

# Building a real-time analytics dashboard with Streamlit, Apache Pinot, and Apache Kafka



Mark Needham  
Developer Relations Engineer at StarTree

# What is real-time analytics?

Real-time analytics is the discipline that applies logic and mathematics to data to **provide insights for making better decisions quickly**.

On-demand real-time analytics waits for users or systems to request a query and then delivers the analytic results.

Continuous real-time analytics is more proactive and alerts users or triggers responses as events happen.

# What is real-time user-facing analytics?

Giving our end-users analytical querying capabilities on fresh data.



# What is real-time user-facing analytics?

## *Real-time*

The data should be available for querying as soon as possible.

# What is real-time user-facing analytics?

## *Real-time*

The data should be available for querying as soon as possible.

## *User-facing*

Queries times should be similar to what you'd get with an OLTP database.

A lot of queries will be executed at the same time.

# UberEats - Restaurant Manager

The screenshot displays the UberEats Restaurant Manager interface for 'The Pizza Place'. The dashboard includes a sidebar with navigation options, a main content area with a sales performance chart, and several summary cards on the right. Five orange callout boxes with numbers 1 through 5 highlight specific features: 1. The 'Last 12 weeks' dropdown menu; 2. The sales comparison between 'Last 12 weeks' and 'Last 12-24 weeks'; 3. The 'Missed orders' card showing 7 items that need attention; 4. The 'Top-selling items' card; 5. The 'Inaccurate orders' card showing 4 items that need attention.

**Uber Eats** **The Pizza Place**

Home

1 Last 12 weeks ▾

2 Sales (\$4.1k) Order volume (240) Ticket size (\$12.63)

\$4,127 Last 12 weeks

\$4,992 Last 12-24 weeks

Mon 2/2 - Sun 9/2 2.5% \$494 5% ▾

Mon 5/12 - Sun 12/12 \$338

3 Missed orders 7 Needs attention View more →

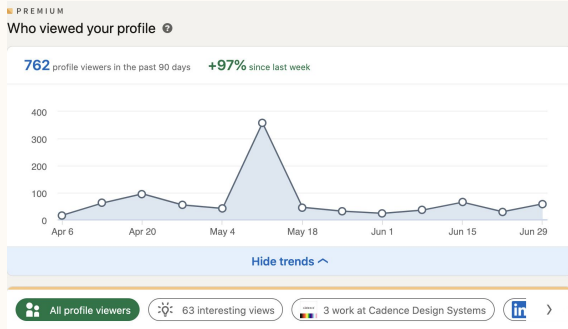
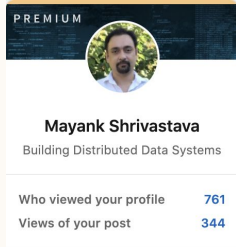
4 Inaccurate orders 4 Needs attention View more →

4 Top-selling items Ranked by order volume

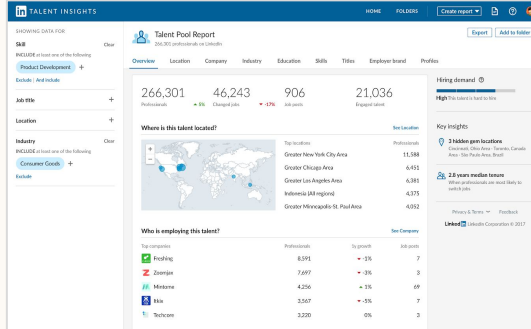
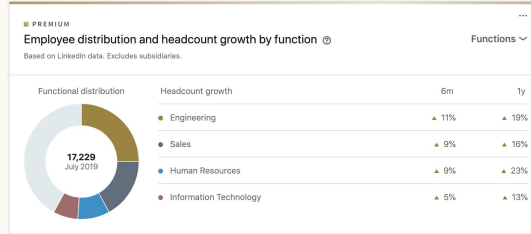
5 Menu item feedback 120 ratings

6 Downtime 7 Needs attention

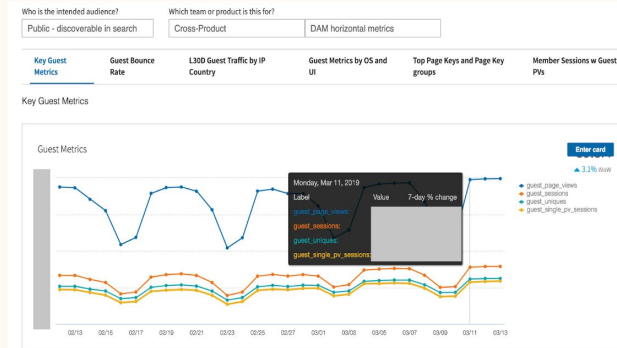
# LinkedIn



Member Facing Analytics



Enterprise Analytics Products



Internal Products and BI Tools

Let's build our  
own dashboard!



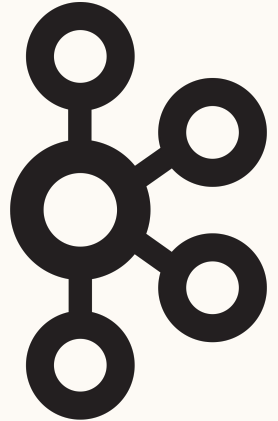
# What tools are we going to use?



Streamlit



pinot



# What is Streamlit?



# Streamlit

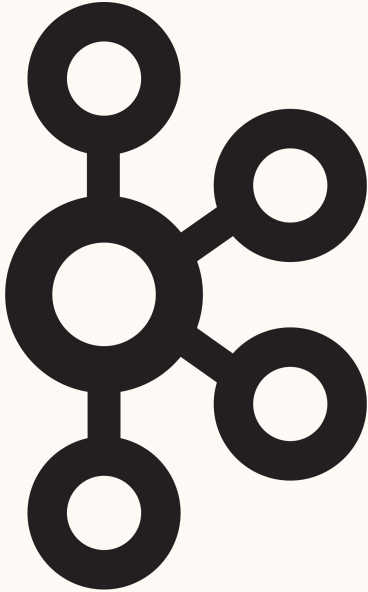
- Python based web application framework
- Makes it easy to create custom data apps
- Integrates with plot.ly, PyTorch, TensorFlow, scikit-learn, and more.
- Apps are shareable using Streamlit Cloud

# What is Apache Pinot?



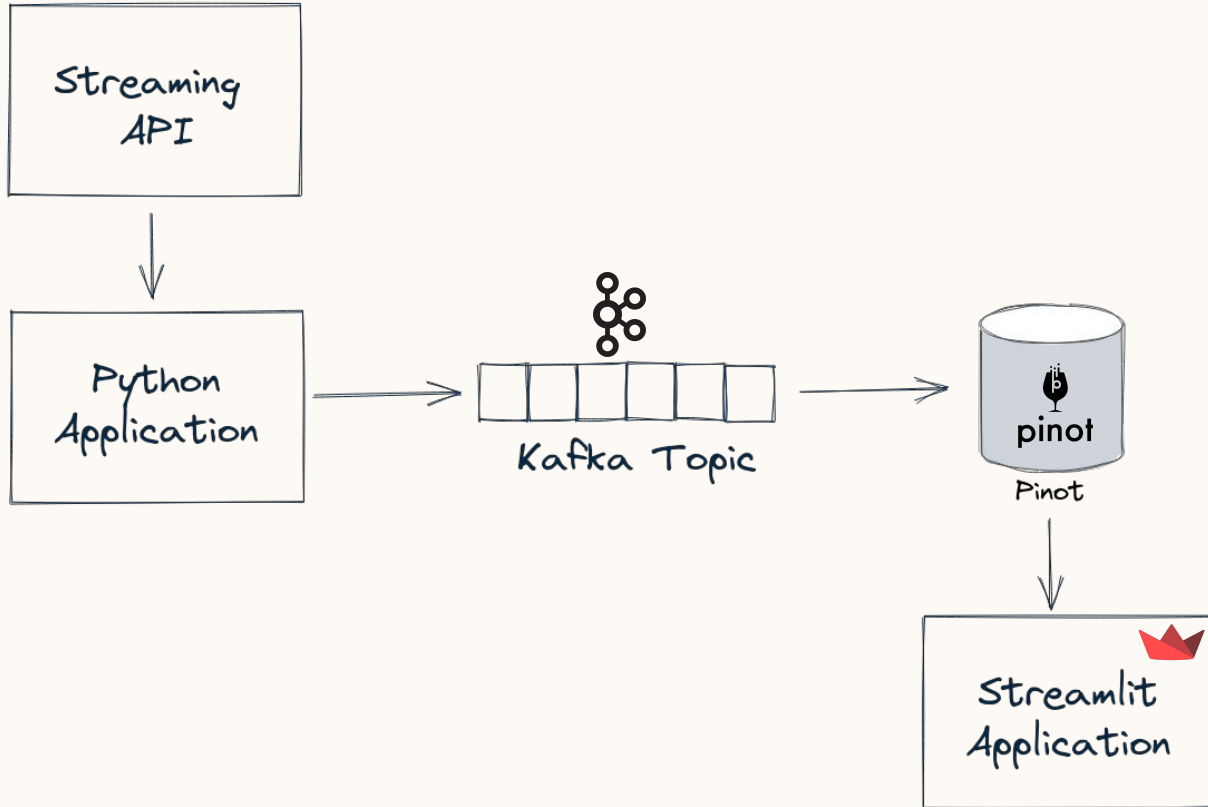
- Distributed OLAP Datastore
- Columnar, Indexed storage
- Queryable using SQL
- Real-time low latency analytics

# What is Apache Kafka?



- A distributed event streaming platform
- Publish and subscribe to streams of records
- Store streams of records in a fault-tolerant durable way
- Process streams of records as they occur

# Real-time analytics dashboard architecture



# Our data set: Wikimedia Recent Changes Feed

- A continuous stream of structured event data describing changes made to various Wikimedia properties.
- Published over HTTP using the Server-Side Events (SSE) Protocol

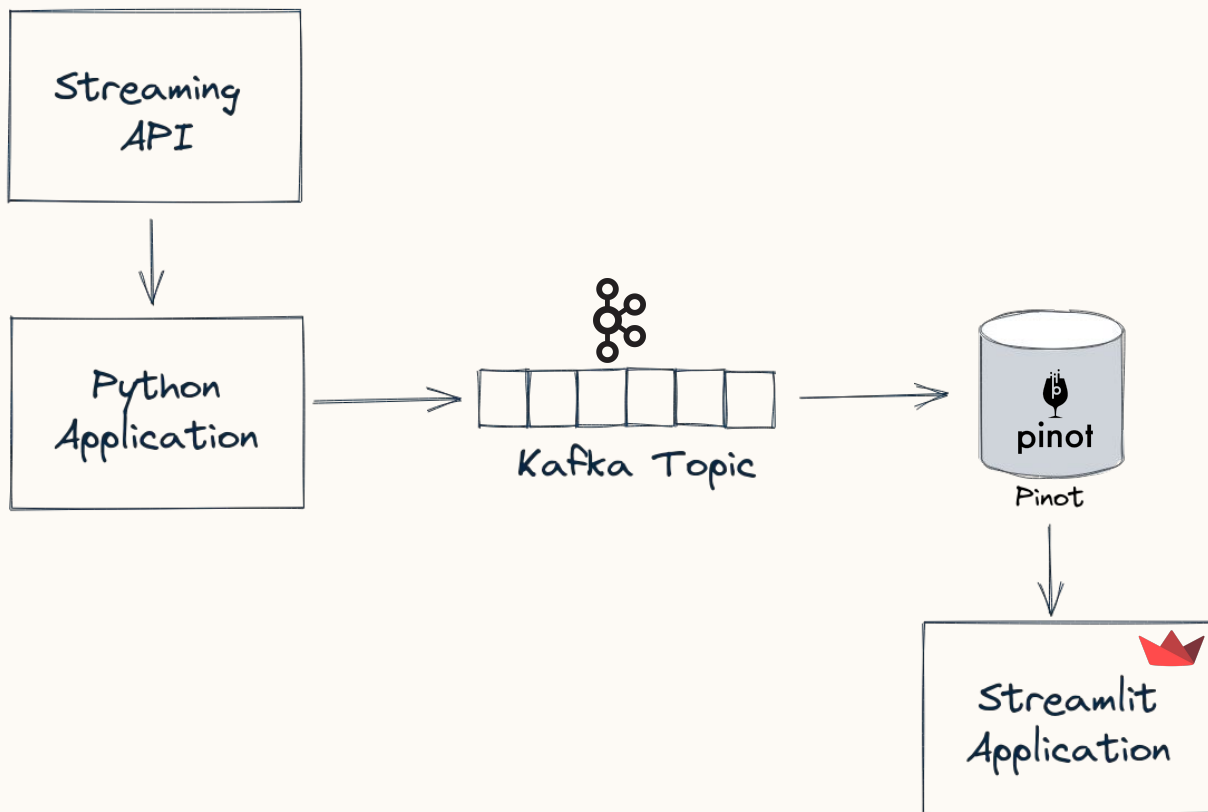
```
event: message
```

```
id:
[{"topic":"eqiad.mediawiki.recentchange","partition":0,"timestamp":1641388001001}, {"topic":"codfw.mediawiki.recentchange","partition":0,"offset":-1}]
data:
{"$schema":"/mediawiki/recentchange/1.0.0","meta":{"uri":"https://tt.wikipedia.org/wiki/%D0%A2%D3%A9%D1%80%D0%BA%D0%B5%D0%BC:%D0%92%D0%B5%D1%80%D0%BC%D0%BE%D0%BD%D1%82_%D1%88%D3%99%D2%BB%D3%99%D1%80%D0%BB%D3%99%D1%80%D0%B5","request_id":"946c08cb-be82-42d2-bb56-575cf3886a31","id":"806ff73f-b3b5-4174-8210-4950b5ee302b","dt":"2022-01-05T13:06:41Z","domain":"tt.wikipedia.org","stream":"mediawiki.recentchange","topic":"eqiad.mediawiki.recentchange","partition":0,"offset":3549434838},"id":24972084,"type":"category","category_id":14,"title":"Төркөм:Вермонт шәһәрләрe","comment":"[[:Бриджвотер (Вермонт)]] төркөмгә өстәлгән","timestamp":1641388001,"user":"Umarbot","bot":true,"server_url":"https://tt.wikipedia.org","server_name":"tt.wikipedia.org","server_script_path":"/w","wiki":"ttwiki","parsedcomment":"<a href=\"/wiki/%D0%91%D1%80%D0%B8%D0%B4%D0%B6%D0%B2%D0%BE%D1%82%D0%B5%D1%80_(%D0%92%D0%B5%D1%80%D0%BD%D1%82)\" title=\"Бриджвотер (Вермонт)\">>Бриджвотер (Вермонт)</a> төркөмгә өстәлгән"}
```



Demo Time

# Recap: Real-time analytics dashboard





## Resources

- Streamlit - [streamlit.io](https://streamlit.io)
- Apache Pinot - [pinot.apache.org](https://pinot.apache.org)
- Apache Kafka - [kafka.apache.org](https://kafka.apache.org)
- Code - [github.com/mneedham/pinot-wiki](https://github.com/mneedham/pinot-wiki)

**@markneedham**



The End