DSE 584

Summer 2025

47509

CPT Internship Report

Amazon Web Services Inc.

Shreya Sett

ASU ID: 1232650733

M.S. in Data Science, Analytics and Engineering (Computing and Decision Analytics)

Saransh Burman

Sr. Solutions Architect

saranshb@amazon.com

+1-818-943-7775

--------------------------- ----------------

**Industry Mentor’s Signature**  Date

--------------------------- ----------------

**Faculty Chair’s Signature** (Thesis & PhD students) Date

--------------------------- ----------------

**Student’s Signature** Date

Student: By signing you certify that you have not plagiarized any of this document.

--------------------------- ----------------

Program Chair’s Signature (Internal) Date

1. **Company information**

Amazon Web Services is a leading cloud computing platform and a subsidiary of Amazon, launched in 2006. It offers a wide range of services, such as compute (EC2), storage (s3), databases (RDS, DynamoDB), machine learning (SageMaker), analytics, security and more. These services are delivered on a pay-as-you-go allowing businesses of all sizes to access scalable, reliable and cost-effective infrastructure without investing in physical hardware. AWS serves millions of customers globally, including startups, large enterprises, government agencies, and educational institutions.

One of the key strengths of AWS is it’s global infrastructure, with data centers in multiple geographic regions to ensure high availability and low latency. The platform considers security as it’s top priority for all endeavors. The platform also emphasizes innovation offering tools that help businesses meet complex regulatory requirements while rapidly deploying new applications. AWS leads the cloud market, enabling digital transformation across industries such as healthcare, finance, media and retail, and supporting emerging technologies like artificial intelligence, Internet of Things, and serverless computing.

Cloud Sales Center (CSC) inside AWS is a centralized team that helps customers especially small to mid-sized businesses adopt AWS cloud solutions. It is designed to support scalable customer outreach, provide technical guidance, and drive cloud adoption through inside sales, solution architecture support, and customer success management. The CSC usually works remotely or from regional hubs to assist customers through their cloud journey including onboarding, optimization and expansion. This team includes Account Managers, Cloud Sales Representatives, Solution Architects, and Customer Success Managers, who collaborate to help customers understand AWS services, make cost effective choices, and scale their cloud usage efficiently.

A Solutions Architect is a pivotal role within CSC, responsible for designing, planning, and overseeing the implementation of complex IT systems and infrastructure. They bridge the gap between business needs and technical solutions, working closely with stakeholders to understand requirements and translate them into robust, scalable architectures. Solutions Architects must possess a broad knowledge of various technologies, platforms, and best practices, allowing them to make informed decisions about system design, integration, and optimization. They often lead technical discussions, evaluate new technologies, and provide guidance to development teams. The role requires a balance of technical expertise, strategic thinking, and communication skills, as Solutions Architects must effectively convey complex technical concepts to both technical and non-technical audiences.

A Solutions Architect intern is expected to assist in the design and implementation of IT solutions under the guidance of experienced professionals. This internship is designed to provide hands-on experience in cloud computing, particularly with AWS services, and exposure to real-world business challenges. The intern will be involved in creating proof-of-concept projects, documenting system designs, and participating in client-facing presentations. The role requires a strong foundation in computer science principles, enthusiasm for learning new technologies, and the ability to think critically about complex problems. The intern is expected to contribute ideas, ask insightful questions, and gradually take on more responsibilities as they grow in the role. This internship aims to develop the intern's technical skills, business acumen, and professional network, preparing them for a potential career as a full-fledged Solutions Architect. The intern will gain valuable experience in translating business requirements into technical solutions, understanding cloud architecture best practices, and developing the communication skills necessary to convey complex technical concepts to various stakeholders.

1. **Background about the problem**

In today's digital age, social media has become a vital source of market intelligence and consumer insights. Social media platforms generate massive amounts of user-generated content that contains valuable information about brand perception, consumer sentiment, and market trends. However, the challenge lies in efficiently processing and analyzing this huge volume of data in real-time to extract actionable business insights.

The problem arose from a social media company's need to maintain competitive advantage and make data-driven decisions in an increasingly dynamic market environment. Traditional methods of market analysis often rely on historical data and can take days or weeks to generate insights, by which time the information may become outdated. The company needed a solution that could provide immediate insights from social media conversations, allowing them to respond quickly to market changes, customer concerns, and emerging trends.

For the social media platform with 50 million monthly active users, it faced significant challenges in monetizing its platform and providing value to business customers. Despite its growing user base, the company struggled to compete with larger platforms that offer advanced analytics tools.

Key Issues:

1. Limited Real-time Insights: The company’s traditional analytics system operated on batch processing, causing delays in identifying emerging trends and brand sentiment shifts. This lag time results in missed opportunities for both the platform and its business clients.
2. Manual Analysis Bottleneck: The analytics team spent considerable time manually processing data, leading to delays in reporting and inefficient use of resources. This manual approach couldn’t keep pace with the volume and velocity of data generated by the platform's users.
3. Lack of Competitive Intelligence: Without real-time analysis capabilities, the company could not provide its business customers with timely insights about their competitors or industry trends, putting them at a disadvantage in the fast-paced social media landscape.
4. Ineffective Monetization: The platform's inability to offer advanced, real-time analytics had hindered its ability to justify premium pricing for its services, resulting in lower revenue per business customer compared to competitors.
5. Security Concerns: As data volumes grow, ensuring proper security measures, including access controls and encryption, which became increasingly challenging with the current system.

The project required a proof-of-concept analytics platform that revolutionizes the social media platform’s data capabilities through cloud services. The solution was expected to capture streaming social media data in real-time, process millions of posts using serverless computing, and store data efficiently using both high-performance and cost-effective storage solutions. By integrating AI-powered sentiment analysis and advanced models for trend detection, the platform was intended to deliver instant insights through interactive business dashboards. The architecture must aim to ensure near-instantaneous data processing while maintaining robust security through encryption and strict access controls. This comprehensive solution was meant to enable the social media company to provide real-time campaign metrics, competitive intelligence, and automated trend analysis to business clients, transforming the platform's analytics offering from reactive to proactive. The enhanced capabilities was expected to strengthen the company's market position and create new revenue opportunities through advanced business intelligence services, directly addressing the platform's current monetization challenges.

1. **Focus of the Effort**

To address the solution, I began by identifying the business problem and aligning it with technical goals, such as scalability, cost-efficiency, low latency, and ease of deployment. I broke the overall challenge into smaller components: data ingestion, real-time processing, sentiment and trend analysis, summarization, and dashboard visualization. I took a modular approach to system design, which allowed for flexibility and independent testing of each part of the pipeline. Each stage was iteratively validated, tested, and optimized for performance in the AWS cloud environment.

A significant foundation for this project was my academic background, particularly the course **CSE 511: Data Processing at Scale**, which I took in my first semesterI drew upon the foundational knowledge I gained from the course CSE 511: Data Processing at Scale, which I took during my first semester. This course provided in-depth exposure to building and managing scalable data pipelines, which was directly relevant to the project I was involved in during my internship. It emphasized practical applications of data engineering in cloud environments, with a particular focus on tools such as Amazon Kinesis, AWS Lambda, and Amazon S3 — many of which were central to my internship tasks.

CSE 511 taught me how to design distributed data processing systems capable of handling high-throughput and low-latency requirements, a critical need for real-time analytics and large language model (LLM) support. I applied these learnings by working on real-time ingestion pipelines, implementing transformation logic, and integrating the data flow with analytics and AI services.

In designing the dashboard for, I applied a variety of visualization techniques learned in the Data Visualization class (CSE 578). For location-based trend analysis, I implemented a Sankey diagram, which effectively illustrated the flow of trending topics across different countries or regions. This allowed users to easily trace the origin and spread of viral content. To represent user interests across geographical areas, I employed a choropleth map. This visualization technique used color gradients to show the intensity of interest in specific topics or brands across different locations, providing an intuitive geographical overview of user engagement. For demographic analysis, I utilized donut charts and histograms which effectively showcases age based and gender based engagement metrics.

To implement the solution, I explored and applied various AWS services for streaming analytics, LLM integration, and data visualization. I learned how to design near real-time systems with sub-second latency using event-driven architectures. I investigated optimal data formats (like JSONL) for LLM readiness and how to configure Kinesis shards and retention policies to meet specific load requirements. I also gained practical experience with Bedrock Knowledge Bases and Bedrock Agents for Retrieval-Augmented Generation (RAG) to integrate generative AI in the pipeline.

In order to ensure the system was functioning as expected, I performed several measurements. These included latency tracking from data ingestion to dashboard rendering, monitoring throughput on Kinesis shards, assessing summarization quality, and measuring dashboard refresh intervals in Amazon QuickSight. These metrics helped evaluate system responsiveness, scalability, and user experience, leading to further fine-tuning.

Overall, this project allowed me to combine academic knowledge with hands-on cloud architecture skills, bridging theory and practice. It provided me with a deeper understanding of building production-ready AI-driven data platforms in a real-world cloud setting.

1. **Results**

The solution delivered during the internship provided a centralized interface where users could explore, monitor, and interact with real-time social media data relevant to their brand or domain. With the ability to access insights directly after logging in, users were empowered to take immediate action based on evolving discussions, customer sentiment, and thematic shifts. The platform was adopted by both technical and non-technical stakeholders across marketing, customer success, and leadership teams to track online brand performance and perception.

A key outcome was the simplification of trend discovery. Users could now identify what topics were gaining attention across platforms without having to manually parse data or rely on lagging reports. The addition of concise summaries per topic allowed teams to quickly align on messaging and response strategies during fast-changing social cycles.

Another significant result was the introduction of real-time sentiment analysis of public conversations. This functionality enabled teams to gauge sentiment across different regions and timeframes, contributing directly to campaign adjustments, brand crisis management, and proactive community engagement. Feedback showed that marketing and PR professionals were able to tailor communication strategies with more precision and responsiveness.

The data visualization capability also supported decision-making for executives. Users could interact with visual insights on demographics, engagement trends, and behavioural shifts—facilitating rapid performance reviews and quarterly reporting. This reduced dependency on manual data aggregation and improved stakeholder communication.

Lastly, the solution's conversational interface offered a contextual assistant experience that was integrated with internal data. This proved particularly useful for exploratory analysis, generating instant reports, and answering domain-specific questions, thereby reducing the time-to-insight for everyday business users.

In sum, the solution streamlined access to high-quality, actionable insights, bridging data teams with business teams and transforming raw streaming data into decision-ready intelligence.

1. **Learning experience**