

Anomaly Detection with Apache Pinot & ThirdEye

Yoav Nordmann



WhoAmI

Technology Enthusiast

Nerd & Geek

Tech Lead & Architect

Group Lead & Mentor



twitter.com/YoavNordmann



linkedin.com/in/yoavnordmann

What is Anomaly Detection



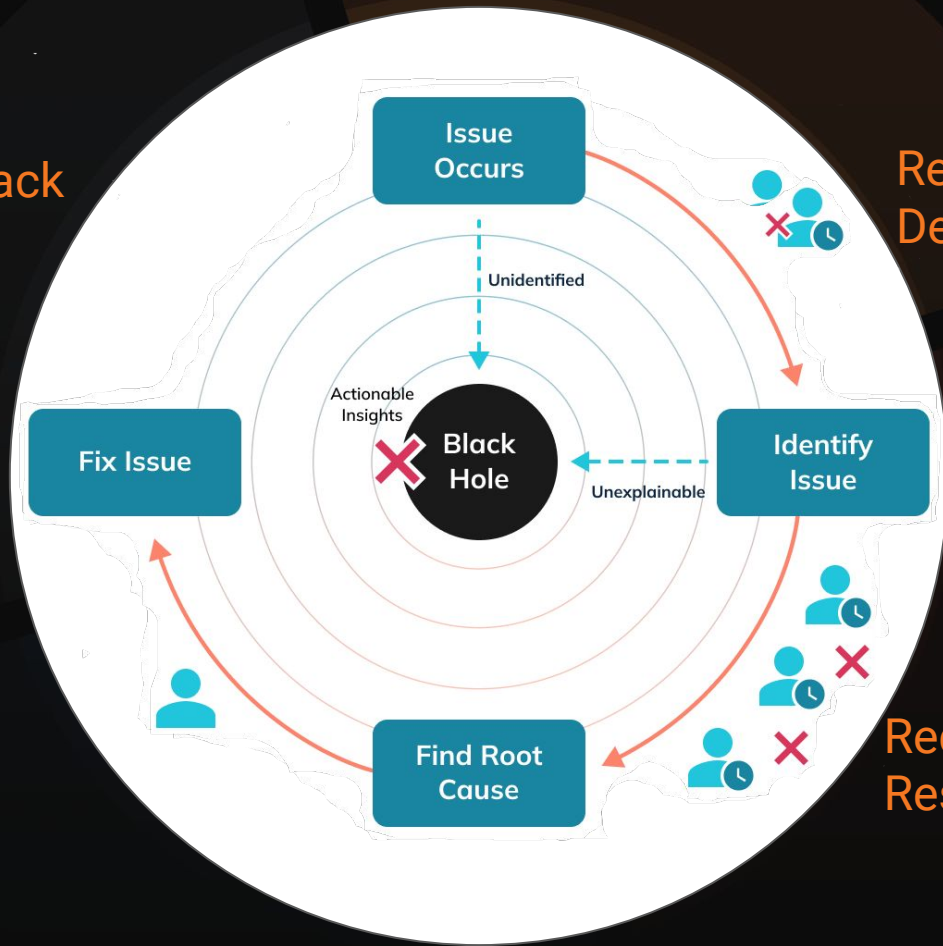
What is an Anomaly



Anomaly detection is understood to be the **identification** and/or **observation** of **data points** and events that **deviate** from a dataset's **normal behavior**

Eliminate Black Hole

Reduce Time To Detect



Reduce Time To Resolution

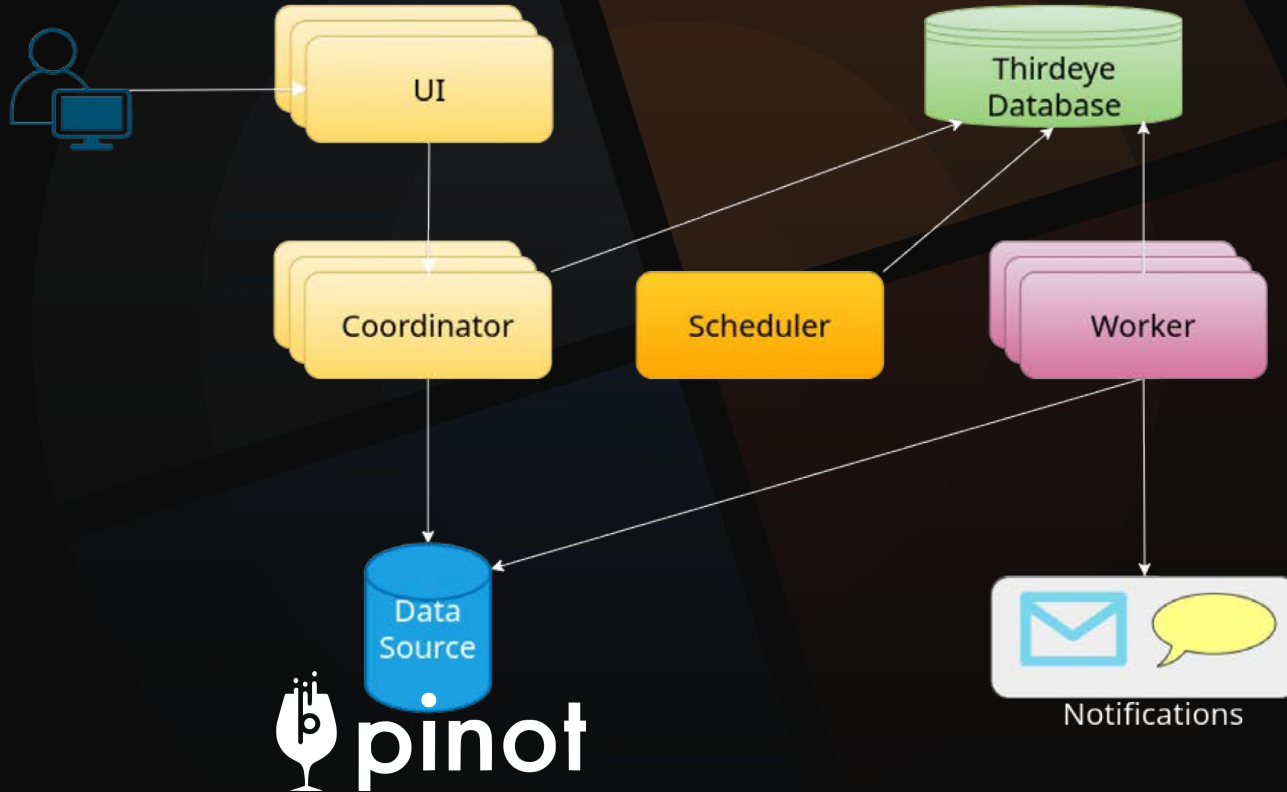
The Star of the Third Eye

ZODIAC

ThirdEye is an **anomaly detection,**
monitoring and interactive
root-cause analysis platform

star⚡tree

Architecture





Apache Pinot is a **real-time distributed OLAP datastore**, purpose-built to provide **ultra low-latency analytics**, even at extremely high throughput

DIMENSIONS

dimensions are the labels used to describe data

DeviceType	Android/Iphone
Country	IL/US

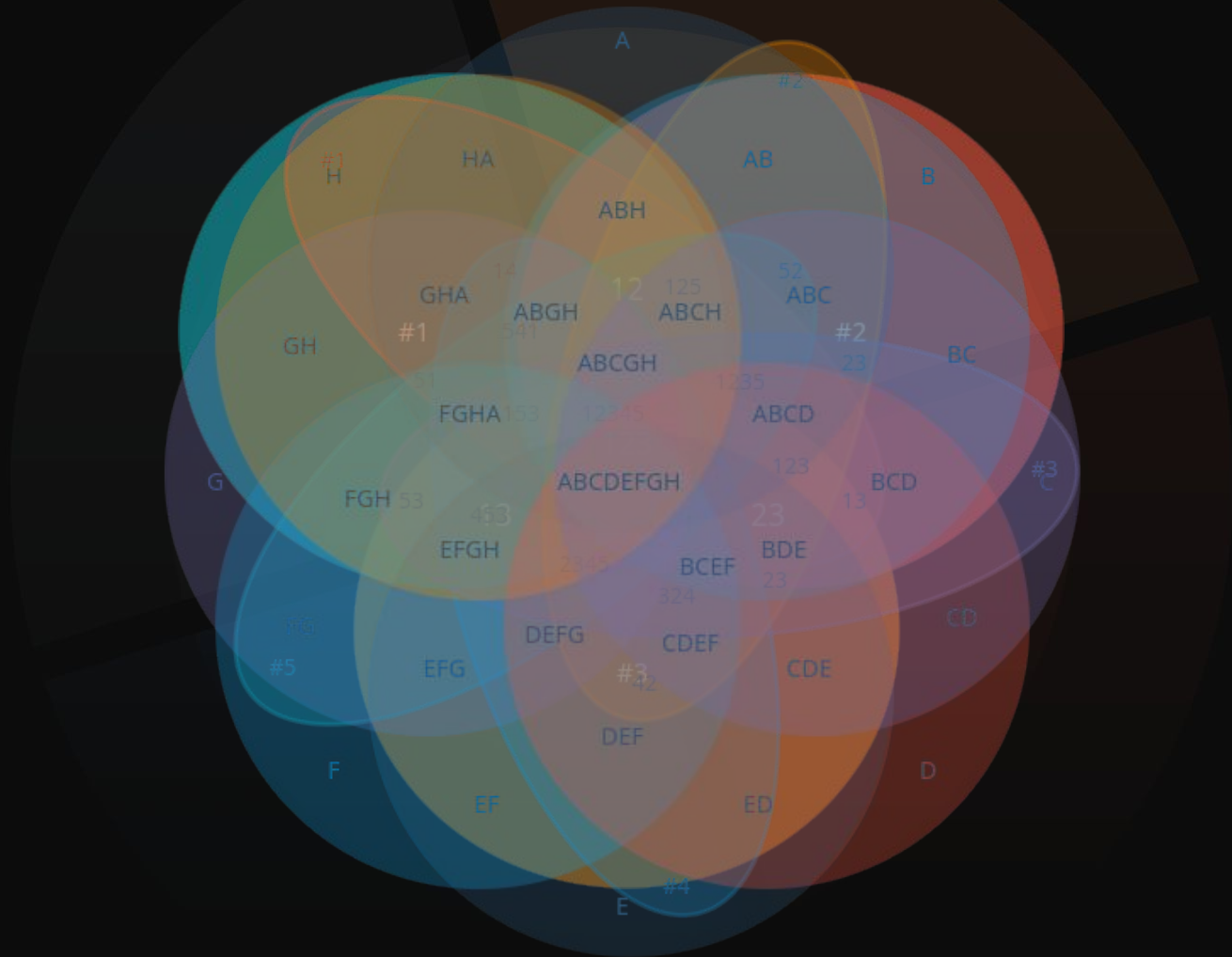
METRICS

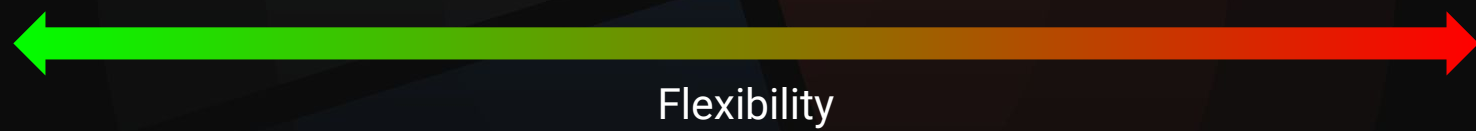
metrics are the quantitative measurements of data

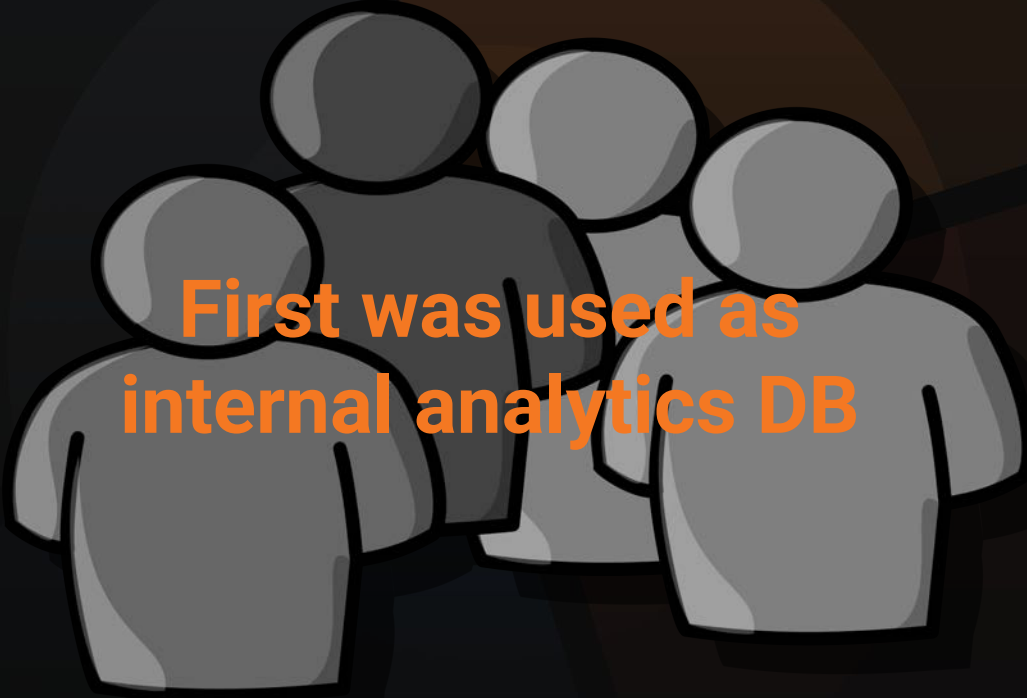
temperature	78.6
views	24

Slice & Dice









**First was used as
internal analytics DB**

A dense crowd of stylized human figures in various colors (blue, green, purple, orange, grey) against a black background. The figures are simple, rounded shapes with a single vertical line for a neck. They are packed closely together, creating a sense of a large group or crowd. The colors are vibrant and saturated. In the center of the crowd, there is a black rectangular box containing the text "ALL THE 500 MILLION LINKEDIN USERS" in a bold, orange, sans-serif font.

**ALL THE 500 MILLION
LINKEDIN USERS**



Search



Home



My Network



Jobs



Messaging



Notifications



Me



Work

Try Premium for free



Yoav Nordmann

Tech Lead | Architect

Who's viewed your profile 112

Impressions of your post 1,129

Access exclusive tools & insights

Try Premium for free

My items

Recent

- Fullstack Developers Israel
- # futurism
- # innovation
- # technology
- High Tech Dream Jobs

Groups

- Fullstack Developers Israel
- High Tech Dream Jobs
- High Tech Community in Jer...



Start a post



Photo



Video



Event



Write article

Sort by: Recent



Felix Palacci likes this

New posts



Christopher Tighe • 2nd

General Manager, at Cisco Switzerland | Powering an Inclusive... + Follow

It's no secret that enterprises are facing a huge cybersecurity talent shortage. It's concerning to hear that, according to the World Economic Forum, 59% of businesses failed to respond to cybersecurity inc...

https://www.devdiscourse.com/article/technology/2147653-the-growing-cybersecurity-skills-gap-whats-causing-i

devdiscourse.com • 5 min read



Felix Palacci and 22 others



Like



Comment



Share



Send



Google Cloud

Promoted

Get insights on how to solve challenges, innovate faster & grow more, at Google Cloud's biggest event of the year.

Add to your feed



GitHub

Company • Computer Software

+ Follow



Stack Overflow

Company • Computer Software

+ Follow



משרות "חבר מביא חבר" בהייטק - רשת פנימית

Company • Human Resources

+ Follow

View all recommendations

About Accessibility Help Center

Privacy & Terms Ad Choices

Advertising Business Services

Get the LinkedIn app More

LinkedIn LinkedIn Corporation © 2022



Messaging

LinkedIn Stats

200k+

QPS

1M+

Max Ingestion
Rate

20B+

Records Scanned
Per Second

Speed & Efficiency



Pluggable indexing technologies

Timestamp Index

JSON Index

Geospatial

Inverted Index

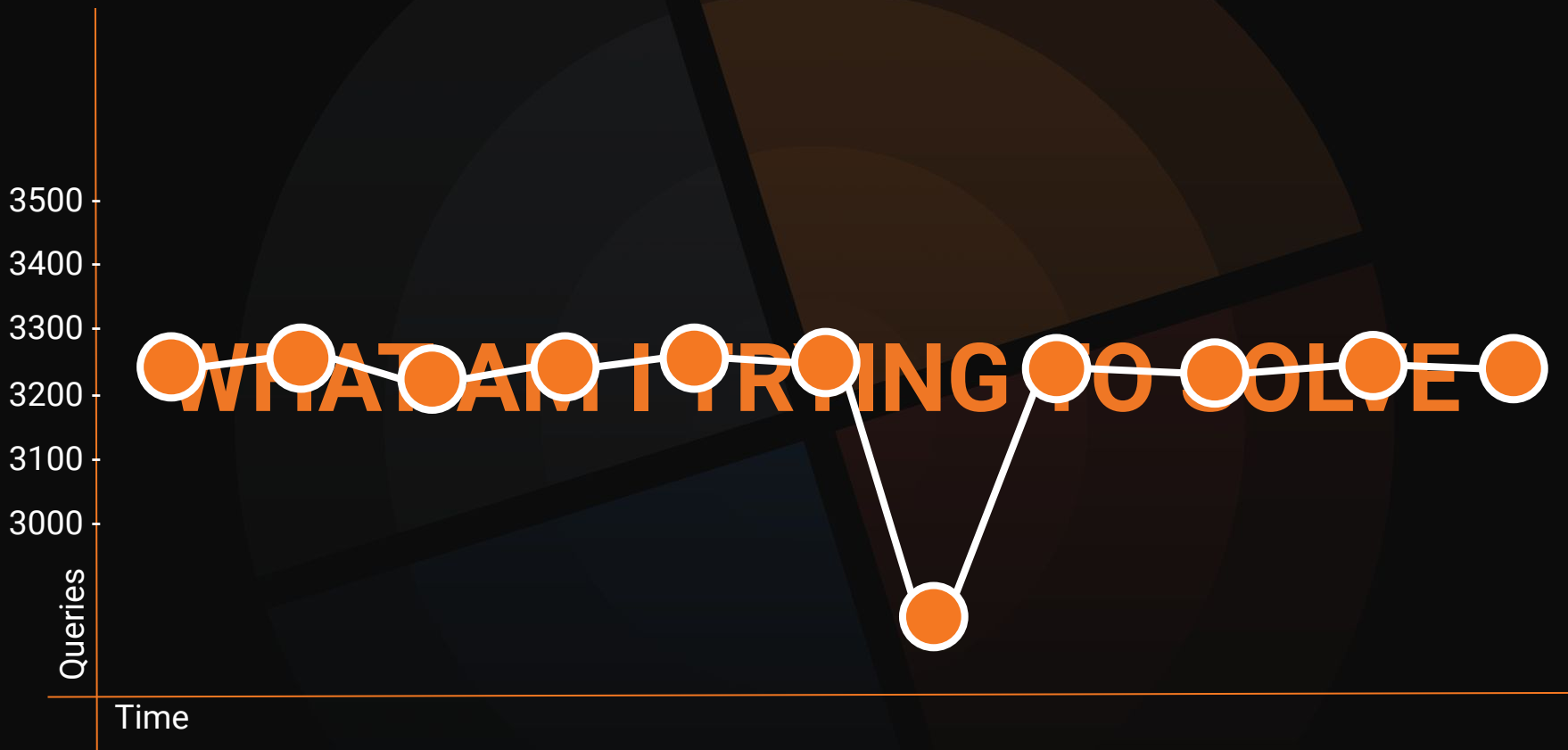
Range Index

Bloom Filter

Forward Index

Text Index

Star-tree Index



Alert template

A detection logic boilerplate that can be used to create Alert.

Example:
create an anomaly if $\${metric}$ is bigger than $\{max_value\}$.

Alert

A complete anomaly detection rule configuration.

Example:
create an anomaly if revenue is bigger than 20000. Check every hour.

Anomaly

A problem detected by
a detection pipeline

Example:
revenue was 30000, above the
threshold of 20000, on Thursday
3, between 9pm and 10pm.

Display Window ?

Feb 26 - Mar 4

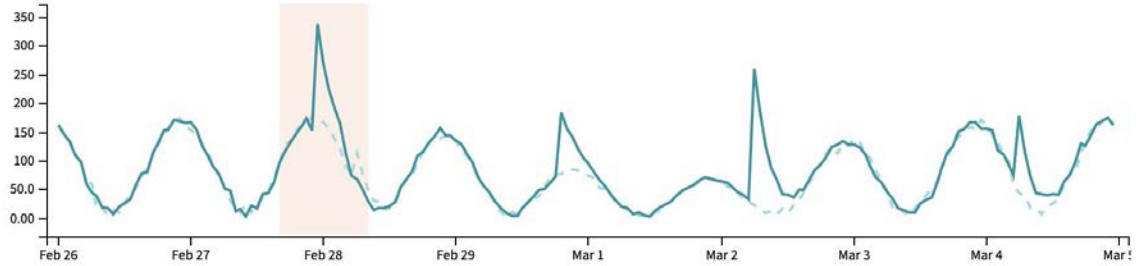
Granularity ?

1 Hour

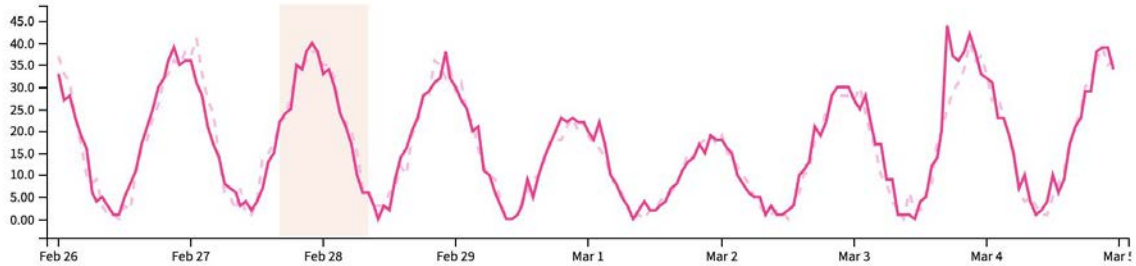
Compare by ?

split

views (chrome, mobile, us)



clicks (chrome, mobile, us)



Investigation Period ?

Feb 27, 04:00 pm

to Feb 28, 08:00 am

Feb 27, 04:00 pm PST — Feb 28, 08:00 am PST

Detector algorithms

- Threshold Rule
- Mean Variance Rule
- Percentage Rule
- Absolute Change Rule
- Holt-Winters Rule (Proprietary)

Want to write your own ?



Root Cause Analysis

Heatmap of Change in Contribution

Tooltip Reference

"Current" Data Date Range
Aug 25, 2021, 11:00 AM to Aug 25, 2021, 01:00 PM

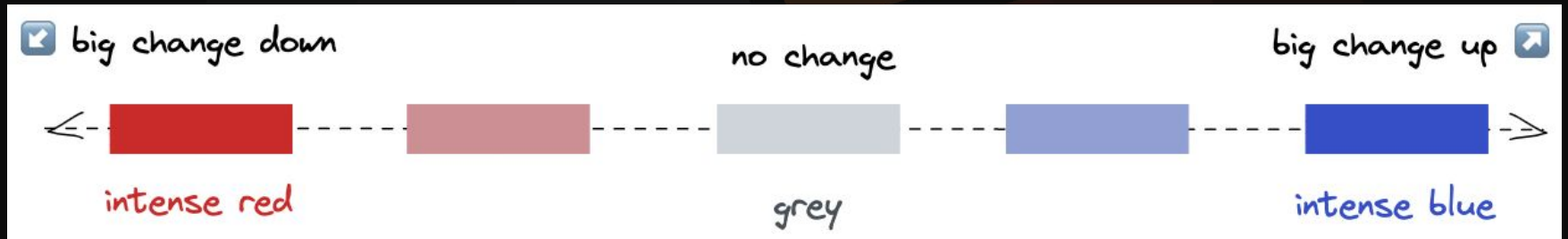
"Baseline" Data Date Range (One Week Ago)
Aug 18, 2021, 11:00 AM to Aug 18, 2021, 01:00 PM

Filter Data Controls

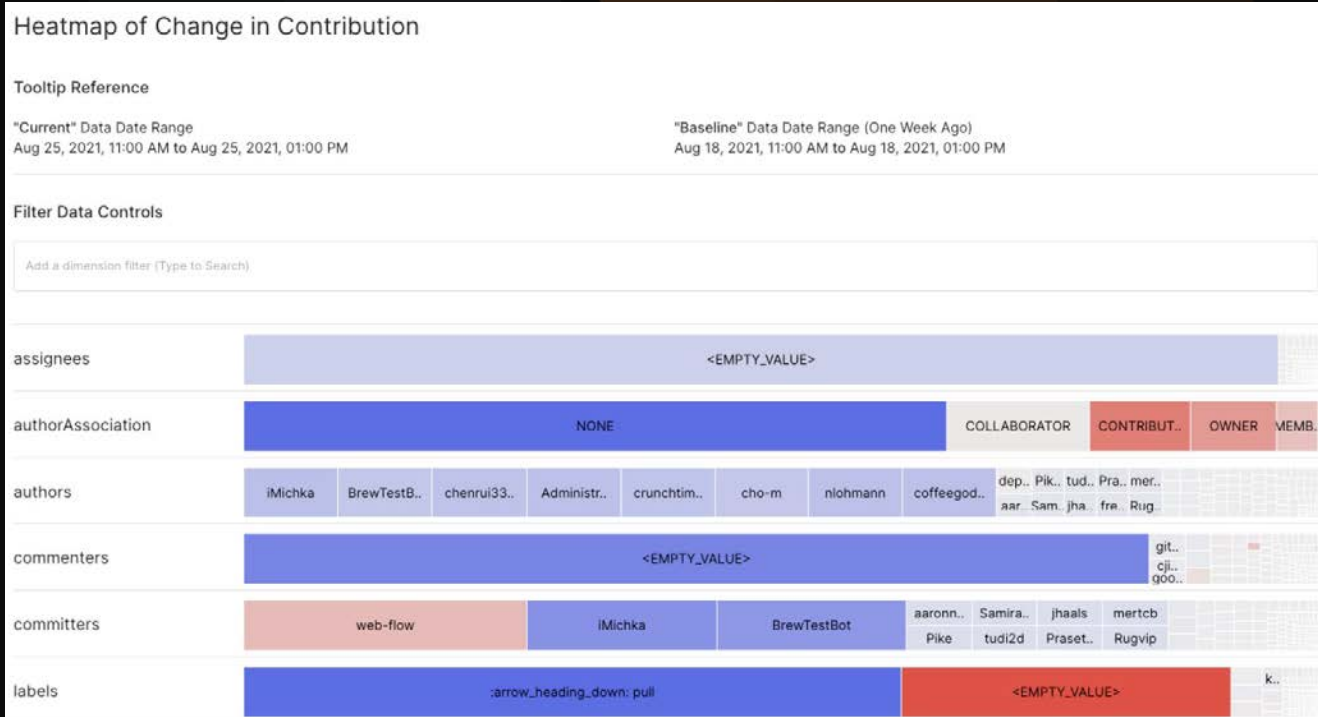
Add a dimension filter (Type to Search)

assignees	<EMPTY_VALUE>														
authorAssociation	NONE						COLLABORATOR	CONTRIBUT..	OWNER	MEMB..					
authors	iMichka	BrewTestB..	chenrui33..	Administr..	crunchtim..	cho-m	nlohmann	coffeegod..	dep..	Pik..	tud..	Pra..	mer..		
									aar	Sam	jha	fre	Rug		

Root Cause Analysis



Root Cause Analysis



AHUM
AHUM


WHAT ABOUT
ALERTS?










Subscriptions

The screenshot shows a web interface for configuration. On the left is a dark blue sidebar with icons for home, search, notifications, settings, and help. The main header is white and contains the word 'Configuration' on the left and a blue 'CREATE' button with a dropdown arrow on the right. Below the header is a horizontal navigation bar with tabs: 'Datasources', 'Datasets', 'Metrics', 'Alert Templates', 'Subscription Groups' (highlighted with a red circle), and 'Events'. The main content area is white and contains a blue icon of a hexagon with a plus sign, the text 'No Subscription Groups created.', and a blue button labeled 'CREATE SUBSCRIPTION GROUP'.

Subscriptions



Create Subscription Group



Subscription Group Properties


Basic information about the group

Name


Schedule
Every 5 minutes, every hour, every day

Channels


Setup the channels and the people you want to receive notifications from this group



EMAIL



SLACK



WEBHOOK

WAIT
WAIT
WAIT

WHAT IF THE
BASELINE WAS
A HOLIDAY



Events

The screenshot shows the 'Configuration' page in the TIKAL interface. The 'Events' tab is highlighted with a red circle. The interface includes a sidebar with navigation icons (lightning bolt, home, refresh, warning, settings) and a main content area with a 'CREATE' button and a table of events.

Configuration CREATE ▾

Datasources Datasets Metrics Alert Templates Subscription Groups **Events**

DELETE < JUL 01, 2022, 12:00 AM - OCT 02, 2022, 11:59 PM > Search Event

<input type="checkbox"/>	Name	Type	Start	End
--------------------------	------	------	-------	-----



Create Event



Event Properties

Name

Type

Start time

End time

Event Metadata

Optionally create custom event properties and corresponding values

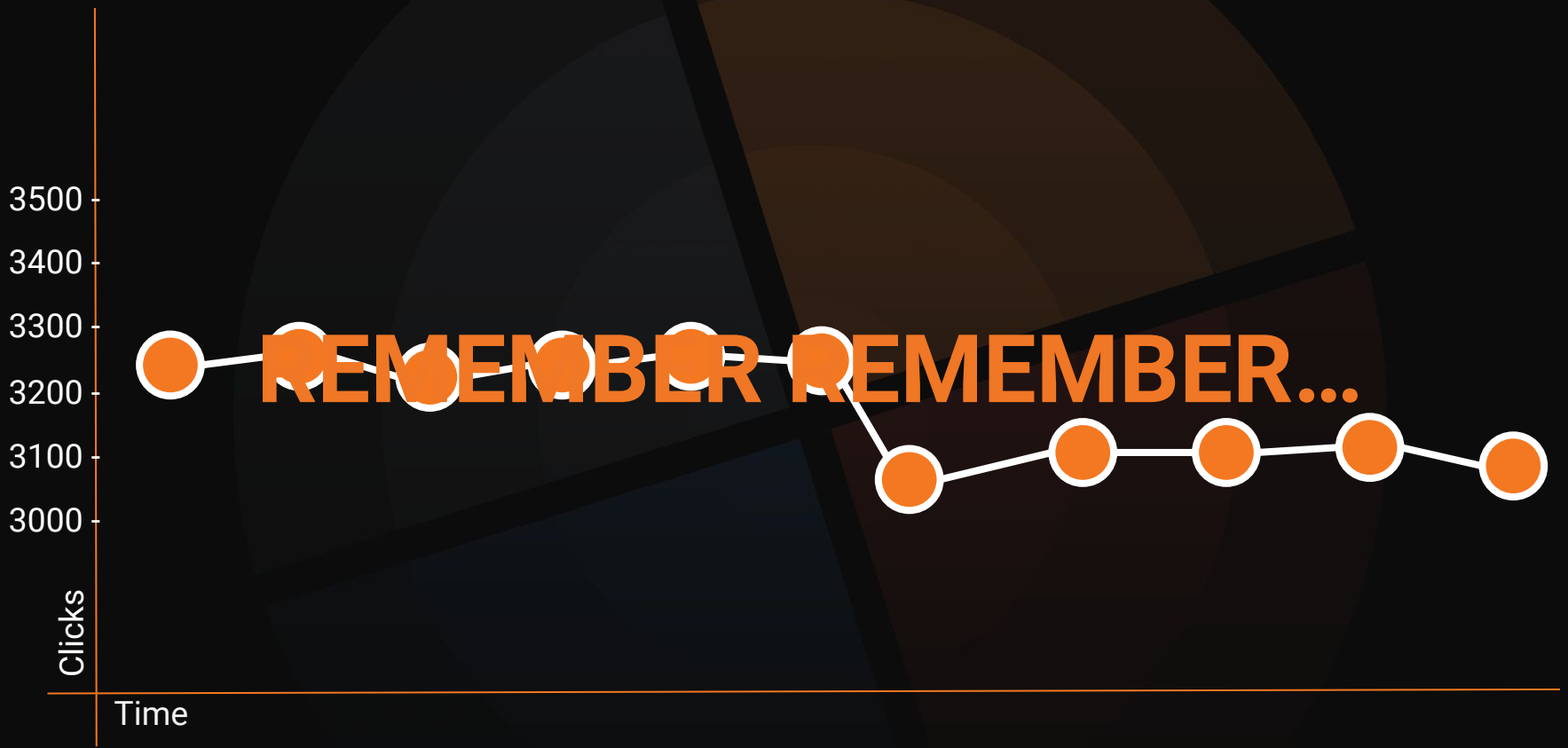
Property name

Property value

ADD METADATA ENTRY

CANCEL

CREATE EVENT



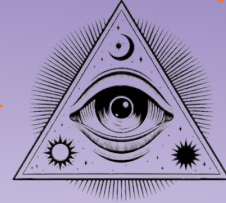
DEMO



kafka



pinot



telegraf



THANK YOU



twitter.com/YoavNordmann



linkedin.com/in/yoavnordmann

