Powered By

Want to appear on this page? Send a quick description of your organization and usage to the mailing list or to @apachekafka or @jaykreps on twitter and we'll add you.

Companies

- **LinkedIn** - Apache Kafka is used at LinkedIn for activity stream data and operational metrics. This powers various products like LinkedIn Newsfeed, LinkedIn Today in addition to our offline analytics systems like Hadoop.
- **Yahoo** - See this.
- **Twitter** - As part of their Storm stream processing infrastructure, e.g. this and this.
- **Netflix** - Real-time monitoring and event-processing pipeline.
- **Square** - We use Kafka as a bus to move all systems events through our various datacenters. This includes metrics, logs, custom events etc. On the consumer side, we output into Splunk, Graphite, Esper-like real-time alerting.
- **Spotify** - Kafka is used at Spotify as part of their log delivery system.
- **Pinterest** - Kafka is used with Secor as part of their log collection pipeline.
- **Uber**
- **Goldman Sachs**
- **Tumblr** - See this
- **PayPal** - See this.
- **Box** - At Box, Kafka is used for the production analytics pipeline & real time monitoring infrastructure. We are planning to use Kafka for some of the new products & features
- **Airbnb** - Used in our event pipeline, exception tracking & more to come.
- **Mozilla** - Kafka will soon be replacing part of our current production system to collect performance and usage data from the end-users browser for projects like Telemetry, Test Pilot, etc. Downstream consumers usually persist to either HDFS or HBase.
- **Cisco** - Cisco is using Kafka as part of their OpenSOC (Security Operations Center). More detail here.
- **Etsy** - See this article.
- **Tagged** - Apache Kafka drives our new pub sub system which delivers real-time events for users in our latest game - Deckadence. It will soon be used in a host of new use cases including group chat and back end stats and log collection.
- **Foursquare** - Kafka powers online to online messaging, and online to offline messaging at Foursquare. We integrate with monitoring, production systems, and our offline infrastructure, including hadoop.
- **StumbleUpon** - Data collection platform for analytics.
- **Coursera** - At Coursera, Kafka powers education at scale, serving as the data pipeline for realtime learning analytics/dashboards.
- **Shopify** - Access logs, A/B testing events, domain events ("a checkout happened", etc.), metrics, delivery to HDFS, and customer reporting. We are now focusing on consumers: analytics, support tools, and fraud analysis.
- **Cerner** - Kafka is used with HBase and Storm as described here.
- **Oracle** - Oracle provides native connectivity to Kafka from its Enterprise Service Bus product called OSB (Oracle Service Bus) which allows developers to leverage OSB built-in mediation capabilities to implement staged data pipelines.
- **Oracle Golden Gate** - GoldenGate offers a comprehensive solution that streams transactional data from various sources into various big data targets including Kafka in real-time, enabling organizations to build fault -tolerant, highly reliable, and extensible analytical applications.
- **CloudFlare** - CloudFlare uses Kafka for our log processing and analytics pipeline, collecting hundreds of billions of events/day data from a thousands of servers.
- **Mate1.com Inc.** - Apache kafka is used at Mate1 as our main event bus that powers our news and activity feeds, automated review systems, and will soon power real time notifications and log distribution.
- **Boundary** - Apache Kafka aggregates high-flow message streams into a unified distributed pubsub service, brokering the data for other internal systems as part of Boundary's real-time network analytics infrastructure.
- **Ancestry.com** - Kafka is used as the event log processing pipeline for delivering better personalized product and service to our customers.
- **DataSift** - Apache Kafka is used at DataSift as a collector of monitoring events and to track user's consumption of data streams in real time. http://highscalability.com/blog/2011/11/29/datasift-architecture-realtime-datamining-at-120000-tweets-p.html
- **Spongecell** - We use Kafka to run our entire analytics and monitoring pipeline driving both real-time and ETL applications for our customers.
- **Wooga** - We use Kafka to aggregate and process tracking data from all our facebook games (which are hosted at various providers) in a central location.
- **AddThis** - Apache Kafka is used at AddThis to collect events generated by our data network and broker that data to our analytics clusters and real-time web analytics platform.
- **Urban Airship** - At Urban Airship we use Kafka to buffer incoming data points from mobile devices for processing by our analytics infrastructure.
- **Metamarkets** - We use Kafka to ingest real-time event data, stream it to Storm and Hadoop, and then serve it from our Druid cluster to feed our interactive analytics dashboards. We've also built connectors for directly ingesting events from Kafka into Druid.
- **Simple** - We use Kafka at Simple for log aggregation and to power our analytics infrastructure.
- **Gnip** - Kafka is used in their twitter ingestion and processing pipeline.
- **Loggly** - Loggly is the world's most popular cloud-based log management. Our cloud-based log management service helps DevOps and technical teams make sense of the the massive quantity of logs. Kafka is used as part of our log collection and processing infrastructure.
- **RichRelevance** - Real-time tracking event pipeline.
- **SocialTwist** - We use Kafka internally as part of our reliable email queueing system.
- **Countandra** - We use a hierarchical distributed counting engine, uses Kafka as a primary speedy interface as well as routing events for cascading counting
- **FlyHajj.com** - We use Kafka to collect all metrics and events generated by the users of the website.
- **uSwitch** - See this blog.
- **InfoChimps** - Kafka is part of the InfoChimps real-time data platform.
- **Visual Revenue** - We use Kafka as a distributed queue in front of our web traffic stream processing infrastructure (Storm).
- **Oolya** - Kafka is used as the primary high speed message queue to power Storm and our real-time analytics/event ingestion pipelines.
- **Datadog** - Kafka brokers data to most systems in our metrics and events ingestion pipeline. Different modules contribute and consume data from it, for streaming CEP (homegrown), persistence (at different "temperatures" in Redis, ElasticSearch, Cassandra, S3), or batch analysis (Hadoop).
- **VisualDNA** We use Kafka 1. as an infrastructure that helps us bring continuously the tracking events from various datacenters into our central hadoop cluster for offline processing, 2. as a propagation path for data integration, 3. as a real-time platform for future inference and recommendation engines
- **Sematext** - in SPM (performance monitoring + alerting), Kafka is used for metrics collection and feeds SPM's in-memory data aggregation (OLAP cube creation) as well as our CEP/Alerts servers (see also: SPM for Kafka performance monitoring). In SA (search analytics) Kafka is used in

search and click stream collection before being aggregated and persisted. In Logsene (log analytics) Kafka is used to pass logs and other events from front-end receivers to the persistent backend.

- Wize Commerce - At Wize Commerce (previously, NexTag), Kafka is used as a distributed queue in front of Storm based processing for search index generation. We plan to also use it for collecting user generated data on our web tier, landing the data into various data sinks like Hadoop, HBase, etc.
- Quixey - At Quixey, The Search Engine for Apps, Kafka is an integral part of our eventing, logging and messaging infrastructure.
- LinkSmart - Kafka is used at LinkSmart as an event stream feeding Hadoop and custom real time systems.
- LucidWorks Big Data - We use Kafka for syncing LucidWorks Search (Solr) with incoming data from Hadoop and also for sending LucidWorks Search logs back to Hadoop for analysis.
- Cloud Physics - Kafka is powering our high-flow event pipeline that aggregates over 1.2 billion metric series from 1000+ data centers for near-to-real time data center operational analytics and modeling
- Graylog2 - Graylog2 is a free and open source log management and data analysis system. It's using Kafka as default transport for Graylog2 Radio. The use case is described here.
- Yieldbot - Yieldbot uses kafka for real-time events, camus for batch loading, and mirrormakers for x-region replication.
- LivePerson - Using Kafka as the main data bus for all real time events.
- Retention Science - Click stream ingestion and processing.
- Strava - Powers our analytics pipeline, activity feeds denorm and several other production services.
- Outbrain - We use Kafka in production for real time log collection and processing, and for cross-DC cache propagation.
- SwiftKey - We use Apache Kafka for analytics event processing.
- Yeller - Yeller uses Kafka to process large streams of incoming exception data for it's customers. Rate limiting, throttling and batching are all built on top of Kafka.
- Emerging Threats - Emerging threats uses Kafka in our event pipeline to process billions of malware events for search indices, alerting systems, etc.
- Hotels.com - Hotels.com uses Kafka as pipeline to collect real time events from multiple sources and for sending data to HDFS.
- Helprace - Kafka is used as a distributed high speed message queue in our help desk software as well as our real-time event data aggregation and analytics.
- Exponential is using Kafka in production to power the events ingestion pipeline for real time analytics and log feed consumption.
- Livefyre - uses Kafka for the real time notifications, analytics pipeline and as the primary mechanism for general pub/sub.
- Exoscale - uses Kafka in production.
- Cityzen Data - uses Kafka as well, we provide a platform for collecting, storing and analyzing machine data.
- Criteo - use Kafka in production for over a year for stream processing and log transfer (over 2M messages/s and growing)
- The Wikimedia Foundation - uses Kafka as a log transport for analytics data from production webservers and applications. This data is consumed into Hadoop using Camus and to other processors of analytics data.
- OVH - uses Kafka in production for over a year now using it for event bus, data pipeline for antiddos and more to come.
- Helpshift produces billions of events with Kafka through an erlang based producer ekaf that supports 8.0, and consumes topics primarily with storm and clojure.
- Parsely - Kafka is used for all data integration of analytics event data.
- VividCortex - VividCortex uses Kafka in our SaaS MySQL performance management platform to reliably ingest high-volume 1-second timeseries data.
- Trivago - Trivago uses Kafka for stream processing in Storm as well as processing of application logs.
- Ants.vn - use Kafka in production for stream processing and log transfer (over 5B messages/month and growing)
- IFTTT - We use Kafka to ingest real-time log and tracking data for analytics, dashboards, and machine learning.
- Homeadvisor - We use Kafka for logging and async event processing, among other uses.
- Skyscanner - the world's travel search engine, uses Kafka for real-time log and event ingestion. It is the integration point for of all stream-processing and data transportation services.
- IBM Message Hub - The Message Hub service in our Bluemix PaaS offers Kafka-based messaging in a multi-tenant, pay-as-you-go public cloud. It's intended to provide messaging services for microservices, event-driven processing and streaming data in to analytics systems.
- iPinYou is the largest DSP in China which has its HQ in Beijing and offices in Shanghai, Guangzhou, Silicon Valley and Seattle. Kafka clusters are the central data hub in iPinYou. All kinds of Internet display advertising data, such as bid/no-bid, impression, click, advertiser, conversion and etc., are collected as primary data streams into Kafka brokers in real time, by LogAggregator (a substitute for Apache Flume, which is implemented in C/C++ by iPinYou, has customized functionality, better performance, lower resource-consuming).
- MailChimp - Kafka powers MailChimp's data pipeline that in turn powers MailChimp Pro, as well as an increasing number of other product features. You can read some of the details here.