Cloud based Data Analytics for News Agency

About Customer

Industry: News Agency

Client is based in Ireland and is a leading quality newspaper, delivers news, opinion and analysis to 793,000 readers each week. Company began its online editions last year to reach its global readers.

Key Challenge

Rapidly growing agency was facing challenge of analyzing a large amount of historical data from various data sources and generate the meaningful analytics out of that.

Data Sources:

- Amazon Redshift Cloud DB
- Weblogs: Capture user behavior and browsing patterns on the web.
- Google Analytics: Capture online traffics information.
- Business Application: Capture the master and transactions information.

Some of the key challenges were:

- Consolidate the data from various semi-structured and structured database and populate into a data warehouse.
- Mapping of data from various data sources.
- Generating behavior analytics, authors' performance, subscribers, segment, marketing analytics.
- Define the Subscriber Churn Model to minimize the customer attrition.

BIZDataPro Solution

BIZDataPro implemented a Cloud based BI Analytics Solution using Pentaho community stack with below features:

- Analyze data sources and derive the KPIs and KRAs jointly with client to meet the business requirements.
- Consolidate various data sources like Amazon Redshift Cloud DB, Weblogs (Semi-structured data), Polopoly(Oracle 11g), Google Analytics(Google APIs), Video brightcove(Rest APIs).
- Deployed various complex ETL jobs and transformation to integrate historical data and live data to generate near
- Interface development to monitor the ETL jobs status as well as perform data cleansing.
- Customize the Pentaho open source code to achieve specific business requirements.

Page View/Reader Analysis

- To visualize the number of readers for each of the category like sports/newspaper/politics etc.
- Segment different type of readers i.e. occasional/frequent/Engaged readers

- Linear Regression for future page views
- Separate analysis for the Mobile readers and Website readers
- Customization to add the comments/descriptions for any of the data points.

Content Analysis

- Visualize different types of author articles and viewers of the articles.
- Top 10 authors with option to check the granularity of article and headline.
- Analyze the history of the different articles published by authors and popularity of those articles.
- Hourly traffic analysis of the article viewers
- Top performing articles and sources which attracted viewers to go to the articles

Content Analysis

- Visualize the topline summary and percentage engagement by entitlement
- Subscriber content preferences and breakdown for subscription type
- Subscriber trends analysis
- Analysis on added Versus Cancelled Versus Pending subscribers
- Effective campaign recommendation to retain the subscribers

System GUI Screens

- Pentaho CE Edition, HP Vertica

Business Benefits

BIZDataPro's successful implementation of Cloud based BI Analytics Solution helped news agency in meeting its subscription and digital marketing goals.

- Centralized authors, subscriber, articles and viewers for generating analytics.
- Minimize the subscriber churning using the effective data analytics.
- Get meaningful insights from historical and real time data.
- Effectively design the new digital media marketing campaign strategy.
- Identify performing and non-performing authors, articles and content.
- Saiku Analytics and Self-service Dashboard Plug-in helped end users to generate the adhoc analytics and own
- Centralized authors, subscriber, articles and viewers for generating analytics.
- Minimize the subscriber churning using the effective data analytics.

BIZDataPro

Systm GUI Screens

Key Performance Index

Reader Analysis (Weekly)

| 2.7 | Analysis Split: All Traffic | | Analysis Period: Last 24 weeks | |

Analysis Period : Monday 18 August 2014 to Sunday 1 February 2015

Website Analysis   Mobile Analysis

| 432,180 | Analysis Period ▲ +18.75% | 131,215 | Analysis Period ▲ +43.11 | 8,002,152 | |
| Total Readers (last week) | ☑ | Enagaged readers (+4) articles per week | ☑ | Total PageView (GA) | |

| 311,023 | Analysis Period ▲ +14.54% | 78,259 | Analysis Period ▲ +42.60% | 41,898 | Analysis Period ▲ +50.86% |
| Occasional Readers (1-5 per wk) | ☐ | Frequent Readers (6-10 per wk) | ☐ | Heavy Readers (11+ per wk) | ☐ |

XLS

Total Readers ● Total Readers (+4) — Total Readers (Linear trend) — Total Readers (+4) (Linear trend)

Year-Week	Analysis Split	Comment
2014-52	All Traffic	Christmas Holiday fall

First Previous 1 2 3 4 Next Last

Content Dashboard

Date Range : 2016-01-01 2016-01-01   Site Section : ALL   Measure : Day1   Author : ALL

CopyFlow   Articles   Article Analyzer

Top-N : 20

Bubble Chart - Article Analyzer

Reader Source
- ● No Bubble
- ○ Home
- ○ Article
- ○ Index
- ○ Twitter
- ○ Facebook
- ○ Google
- ○ Others

Reader Next Read
- ○ Home
- ○ Index
- ○ Search
- ○ Article
- ○ Blog

XLS

Article ID	Headline	Author	Page View	Eng %
82138	Oscars	Sunny Pujara	9,614	94%
12345	Poll results	Charushi Pandit	7,865	55%
76252	Films to see in 2016	Rachana Shah	1,019	61%

First Previous 1 Next Last

Content Dashboard

| Date Range : 2016-01-01 | 2016-01-01 | Site Section : ALL | Measure : Day1 | Author : ALL |

CopyFlow | **Articles** | **Article Analyzer**

CopyFlow:
- 36 (24.5%)
- 61 (41.5%)
- 50 (34%)
- Contributor ■ Staff ■ Wire

PageView:
- 36,230 (17.5%)
- 68,661 (33.2%)
- 101,982 (49.3%)
- Contributor ■ Staff ■ Wire

Average:
- 1,006 (24.1%)
- 1,126 (27%)
- 2,040 (48.9%)
- Contributor ■ Staff ■ Wire

Top Ten Authors by Count

Author Name	Count
Sunny Pujara	47
Jinal Thakkar	6
Niha Shaikh	4
Dhslen Jogi	4
Hardie Shah	4
Amisha Patel	
Bhoomika Patel	
Jagat Patel	
Niraj Patel	
Vishal Kamlani	

PageView Performance on Day1

Writer Type	Site Section	Author	Copy ⇕	Page Views ⇕	Avg Page Views ⇕	Engagement ⇕	No. of readers ⇕	Occasional ⇕	Frequent ⇕	Heavy ⇕
Wire	ALL	Sunny Pujara	29	32,735	1,129	80%	32,701	49%	23%	28%
Contributor	ALL	Sunny Pujara	18	19,594	1,089	78%	19,528	56%	21%	23%
Staff	ALL	Jinal Thakkar	6	15,658	2,610	73%	15,646	54%	24%	22%
Staff	ALL	Hardie Shah	4	1,844	461	59%	1,844	35%	28%	37%
Staff	ALL	Niha Shaikh	4	682	171	64%	682	37%	26%	37%

First | Previous | 1 | 2 | Next | Last