AN OPEN SOURCE DATABASE MANAGEMENT SYSTEM FOR TIME SERIES DATA

INFLUXDB

PRESENTED BY

HANNES BÄHR

CONTENT

1. Time Series Data and Databases
2. InfluxData: Basic Facts
3. InfluxDB Products and Licenses
4. IT Architecture Overview
5. Data Models in InfluxDB
6. APIs and Client Libraries
7. Use Cases in Industry
8. Set Up and Application

1. TIME SERIES DATA AND DATABASES

Time Series Data

Time series data is a sequence of data points collected or recorded at time-ordered intervals.

Characteristics

- Each data point has a timestamp
- Often large volumes and high velocity
- Frequent updates or real-time ingestion
- Challenges: Managing large volumes efficiently, ensuring fast writing and querying

Importance in Big Data

- Enables predictive analytics, anomaly detection, forecasting
- Provides detailed insights into patterns and behaviors over time
- Applications/Use Cases: IoT (sensor readings of temperature and humidity), financial markets (stock price and exchange rate movements), health care (vital data)

1. TIME SERIES DATA AND DATABASES

Time Series Databases:

Specialized DBs for storing, retrieving, and managing time-stamped (temporal) data efficiently.

Key Features

- Optimized for time-based queries (e.g., ranges, aggregations)
- High ingestion rate for continuous data streams
- Support for data compression and retention policies (e.g. automatically delete old data)
- Examples: **InfluxDB**, TimescaleDB, Prometheus, OpenTSDB

Advantages over Traditional Databases

- Better performance for time-based data access
- Efficient handling of data lifecycle (downsampling, retention)
- Designed for high write loads and fast reads
- Use cases: IoT telemetry, financial tick data, health care monitoring (also see last slide)

2. INFLUXDATA: BASIC FACTS

Key Dates

- 2012: Paul Dix and Todd Persen founded company *Errplane* in New York City, because they needed a scalable database for handling large amounts of time series data
- 2013: The first open-source version of **InfluxDB** was released
- 2015: Rebranding from *Errplane to InfluxData* and relocation to San Francisco

Mission and Focus

- Help developers to build intelligent, real-time systems with their time series data
- Being the leading platform for time series data management

Growth, Main Products and Community

- Major investor funding: $100M+ by Battery Ventures, Citi Ventures, Bloomberg Beta, and more
- Products: InfluxDB, Telegraf (agent), Chronograf (visualization), Kapacitor (processing)
- 2600+ enterprise customers, 1M+ active instances running on InfluxDB, and large open-source community events like "InfluxDays"

3. INFLUXDB PRODUCTS AND LICENSES

Product	Description	Key Features	License
InfluxDB OSS (Open Source)	Free, self-hosted version for local use	• Time series data storage • InfluxQL/Flux support • Single-node operation	MIT License
InfluxDB Cloud	Fully managed SaaS offering by InfluxData	• Serverless • Auto-scaling (pay-as-you-go) • REST API access • Dashboards • Cloud-native integrations	Proprietary SaaS License (Commercial License)
InfluxDB Enterprise	Self-hosted, on-premise solution for businesses	• High availability • Clustering • Role-Based Access Control • Backup and restore	Commercial License

4. IT ARCHITECTURE OVERVIEW

Design Principles

- Modular services allow for scalability, high availability, and fault isolation
- Suited for cloud-native, on-premise, and hybrid deployments

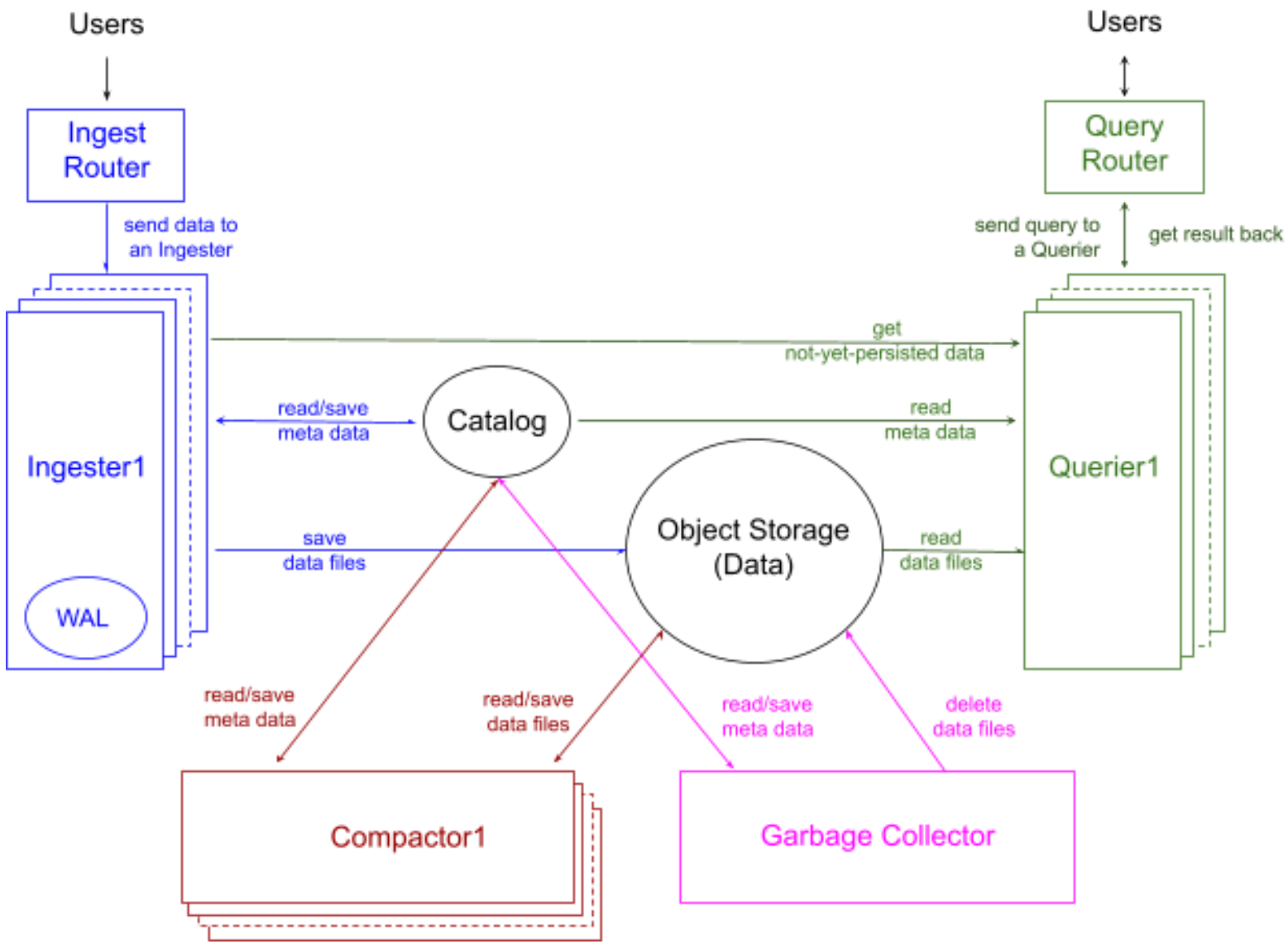
Core Architectural Layers

- Data Ingestion: Handles high-throughput writes from APIs, client libraries, and tools like Telegraf
- Query Engine: Executes InfluxQL and Flux queries, Optimized for time-based operations
- Data Compaction: Merges, compresses, and organizes time series data to optimize storage and query performance
- Garbage Collection: Automatically removes expired or deleted data (retention policies)

Storage System

- Catalog: Small but fast metadata store for measurements, tags, and schema
- Object Storage: Large, distributed system where actual time series data is stored; supports horizontal scaling and high durability

4. IT ARCHITECTURE OVERVIEW



InfluxDB 3.0 Architecture

5. DATA MODELS IN INFLUXDB

The following table with sample data represents time series records stored in InfluxDB and is used for the following slides to explain data elements concepts.

_time	**_measurement**	location	scientist	**_field**	**_value**
2019-08-18T00:00:00Z	census	klamath	anderson	bees	23
2019-08-18T00:00:00Z	census	portland	mullen	ants	30
2019-08-18T00:06:00Z	census	klamath	anderson	bees	28
2019-08-18T00:06:00Z	census	portland	mullen	ants	32

5. DATA MODELS IN INFLUXDB

Timestamps
- All data stored in InfluxDB has a '_time' column that stores timestamps
- InfluxDB formats timestamps: date and time in RFC3339 UTC associated with data

Fields
- Field key: Name of field in '_field' column: String
- Field value: Value of field in '_value' column : String, float, integer, Boolean
- Field set: Collection of field key-value pairs **associated with a specific timestamp**

Tags
- Tag key: Columns of data table, that represent tags: "location" and "scientist" in the table
- Tag value: Values in column of respective tag key
- Tag set: Collection of tag key-value pairs

Measurements
- Strings, that describe the data with a measurement name: "census" in the table
- Act as a container for tags, fields, and timestamps

5. DATA MODELS IN INFLUXDB

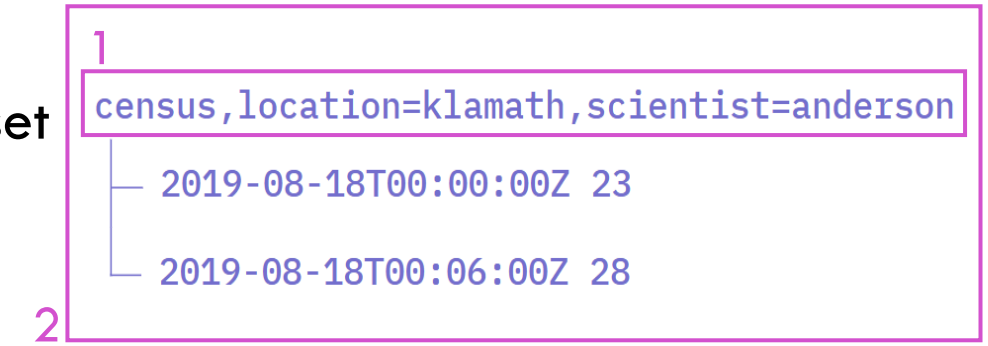
Points

- Includes the series key, a field value, and a timestamp:

```
2019-08-18T00:00:00Z census ants 30 portland mullen
```

Series

- Series key[1]: Unique combination of measurement and tag set
- Series[2]: Timestamps and field values for a given series key



Buckets

- All InfluxDB data is stored in a bucket
- Combines the concept of a database and a retention period
- Belongs to an organization

Organizations

- Workspace for a group of users
- All dashboards, tasks, buckets, and users belong to an organization

6. APIS AND CLIENT LIBRARIES

HTTP API

- Core interface to InfluxDB for reading, writing, querying, and managing data
- RESTful, accessible via */api/v2/* in InfluxDB 2.x+
- Full HTTP API for OSS and additional specific endpoints for Cloud and Enterprise products

Authentication and Access

- Token-based authentication in InfluxDB 2.x and 3.x
- OAuth2 support in InfluxDB Cloud

Query Languages

- InfluxQL (SQL-like): Supported in OSS and partially in Enterprise
- Flux: Main query language in 2.x, functional, supports joins and transformations
- InfluxDB 3.x uses SQL as the default query language (based on Apache DataFusion)

6. APIS AND CLIENT LIBRARIES

Client Libraries

- Libraries simplify writing points, executing queries, managing buckets/auth, etc.
- Official SDKs for Python, Go, Java, JavaScript, C#, and Rust

Visualization and Integrations

- Native dashboards in InfluxDB Cloud and OSS UI
- Integrates with Grafana, Telegraf, Prometheus, Home Assistant, etc.

Deployment Flexibility

- APIs and libraries work across local (OSS), cloud, and hybrid environments

7. USE CASES IN INDUSTRY

IoT and Sensor Data Monitoring

- Collects high-frequency data from sensors (temperature, humidity, GPS, energy meters)
- Used in smart homes, smart cities, and industrial automation (Industry 4.0)

Financial and Business Analytics

- Tick data, foreign exchange rates, and transaction logs over time
- Time-based analysis and anomaly detection in financial systems

DevOps and Infrastructure Monitoring

- Stores CPU, memory, disk, and network metrics
- Used with Telegraf and Grafana for system performance dashboards
- Real-time alerting with low-latency querying

Industry Customers

- Siemens, Tesla, Adidas, Cisco, and Comcast are public users of InfluxDB

8. SET UP AND APPLICATION

Installation Options

- InfluxDB OSS: Download for Linux, macOS, Windows or use via Docker, Homebrew
- InfluxDB Cloud: No install required, create account and use via web UI and API
- Enterprise: Installed on-premise, supports clustering setup

Configuration

- Done via 'influxdb.conf' file in OSS and Enterprise
- Cloud configuration is managed via UI/API settings
- Settings include retention policies, authentication, logging, storage paths

Data Ingestion/Writing

- HTTP API: direct write endpoint
- Telegraf agent: metrics collector with plugins
- Client libraries (Python, Go, etc.)
- Line protocol, JSON, CSV formats

8. SET UP AND APPLICATION

Data Querying

- OSS/Enterprise: InfluxQL or Flux depending on version
- Cloud 3.x: Uses SQL (DataFusion-based engine)
- Queries can filter, aggregate, downsample, and join across series

Dashboarding and Visualization

- Native InfluxDB UI for 2.x and 3.x (Cloud)
- Common integration: Grafana (via API and plugin support)

Automation and Scripting

- Use Flux tasks (in 2.x) to automate data processing
- Use client SDKs to script ingestion, alerts, and analysis

THANKS FOR YOUR ATTENTION!