Migrate a massive on-prem Hadoop data center to cloud with 10x lower TCO

aws partner network

Executive Summary

LatentView Analytics helped in building and executing a strategic plan to migrating massive on-prem data from Hadoop ecosystem to AWS cloud based data lake ecosystem and enabled continuous ingestion of high volume data at 10% lower TCO
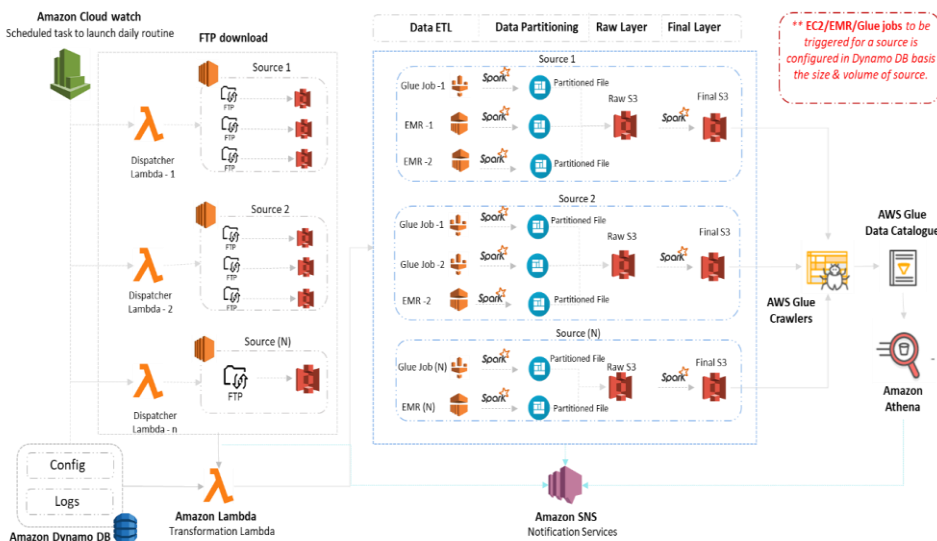
Business Challenge

- Prosieben had to handle 30% MoM growth of its data center data
- No analytical workloads were feasible to due insufficient capacity
- Exponential growth of business and data has led to capacity constraints in On-premise system with Full-load operation

Solution:

We migrated existing Data Warehouse and its pipeline carrying 28 sources of Data with distinct Varieties and Volumes in a configuration driven manner. We were also able to legacy ETL workflows with AWS native ETL components for better integration & cost efficiency
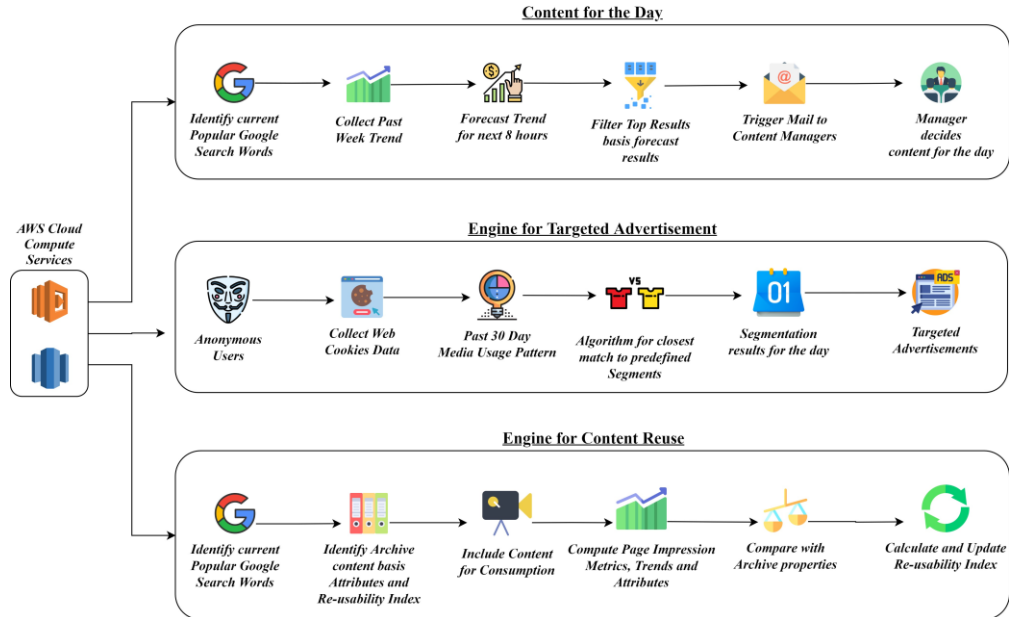
Architecture:

Solution Highlights:

- **Fast Data Availability**
  - 67% Reduction in lead time for Data processing from source to target
- **Cost Savings**
  - 22% savings in cost compared to On-premise maintenance
- **Informed Decision Making**
  - No more uncertainty about Data Consumption methodologies and time to results
- **Business Agility**
  - Reduction in lead time from Months to a Week to onboard new data sources
- **Business Continuity**
  - No more Business process fails due to capacity constraints
- **Future Ready**
  - A robust system capable of handling all four levels of Analytics Maturity

Business Impact:

We were able to enable end to end analytical use cases and supported in accelerated platform adoption for client's data scientists and analysts.

Content for the Day

Identify current Popular Google Search Words → Collect Past Week Trend → Forecast Trend for next 8 hours → Filter Top Results basis forecast results → Trigger Mail to Content Managers → Manager decides content for the day

AWS Cloud Compute Services

Engine for Targeted Advertisement

Anonymous Users → Collect Web Cookies Data → Past 30 Day Media Usage Pattern → Algorithm for closest match to predefined Segments → Segmentation results for the day → Targeted Advertisements

Engine for Content Reuse

Identify current Popular Google Search Words → Identify Archive content basis Attributes and Re-usability Index → Include Content for Consumption → Compute Page Impression Metrics, Trends and Attributes → Compare with Archive properties → Calculate and Update Re-usability Index