June 2024

Building an app with geospatial real-time data in Tinybird

Raquel Barbadillo | rbarbadillo@tinybird.co

Tinybird

- Backend Engineer at Tinybird
- Used to be a Chemical Engineer, loved computers so here I am a couple years later.
- Not a geospatial expert like Ramiro, but I can hold my own building data-intensive apps.

Me grinning like a fool the last time I was in this beautiful region

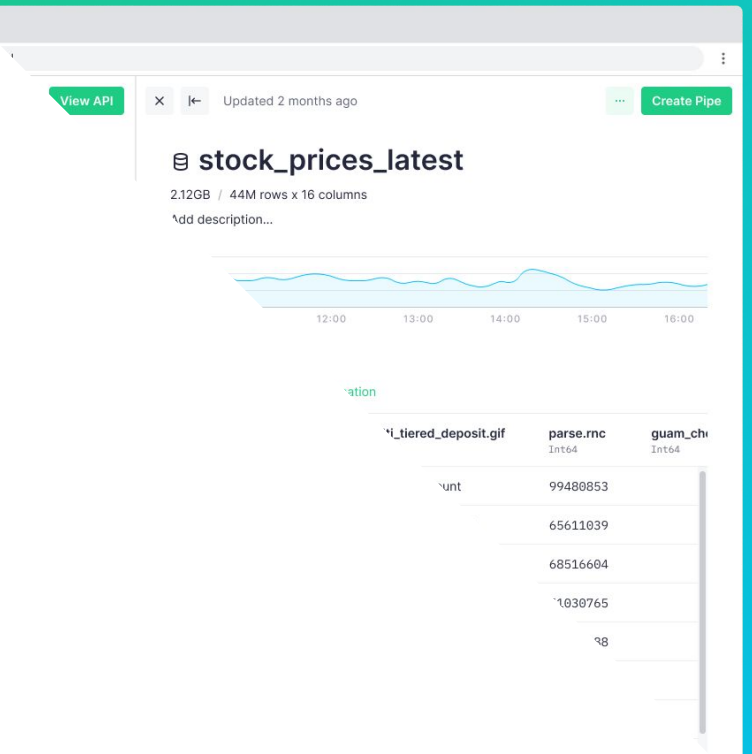
Agenda

- Intro to Tinybird
- What are we going to be building?
- Actually building the thing

Slides at → https://tinyurl.com/tinybird-workshop

Intro to Tinybird

Build fast APIs, faster.



tinybird

Intro to Tinybird

Develop...

Without infrastructure

Only using SQL

Low-latency, high-concurrent APIs over real time data

At any scale

Operational Intelligence

Big Retailers use Tinybird to deliver operational intelligence at scale

With sales in 215 markets, and more than 11 million unique visits to their website each day, one of our clients generates more than 1TB of transactional data every day.

The business team wanted to use data to make real-time decisions around sales, budgets and more. Data complexity was high, with information originating from many different systems and databases. Multiple queries was making it hard to support the high-level of concurrency for this ever-growing volume of data. Difficult to bring data together and use it as a single source of truth for decision-making.

How Tinybird helped

11M
Database queries

211b
Rows written

11.9T
Rows read

+333TB
Processed

situm
INDOOR POSITIONING

In-Product Analytics

Situm relies on Tinybird to accelerate the speed of data processing and querying at scale

Situm turns any mobile phone or BlueTooth low-energy transmitter into a precision tracking or navigation system for more than 6,000 buildings around the world.

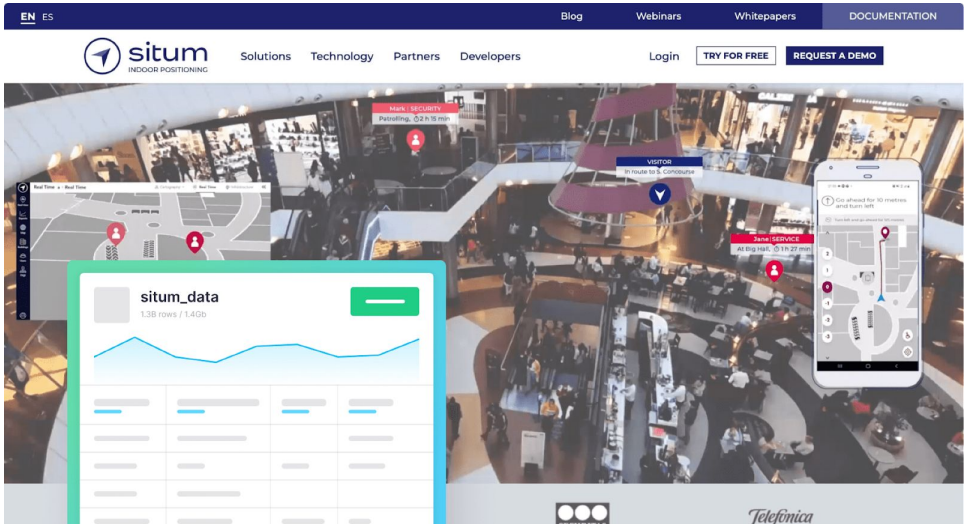
Originally built on PostgreSQL, and later on MongoDB, the engineering team was spending a lot of time fine-tuning and maintaining their database. Taking resources away from product development and in turn slowing down the release of new features.

How Tinybird helped

7TB
Processed data/day

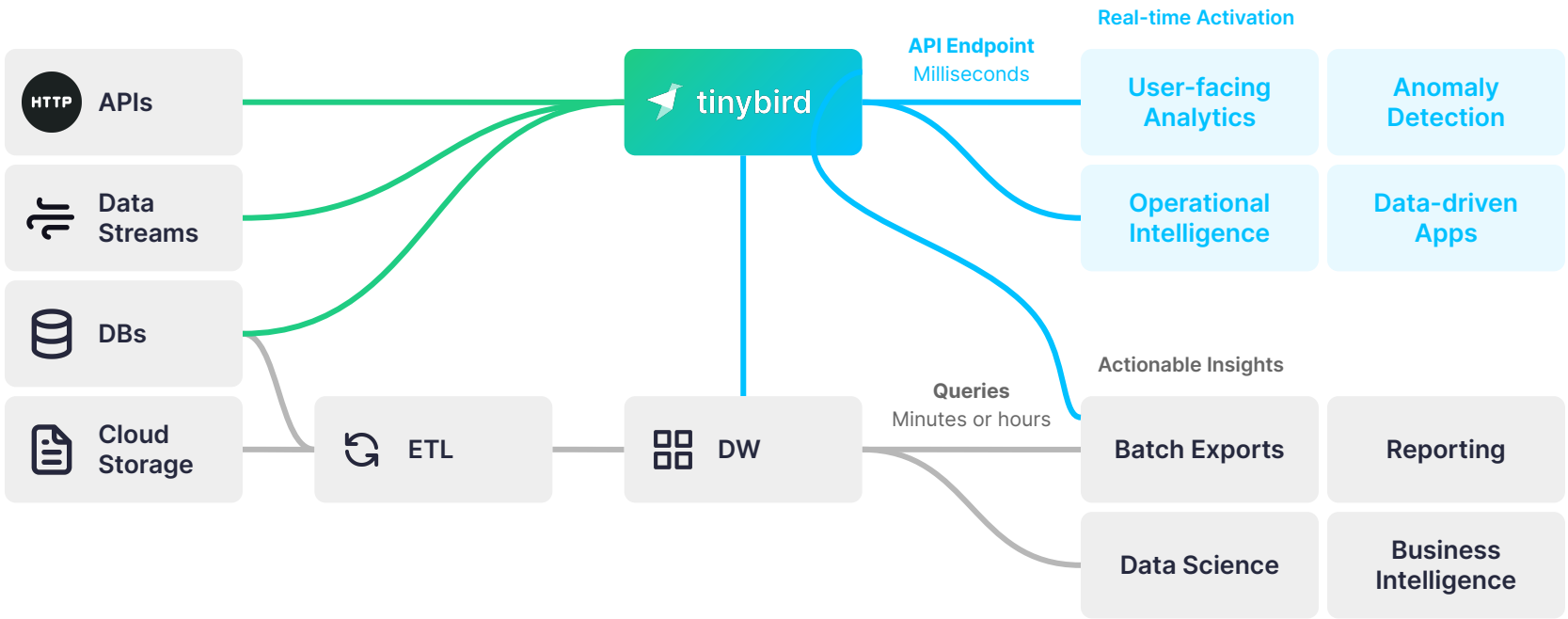
10x
Reduction in engineering hours

<1sec
Data is ingested and transformed



Fast APIs, Faster

The real-time complement to your DW



Ingest

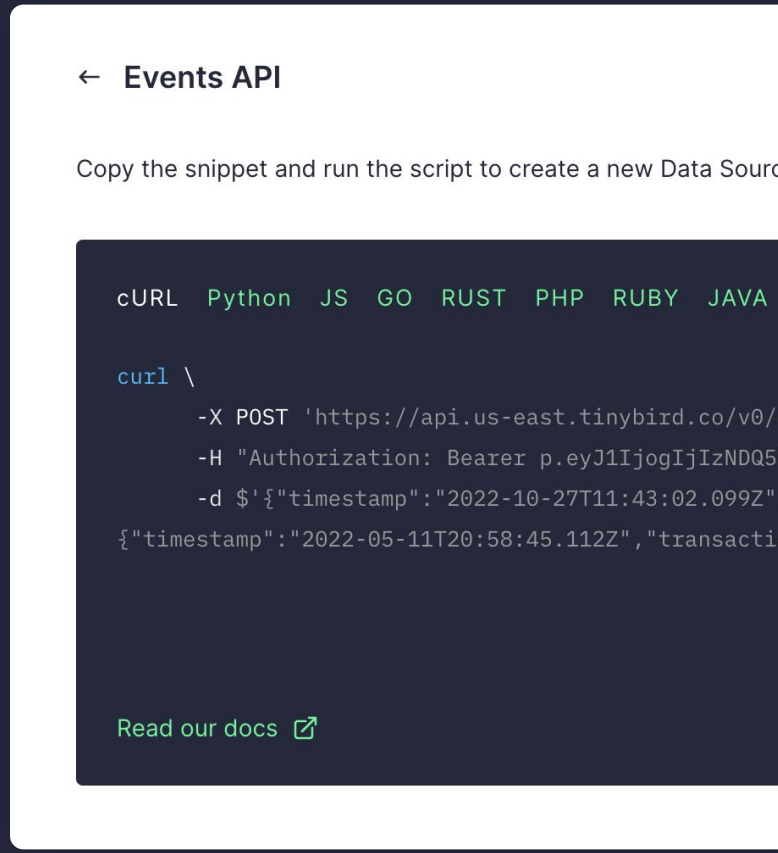
Ingest

Events API

The Events API is designed to enable **high-throughput streaming ingestion** into Tinybird from an easy to use HTTP API.

- ➔ Ingest JSON events with a simple POST request at 1,000+ requests per second.
- ➔ Stream events to Tinybird without the overhead and complexity of externals tools like Kafka.

Read more here.



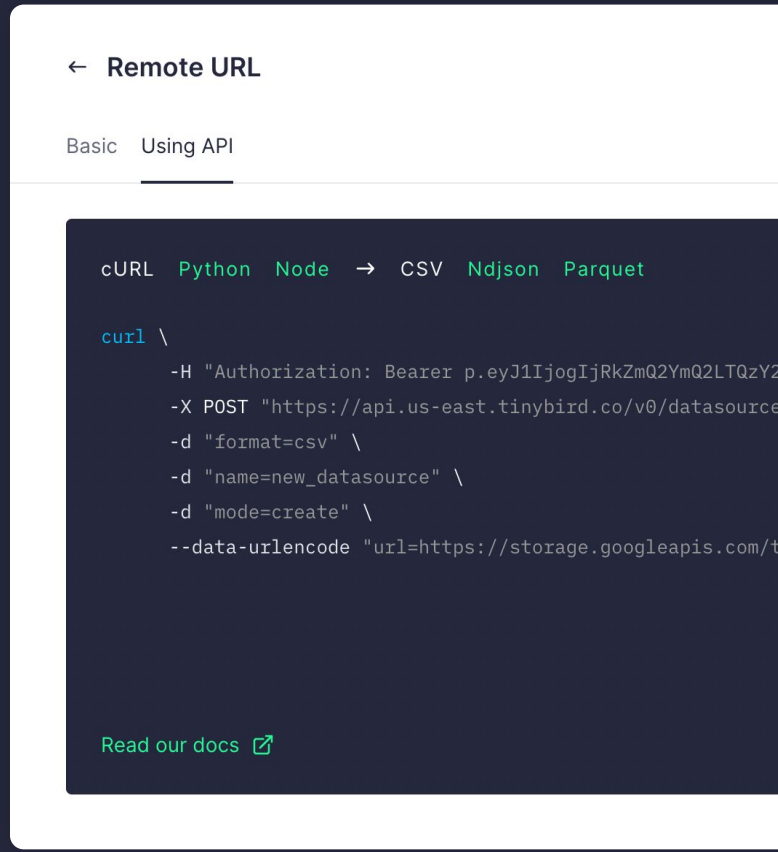
Ingest

Data Sources API

The Data Sources API allows you to **list, create, update, or delete** your Tinybird Data Sources.

➔ Tinybird provides a way to programmatically manage your Data Sources and extends ClickHouse's functionality to enable updates and deletes.

Read more here.



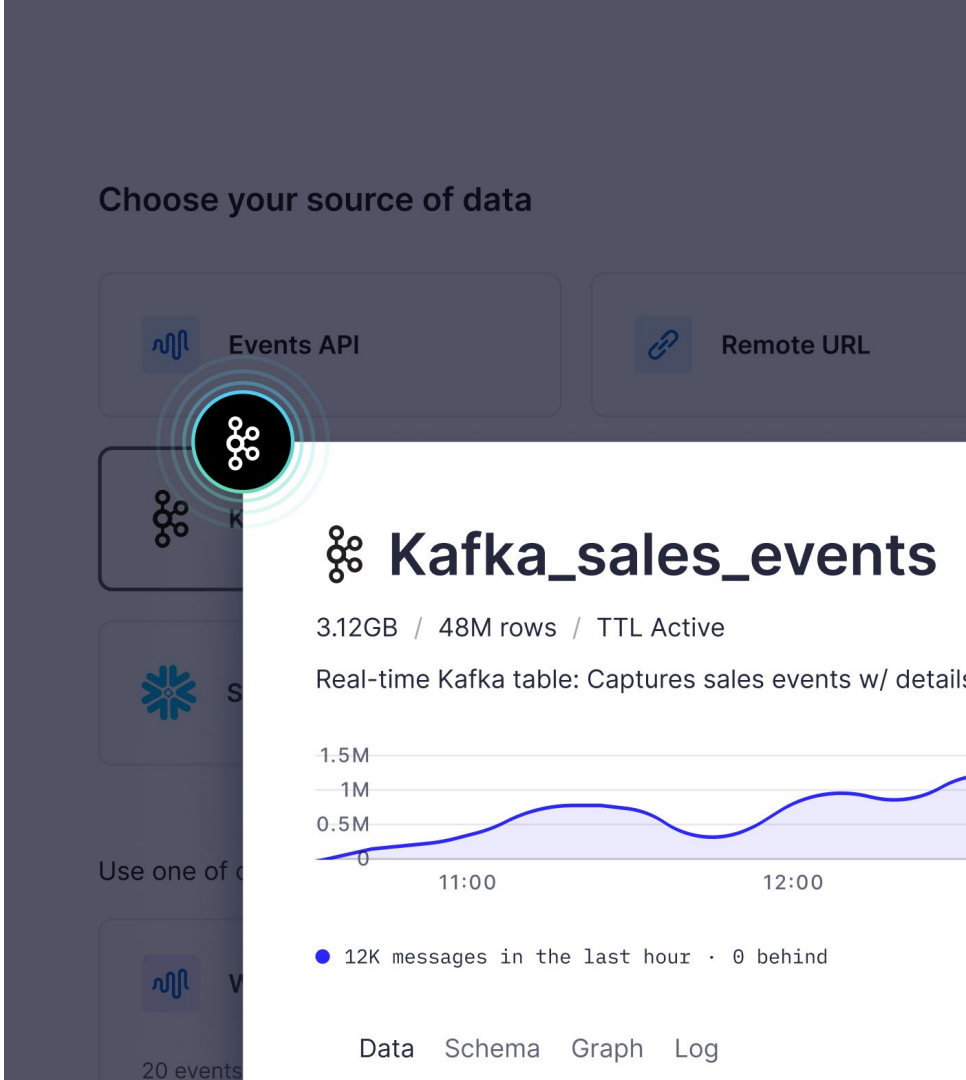
Ingest

Kafka Connector

Native, managed connector that allows you to i**ngest data from your existing Kafka cluster** and load it into Tinybird.

- ➔ Build powerful real-time analytics over streaming data.
- ➔ Gives users a simple means of transforming and enriching data streams and publishing them as high-concurrency, low-latency APIs.
- ➔ Fully managed and requires no additional tooling.
- ➔ Compatible with Confluent, RedPanda, Amazon MSK and open source Kafka.

Read more here.



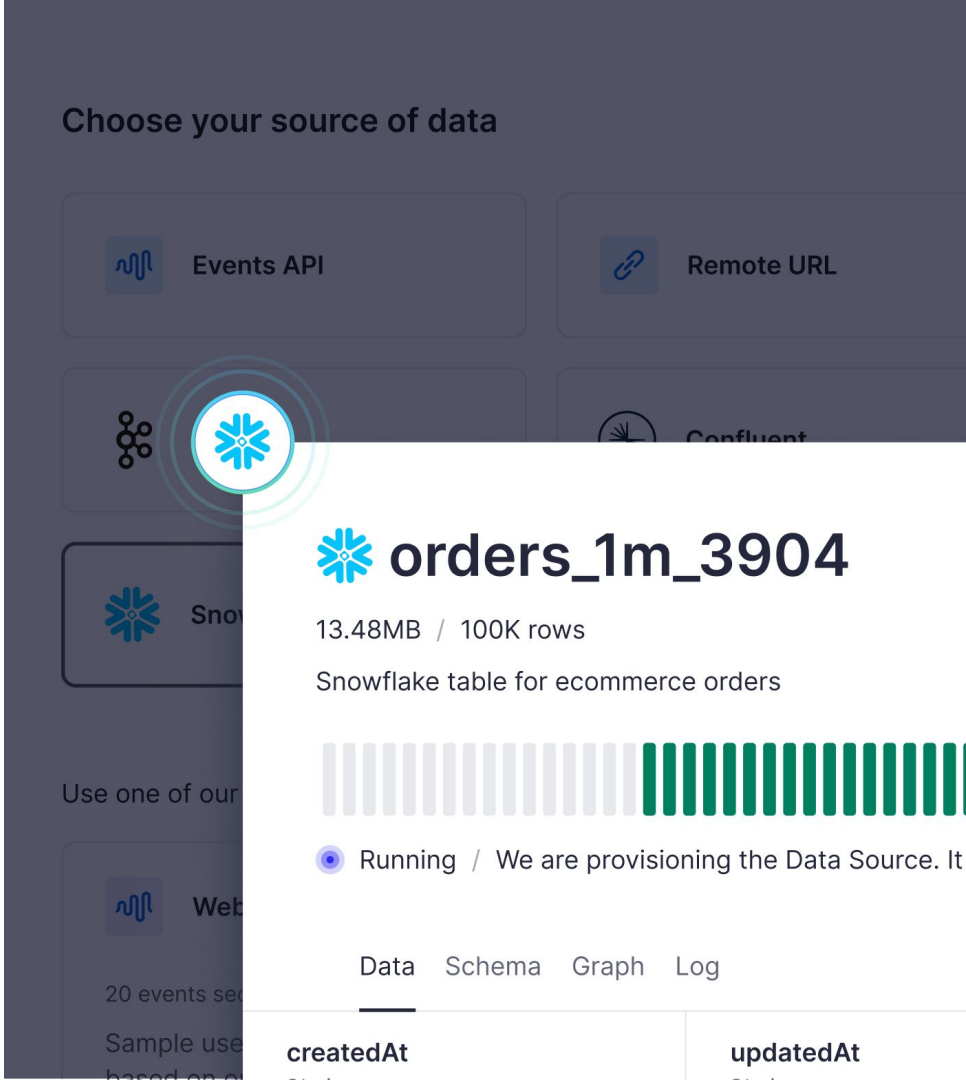
Ingest

Snowflake Connector

Native, managed connector that allows you to **load data from your existing Snowflake account** into Tinybird.

- ➔ Build a low-latency publication layer on top of Snowflake at a fraction of the cost of alternative techniques.
- ➔ Enrich real-time streams with static or batch data from your data warehouse.
- ➔ Fully managed and requires no additional tooling.

Read more here.



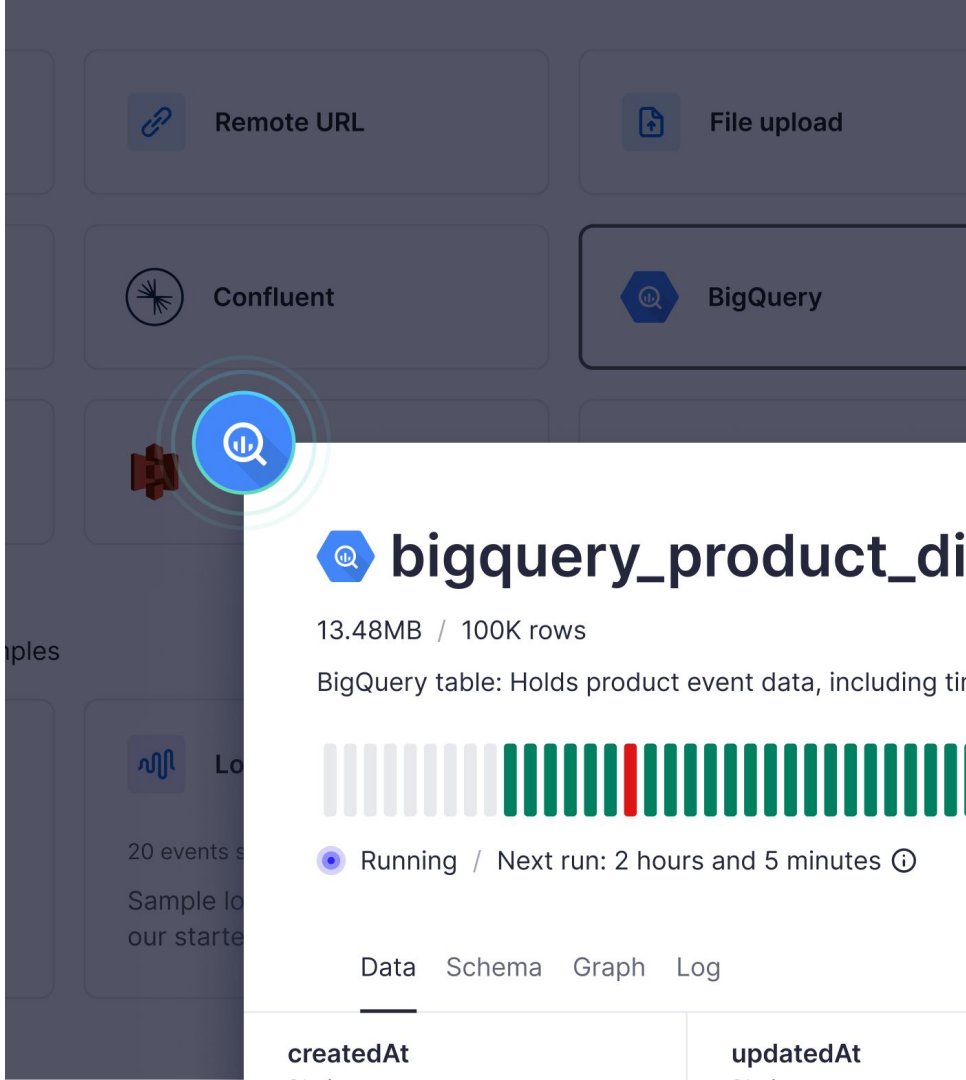
Ingest

BigQuery Connector

Native, managed connector that allows you to **load data from your existing BigQuery account** into Tinybird.

- ➔ Build a low-latency publication layer on top of BigQuery at a fraction of the cost of alternative techniques.
- ➔ Enrich real-time streams with static or batch data from your data warehouse.
- ➔ Fully managed and requires no additional tooling.

Read more here.



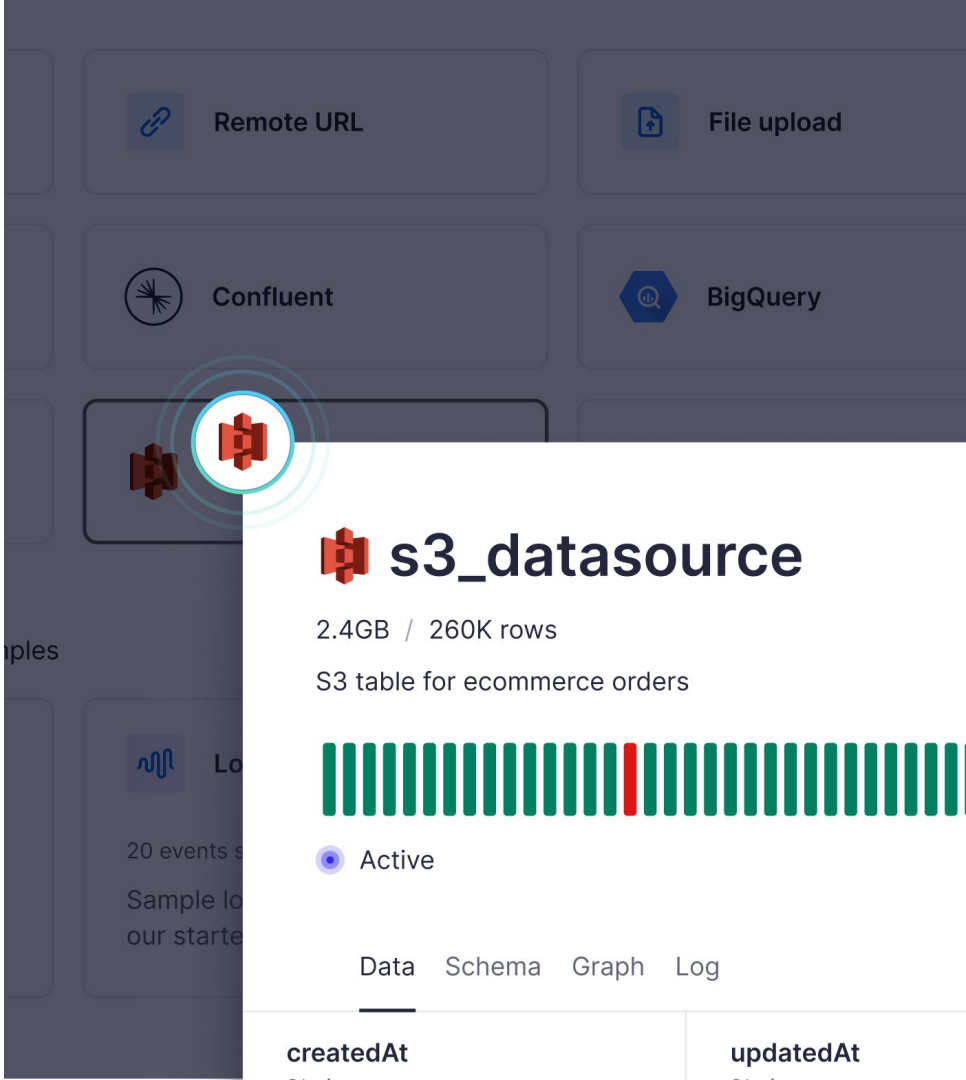
Ingest

S3 Connector

Native, managed connector that allows you to **ingest files from your Amazon S3 bucket** into Tinybird.

- ➔ Easily sync files in S3 to Tinybird, query the data with SQL, and publish low-latency analytics APIs.
- ➔ Provides a better storage layer for user-facing analytics, where data freshness and low latency matter.
- ➔ Fully managed and requires no additional tooling.

Read more here.



Develop

Develop

ClickHouse

The world's fastest open source, columnar, OLAP database **purpose-built for real-time analytics** over fresh data.

- ➔ ClickHouse is the bedrock of Tinybird; we leverage all of the features, speed, and scalability, especially in relation to real-time use cases.
- ➔ The most active open source, real-time analytics database in the world: >500 active contributors, compared to <40 for the leading competitor.
- ➔ Contributors include Intel, Cloudflare, and Yandex.
- ➔ Tinybird employs dedicated ClickHouse contributors, and remains true to the open source project, eliminating the risk of vendor lock-in.
- ➔ Tinybird builds on top of ClickHouse, providing the easiest and most robust platform to take ClickHouse into production.

Develop

SQL & Templating

Develop with the same **SQL you know and love**.

- ➔ SQL is a ubiquitous skill that most engineers in data have experience with, while also being accessible to wider teams in the business.
- ➔ Engineers develop in Tinybird using SQL.
- ➔ Tinybird extends SQL with a templating language that allows for dynamic query parameters to make API responses more flexible at query time.
- ➔ The templating language also enables more advanced Python-style conditional logic.

```
1 %
2 SELECT
3     toStartOfDay(timestamp) AS day
4     sum(sales) AS revenue
4 FROM  kafka_sales_events
5 WHERE
6   country == {{ String(country,
    description="ISO-Alpha2 Country Code",
    required=True) }}
7   AND day >= {{ Date(start_day,
    '2023-07-20', description="YYYY-MM-DD
    Start Date", required=True) }}
8   AND day <= {{ Date(end_day, '2023-07-27',
    description="YYYY-MM-DD End Date",
    required=True) }}
9 GROUP BY day
```

Develop

Pipes

Pipes are **SQL-based notebooks** you use to build features over your data.

- ➔ Pipes contain one or more chained SQL queries that result in either an API Endpoint, Materialized View, or COPY job.
- ➔ Pipes allow you to break your query logic down into multiple, smaller queries, which simplifies and accelerates development over your data.
- ➔ With just SQL, engineers can build real-time features easier and faster. No new skills required.

Read more here.

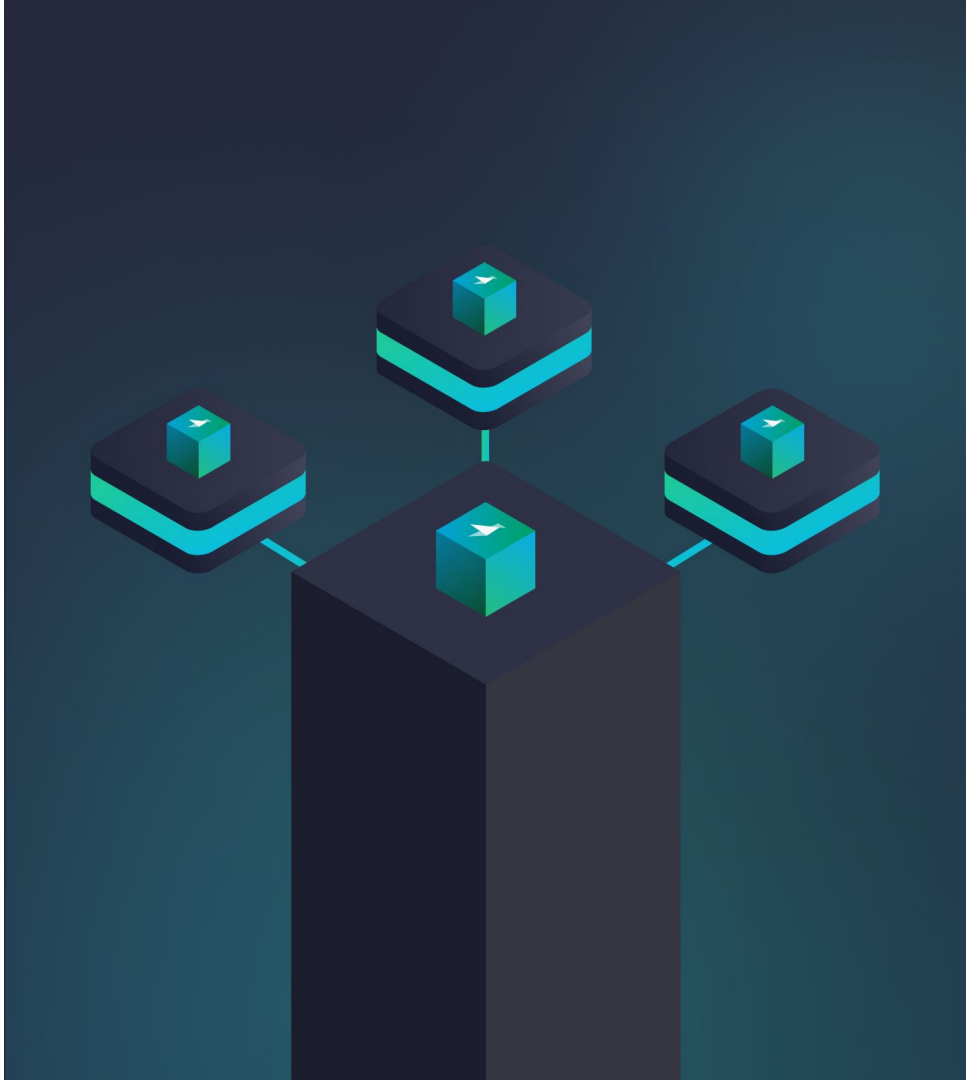
Develop

Workspaces

A Workspace is a **dedicated space** that contains your Tinybird resources.

- ➔ A Workspace is a container for your data, queries, API Endpoints, and tokens.
- ➔ You can have as many Workspaces as needed, organizing them however suits you best: per team, per use case, per environment, and so on.
- ➔ Multiple users can collaborate within a single Workspace.
- ➔ Data can be shared between Workspaces as needed.

Read more here.



Develop

Materialized Views

Pre-aggregate and pre-filter large Data Sources **incrementally upon ingestion**.

- ➔ Complex queries run over large datasets are invariably slow, so you must find ways to recover speed you lose as data size increases.
- ➔ Materialized Views give you a way to pre-aggregate and pre-filter large Data Sources incrementally, adding simple logic using SQL to produce a more relevant Data Source with significantly fewer rows.
- ➔ Shift computational load from query to ingestion time so Endpoints stay blazing fast.

Read more here.

Develop

Extensive Functions

Tinybird supports the **standard SQL functions** every engineer expects, and more.

- ➔ Expanded date and window functions.
- ➔ Extensive functions for URLs.
- ➔ Comprehensive geospatial functions.
- ➔ Many additional functions for IP address, encodings, JSON, arrays and maps.

```
toYear(), toQuarter(), toMonth(),
toStartOfYear(),
toStartOfMonth(), toStartOfDay(),
toStartOfHour(),
toStartOfMinute(), date_add(),
addMinutes(), protocol(),
domain(), domainWithoutWWW(),
pathFull(), geoDistance(),
greatCircleAngle(),
pointInPolygon(), JSONExtract(),
JSONExtractInt(),
JSONExtractKeys(),
JSONExtractString()
...
```

Develop

Geospatial Functions

Perform geospatial operations thanks to:

- ➔ ClickHouse geometric data types to represent geographical objects — locations, lands, etc.
- ➔ Supported data types: **Point**, **Ring**, **Polygon**, and **Multipolygon**.
- ➔ It contains the basic set of geospatial operations such as **greatCircleDistance** or **PointInPolygon**.
- ➔ Create faster spatial indexes using geohash, h3 or S2.

```
greatCircleDistance(),
geoDistance(),
greatCircleAngle(),
pointInEllipses(),
pointInPolygon(),

geohashEncode(),
geohashDecode(),
geohashesInBox()

h3IsValid(),
geoToH3(),
h3Distance(),
…

geoToS2(),
s2GetNeighbors(),
s2CapContains(),
...
```

Publish

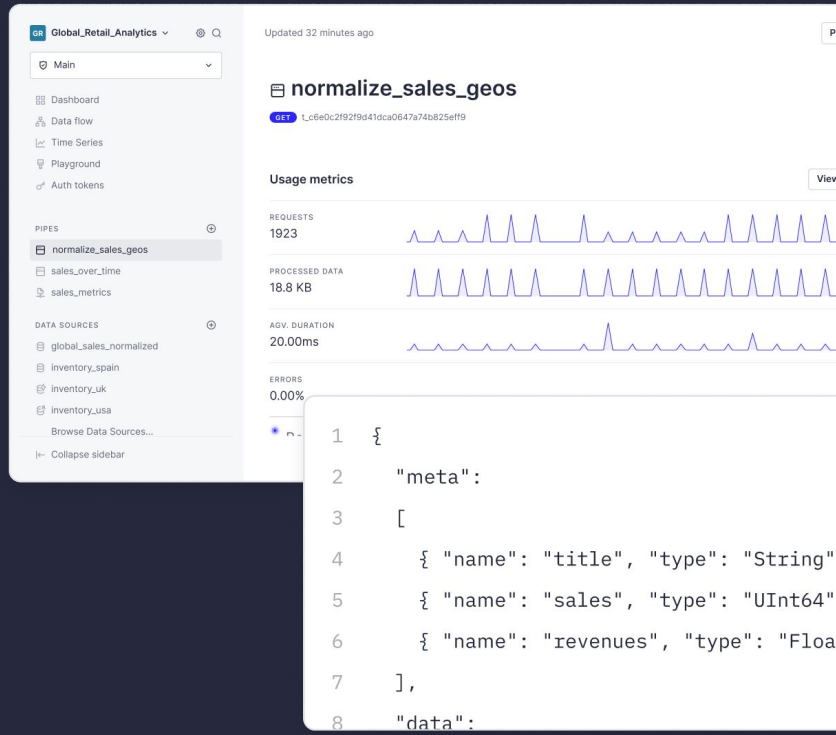
Publish

API Endpoints

Publish the result of a Pipe as an HTTPS API so your application can easily consume analytics over your data.

- ➔ With a single click, you can APIfy your data in a secure, scalable way with just SQL.
- ➔ No additional dev cycles to build an API layer.
- ➔ Engineers can focus on writing query logic and forget about how it will be consumed.
- ➔ Leaders don't need to worry about additional tools like API gateways, load balancers, security, etc. Everything is managed by Tinybird.

Read more here.



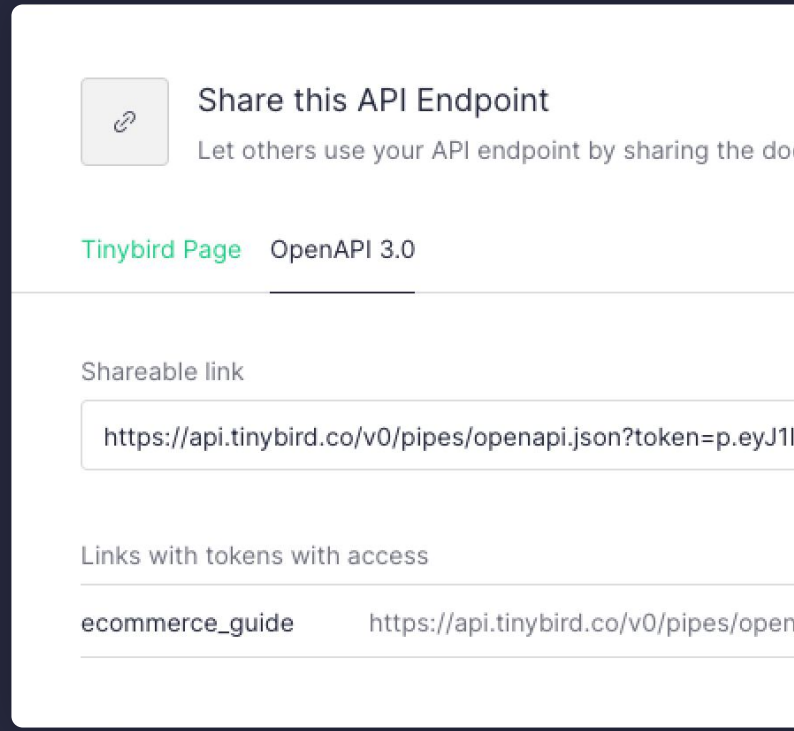
Publish

Endpoint Docs

Tinybird **auto-generates documentation** for your API Endpoints.

- ➔ The docs are Open API 3.0 compatible and contain info about query parameters, code snippets, and sample results.
- ➔ Engineers can focus on developing, not documenting. It's dead-simple to share docs with your team in the format you already know.

Read more here.



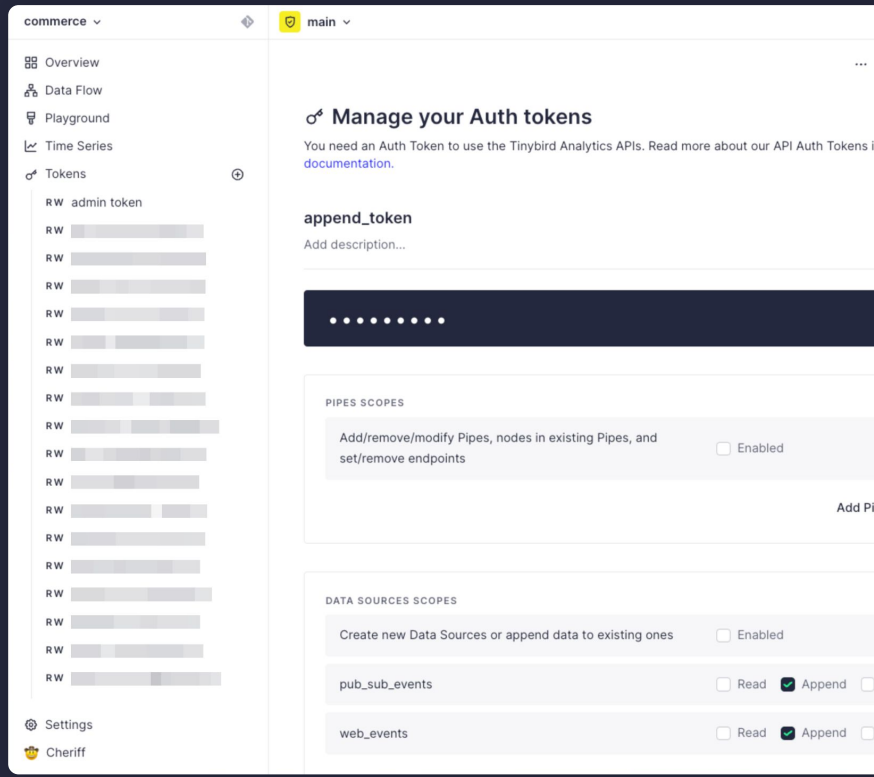
Publish

Auth Tokens

Auth Tokens **protect access** to your Tinybird Data Sources, Pipes, and Endpoints.

- ➔ Any action via the CLI or REST API requires a valid Auth Token.
- ➔ You can assign a custom scope to Auth Tokens to protect your resources from misuse and abuse.

Read more here.



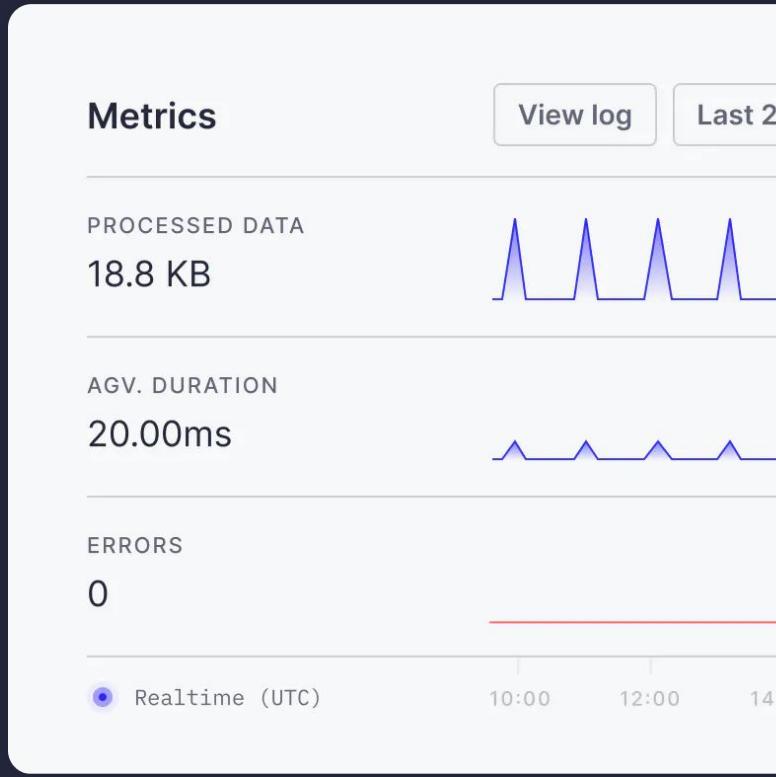
Publish

Service Data Sources

Service Data Sources allow you to **inspect what is going on** in your Tinybird account.

- ➔ Monitor API Endpoint performance and errors, Data Source operations, and more.
- ➔ You can treat Service Data Sources like any other Data Source; build API Endpoints and Time Series to monitor your Data Project in real time.

Read more here.

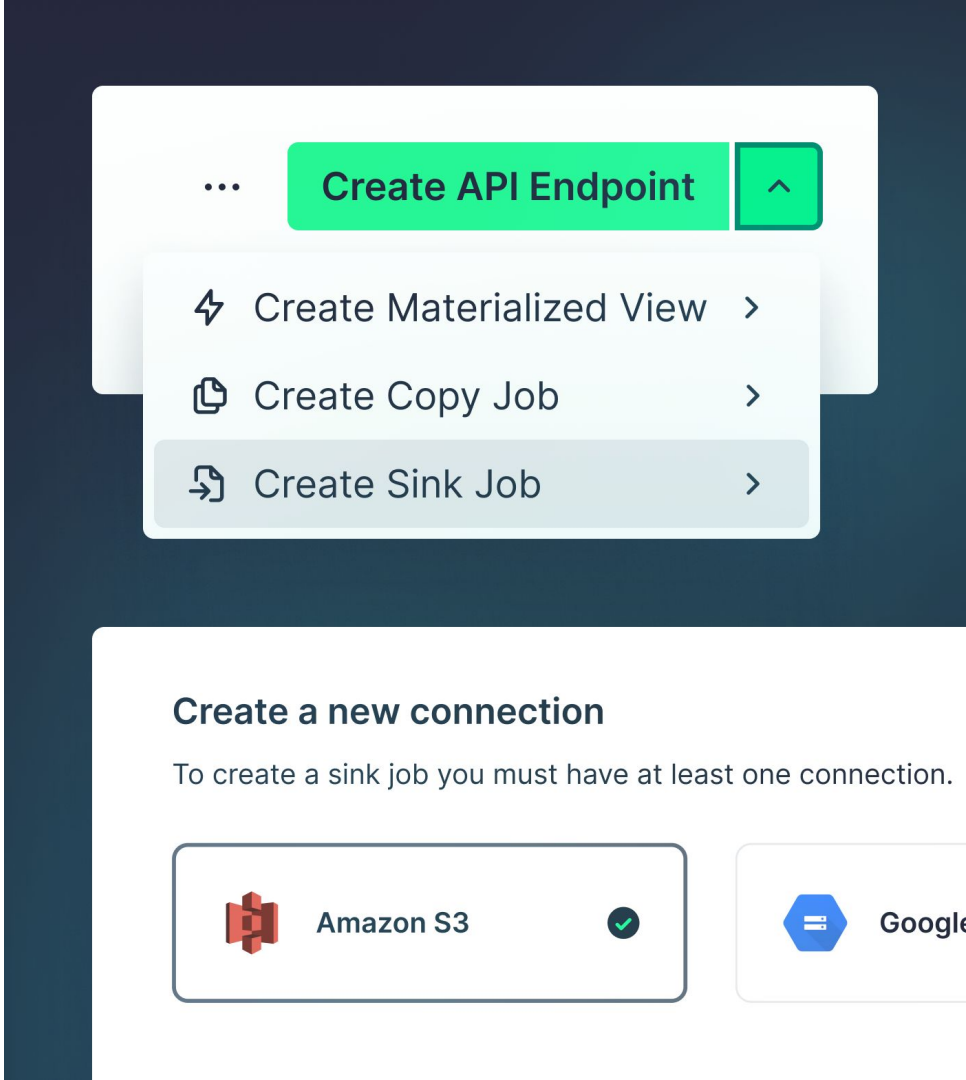


Publish Feature

Sinks

New and comfortable way to export data. Now you can easily get your data into your bucket in S3 or GCS.

- ➔ You can **export data** on your own in a **scheduled manner**.
- ➔ **No third-party orchestration** needed.
- ➔ Export process not constrained by the limits our our realtime endpoints.



Productivity

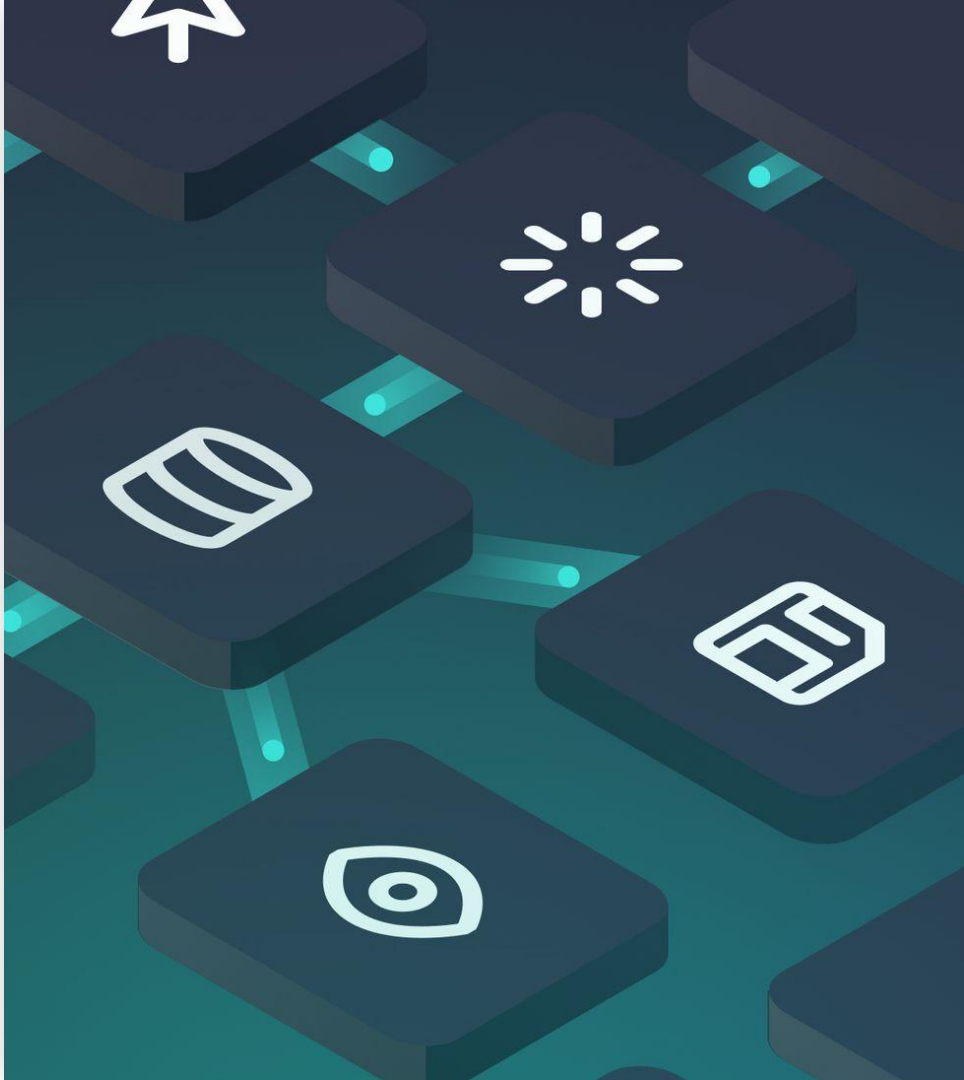
Productivity

REST API

Tinybird is an **API-first** platform.

- ➔ All functionality provided by the platform is accessible via the API to allow for custom, programmatic use.
- ➔ Engineering teams often want to automate a platform before taking it into production. The REST API provides endless flexibility for automation.
- ➔ The Tinybird API is extensively documented with clear examples.

Read more here.

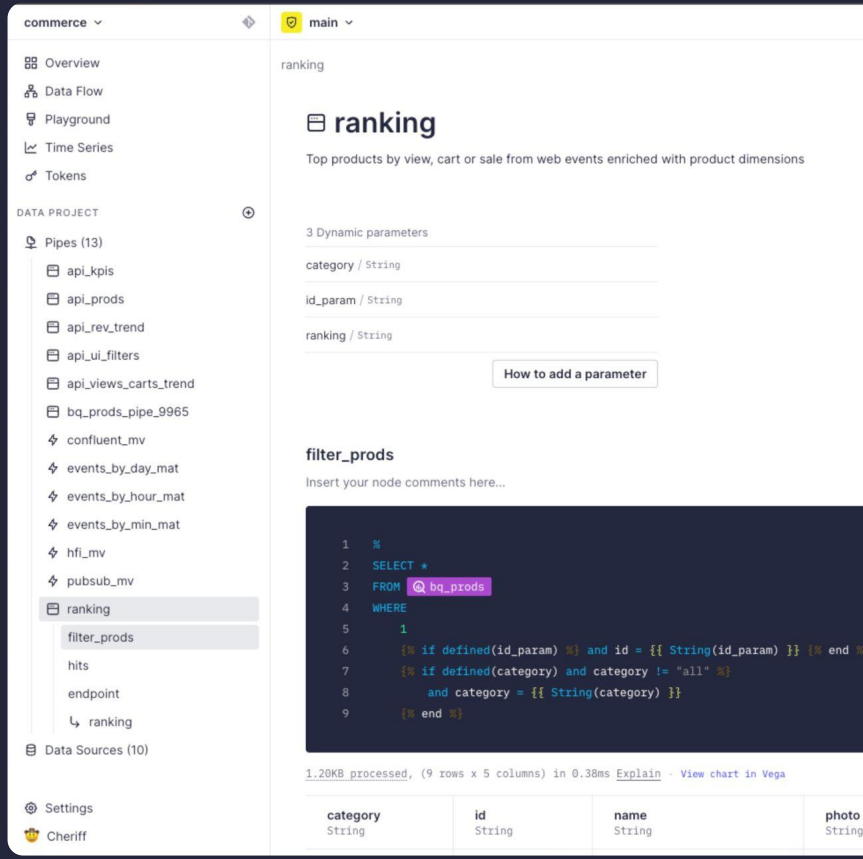


Productivity

UI

The Tinybird UI provides a **delightful experience** to enhance your productivity.

- ➔ Tinybird has a modern, notebook-inspired user interface to simplify and accelerate development.
- ➔ Compose discrete queries and chain them together for readability.
- ➔ Autocomplete and format your queries.
- ➔ Avoid context switching with query and schema split-screen.
- ➔ Queries executed in the UI are free! Preview query results in development and get to production faster.



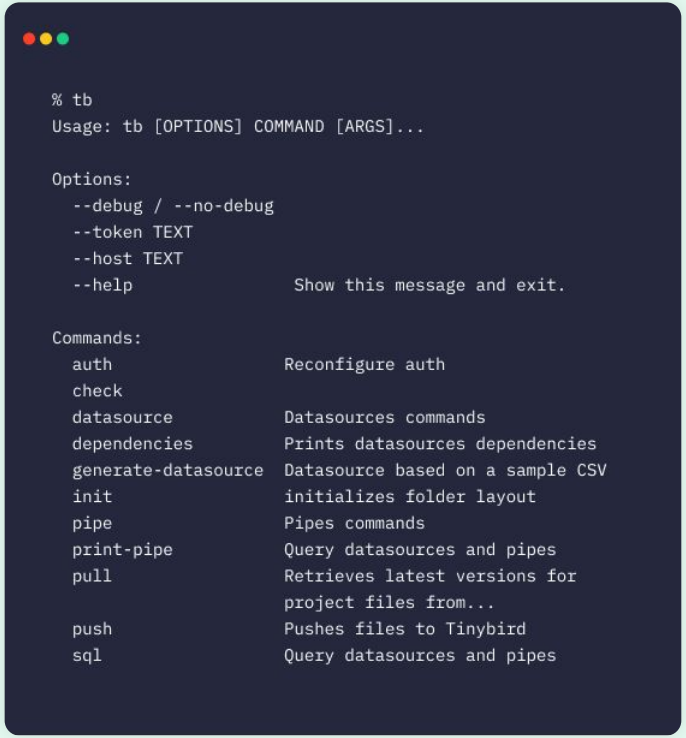
Productivity

CLI

The CLI provides an **easy-to-use programmatic interface** for the Tinybird REST API.

- ➔ The CLI can be used by engineers who prefer to work with files from the terminal.
- ➔ It can also be used in CI/CD pipelines to automate the deployment and testing of Data Projects.

Read more here.

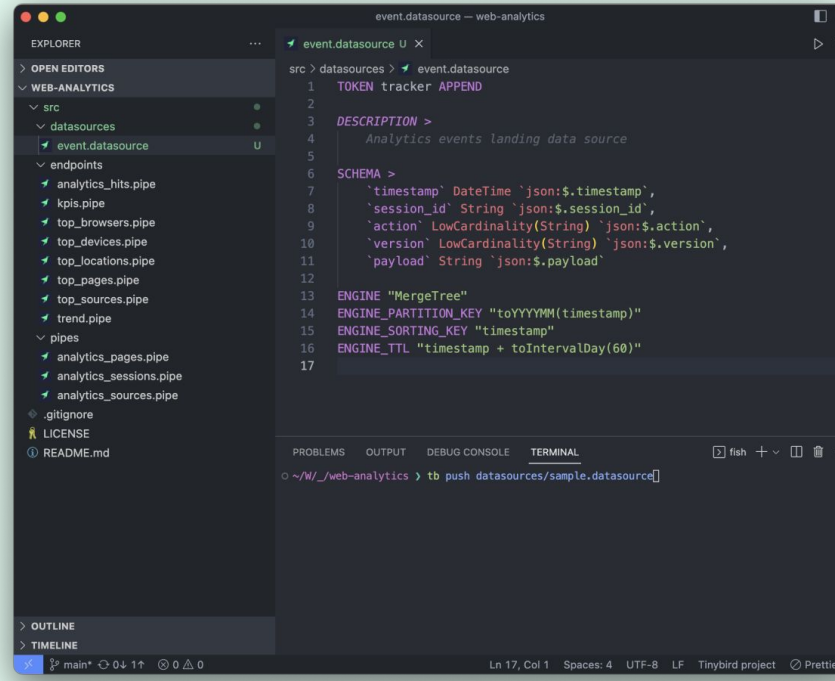


Productivity

Data Project Files

A Data Project is a **set of files** that describes how your data should be stored, processed, and published.

- ➔ Every schema, query, and API created in Tinybird can be represented as configuration files.
- ➔ A Data Project is like the source code of your project; it does not include the data itself.
- ➔ Any work performed in the UI can be pulled down into files using the CLI.
- ➔ These files can be stored in version control, like git, and developed following a standard version control flow.
- ➔ Any work performed in files and git can be pushed to Tinybird, also making it available in the UI.

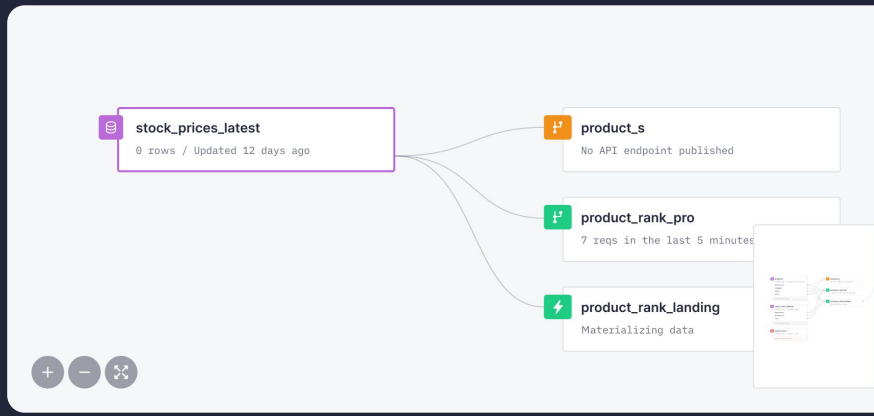


Productivity

Data Flow

Visualize the lineage in your Data Project to see how Data Sources, Pipes, and Endpoints are connected.

➔ View the end-to-end flow of your Data Project and quickly spot dependencies.



Coding time!

Let’s get to work ⚒️

Open two tabs:

→ app.tinybird.co

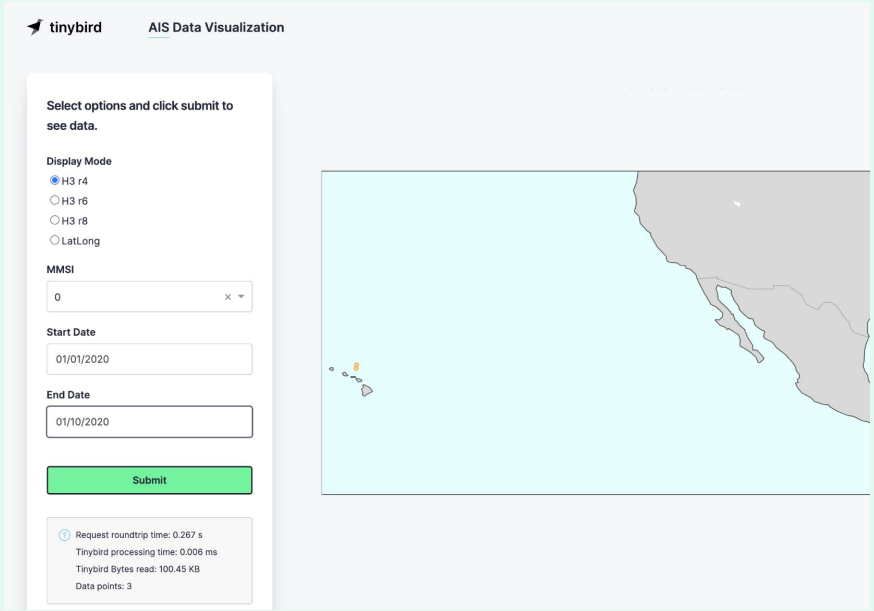
→ tbrd.co/jornadassiglibre

We need to build, at least, two endpoints:

1. One to get the list of unique MMSI
2. Another one to get the location of a MMSI, for a specific time range.

ClickHouse docs:

- Functions for Working with H3 Indexes - `h3ToGeoBoundary` and `geoToH3`
- `ReplacingMergeTree`
- Dynamic column selection



Thank you!

rbarbadillo@tinybird.co