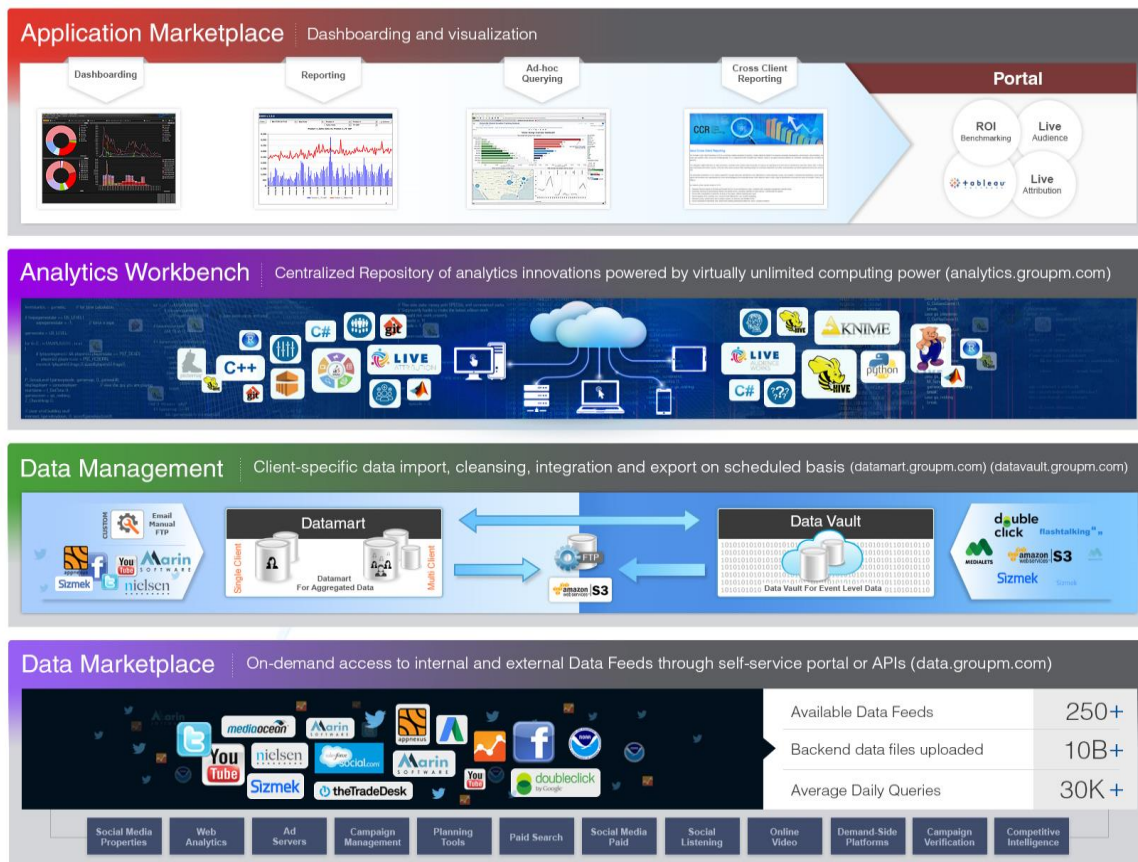


Building client-facing
dashboards



Expanding use of data
science packages

Building blocks of DAP



DAP Customer Example



Use cases:

- Custom Dashboards through LIVE
- Operational reporting for Display & Web Analytics through GroupM Datamart open data connectivity to Excel
- Display Campaign analytics
- Marketing mix modeling using SAS

DAP Sunset

Issues

- Severe performance issues in the on-prem Data Center
- Costly solution with no good option to scale up/down base on usage
- Organizational changes and business requirement to consolidate teams

Opportunities

- Database server-less options
- Clients are open to host their data on the cloud
- Data teams in the agencies are open to use non-Microsoft solutions
- More 3rd party ETL tools are available on the AWS Marketplace

The solution: What the world looks
like today



The new Solution Architecture and benefits

Data sources



Email FTP

Pulls data from file received over email/FTP



3rd party sources

Pulls metrics from planning tools, ad servers, social media and search



Transform & Load



Off-the-shelf product
Support Python script
Easy deployment
Great support
Good ROI

Data Repository



Cloud-based platform
Server-less DWH
Multi format support
Easy to scale when needed
Different account per agency/region
Build in connectors for BI and reporting tools

Reporting

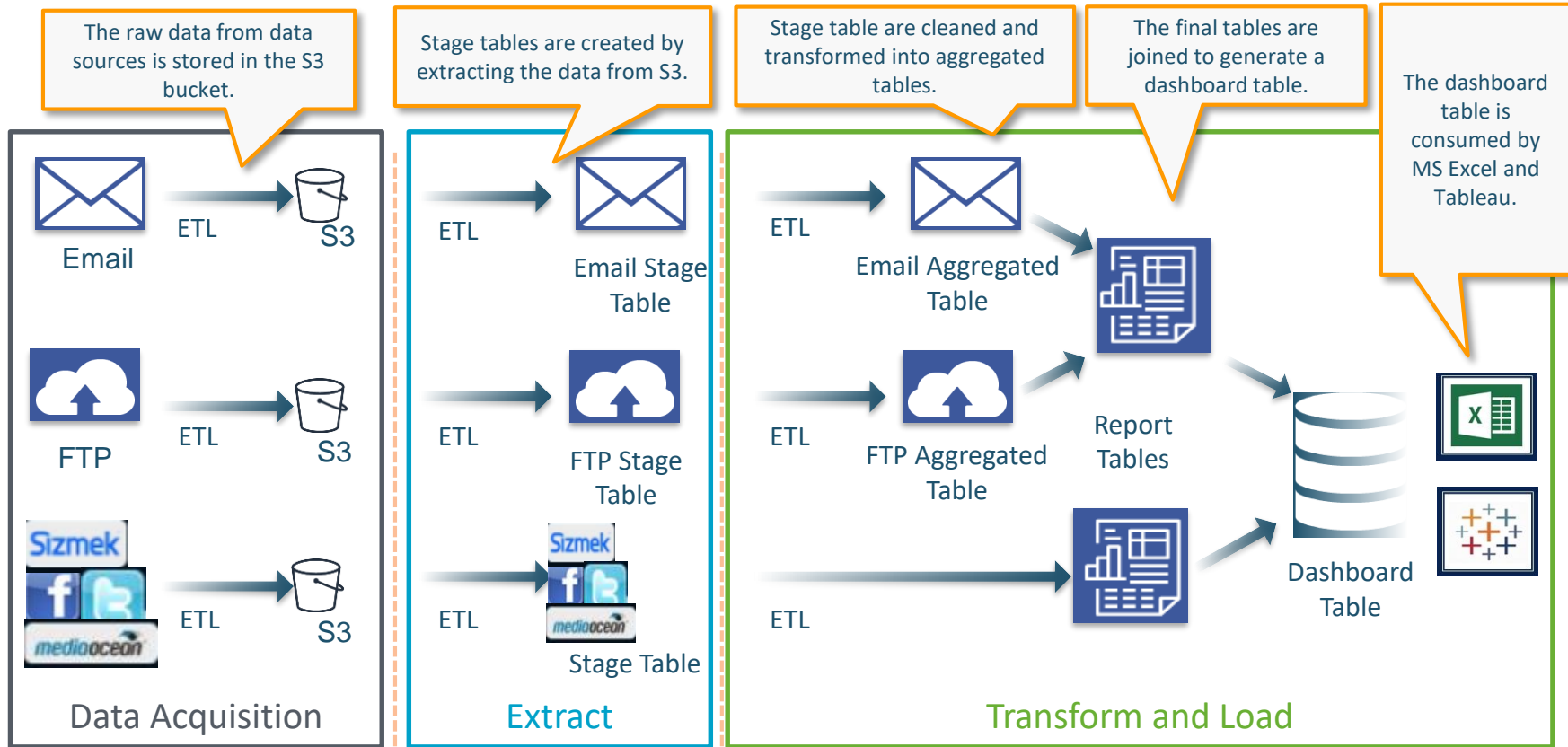


Seamless integration

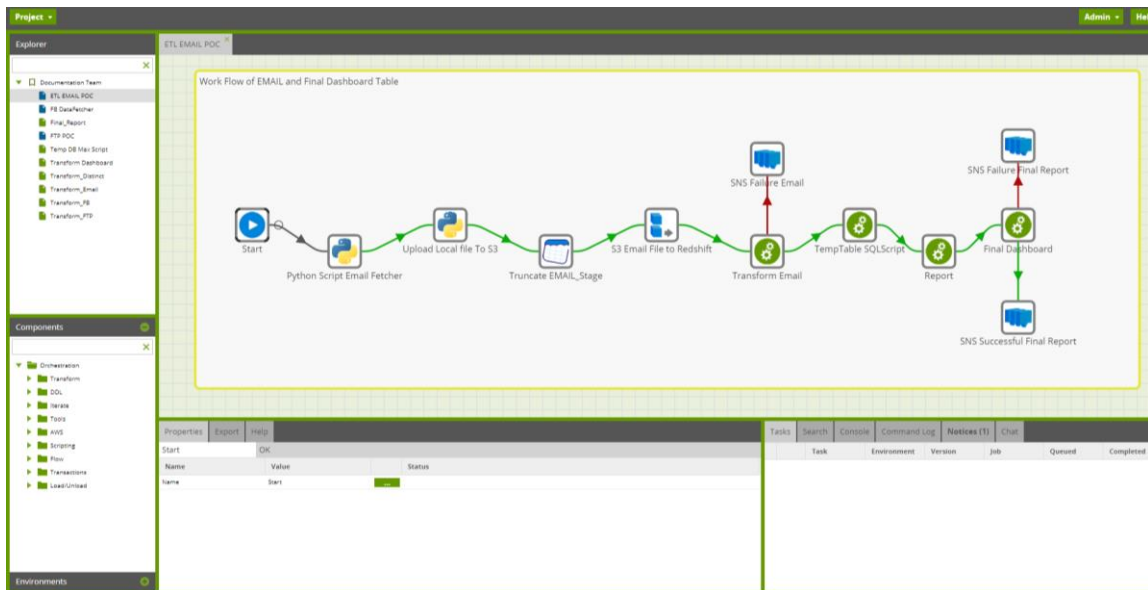
Advanced Analytics

Execute independent R scripts

Data Flow – Example



Matillion and Redshift setup



gt-na-feeds

Database Backup

gt-na-feeds.cv9p7suyrzvd.us-east-1.redshift.amazonaws.com:5439 (authorized)

Properties

Cluster Name gt-na-feeds
Cluster Type Multi Node
Node Type dc2.8xlarge
Nodes 2
Zone us-east-1a
Created Time March 21, 2018 at 11:47:54 AM UTC-4
Maintenance Track Current
Cluster Version 1.0.3906
Up to date
VPC ID [REDACTED]
Subnet Group [REDACTED]
VPC security groups default (sg-0b71be7a) (active)
Cluster Parameter Group default.redshift-1.0 (in-sync)
Enhanced VPC Routing No
IAM Roles See IAM roles

Cluster Status

Cluster Status available
Database Health healthy
In Maintenance Mode no
Parameter Group Apply Status in-sync
Pending Modified Values None

Cluster Database Properties

Port 5439

Backup, Audit Logging, and Maintenance

Automated Snapshot Retention Period 1

Reasons for our choices

Amazon Redshift

- Server-less and Scalable
- Adjust performance to cost based on node type (DS vs DC)
- Option to work with files (Spectrum)
- Copy and Unload commands makes it easier to work with the data
- Good for both aggregated and big data

Matillion

- Integrates well with AWS (S3/Redshift/CloudWatch/SNS)
- GUI-based ETL flow Designer
- Supports Python scripts & variables
- Dependency-based and asynchronous execution
- Wide array of connectors
- Push-down transformation

How we got there

Phase 1: Research

- Self and public AWS training
- Learn from webinars and AWS best practices
- Create sandboxes and use test drive environment from Matillion

Phase 2: POC

- Based on simple but real use-cases
- Compare 3-4 DB and ETL tool
- Create detailed documentation

Phase 3: Implementation

- Find beta users to run pilots
- Minimize the impact on production and clients deliveries
- Run as SW projects with SOW, requirements and project plan

Biggest hurdles



Hurdles

Switching from Microsoft products

- Templates to help with implementation
- Programmatic creation of new configuration and DB schema
- Pre-made Python libraries

Adoption

- Training
- Work with management to incentivize adaption and find dedicated beta users
- Client perception for storing data in the cloud

Move from fixed cost structure to a granular model

- Simplified cost calculator adjust to more specific needs
- Create set of examples
- Compression matrix

Summary



Learn More

- Sign Up for a Demo with a Matillion Solution Architect - info.matillion.com/awstechtalk
- Learn More About Amazon Redshift - <https://aws.amazon.com/redshift/>
- Learn about how AWS can help with Digital Marketing Analytics – <https://aws.amazon.com/digital-marketing/analytics/>

Thank you!