Abhishek Satpathy
PES2202400893

Big Data Architecture for Digital Marketing Campaign Analytics

Abhishek Satpathy

BDA - Mini Project (2025)

Abhishek Satpathy
PES2202400893

Use Case Overview

An advertiser wants to build a campaign analytics platform to process high-volume user activity logs generated in real time. Each campaign produces up to 1 GB/hour, with peaks of 32 GB/hour across 50 concurrent campaigns. The platform should deliver insights like total hits, unique users, and trends by location and device—requiring a scalable big data solution for both real-time and batch processing.

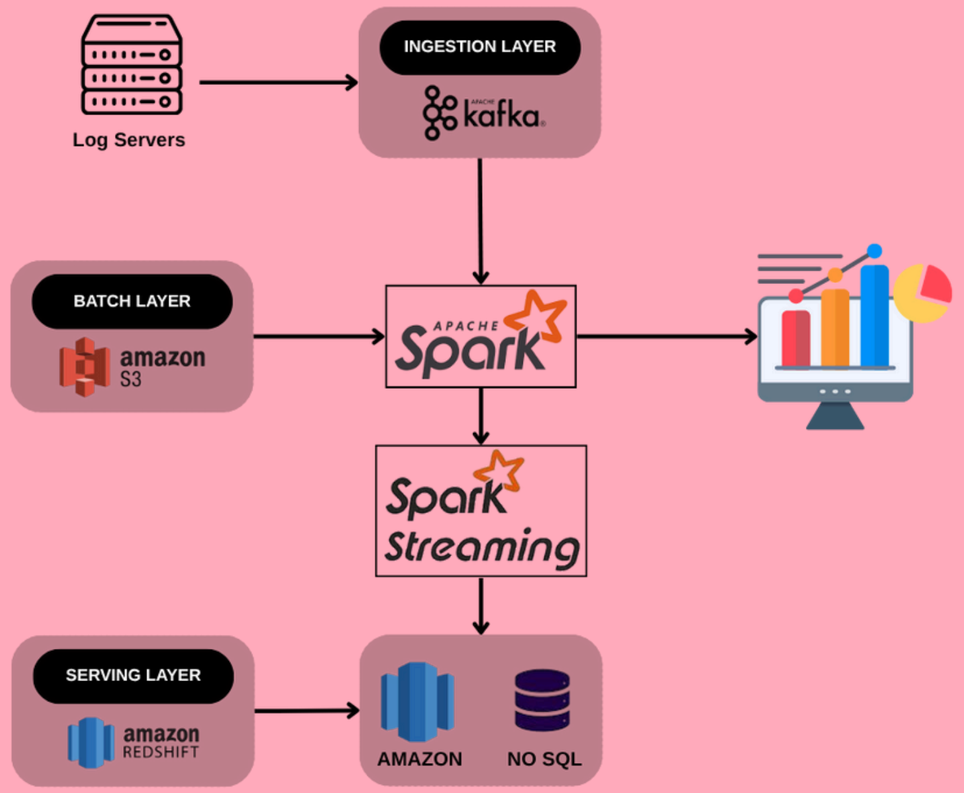
Abhishek Satpathy
PES2202400893

Proposed System Architecture

1. Ingestion Layer: **KAFKA**
2. Batch Layer: **SPARK on S3**
3. Speed Layer: **SPARK STREAMING**
4. Serving Layer: **REDSHIFT or NOSQL**
5. Visualization: **POWER BI or TABLEAU**

Abhishek Satpathy
PES2202400893

Proposed System Architecture



Abhishek Satpathy
PES2202400893

Architecture Type & Justification

The proposed architecture is a cloud-based Lambda Architecture, combining batch and real-time processing to handle large-scale, fast-moving campaign data. It ensures scalability, flexibility, and timely insights, making it ideal for dynamic digital marketing environments.

Abhishek Satpathy
PES2202400893

Data Ingestion Layer

- Tool: Apache Kafka
- Purpose: Real-time streaming of JSON logs
- Capacity: Handles high-throughput data spikes
- Alternative: AWS Kinesis for managed streaming

APACHE kafka®

Abhishek Satpathy
PES2202400893

Data Storage Layer

- Raw Storage: Amazon S3
- Processed Storage:
- NoSQL (DynamoDB): Fast lookup
- SQL (Redshift): Historical querying
- Format: Store as JSON, convert to Parquet for optimization

amazon S3

amazon REDSHIFT

Amazon DynamoDB

Abhishek Satpathy
PES2202400893

Data Processing Layer

- Batch Processing: Apache Spark for deep analytics
- Real-time Processing: Spark Streaming / Flink for live metrics
- Serving: Output pushed to storage & visualization layers

APACHE Spark

Spark Streaming

Abhishek Satpathy
PES2202400893

Analytics & Visualization

- Insights Generated: Total hits & Unique users
- Location- and device-based metrics
- Tools: Power BI, Tableau
- Use Case: Dashboards for campaign performance

Power BI

tableau

Abhishek Satpathy
PES2202400893

Security & Data Governance

Security is ensured through role-based access control using IAM policies, along with encryption of data both at rest and in transit. Audit logs and access monitoring help maintain data integrity and compliance. The architecture supports secure handling of sensitive campaign data across all layers.

Abhishek Satpathy
PES2202400893

Advantages of Proposed Architecture

The proposed architecture provides a balanced approach to handling high-volume, real-time marketing data while ensuring flexibility and scalability. It supports quick insights, secure data handling, and easy integration with visualization tools.

- Scalable & Efficient: Handles high data volume from multiple campaigns with ease.
- Real-time & Batch Insights: Enables quick decisions and deep historical analysis.
- Secure & Modular: Ensures data protection with a flexible, cloud-based design.

Abhishek Satpathy
PES2202400893

Thank You