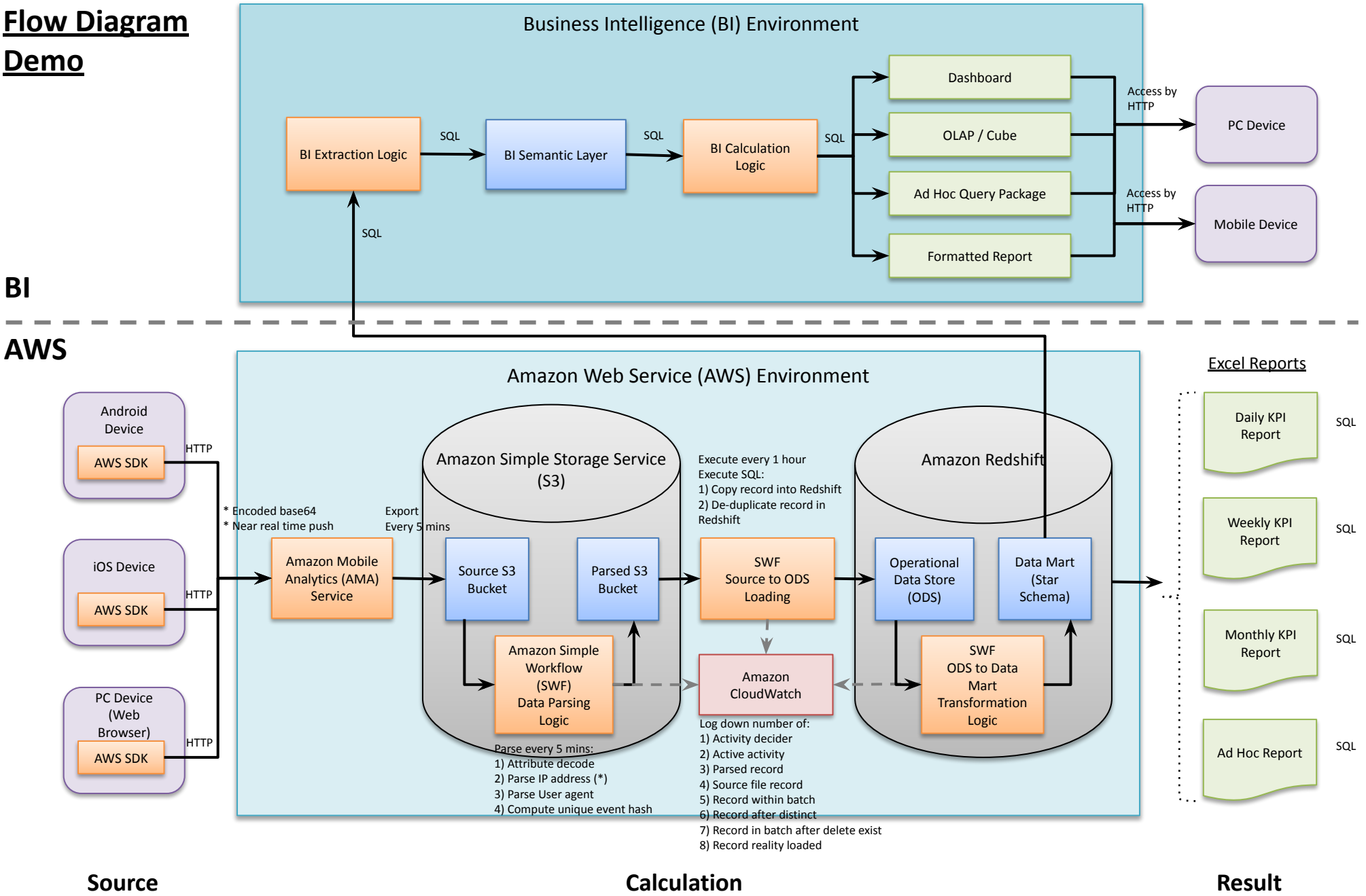
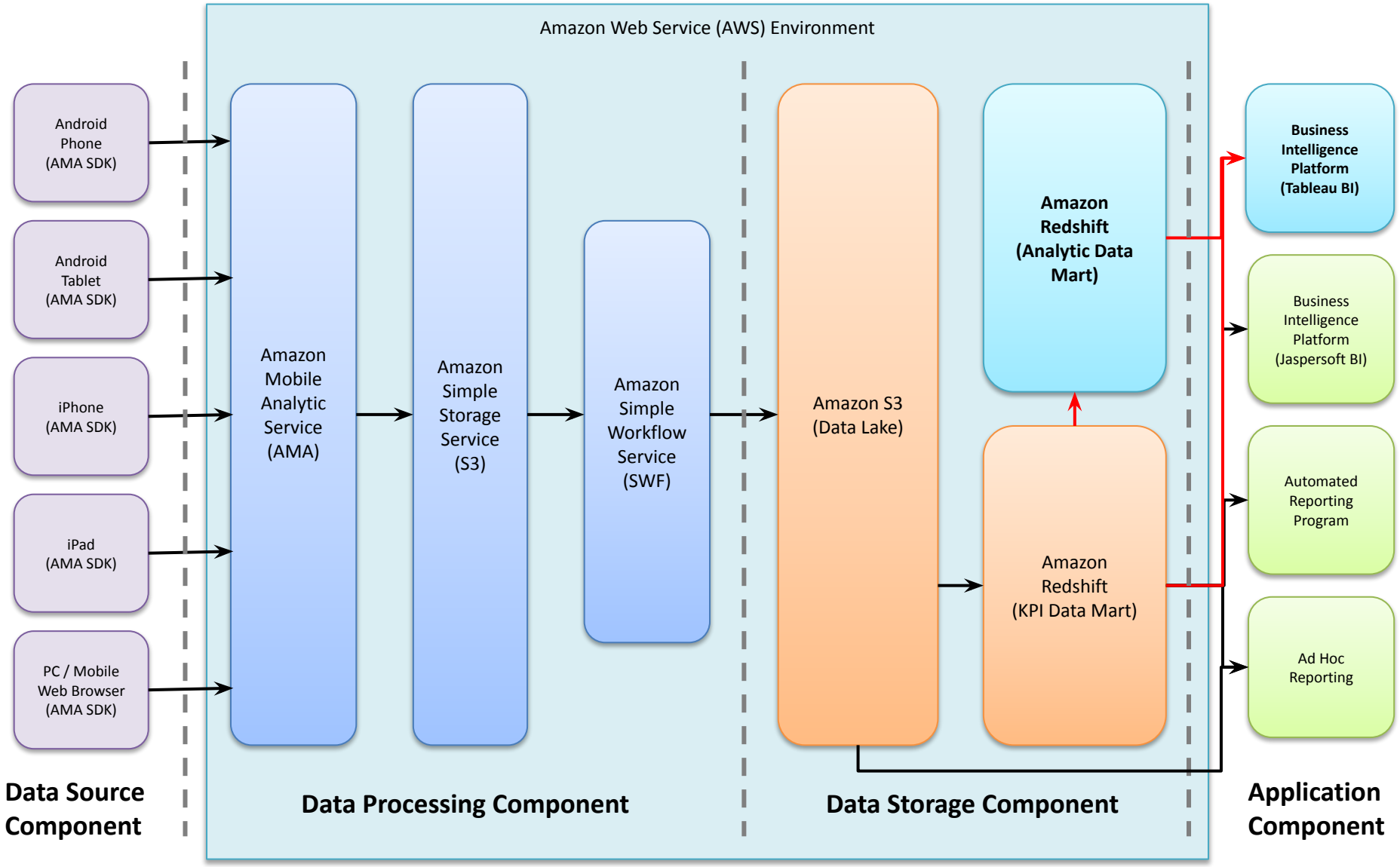
AWS & BI Data Flow Diagram Demo



Data Infrastructure Architecture Demo



AWS System Design Document Demo - Oscar Lee

Table of Contents

External Data Overview **2**

External data sources **2**

AWS Starters **3**

Related Reports **4**

Data source reload **5**

Specific format report **5**

Subscription report 5

Fact Table related 5

Useful commands **7**

External Data Overview

External data sources

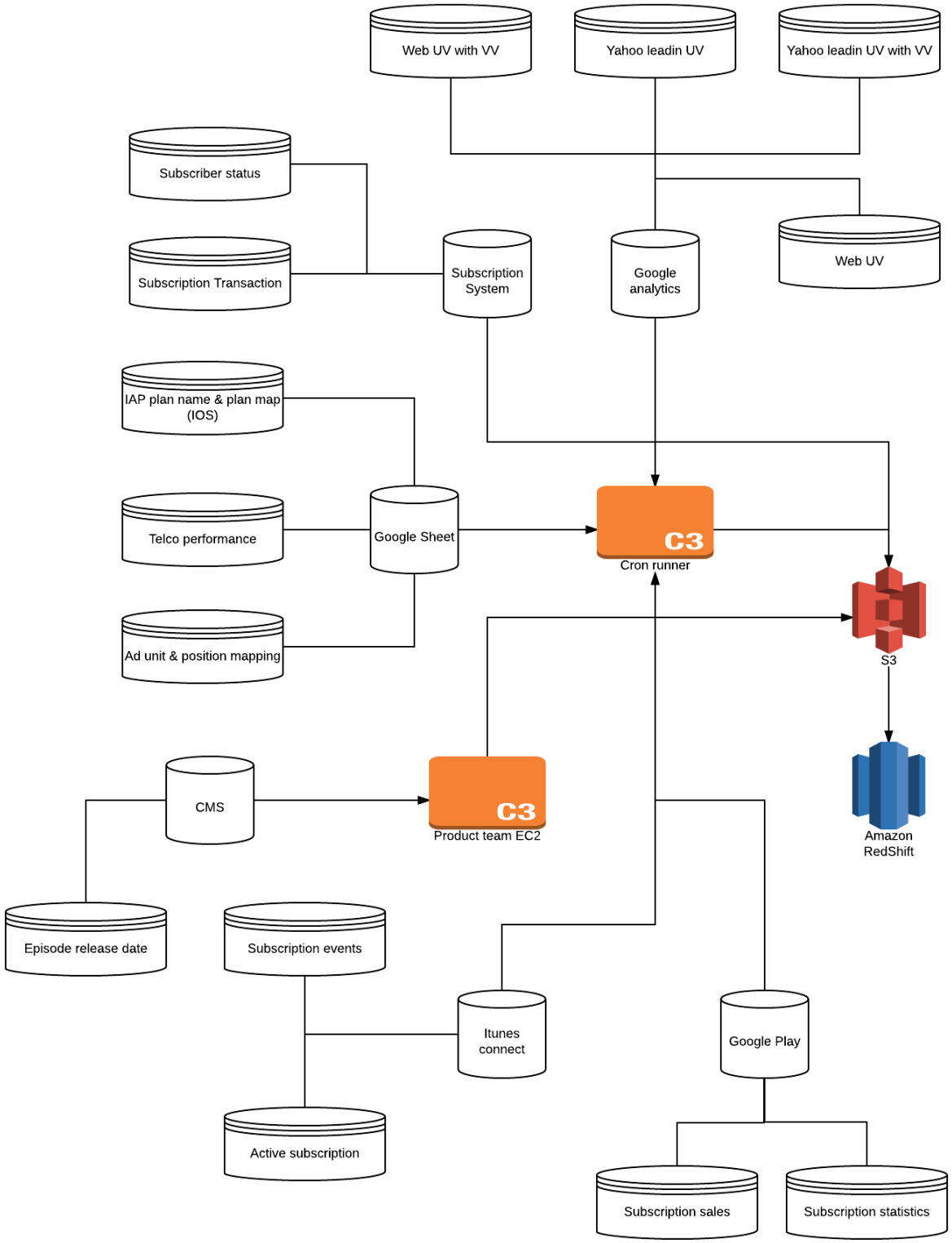


Figure 1

AWS Starters

ID	Script location	Direct execute	Invoker / Module	Data Source	Description
0	cli/cms-export/export.sh	True	N/A	CMS	Export Production Team CMS data to Data Team S3
1	cli/external/itunes-connect/download-active-subscription.sh	False	operationscript-common/operationscript/utils/thirdparty/itunes/loader.py	Itune Connect	Download Itune Connect Subscription Number
2	cli/external/itunes-connect/download-subscription-events.sh	False	operationscript-common/operationscript/utils/thirdparty/itunes/loader.py	Itune Connect	Download Itune Connect Subscription Event
3	cli/external/google-play/download-subscriptions-statistics.sh	False	operationscript-common/operationscript/utils/thirdparty/googleplay/loader.py	Google Cloud Storage(Google Play)	Download Google Play Report
4	N/A	False	operationscript-common/operationscript/utils/thirdparty/googledrive/loader.py	Google Drive	Download Google Sheet as CSV format
5		False	operationscript-common/operationscript/utils/thirdparty/ga/query.py	Google Analytics	Download GA data
6		False	operationscript-common/operationscript/utils/thirdparty/viu/cms/loader.py	CMS	Load CMS Data into Redshift
7			operationscript-common/operationscript/utils/thirdparty/viu/loader.py	Subscription report	Load Subscription Data into Redshift

Related Reports

Owner	Source	S3 prefix example	Report
CMS	Episode release date	ott-csv-unload/thirdparty/internal/viu-cms/2016/10/26/cms-episode-2016-10-26.csv.gz	Correlation between content window and subscription
Itunes connect	Subscription events	ott-csv-unload/thirdparty/external/itunes/subscription-events/2016/10/23/Subscription_Event_85082573_20161023.txt.gz	Subscription report (Weekly, Daily, Monthly)
Itunes connect	Active subscription	ott-csv-unload/thirdparty/external/itunes/active-subscription/2016/10/23/Subscription_85082573_20161023.txt.gz	Subscription report (Weekly, Daily, Monthly)
Google Play	Subscription sales	ott-csv-unload/thirdparty/external/google-play/subscription-sales/2016/10/23/sales_20161023.csv	Subscription report (Weekly, Daily, Monthly)
Google Play	Subscription statistics	ott-csv-unload/thirdparty/external/google-play/subscription-statistics/2016/10/23/	Subscription report (Weekly, Daily, Monthly)
manual google sheet input	Ad Unit & position map	ott-csv-unload/thirdparty/internal/google-drive/2016/10/26/ad-unit-id-mapping-2016-10-26.csv.gz	Report related to AD
Telco (manual google sheet input)	Telco performance	ott-csv-unload/thirdparty/internal/google-drive/2016/10/26/telco-performance-2016-10-26.csv.gz	Subscription report (Weekly, Daily, Monthly)
IAP (manual google sheet input)	IAP plan name & code map	ott-csv-unload/thirdparty/internal/google-drive/2016/10/26/iap-plan-map-2016-10-26.csv.gz	Used for import Active subscription / Subscription events
Subscription	Subscriber status	ott-csv-unload/thirdparty/internal/viu-premium/2016/10/26/viu-premium-20161026.csv.gz	Subscription report (Weekly, Daily, Monthly)
Subscription	Subscription transaction	ott-csv-unload/thirdparty/internal/viu-premium/2016/10/26/viu-premium-20161026_transaction_subscription_ios_iap.csv.gz ott-csv-unload/thirdparty/internal/viu-premium/2016/10/26/viu-premium-20161026_transaction_subscription_android_iap.csv.gz	Subscription report (Weekly, Daily, Monthly)

		ott-csv-unload/thirdparty/internal/viu-premium/2016/10/26/viu-premium-20161026_transaction_subscription_csl.csv.gz	
Google analytics	Web UV	ott-csv-unload/thirdparty/external/google-analytics/2016/10/26/third-party-data.gz	Daily/Weekly/Monthly yellow report
Google analytics	Web UV with VV	ott-csv-unload/thirdparty/external/google-analytics/2016/10/26/third-party-data.gz	Daily/Weekly/Monthly yellow report
Google analytics	Yahoo leadin UV with VV	ott-csv-unload/thirdparty/external/google-analytics/2016/10/26/third-party-data.gz	Daily/Weekly/Monthly yellow report
Google analytics	Yahoo leadin UV	ott-csv-unload/thirdparty/external/google-analytics/2016/10/26/third-party-data.gz	Daily/Weekly/Monthly yellow report

Data source reload

Reload subscription data example:

1. Open terminal
2. Active virtualenv (source ./env/bin/active)
3. cd ./report-runner
4. ipython -i console.py
5. Enter below command:

```
from dm.dm_builder import build_external_subscription_data
sql_batchs+=build_external_subscription_data(args)
sql.batch_execute_sql(conn,sql_batchs)
```

Specific format report

Subscription report

Fact Table related

Name	Description
procedures/data-marts/subscribers/subscriber-daily.txt	Daily to sum up the the weekly & monthly subscription number
procedures/data-marts/subscribers/subscriber-weekly.txt	Calculate by business logic to get the weekly subscription number
procedures/data-marts/subscribers/subscriber-monthly.txt	Calculate by business logic to get the monthly subscription number
procedures/data-marts/subscribers/subscriber-total.txt	Recurring calculation for daily, weekly and monthly active subscriber

Layout

Country	Period	檔案
Hong Kong	Daily	assets/templates/subscription-report/daily/subscription-hk.xlsb
Hong Kong	Weekly	assets/templates/subscription-report/weekly/subscription-hk.xlsb
Hong Kong	Month	assets/templates/subscription-report/monthly/subscription-hk.xlsb
Singapore	Daily	assets/templates/subscription-report/daily/subscription-sg.xlsb
Singapore	Weekly	assets/templates/subscription-report/weekly/subscription-sg.xlsb
Singapore	Monthly	assets/templates/subscription-report/monthly/subscription-sg.xlsb

Data structure

Columns Name	Description
Type	Metric type (Churn, Active subscriber..etc.)
Resolution Type	Daily, Weekly, Monthly
Country	Hong Kong, Singapore
Platform	IOS IAP, Redemption, Android IAP ..etc.
Start Date	
Meta 1	Plan Code
Meta 2	Channel Name (equal to platform)
Meta 3	Channel sector
Meta 4	Payment status
Meta 5	N/A
Meta 6	N/A
Meta 7	N/A
Meta 8	N/A
value	

Excel formula (GETPIVOTDATA)

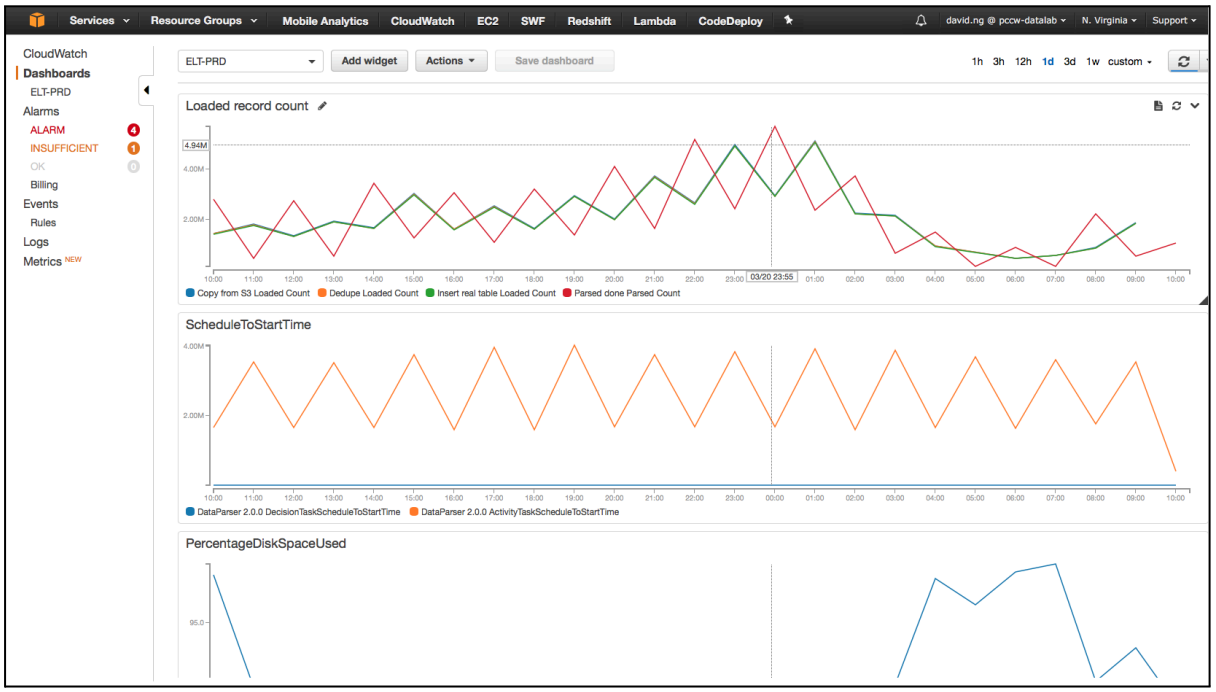
GETPIVOTDATA ("value",'Active subscriber'!$A$5,"Resolution Type","daily","Platform","Redemption","Start Date",”1/25/2017”,"Meta 1",”Samsung Galaxy Tablet Promotion”,"Meta 2","Redemption Campaign","Meta 4","Paying")

SQL expression:

```
SELECT
        value
FROM
        'Active subscriber'!$A$5
WHERE
        “Resolution Type” = ‘daily’
AND
        Platform = ‘Redemption’
AND
        “Start Date” = ‘1/25/2017’
AND
        “Meta 1” = ‘Samsung Galaxy Tablet Promotion’
AND
        “Meta 2” = ‘Redemption Campaign’
AND
        “Meta 4” = ‘Paying’;
```

ELT Monitoring

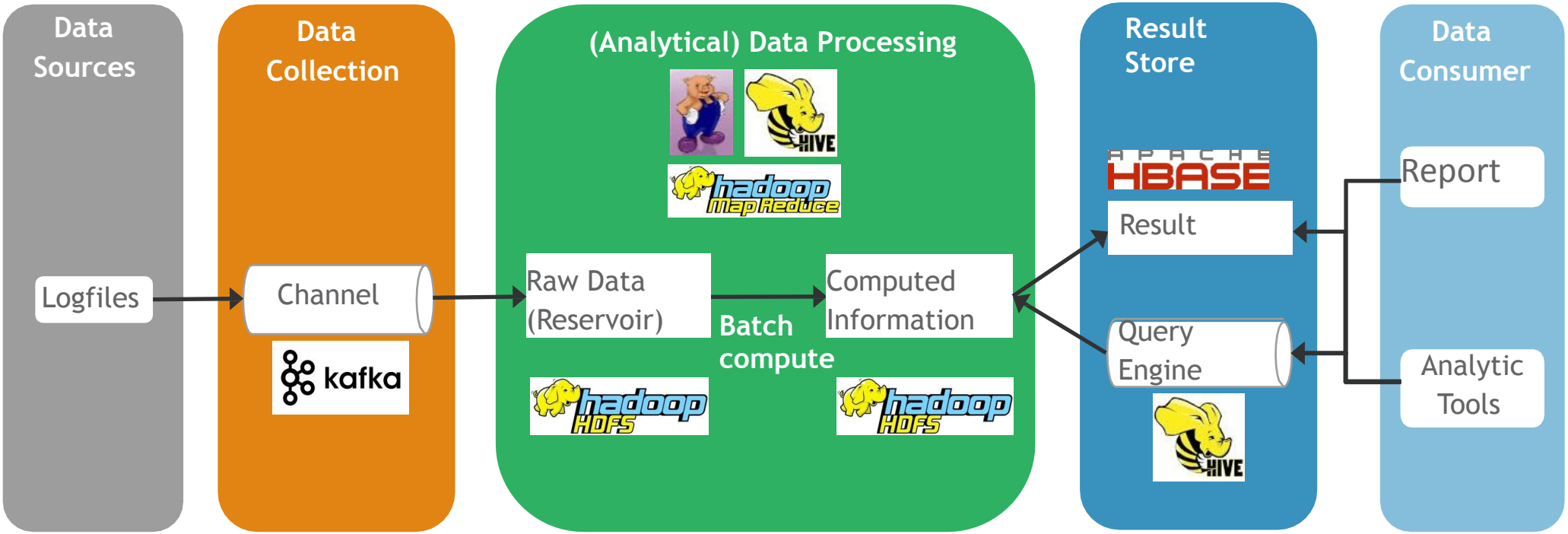
AWS Cloud Watch -> Dashboard -> ELT-PRD



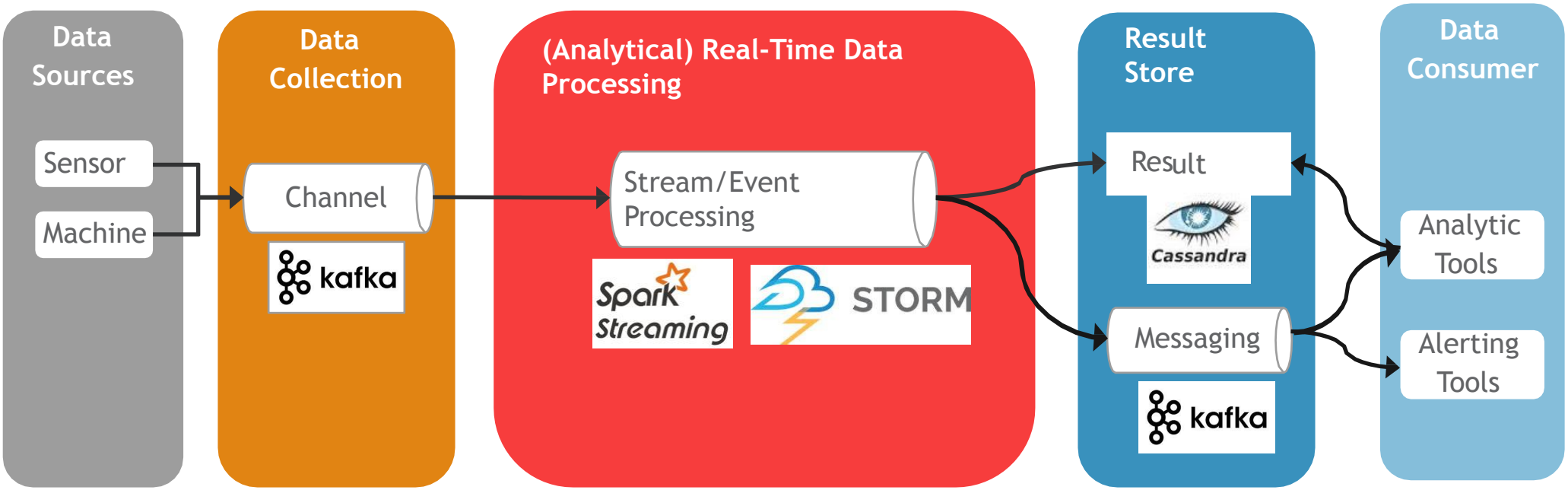
Key Factors to Consider in a Big Data Analytic Solution

- Latency
- Keep raw and un-interpreted data “forever” ?
- Volume, Velocity, Variety
- Ad-Hoc Query Capabilities needed ?
- Robustness & Fault Tolerance
- Scalability

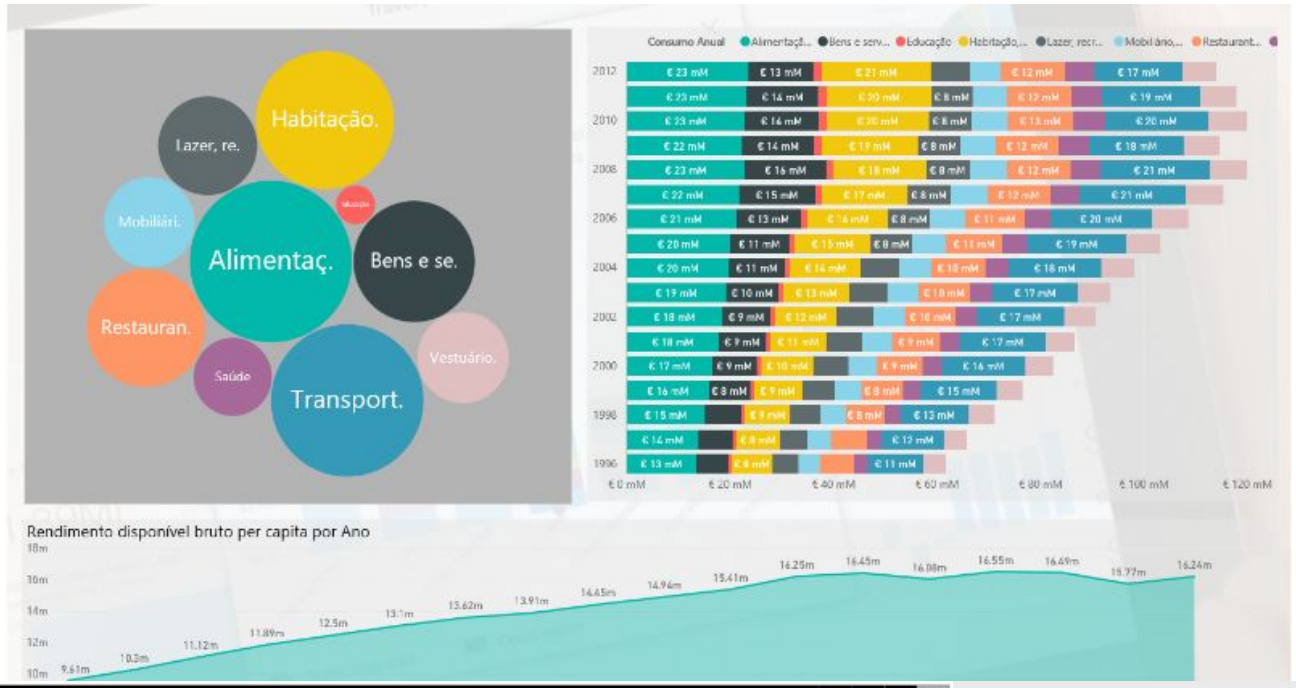
Batch Computing Big Data Analytic Process

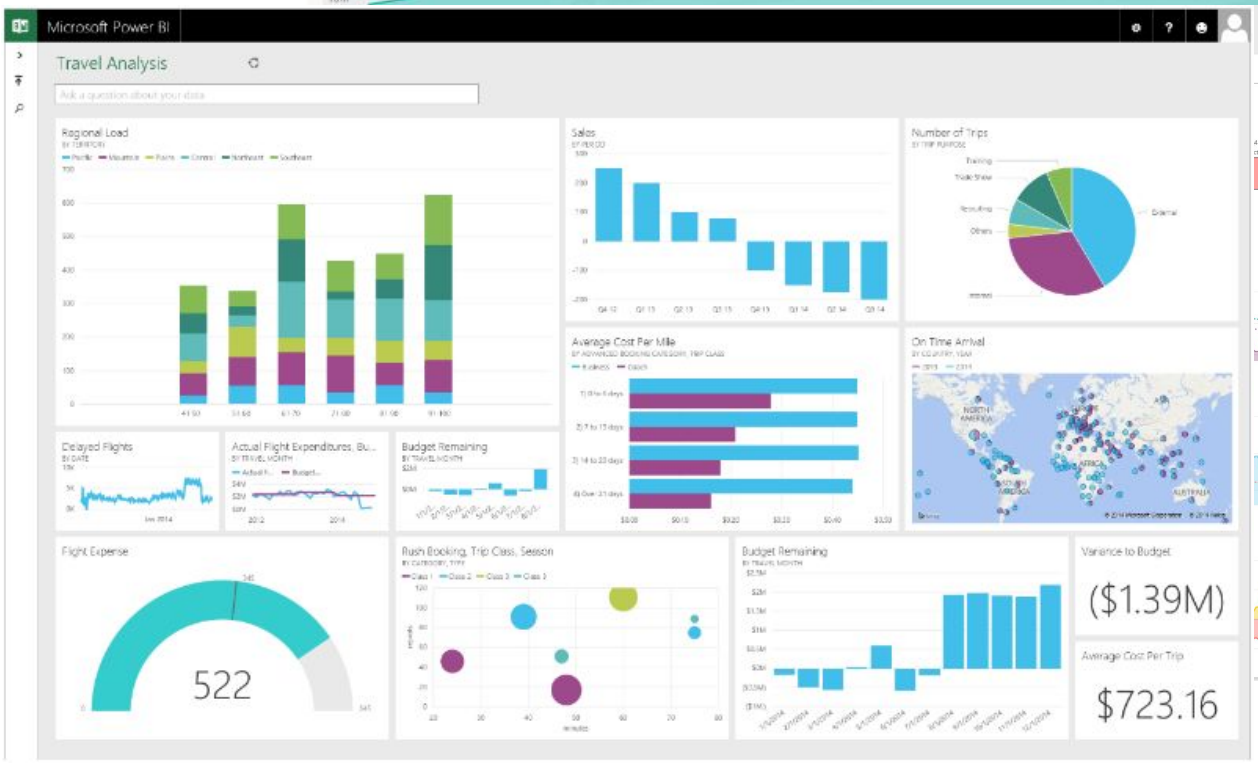


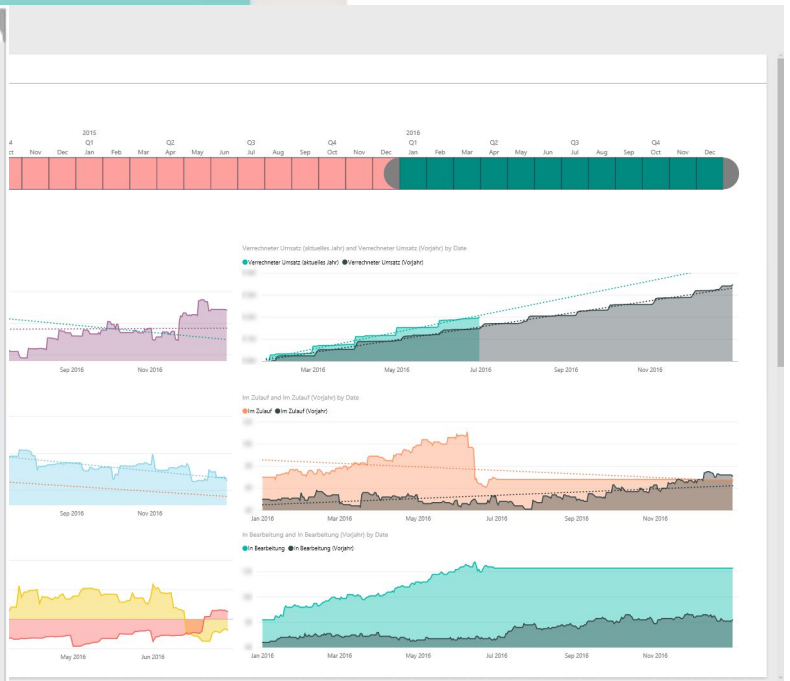
Real-Time Analytics and Alerting on Sensor Event Big Data



Big Data Dashboard Examples



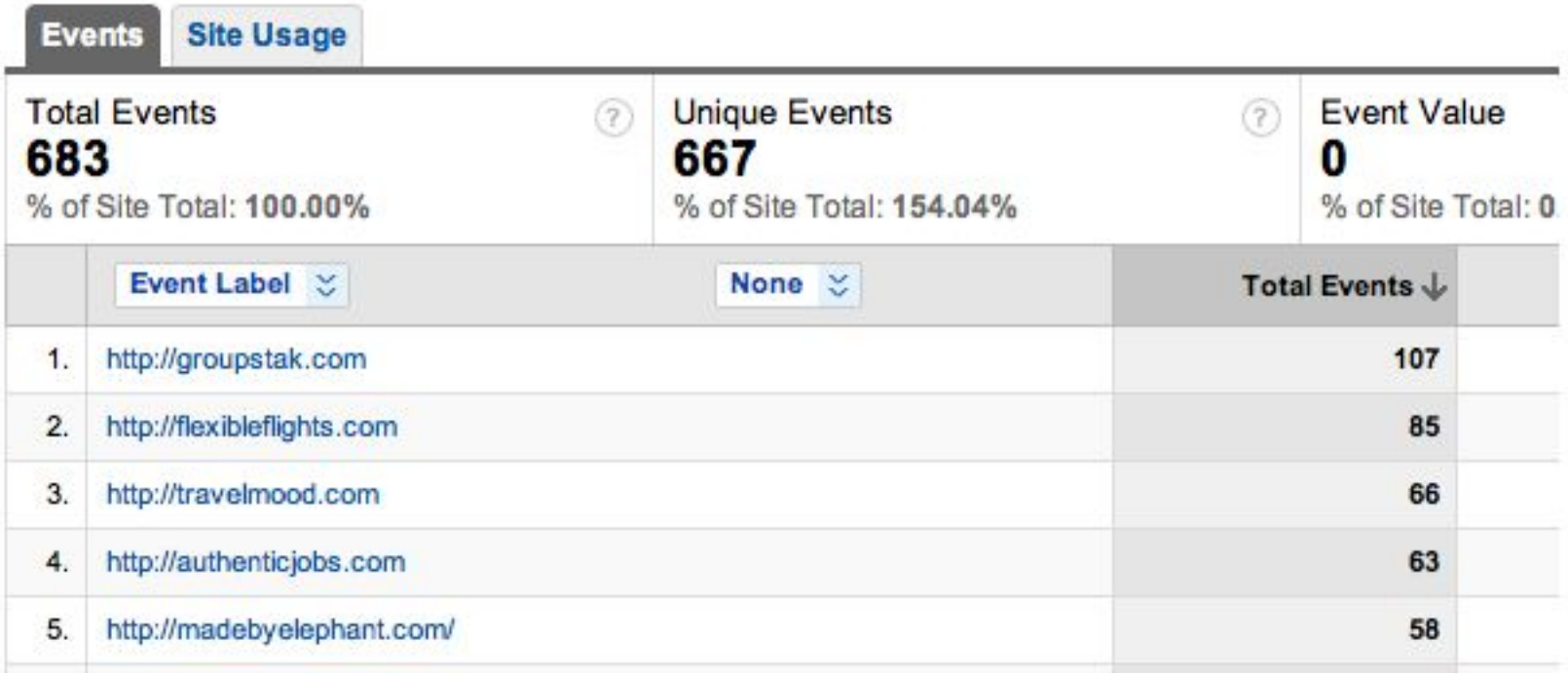




Track exiting traffic of website and app

Adding event tracking by using JQuery

```
$.expr[':'].external = function(obj){
  return !obj.href.match(/^mailto:/)
    && (obj.hostname != document.location.hostname);
};
```

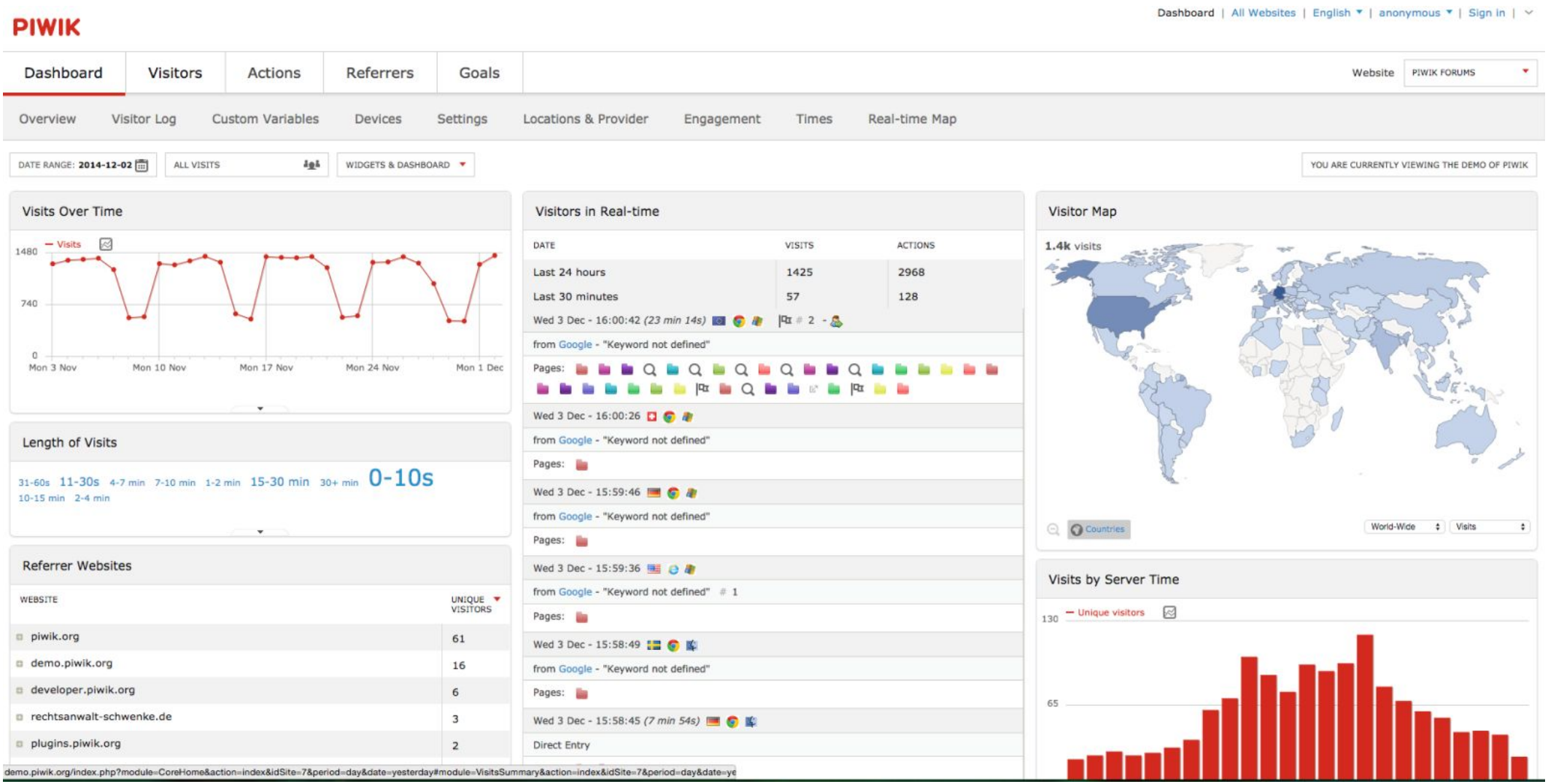


Events | Site Usage

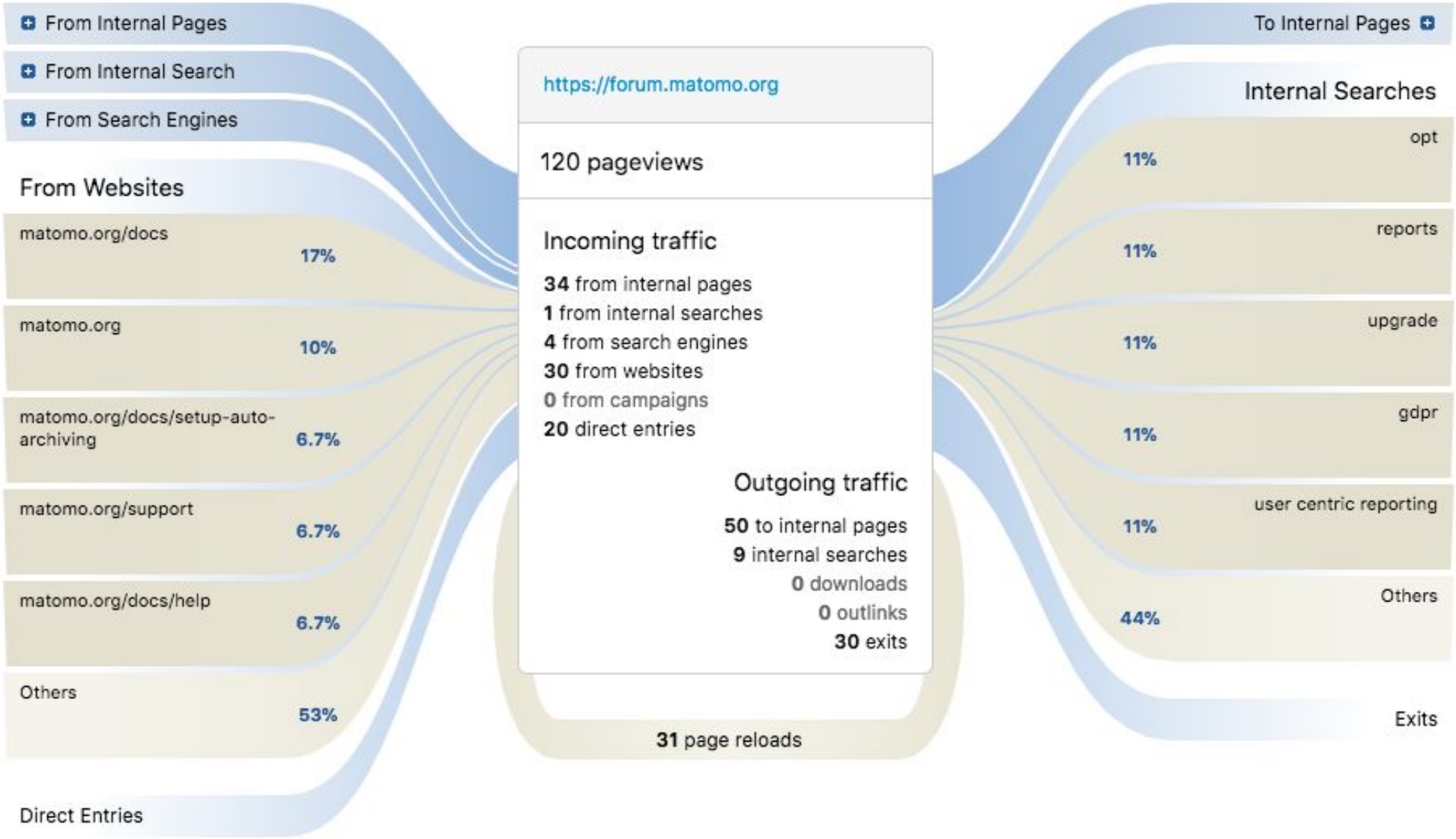
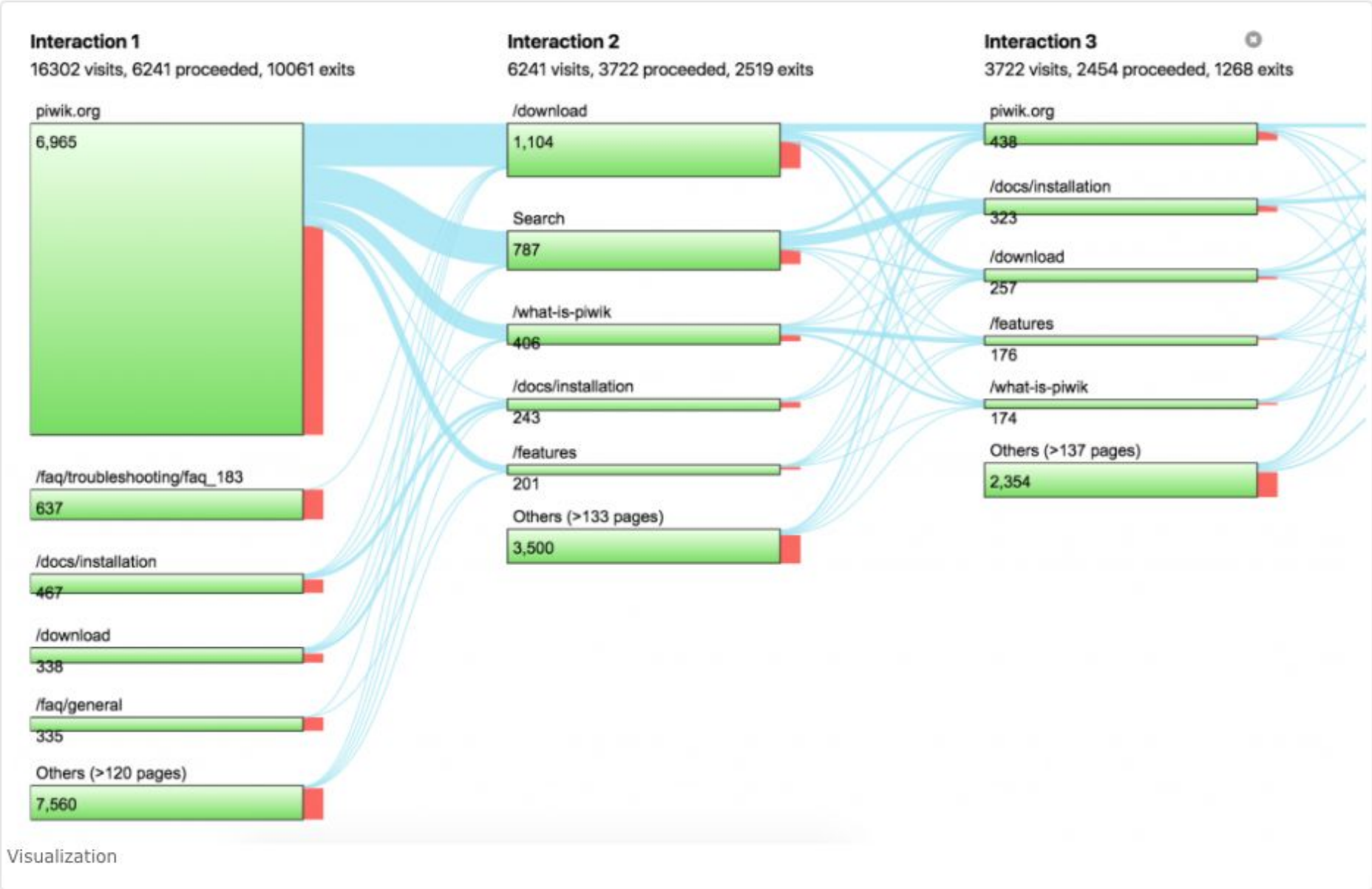
Total Events	Unique Events	Event Value
683	**667**	**0**
% of Site Total: **100.00%**	% of Site Total: **154.04%**	% of Site Total: **0**

	Event Label	None	Total Events ↓
1.	http://groupstak.com		107
2.	http://flexibleflights.com		85
3.	http://travelmood.com		66
4.	http://authenticjobs.com		63
5.	http://madebyelephant.com/		58

Analytic Platform



Transition Report



Users Flow Visualization