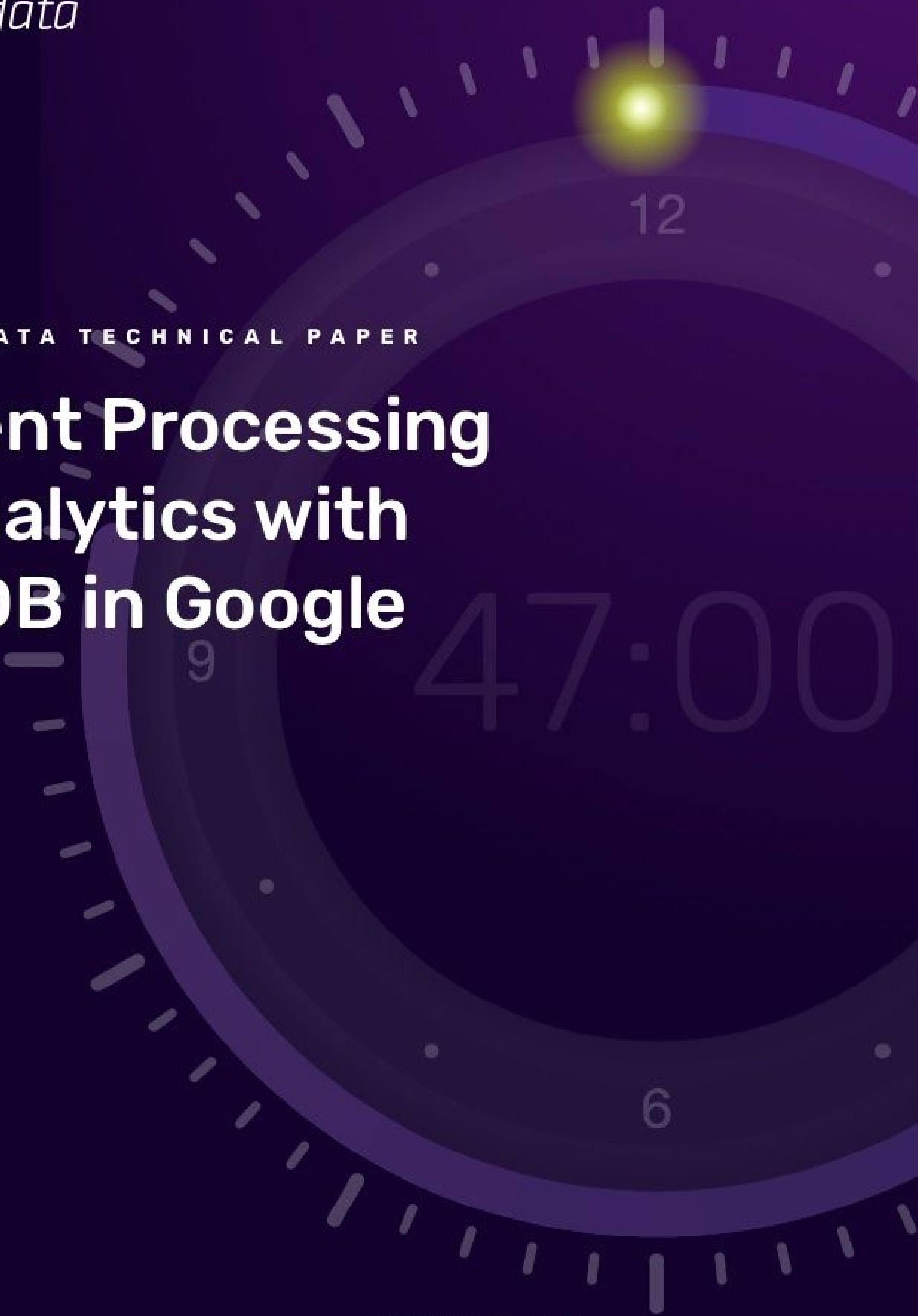
AN INFLUXDATA TECHNICAL PAPER

IoT Event Processing and Analytics with InfluxDB in Google Cloud



External Contributors
Christoph Bussler
Solutions Architect, Google

June 2020

Company in brief

Google Cloud Platform, offered by Google, is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products, such as Google Search, Gmail, file storage, and YouTube.

Google Cloud is a global, enterprise-grade supported infrastructure of computational assets available as an on-demand service. The Google Cloud platform reduces the burden on IT by helping customers modernize workloads on world-class infrastructure, protect data with multilayered security, drive decision-making with intelligent analytics and adopt hybrid and multi-cloud environments without vendor lock-in.

Case overview

Google Cloud has a native architecture for collecting, processing, analyzing and archiving of events from IoT devices, vehicles as well as upstream software systems. At the center of the architecture is InfluxDB and its connection to global native Google Cloud services like BigQuery, Cloud Machine Learning Engine and Kubernetes. The architecture demonstrates how access to global scaling cloud services addresses use cases from the energy sector.

> *“From a customer perspective, if you ask "How do I solve my problem with all that's out there?" That's where we come into play: how to connect all of this.”*
>
> ***Christoph Bussler**, Solutions Architect, Google*

The business problem

Many Google Cloud customers have IoT event processing and time-based processing use cases, which require a time series database. The solution described here is a use case of IoT event processing in the energy sector. The solution’s architecture is complex because it addresses several sub-use cases that exemplify what you can do with time series databases as well as what Google Cloud and its services can contribute to your architecture.

The solution answers common customer questions of how to put Google Cloud services to work for IoT event processing and analytics. In the context of IoT events in the energy sector, InfluxDB is the time series component within this solution's architecture,

Overview of energy sector use cases

IoT is one way of monitoring energy systems, whose use cases can be classified into three categories:

1. **Production**:
   - Energy production systems monitoring and anomaly detection (oil, gas, wind. hydro, solar and others)

2. **Distribution**:
   - Smart grid: Maintaining an equilibrium across energy supply and demand
     - Renewable and non-renewable energy demand forecasting
     - Using machine learning to predict future behavior based on past behavior (synchrophasor technology uses monitoring devices, which take high-speed measurements of phase angles, voltage and frequency that are time-stamped with high-precision clocks)

3. **Consumption**:
   - Fleet performance optimization (cars, trucks, planes, etc.)
   - Commercial manufacturing sites, office buildings
   - Public infrastructure
   - Private households

From production to distribution to consumption, there are many phases, technologies and aspects that need to be simultaneously monitored, maintained, controlled, and forecast. Time series databases are one mechanism to put it all into perspective.

How energy environments can be analyzed

Energy use case analysis encompasses:

- **Event collection** (from production systems, consumption locations, and absence check which requires static equipment inventory data)
- **Event monitoring** (outages, trends, anomaly detection)
- **Forecasting/prediction** (combination of current events, historic events, non-event data like models, weather data, road conditions, etc.)

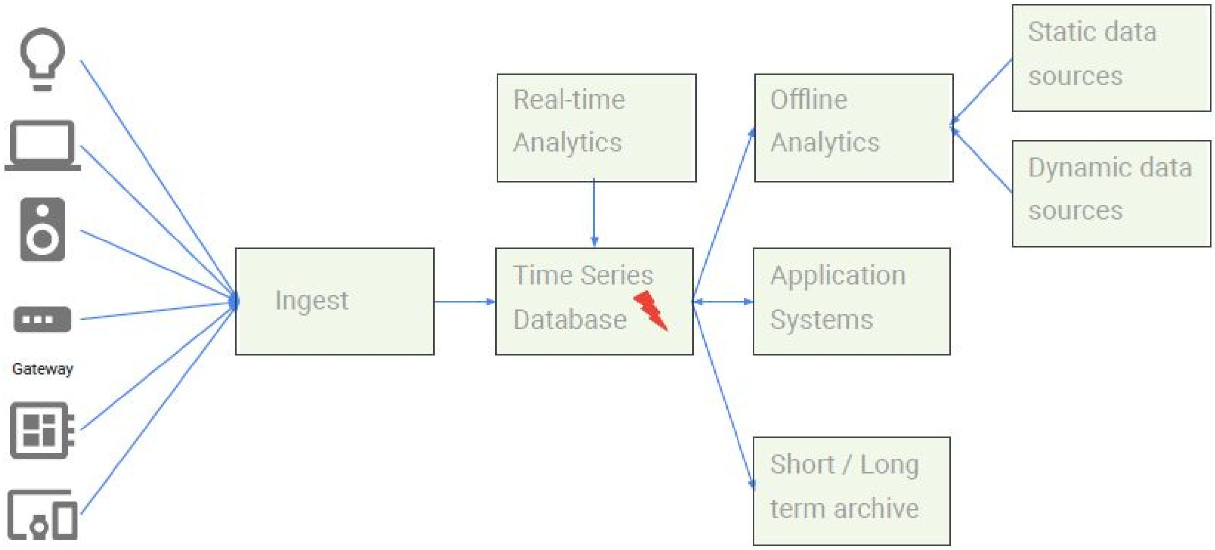
- **Off-line analysis** (combination with non-event data)
- **Archiving** (a time horizon of event collection to enable long-duration analysis and predictions as well as data storage to meet regulatory requirements)

One solution that can cover all the above energy use cases would be ideal — this is what the IoT event processing and analysis solution with InfluxDB in Google Cloud does.

The technical problem

From a technical perspective, the solution had to fit the life cycle of an IoT event, shown below.

Life Cycle of an IoT Event



The above life cycle architecture diagram addresses, to a large extent, all the energy use cases discussed above. (Some use cases might require adding to or modifying this architecture to meet specific needs.) The icons on the left represent various input types, such as sensors and devices. These inputs produce events that need to be ingested at speed since IoT events are often produced at a high rate. The dataflow is as follows:

- An event is issued by some type of device and **ingested by a time series database**. The database might have real-time analytics capability enabling analysis of events as they stream in, and enabling a view of all events and event subsets to observe change in the value of metrics they represent. The source of each event is important to understand, and a time series database allows such events to be stored and accessible for monitoring, As events are continuously added, real-time analytics enables a real-time view of performance.

- **Offline analytics** can be done, such as in the case of a fleet management system by analyzing what's happening to the fleet (where vehicles are located at a given point in time and where they might be going). In the case of vehicle monitoring, for example, events are sent to separate analytics subsystems which might be based on:
    - **Static** data like an inventory of all fleet vehicles
    - **Dynamic** data sources like the allocation and maintenance state

  A complex environment could have a large number of static and dynamic data sources that have to be combined in order to accomplish offline analytics.

- Various **application systems** might depend on events. Among them might be an event monitoring system which was established at a previous time and which was using a different technology. The two-way blue arrow indicates that there might be events in the infrastructure that should be sent *to* the time series database for correlation with the current data set, and not just *from* the database to other systems.

- **Short-term and long-term archiving** of events allows re-examining events as needed, years or decades later.

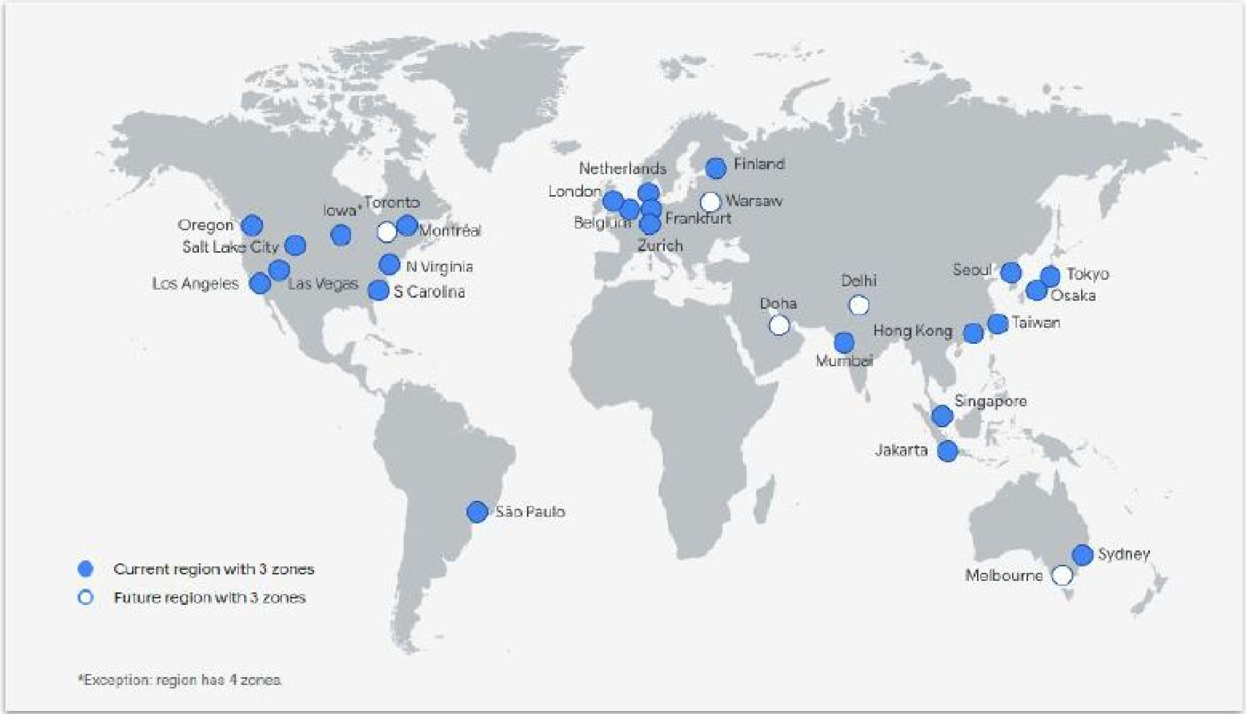
Let's consider the above life cycle in the context of Google Cloud's global network and services.

Google Cloud's global network and services

Google Cloud's planet-scale infrastructure delivers the highest level of performance and availability in a secure, sustainable way:

- Google Cloud's computers, hard disk drives and virtual resources, such as virtual machines (VMs), are contained in Google's data centers around the globe. Each data center location is in a region. Regions are available in Asia, Australia, Europe, North America, and South America. Each region is a collection of zones, which are isolated from each other within the region.
- Google Cloud's global data center distribution allows you to write global applications and put devices everywhere feeding into the same system no matter where they are. This distribution of resources provides several benefits, including redundancy in case of failure and reduced latency by locating resources closer to clients.

Google Cloud: Regions and Zones

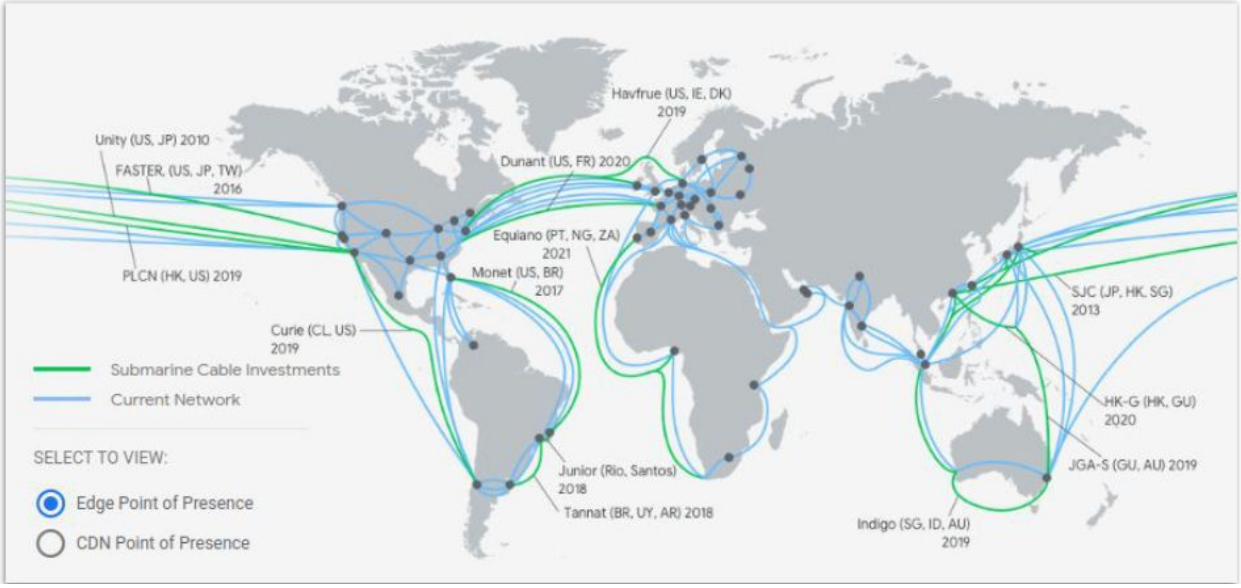


See the latest Google Cloud region and zone map.

- Each blue dot and white dot is a geographic region where Google Cloud is represented.
- Each region has at least three zones (data centers).
- Each data center has computing and networking infrastructure that allows Google Cloud services to run and execute.

The viewpoint you can take with Google Cloud is global. All the regions are connected by Google's private, software-defined network, which provides fast and reliable connections to users around the world.

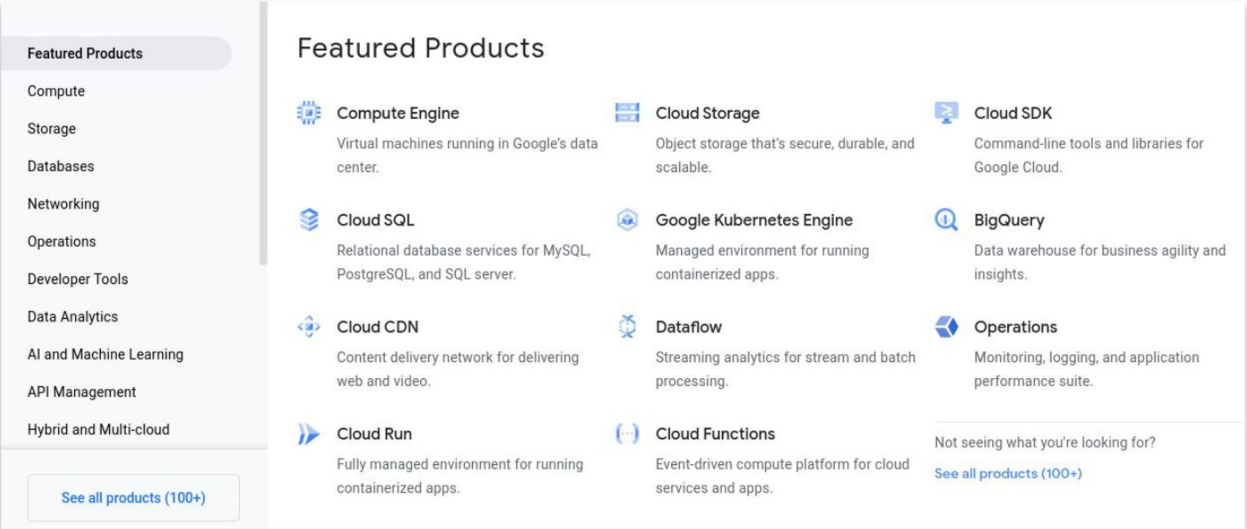
Google Cloud: Global Network



See the latest Google Cloud global network map.

If all your services are within the Google network, they remain on its private network. Because Google owns the network, it can control admission to it, control throughput latency, and put out certain optimizations that would not be possible on the public internet. That's important for the scale and the low latency needed for IoT event processing on a global scale. This network is constantly being expanded since the number of services, as well as data volumes, are increasing. The network powers the services available on the Google Cloud environment. The list of available Google Cloud services is long and keeps growing. A small selection of these services is shown in the screenshot below,

Google Cloud: Cloud Services



Featured Products
Compute
Storage
Databases
Networking
Operations
Developer Tools
Data Analytics
AI and Machine Learning
API Management
Hybrid and Multi-cloud

See all products (100+)

Featured Products

Compute Engine
Virtual machines running in Google's data center.

Cloud Storage
Object storage that's secure, durable, and scalable.

Cloud SDK
Command-line tools and libraries for Google Cloud.

Cloud SQL
Relational database services for MySQL, PostgreSQL, and SQL server.

Google Kubernetes Engine
Managed environment for running containerized apps.

BigQuery
Data warehouse for business agility and insights.

Cloud CDN
Content delivery network for delivering web and video.

Dataflow
Streaming analytics for stream and batch processing.

Operations
Monitoring, logging, and application performance suite.

Cloud Run
Fully managed environment for running containerized apps.

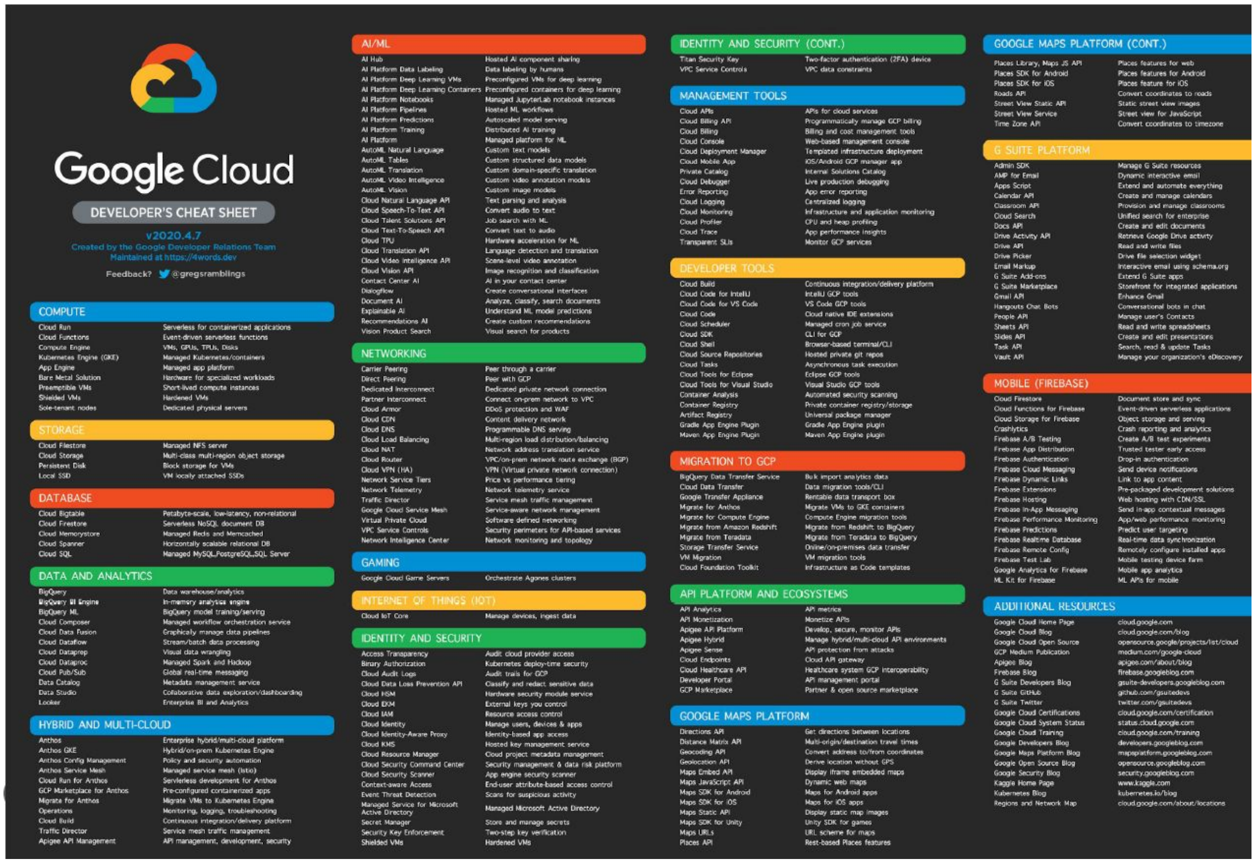
Cloud Functions
Event-driven compute platform for cloud services and apps.

Not seeing what you're looking for?
See all products (100+)

See the latest list of Google Cloud products.

A full list of Google Cloud services can be found in the below Google Cloud Developers cheat sheet, which describes every product in the Google Cloud family in 4 words or less.

Google Cloud: Cloud Services Cheat Sheet



Google Cloud

DEVELOPER'S CHEAT SHEET

v2020.4.7
Created by the Google Developer Relations Team
Maintained at https://4words.dev

Feedback? @gregsramblings

COMPUTE

Cloud Run	Serverless for containerized applications
Cloud Functions	Event-driven serverless functions
Compute Engine	VMs, GPUs, TPUs, Disks
Kubernetes Engine (GKE)	Managed Kubernetes/containers
App Engine	Managed app platform
Bare Metal Solution	Hardware for specialized workloads
Preemptible VMs	Short-lived compute instances
Shielded VMs	Hardened VMs
Sole-tenant nodes	Dedicated physical servers

STORAGE

Cloud Filestore	Managed NFS server
Cloud Storage	Multi-class multi-region object storage
Persistent Disk	Block storage for VMs
Local SSD	VM locally attached SSDs

DATABASE

Cloud Bigtable	Petabyte-scale, low-latency, non-relational
Cloud Firestore	Serverless NoSQL document DB
Cloud Memorystore	Managed Redis and Memcached
Cloud Spanner	Horizontally scalable relational DB
Cloud SQL	Managed MySQL,PostgreSQL,SQL Server

DATA AND ANALYTICS

BigQuery	Data warehouse/analytics
BigQuery BI Engine	In-memory analytics engine
BigQuery ML	BigQuery model training/serving
Cloud Composer	Managed workflow orchestration service
Cloud Data Fusion	Graphically manage data pipelines
Cloud Dataflow	Stream/batch data processing
Cloud Dataprep	Visual data wrangling
Cloud Dataproc	Managed Spark and Hadoop
Cloud Pub/Sub	Global real-time messaging
Data Catalog	Metadata management service
Data Studio	Collaborative data exploration/dashboarding
Looker	Enterprise BI and Analytics

HYBRID AND MULTI-CLOUD

Anthos	Enterprise hybrid/multi-cloud platform
Anthos GKE	Hybrid/on-prem Kubernetes Engine
Anthos Config Management	Policy and security automation
Anthos Service Mesh	Managed service mesh (Istio)
Cloud Run for Anthos	Serverless development for Anthos
GCP Marketplace for Anthos	Pre-configured containerized apps
Migrate for Anthos	Migrate VMs to Kubernetes Engine
Operations	Monitoring, logging, troubleshooting
Cloud Build	Continuous integration/delivery platform
Traffic Director	Service mesh traffic management
Apigee API Management	API management, development, security

AI/ML

AI Hub	Hosted AI component sharing
AI Platform Data Labeling	Data labeling by humans
AI Platform Deep Learning VMs	Preconfigured VMs for deep learning
AI Platform Deep Learning Containers	Preconfigured containers for deep learning
AI Platform Notebooks	Managed JupyterLab notebook instances
AI Platform Pipelines	Hosted ML workflows
AI Platform Predictions	Autoscaled model serving
AI Platform Training	Distributed AI training
AI Platform	Managed platform for ML
AutoML Natural Language	Custom text models
AutoML Tables	Custom structured data models
AutoML Translation	Custom domain-specific translation
AutoML Video Intelligence	Custom video annotation models
AutoML Vision	Custom image models
Cloud Natural Language API	Text parsing and analysis
Cloud Speech-To-Text API	Convert audio to text
Cloud Talent Solutions API	Job search with ML
Cloud Text-To-Speech API	Convert text to audio
Cloud TPU	Hardware acceleration for ML
Cloud Translation API	Language detection and translation
Cloud Video Intelligence API	Scene-level video annotation
Cloud Vision API	Image recognition and classification
Contact Center AI	AI in your contact center
Dialogflow	Create conversational interfaces
Document AI	Analyze, classify, search documents
Explainable AI	Understand ML model predictions
Recommendations AI	Create custom recommendations
Vision Product Search	Visual search for products

NETWORKING

Carrier Peering	Peer through a carrier
Direct Peering	Peer with GCP
Dedicated Interconnect	Dedicated private network connection
Partner Interconnect	Connect on-prem network to VPC
Cloud Armor	DDoS protection and WAF
Cloud CDN	Content delivery network
Cloud DNS	Programmable DNS serving
Cloud Load Balancing	Multi-region load distribution/balancing
Cloud NAT	Network address translation service
Cloud Router	VPC/on-prem network route exchange (BGP)
Cloud VPN (HA)	VPN (Virtual private network connection)
Network Service Tiers	Price vs performance tiering
Network Telemetry	Network telemetry service
Traffic Director	Service mesh traffic management
Google Cloud Service Mesh	Service-aware network management
Virtual Private Cloud	Software defined networking
VPC Service Controls	Security perimeters for API-based services
Network Intelligence Center	Network monitoring and topology

GAMING

Google Cloud Game Servers	Orchestrate Agones clusters

INTERNET OF THINGS (IOT)

Cloud IoT Core	Manage devices, ingest data

IDENTITY AND SECURITY

Access Transparency	Audit cloud provider access
Binary Authorization	Kubernetes deploy-time security
Cloud Audit Logs	Audit trails for GCP
Cloud Data Loss Prevention API	Classify and redact sensitive data
Cloud HSM	Hardware security module service
Cloud EKM	External keys you control
Cloud IAM	Resource access control
Cloud Identity	Manage users, devices & apps
Cloud Identity-Aware Proxy	Identity-based app access
Cloud KMS	Hosted key management service
Cloud Resource Manager	Cloud project metadata management
Cloud Security Command Center	Security management & data risk platform
Cloud Security Scanner	App engine security scanner
Context-aware Access	End-user attribute-based access control
Event Threat Detection	Scans for suspicious activity
Managed Service for Microsoft Active Directory	Managed Microsoft Active Directory
Secret Manager	Store and manage secrets
Security Key Enforcement	Two-step key verification
Shielded VMs	Hardened VMs

IDENTITY AND SECURITY (CONT.)

Titan Security Key	Two-factor authentication (2FA) device
VPC Service Controls	VPC data constraints

MANAGEMENT TOOLS

Cloud APIs	APIs for cloud services
Cloud Billing API	Programmatically manage GCP billing
Cloud Billing	Billing and cost management tools
Cloud Console	Web-based management console
Cloud Deployment Manager	Templated infrastructure deployment
Cloud Mobile App	iOS/Android GCP manager app
Private Catalog	Internal Solutions Catalog
Cloud Debugger	Live production debugging
Error Reporting	App error reporting
Cloud Logging	Centralized logging
Cloud Monitoring	Infrastructure and application monitoring
Cloud Profiler	CPU and heap profiling
Cloud Trace	App performance insights
Transparent SLIs	Monitor GCP services

DEVELOPER TOOLS

Cloud Build	Continuous integration/delivery platform
Cloud Code for IntelliJ	IntelliJ GCP tools
Cloud Code for VS Code	VS Code GCP tools
Cloud Code	Cloud native IDE extensions
Cloud Scheduler	Managed cron job service
Cloud SDK	CLI for GCP
Cloud Shell	Browser-based terminal/CLI
Cloud Source Repositories	Hosted private git repos
Cloud Tasks	Asynchronous task execution
Cloud Tools for Eclipse	Eclipse GCP tools
Cloud Tools for Visual Studio	Visual Studio GCP tools
Container Analysis	Automated security scanning
Container Registry	Private container registry/storage
Artifact Registry	Universal package manager
Gradle App Engine Plugin	Gradle App Engine plugin
Maven App Engine Plugin	Maven App Engine plugin

MIGRATION TO GCP

BigQuery Data Transfer Service	Bulk import analytics data
Cloud Data Transfer	Data migration tools/CLI
Google Transfer Appliance	Rentable data transport box
Migrate for Anthos	Migrate VMs to GKE containers
Migrate for Compute Engine	Compute Engine migration tools
Migrate from Amazon Redshift	Migrate from Redshift to BigQuery
Migrate from Teradata	Migrate from Teradata to BigQuery
Storage Transfer Service	Online/on-premises data transfer
VM Migration	VM migration tools
Cloud Foundation Toolkit	Infrastructure as Code templates

API PLATFORM AND ECOSYSTEMS

API Analytics	API metrics
API Monetization	Monetize APIs
Apigee API Platform	Develop, secure, monitor APIs
Apigee Hybrid	Manage hybrid/multi-cloud API environments
Apigee Sense	API protection from attacks
Cloud Endpoints	Cloud API gateway
Cloud Healthcare API	Healthcare system GCP interoperability
Developer Portal	API management portal
GCP Marketplace	Partner & open source marketplace

GOOGLE MAPS PLATFORM

Directions API	Get directions between locations
Distance Matrix API	Multi-origin/destination travel times
Geocoding API	Convert address to/from coordinates
Geolocation API	Derive location without GPS
Maps Embed API	Display iframe embedded maps
Maps JavaScript API	Dynamic web maps
Maps SDK for Android	Maps for Android apps
Maps SDK for iOS	Maps for iOS apps
Maps Static API	Display static map images
Maps SDK for Unity	Unity SDK for games
Maps URLs	URL scheme for maps
Places API	Rest-based Places features

GOOGLE MAPS PLATFORM (CONT.)

Places Library, Maps JS API	Places features for web
Places SDK for Android	Places features for Android
Places SDK for iOS	Places feature for iOS
Roads API	Convert coordinates to roads
Street View Static API	Static street view images
Street View Service	Street view for JavaScript
Time Zone API	Convert coordinates to timezone

G SUITE PLATFORM

Admin SDK	Manage G Suite resources
AMP for Email	Dynamic interactive email
Apps Script	Extend and automate everything
Calendar API	Create and manage calendars
Classroom API	Provision and manage classrooms
Cloud Search	Unified search for enterprise
Docs API	Create and edit documents
Drive Activity API	Retrieve Google Drive activity
Drive API	Read and write files
Drive Picker	Drive file selection widget
Email Markup	Interactive email using schema.org
G Suite Add-ons	Extend G Suite apps
G Suite Marketplace	Storefront for integrated applications
Gmail API	Enhance Gmail
Hangouts Chat Bots	Conversational bots in chat
People API	Manage user's Contacts
Sheets API	Read and write spreadsheets
Slides API	Create and edit presentations
Task API	Search, read & update Tasks
Vault API	Manage your organization's eDiscovery

MOBILE (FIREBASE)

Cloud Firestore	Document store and sync
Cloud Functions for Firebase	Event-driven serverless applications
Cloud Storage for Firebase	Object storage and serving
Crashlytics	Crash reporting and analytics
Firebase A/B Testing	Create A/B test experiments
Firebase App Distribution	Trusted tester early access
Firebase Authentication	Drop-in authentication
Firebase Cloud Messaging	Send device notifications
Firebase Dynamic Links	Link to app content
Firebase Extensions	Pre-packaged development solutions
Firebase Hosting	Web hosting with CDN/SSL
Firebase In-App Messaging	Send in-app contextual messages
Firebase Performance Monitoring	App/web performance monitoring
Firebase Predictions	Predict user targeting
Firebase Realtime Database	Real-time data synchronization
Firebase Remote Config	Remotely configure installed apps
Firebase Test Lab	Mobile testing device farm
Google Analytics for Firebase	Mobile app analytics
ML Kit for Firebase	ML APIs for mobile

ADDITIONAL RESOURCES

Google Cloud Home Page	cloud.google.com
Google Cloud Blog	cloud.google.com/blog
Google Cloud Open Source	opensource.google/projects/list/cloud
GCP Medium Publication	medium.com/google-cloud
Apigee Blog	apigee.com/about/blog
Firebase Blog	firebase.googleblog.com
G Suite Developers Blog	gsuite-developers.googleblog.com
G Suite GitHub	github.com/gsuitedevs
G Suite Twitter	twitter.com/gsuitedevs
Google Cloud Certifications	cloud.google.com/certification
Google Cloud System Status	status.cloud.google.com
Google Cloud Training	cloud.google.com/training
Google Developers Blog	developers.googleblog.com
Google Maps Platform Blog	mapsplatform.googleblog.com
Google Open Source Blog	opensource.googleblog.com
Google Security Blog	security.googleblog.com
Kaggle Home Page	www.kaggle.com
Kubernetes Blog	kubernetes.io/blog
Regions and Network Map	cloud.google.com/about/locations

Given Google Cloud's global footprint and private network, it might make more sense to use existing technologies than to build services up from scratch. Before launching any development activity on your own, it's a good strategy to explore available services. When you develop your website or application on Google Cloud, you can mix and match services into combinations that provide the infrastructure you need, and then add your code to enable the scenarios you want to build.

Google Cloud: select services

Google Cloud services that are used in the IoT event processing solution discussed here include:

BigQuery caters to Google Cloud customers who need to store and analyze three-figure terabyte data volumes and perform a broad-range total data set analysis:

- Cloud native analytics database - managed service
- Large-scale columnar SQL database (PB) providing multi-regional service

Cloud Spanner makes any access you have anywhere on the planet always consistent on a global scale:

- Cloud native relational online transaction database providing linear scaling and multi-regional and inter-continental service

Coldline Storage allows you to store forever any amount of data (without the need for event compression or for taking aggregates for storage reasons) and inspect the raw data in the future for analysis:

- Long-term, cost-effective cloud native storage with high availability

Cloud Machine Learning Engine makes predictions possible to enable you to feed in data and then run the learning and analysis as needed.

- Cloud native ML engine providing online and batch predictions

Now let's examine how such services play into an IoT event processing and analytics architecture using InfluxDB in Google Cloud.

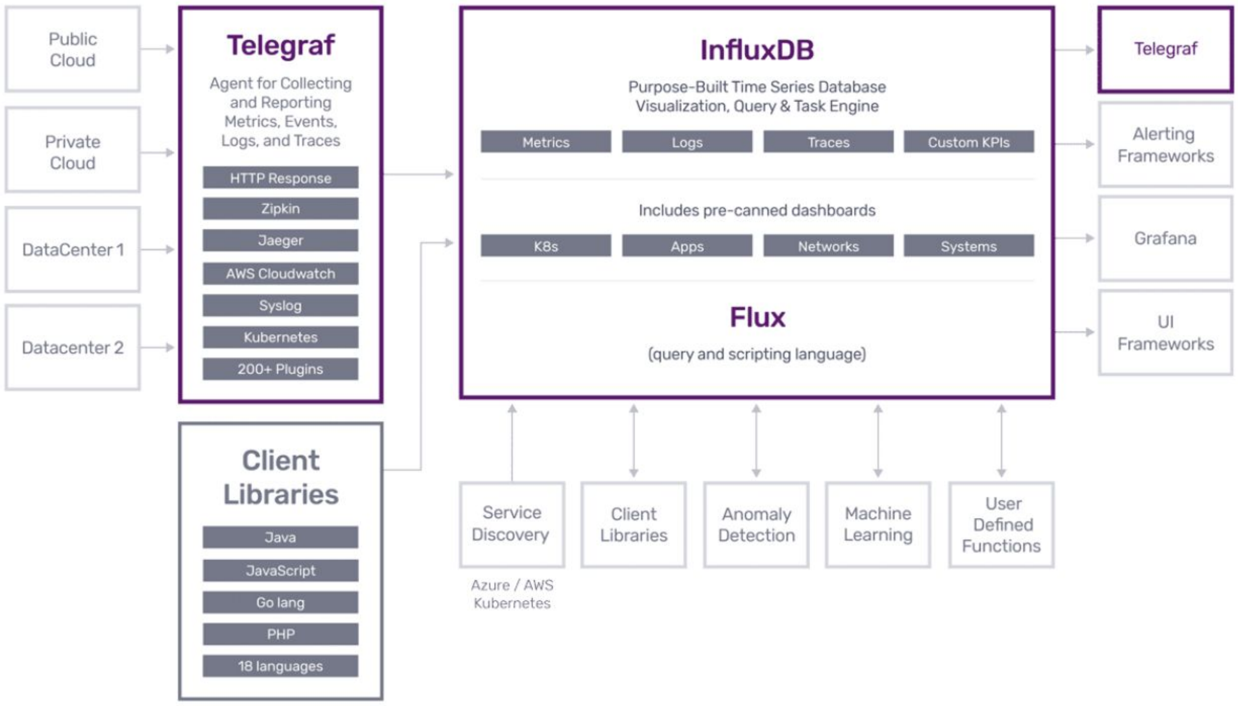
The recommended solution

> *"The IoT solution embeds the event life cycle into Google Cloud and InfluxDB platform residing into a larger architecture representation."*

Why InfluxDB?

IoT data is time series data, and the InfluxDB Platform works well with the other Google services in an IoT architecture because it fulfills the requirement for a separate yet complete time series platform providing ingestion, collection, storage and analysis.

InfluxDB Platform Architecture



The InfluxDB Platform is the essential time series toolkit — dashboards, queries, tasks and agents all in one place. A unified platform making it faster and easier than ever to develop and deploy modern time-based applications, it is available in three versions:

InfluxDB Open Source

InfluxDB Open Source incorporates everything you need in a time series platform into a single binary. Core capabilities include

- Time series data storage and querying
- Processing in the background
- Integrations with third-party services (including those from Google)
- Collection agent configuration
- Highly configurable dashboards and alert processing

InfluxDB Cloud

InfluxDB Cloud is a fast, elastic, serverless time series platform as a service — easy to use with usage-based pricing. Available on Google Cloud, it is a serverless platform that is purpose-built for time series data. This allows it to handle the relentless scale of time-stamped metrics and events generated by modern microservices, devices, and sensors — something that general-purpose databases can't do.

InfluxDB Enterprise

The InfluxDB Enterprise subscription turns any InfluxData instance into a production-ready cluster that can run anywhere. Available in the Google marketplace, this is ideal for the IoT implementation where an on-prem version may be important due to compliance requirements.

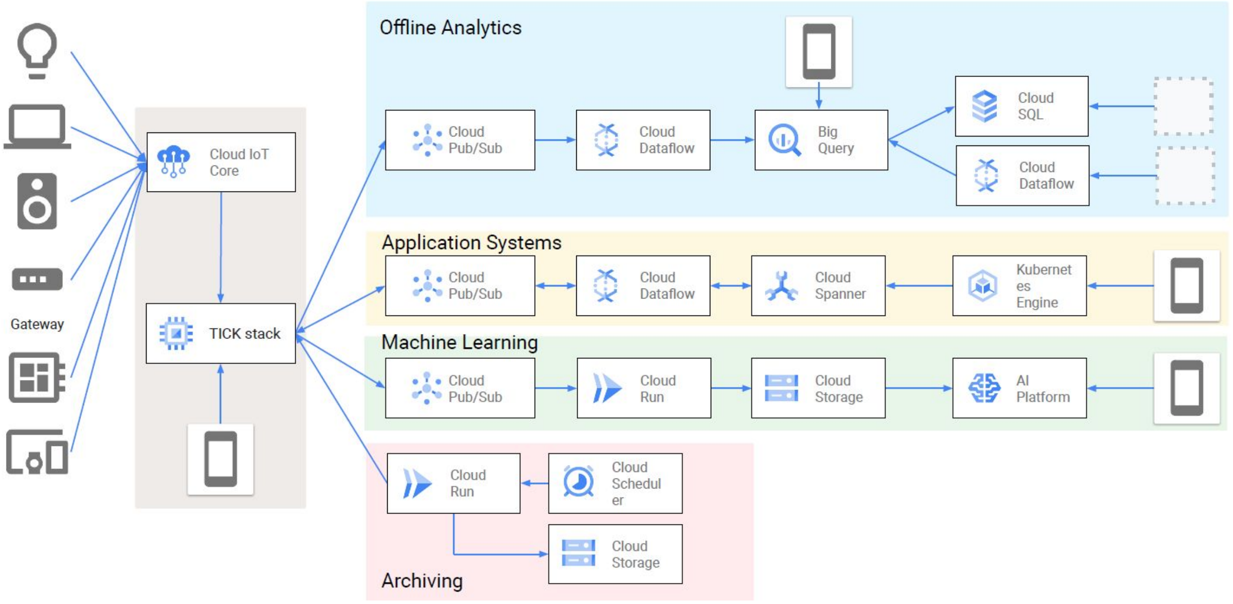
InfluxDB Enterprise and InfluxDB Cloud are readily available on the Google Cloud Marketplace, which offers ready-to-go development stacks, solutions and services to accelerate development. Users have one-click access to the industry's leading time series platform for the collection, storage, analysis and visualization of metrics and events for real-time decision making. Integrated billing makes it easy to use your Google Cloud Marketplace credits toward InfluxDB.

Google Cloud's IoT event processing and analysis solution integrates InfluxDB platform as a building block into a global cloud architecture that can process all the different event life cycles discussed above.

Technical architecture

> *"InfluxDB really implements the ability to store all your events, to query them, and also to interface with other systems such as Google Cloud services."*

Cloud Native IoT Event Processing Architecture



Every component in the above figure represents a cloud-native service, which eliminates the need to maintain VMs, monitor machines, or worry about the underlying infrastructure. All the solutions' components are available as a service from a service UI. All services are global, and nothing is restricted to a certain geography. An overview of each architecture layer is provided below.

Global IoT event collection

Google Cloud has a globally available service called Cloud IoT Core that allows you to register any type of IoT device as long as it's compliant with certain protocols. Cloud IoT Core has an integration with the InfluxDB platform. Any events coming in can be fed into InfluxDB, and the interface and toolset can be used to analyze event data. This provides a first set of entry points for IoT event collection processing on a global scale.

Offline analytics

A Pub/Sub connector enables offline analytics. Cloud Pub/Sub is a topic-based publish/subscribe queueing system whereby you create a topic, push it through, or subscribe to it. The system is global, which means that any created topic is available worldwide; that anywhere on the planet that you have services running, you can subscribe to this topic; and that any data fed into this topic is available everywhere.

The InfluxDB platform has a Cloud Pub/Sub integration, which makes the events ingested into InfluxDB available to the Pub/Sub system. You can write, for example, a data flow that feeds the events into BigQuery. Then, using the user interface, you can access and run the queries you write,

BigQuery, as shown in the diagram, allows a predicate push down into other databases, such as a Cloud SQL database. This database acts like a MySQL variant. You can write analytics queries that also access the relational system, MySQL or Postgres, in order to combine data into analytics results. Or you can have other systems feeding into BigQuery that come from weather reports. This would be a way to integrate the events that you collected from all these devices into an analytics system that might also be taking advantage of data from other sources.

Application systems

On the diagram's right side is an application running in Kubernetes. Kubernetes is available as a managed service, Google Kubernetes Engine (GKE) on Google Cloud, that can implement application logic, such as fleet management. GKE runs on Cloud Spanner. (The two respective boxes in the diagram together constitute the application.)

This architecture layer allows for bi-directional data transfer:

1. Events can stream from InfluxDB through Cloud Pub/Sub into Cloud Dataflow, which then feeds it *into* Cloud Spanner. These events are then available as transactional data to, for example, an inventory application.
2. Data can also be sent *from* Cloud Spanner to InfluxDB if needed for analysis at that point in time, to integrate data that may be streaming from devices managed elsewhere or from other datasets that you want to have available.

Bi-directional data transfer is important because it forms an ecosystem in which IoT events can feed other systems, and other systems can feed the events. Datasets or insights over time can be combined.

Machine learning

The next layer is machine learning. On the right side of the diagram is the AI platform — a Google Cloud service that provides machine learning capabilities, including online and batch predictions. Events are fed into the AI engine:

- Collected events are extracted and sent to Cloud Run, a Google Cloud service which allows implementing computing logic.
- Events are then stored in Cloud Storage for use as input for the AI Platform.
- This data flow is a mechanism to model analyses and predictions based on events imported to the AI platform.

Archiving

The final layer is a mechanism to archive events. Cloud Run can submit code implemented in a Docker container. Once you submit the docker image to Google Cloud, it provides an entry point, an HTTP REST endpoint which you can invoke and which tracks whatever your code is doing. (This is a means to implement cloud functions in any workable way in Docker, so you're not restricted to the languages supported by cloud functions, but can instead use any language.)

Scaling occurs on the spot as demand increases when invocations come in:

- Managing invocations can be done using Cloud Scheduler: a configuration interface that allows setting invocations on a regular basis, based on desired number, timing, and frequency. One entrypoint is a Cloud Run docker image. On this basis, you control what data to extract and delete from the time series database and then feed it into a long-term storage environment, resulting in a sliding window of remaining actual data.

- Since any logic can be implemented in the Cloud Run docker image, the same data management mechanism described above can also be used to feed data into another storage environment, such as an operation database, analytics database or queueing system.

Results

> *"All the components of this solution can function at the same time in various world regions."*

Google Cloud's native IoT event processing architecture, deploying InfluxDB, covers virtually all energy industry use cases as it addresses IoT events' complete life cycle and deploys native Google Cloud services which have unmatched global reach and scalability.

This architecture frees developers worldwide from managing underlying infrastructure. It empowers them to easily set up an IoT monitoring system and extract IoT-data-driven insights that can solve business problems and translate into time and cost efficiencies for energy industry enterprises.

This use case was presented at InfluxDays London 2020.

About InfluxData

InfluxData is the creator of InfluxDB, the open source time series database. Our technology is purpose-built to handle the massive volumes of time-stamped data produced by IoT devices, applications, networks, containers and computers. We are on a mission to help developers and organizations, such as Cisco, IBM, PayPal, and Tesla, store and analyze real-time data, empowering them to build transformative monitoring, analytics, and IoT applications quicker and to scale. InfluxData is headquartered in San Francisco with a workforce distributed throughout the U.S. and across Europe.

Learn more.

InfluxDB documentation, downloads & guides

Download InfluxDB
Get documentation

Additional case studies
Join the InfluxDB community

influxdata
799 Market Street
San Francisco, CA 94103
(415) 295-1901
www.InfluxData.com
Twitter: @InfluxDB
Facebook: @InfluxDB