Modernizing the Telecom Invoice Generation Process

Transforming telecom invoice generation with a scalable Spark-based mediation layer that processes 30 million records in minutes instead of hours.

Overview

- Built a robust ETL pipeline using Apache Spark and Amazon EMR to automate the extraction, validation, and enrichment of telecom usage data.

- Developed a high-performance mediation layer that integrates with SureTax to generate accurate, tax-included invoices for global customers.

- The client achieved a 90% improvement in process efficiency, slashing daily billing cycles from 8 hours to just 14 minutes.

Client Profile

US-based global telecom solutions expert providing customer experience and call center technology to mid-sized and large businesses. The company sells cloud-based and on-premises software, contact center applications, and omnichannel customer engagement solutions.

Challenges: Processing Millions of Records

- **Legacy Bottlenecks:** The existing billing process took over 8 hours daily, delaying financial reporting and customer invoicing.

- **Data Quality Issues:** High volumes of raw data from multiple sources required rigorous auditing and validation to ensure billing accuracy.

- **Complex Tax Integration:** The need to calculate precise taxation across various jurisdictions required seamless integration with SureTax.

- **Scale Requirements:** The system needed to handle over 30 million records from 30,000+ files daily without performance degradation.

QBurst Solution: Scalable Spark-Based Mediation

We delivered a solution that leverages the distributed computing power of Apache Spark running on Amazon EMR to create a unified information base for the client's billing system.

- **Dual-Stage Aggregation:**

  - **Daily Aggregation:** Python scripts ingest CSV data into Amazon S3, where Spark applications perform validation, derivation, and enrichment before storing data in optimized Parquet format.

  - **Monthly Aggregation:** A secondary Spark process executes complex rating and taxation logic, saving finalized billed data into the production database.

- **Data Audit Model:** Implemented a custom audit layer within the ETL process to ensure data integrity and fix existing system errors during transit.

- **Cloud-Native Scalability:** Utilized Amazon EMR to dynamically scale compute resources, allowing the system to handle sudden spikes in call volume and data traffic.

Technical Highlights

- **High-Performance Storage:** Transitioned to Parquet format to enable faster data retrieval and significantly reduce storage costs.

- **Automated Rating & Taxation:** Integrated SureTax via APIs to automate tax-inclusive invoice generation.

- **Standardized Formatting:** Provided a unified data format across the mediation layer, improving compatibility with downstream financial systems.

- **Fault-Tolerant Processing:** Fixed long-standing issues in the legacy data flow through rigorous ETL testing and error-handling logic.

Impact

- **97% Faster Daily Billing:** Reduced processing time from 8 hours to 14 minutes for 30 million records.

- **Ultra-Fast Monthly Runs:** Final monthly billing cycles now complete in just 2.2 minutes.

- **90% Process Efficiency:** Dramatically improved the speed and accuracy of the entire invoice generation workflow.

- **Improved Profit Margins:** Achieved through significant cost savings in infrastructure, time, and manual effort.

- **25% Faster Data Retrieval:** Standardized billing formats enabled rapid analysis and reporting for stakeholders.