Powered By

A list of companies that are using Samza.

Company	Description
LinkedIn	Samza was originally developed at LinkedIn. It's currently used to process tracking data, service log data, and for data ingestion pipelines for realtime services.
Intuit	At Intuit, we use Samza to enrich events with more contextual data from various sources (CMDB, Change Management, Incident Management, Problem Management). This gives us more meaningful events that an operations centre person can act on.
MobileAware	MobileAware uses Samza as a foundation for two products: real time analytics and multi channel notification (push, text message and html5).
Fortscale	Fortscale is using Samza to processes security events log as part of the data ingestion pipelines and on-line machine learning models creation process.
Project Florida	Project Florida uses Samza for real-time monitoring of data streams from wearable sensors, for preventative healthcare purposes.
Improve Digital	Improve Digital is using Samza as the foundation of a realtime processing capability performing data analytics and as the basis for an alerting system.
Quantiply	Quantiply uses Samza to bring together user event, system performance, and business operational data for real-time visibility and decision support.
Happy Pancake	Happy Pancake, Northern Europe's largest internet dating service, is using Samza for all event handlers and data replication.
VinTank	VinTank uses Samza to power our social media analysis and NLP pipeline.
DoubleDutch	DoubleDutch uses Samza to power their analytics platform and stream data live into an event dashboard for real-time insights.
Banno	Jack Henry & Associates uses Samza to process user activity data across its Banno suite of products for financial institutions.
Metamarkets	Metamarkets uses Samza to transform and join real-time event streams, then forward them into a Druid cluster for interactive querying.
Digitalsmiths	Digitalsmiths, a TiVo company, is leveraging Samza to do online processing of views and ratings to help power personalized content recommendations and analytics.
State.com	State uses Samza to process and join streams of changes from MongoDB to update a wide range of realtime services that support the website and mobile apps. These include search, user recommendations, opinion metrics and lots more.
Uber	Uber uses Samza to provide stream processing as a service. Currently, this platform supports two categories of use cases: metrics aggregation and near real time state machine for doing database updates. We're planning to add more use cases related to fraud detection and root cause analysis in the near future.
Netflix	Netflix uses single-stage Samza jobs to route over 700 billion events / 1 peta byte per day from fronting Kafka clusters to s3/hive. A portion of these events are routed to Kafka and ElasticSearch with support for custom index creation, basic filtering and projection. We run over 10,000 samza jobs in that many docker containers.
Optimizely	Optimizely uses Apache Samza to aggregate and enrich billions of events per day to power real-time analytics of Experiments and Personalization experiences.
TripAdvisor	TripAdvisor uses Samza to process billions of events daily for analytics, machine learning, and site improvement.
Redfin	Redfin uses Samza and Kafka for sending millions of email and push notifications to our customers everyday. We chose Samza for distributed processing because it integrates really well with Kafka. Samza also provides managed state and a resilient local store which we found to be very useful features.
VMWare	vRealize Network Insight (vRNI) [http://www.vmware.com/products/vrealize-network-insight.html] is VMware's flagship product for delivering intelligent operations for software defined network environments (e.g. NSX). It can optimize network performance and availability by providing visibility and analytics across virtual and physical networks, provide planning and recommendations for implementing micro-segmentation security, and operational views to quickly and deterministically manage and scale VMware NSX deployments. At the heart of the vRNI architecture are a set of distributed processing and analytics modules that crunch large amounts of streaming data on a cluster of multiple machines. It is critical that these operations are carried out in a way that is reliable, efficient and robust - even in the face of dynamic faults in the underlying infrastructure layers. We have been successfully using Apache Samza as a distributed streaming data processing framework for executing these analytical modules reliably and efficiently at a very large scale, thus helping us focus on our core business problems.
Ntent	At Ntent, we use Samza to power our streaming content ingestion system. We take crawled web pages and news articles, and pass them through a multi-stage processing pipeline that cleanses, classifies, extracts features that power other learning models, stores, and indexes the content for search.
Movio	At Movio, we use Samza to process and enrich billions of change data capture events on all our databases in real-time.