gathr

Real-time business activity monitoring using Gathr

Business activity monitoring (BAM) provides end-to-end visibility of business processes, enabling enterprises to make better-informed decisions and quickly address problem areas.

A BAM framework is used for continuous event correlation, real-time alerts, and monitoring KPI statistics for business activities in real-time.

As business processes become increasingly complex, latency-sensitive, and data-intensive, KPI monitoring gets more challenging. With new metrics emerging, IT teams often create siloed solutions for specific business units or lines of business (LOBs), leading to poor visibility across multiple systems.

About the Customer

A US-based Fortune 500 mortgage lender that provides access to affordable mortgage finance to millions of Americans.

This mortgage lender wanted a BAM solution that could apply data reconciliation checks, aggregate data anomalies for reporting, and perform rule-based data quality checks in real-time. They were looking to create a solution for their existing

multi-stage data collection and processing system, which was loosely coupled and based on discrete technologies in a Staged Event-Driven Architecture (SEDA).

Business needs

Power real-time ingestion, event processing, and KPI computing

The customer needed to ingest data at varying rates and in different formats from multiple upstream applications. They wanted a solution to customize data aggregation rules and manage temporal event processing, late data arrival, and KPI computation.

Detect data issues with continuous matching and monitoring

The customer was looking for real-time data quality management to fix missing data issues, along with the ability to create gatekeeper rules for data attributes. They wanted to route bad data to error sinks for further analysis and ensure early detection of corrupt data feeds to avoid rework.

Ensure high performance, alerts, and customization

The customer needed a high-performance platform that supported parallel processing of data from multiple sources and enabled the addition of complex events in real-time. They wanted to seamlessly update processing logic at runtime, add/modify alerting scenarios as necessary, and leverage an enriched UI for better monitoring.

Solution

Gathr enabled the client to implement an end-to-end continuous monitoring solution leveraging out-of-the-box drag-and-drop operators and a custom development framework. Using a data-first, code-free approach, Gathr allowed users to effortlessly design and manage complex data flows on a visual canvas.

Solution highlights

- Provided ready-to-use capabilities for:
  - Data ingestion and quality management
  - Rule-based alerts
  - In-memory pattern matching
  - Event correlation
  - Data aggregation and profiling
  - Schema detection
  - Pipeline chaining and orchestration

- Deployed and managed data pipelines on Apache Spark-based infrastructure

- Created streaming and batch ETL pipelines for real-time alerting and historical data processing to generate SLA reports

- Enabled real-time data summary generation at different levels (including historical data) and provided an aggregated view of these

- Provided 200+ operators for data wrangling and transformation with support for custom extensions and late data arrival

- Enabled rapid, efficient prototyping, and operationalization of ETL frameworks

We delivered an SLA monitoring solution in 3 modules based on different types of pipelines:

1. Data ingestion, quality management and landing	2. Data correlation and SLA monitoring	3. End of day and month views of ETL processing data
• Ingested data from Enterprise Service Bus (ESB) and archived it in a common landing area (HDFS) • Used data quality processors to discard/normalize data failing business rules • Implemented polyglot architecture by landing validated data in HDFS and storing key data elements in an Elasticsearch index	• Forked streaming data in a pipeline for event processing and correlation • Handled SLA rules via in-built aggregation components • Accessed external reference data on-demand using built-in lookup functions • Monitored for unusual activity patterns, exceptionally high/low data rates, and data processed outside business hours	• Compared processed data from multiple data marts with data aggregated in the landing zone to identify dropouts • Built an index store to allow admin users to drill through raw, mismatched data • Scheduled an in-built workflow to trigger and manage batch ETL flows

Data processing, enrichment, and analysis for real-time BAM
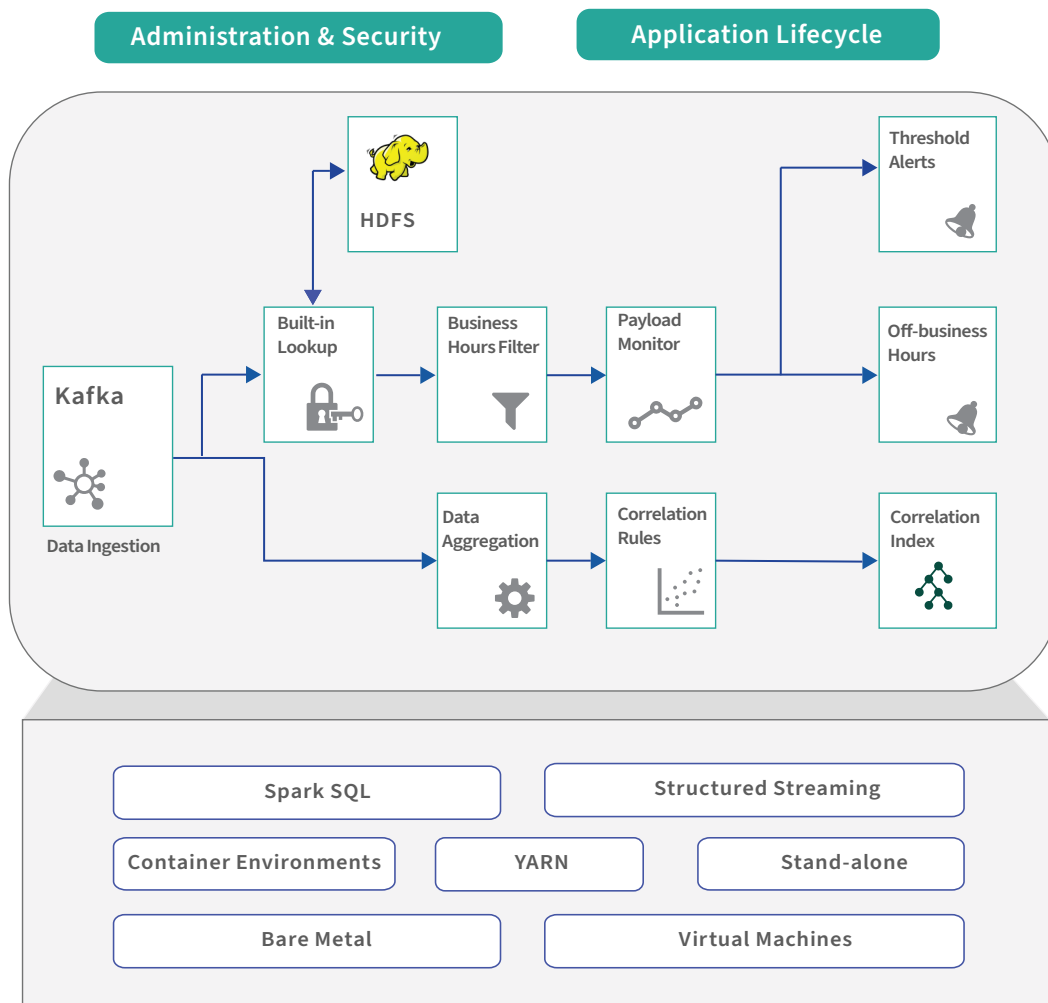
Business benefits

Gathr aggregated data from multiple sources, enhanced business rules, and provided real-time results for better monitoring, decision-making, and data quality management. This, in turn, helped the mortgage lender realize several strategic benefits:

- Enabled ingestion and processing of 1 TB data on a daily basis

- 10x faster data processing compared to the existing system

- Efficient near real-time KPI tracking with code-free business rule updates at runtime

- Improved visibility through correlation of real-time and historical data

- Better decision-making with instant SLA status checks

Easily build fast and reliable data pipelines using Gathr

Start for free

Gathr is a next-gen, cloud-native, fully-managed, no-code data pipeline platform. It's the only all-in-one platform for all your data integration and engineering needs – batch and streaming ingestion, CDC, ETL, ELT, data preparation, machine learning, and analytics. The Spark-based platform brings unmatched speed, performance and flexibility required to handle all types of data and analytics approaches, in ways that traditional ETL tools cannot. With Gathr's visual drag-and-drop interface, native integration for all popular data sources and destinations, an exhaustive set of pre-built operators, and a rich pipeline template gallery, anyone can build and deploy data pipelines, quickly and easily.

Visit www.gathr.one or write to us at contact@gathr.one