OVH
Innovation is Freedom

Code
d'Armor

Introducing
Time Series & Warp 10
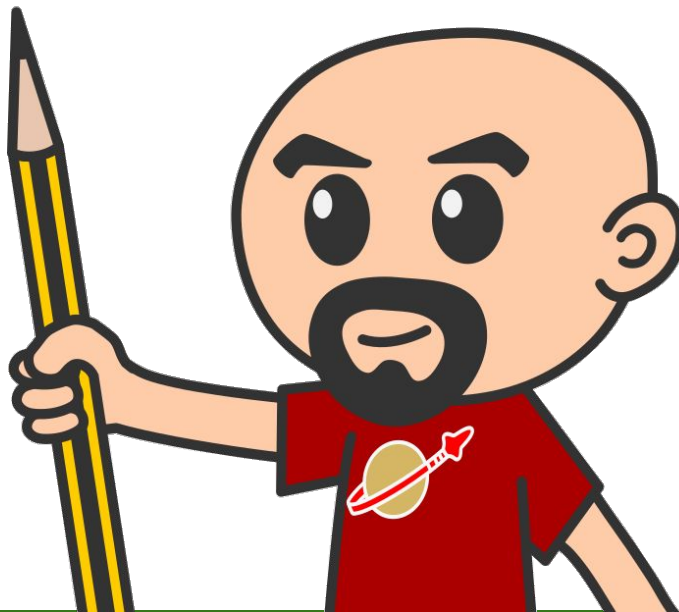
Horacio Gonzalez

`@LostInBrittany`

Horacio Gonzalez

@LostInBrittany

Spaniard lost in Brittany, developer, dreamer and all-around geek

Code d'Armor

Breizh C@mp
Mix de technologies

Finist Devs

GDG Finistère

WARP 10 MEETUP

Google Developers Experts 2015
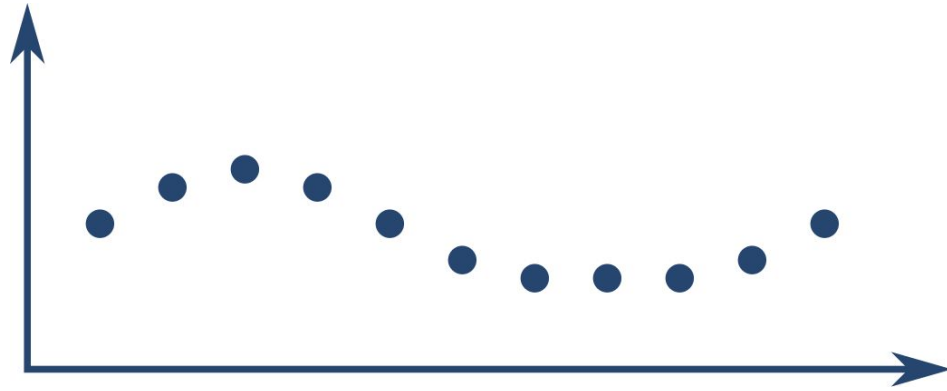Web Technologies GDE

OVH
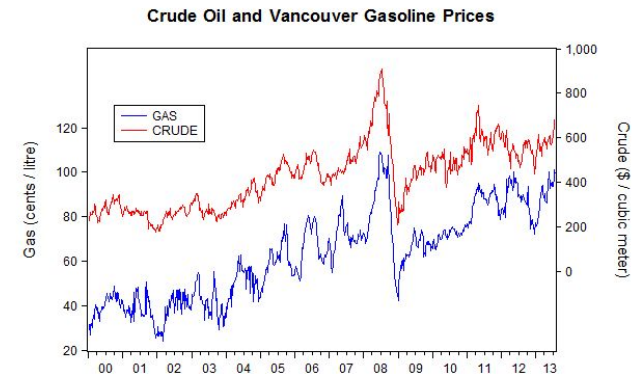Team DevRel

Time Series

What are they?

Time Series

Definition of Time Series:

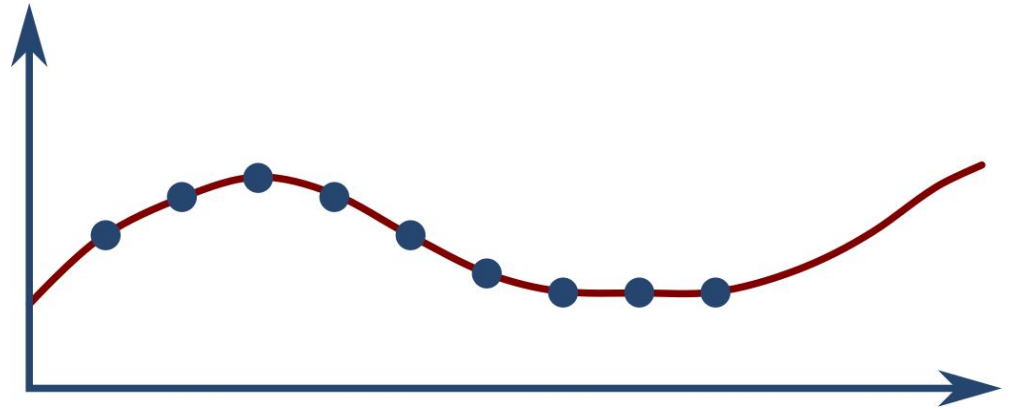An ordered sequence of values of a variable at equally spaced time intervals.

Time Series

- Stock Market Analysis
- Economic Forecasting
- Budgetary Analysis
- Process and Quality Control
- Workload Projections
- Census Analysis
- ...

Crude Oil and Vancouver Gasoline Prices

Time Series

Applications:

- ### Understanding the data

- ### Fit a model
  - Monitoring
  - Forecasting

Time Series

Stock market Analytics

Economic Forecasting

↓

$$$

↓

Study & Research

Time Series

Many specific analytical tools:

- Moving average

- ARMA (AutoRegressive Moving Average)

- Multivariate ARMA models

- ARCH (AutoRegressive Conditional Heteroscedasticity)

- Dynamic time warping

- ...

Time Series

Specific application of general tools

- Artificial neural networks
- Hidden Markov model
- Fourier & Wavelets transforms
- Entropy encoding
- ...

Time Series Databases

What tools do I use?

Time Series Databases

General purpose data analytics tools:

- Matlab

- Python

- R

- …

General purpose relational database engine

Time Series Databases

Does not scale!

OVH
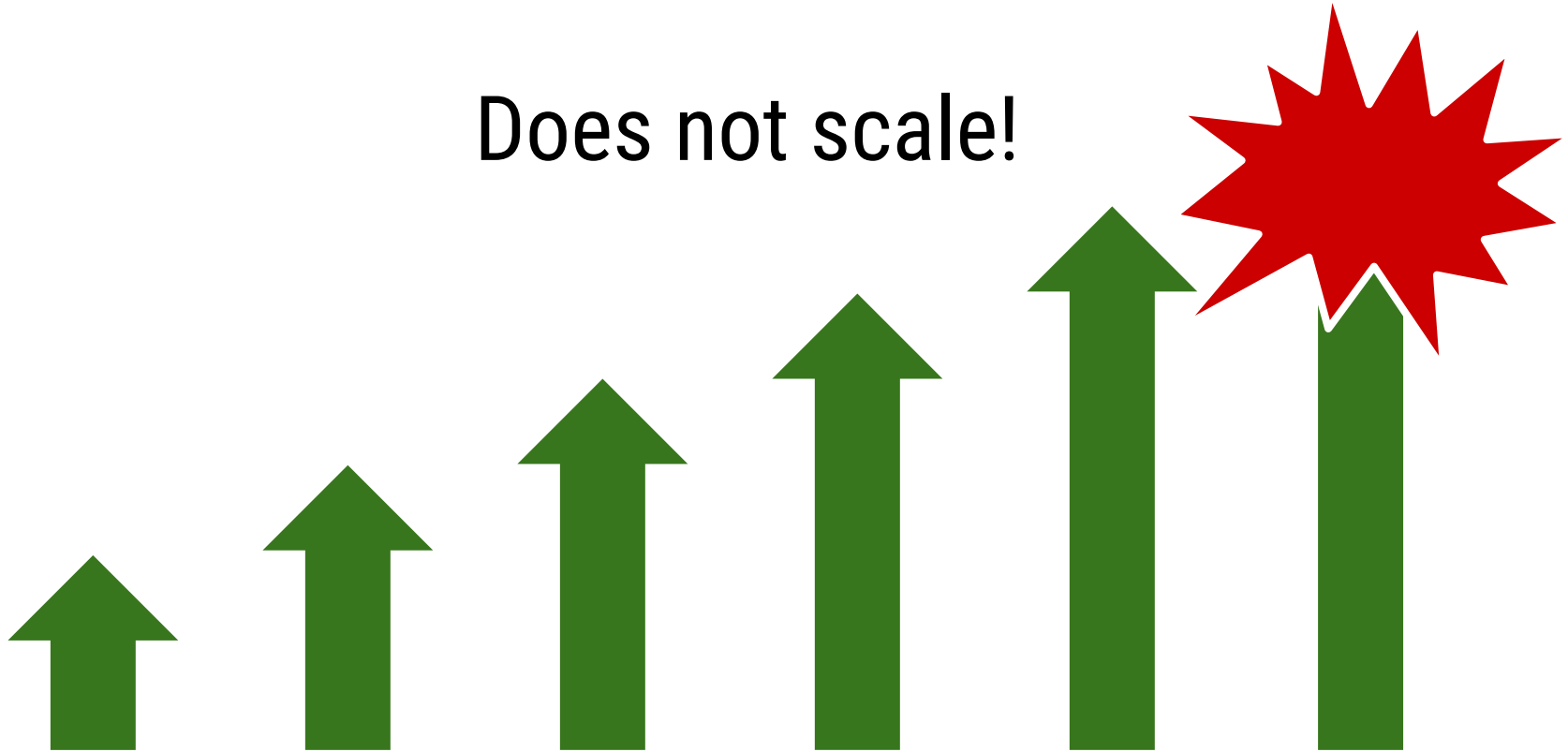
Time Series Databases

Many type of database engine

- Relational databases
- Key-value databases
- Document databases
- Graph databases
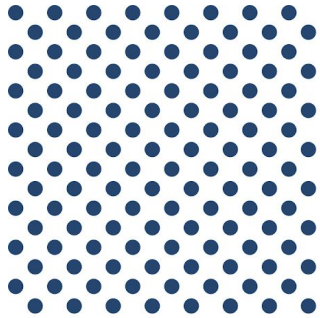- ...

What about Time Series?

Time Series databases!

Time Series Databases

Data model: time series

Long / Double

Boolean

String
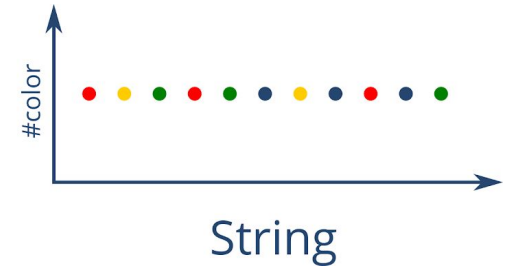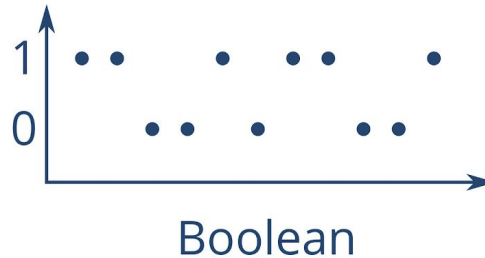
Time Series Databases

The 3 'v'

Volume
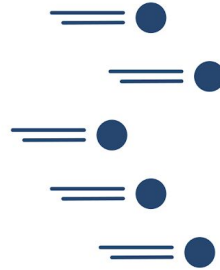
Velocity
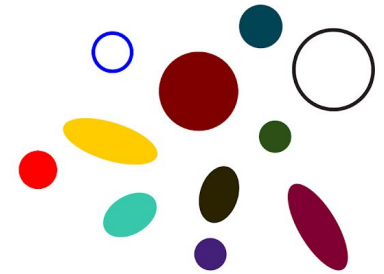
Variety

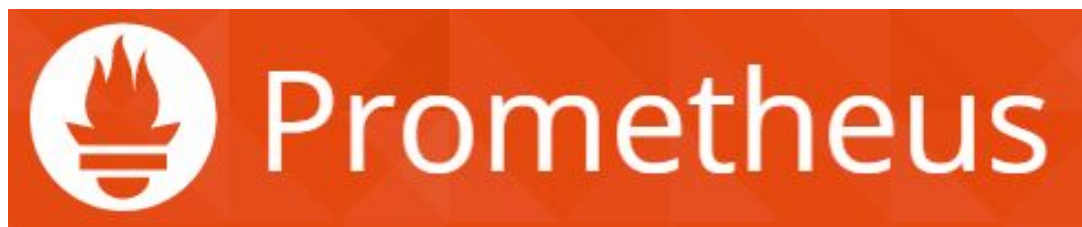# **Time Series Databases**

Many options

Time Series Databases

Time Series Databases
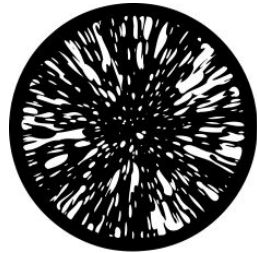
Time Series Databases

Time Series Databases

Time Series Databases

OVH Metrics

What did we choose?

What's OVH Metrics

Managed Cloud Platform for Geo Time Series®

What's a metric?

[**me**-trik] : the science of measuring

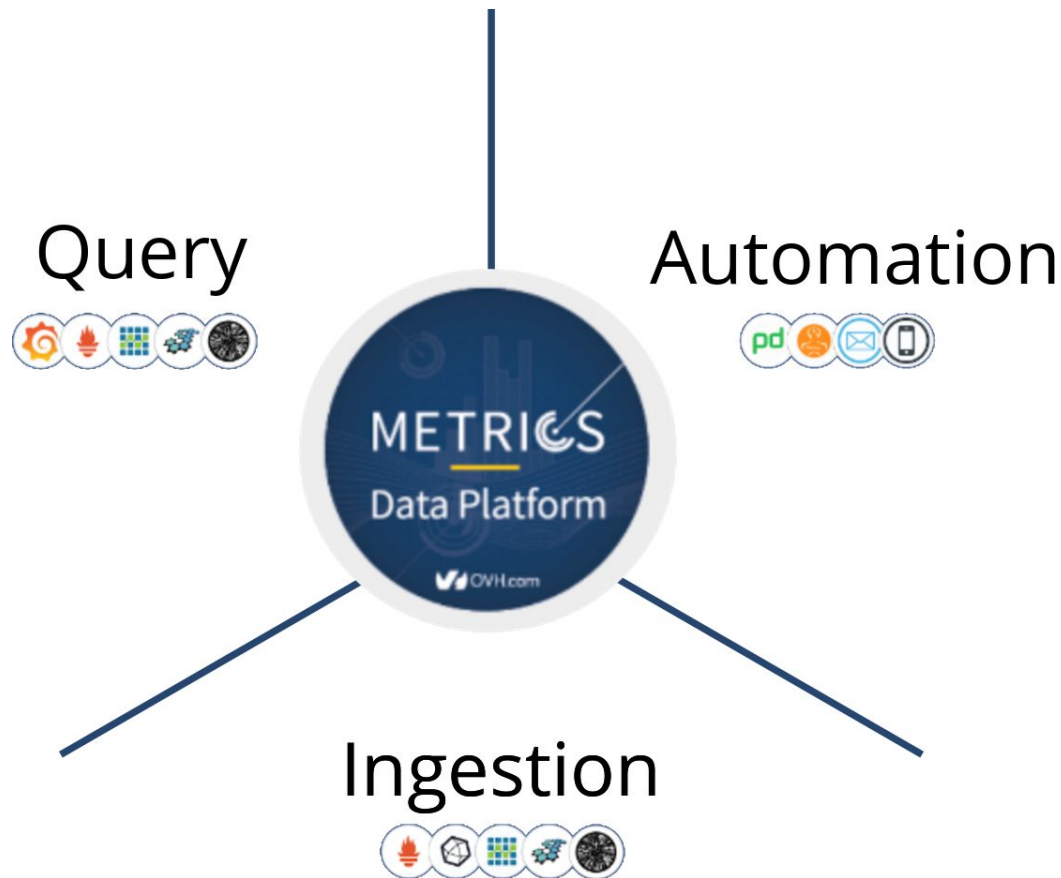# What is a metric?

Metrics are Time Series!

How do we deal with metrics?
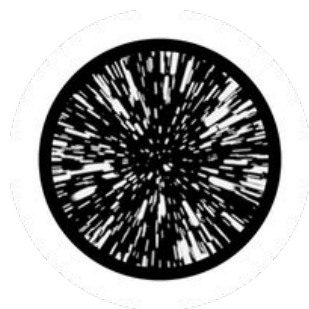
Using a Time Series Database!

But… which one?

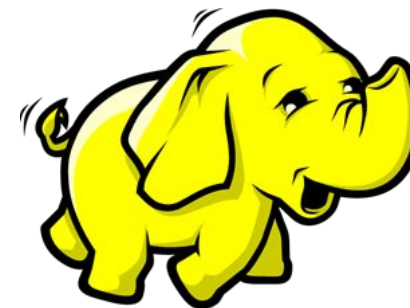Why choose? Let's support all of them!

Metrics Data Platform
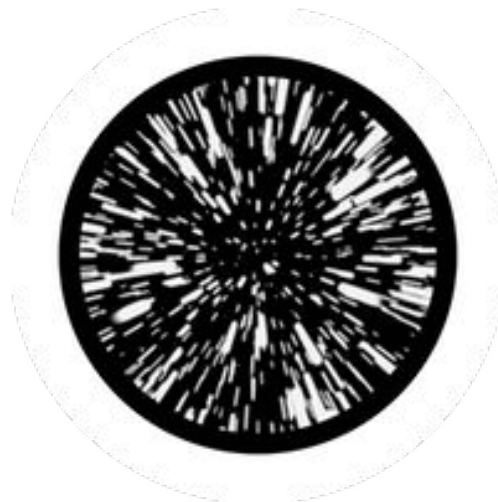

Query

Automation

Ingestion
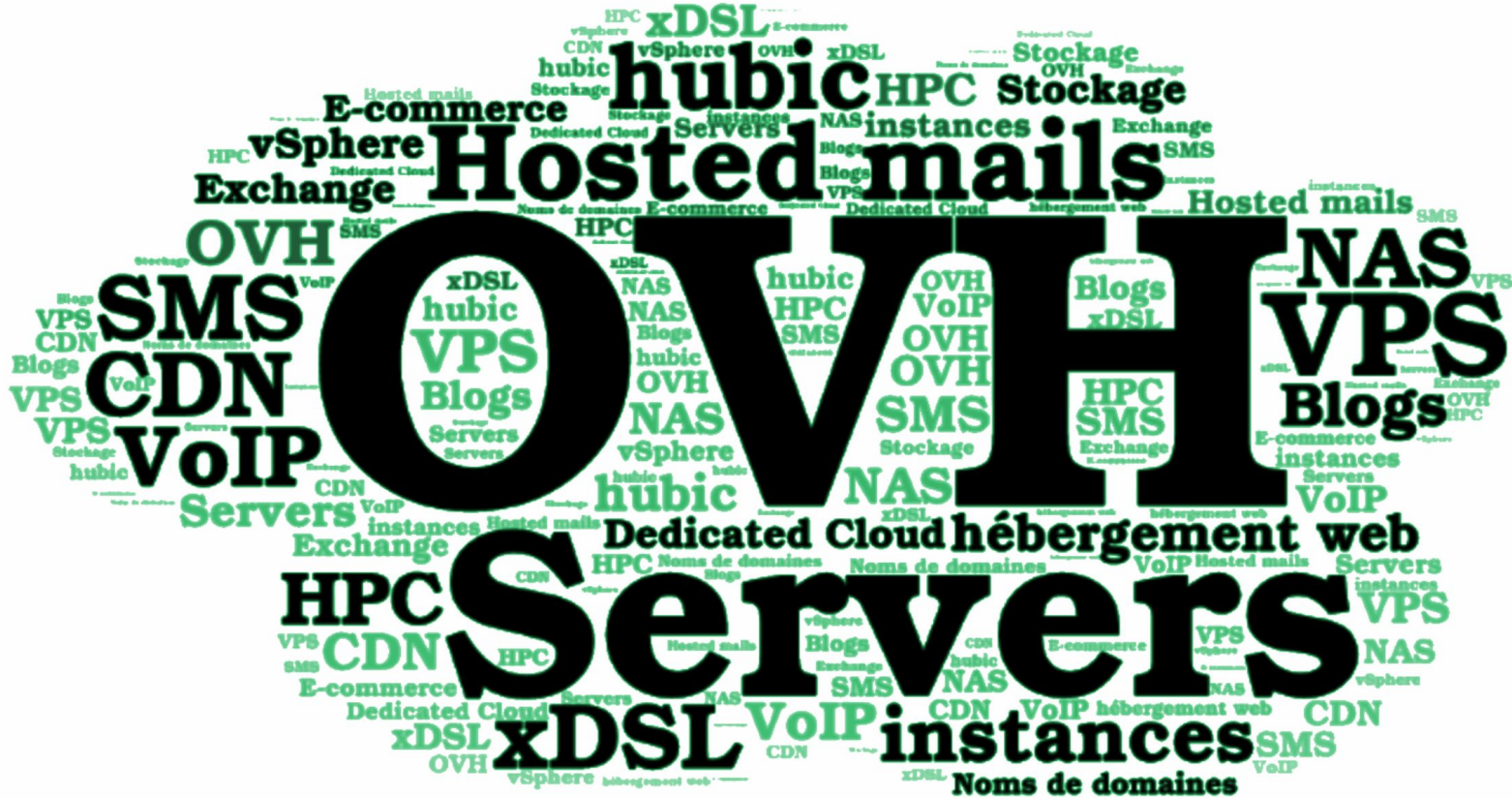
Metrics Data Platform

Metrics Data Platform

And why Warp 10?

Monitoring OVH

Metrics Data Platform

Some Metrics's metrics:

- 1.5M datapoints/s, 24h/7

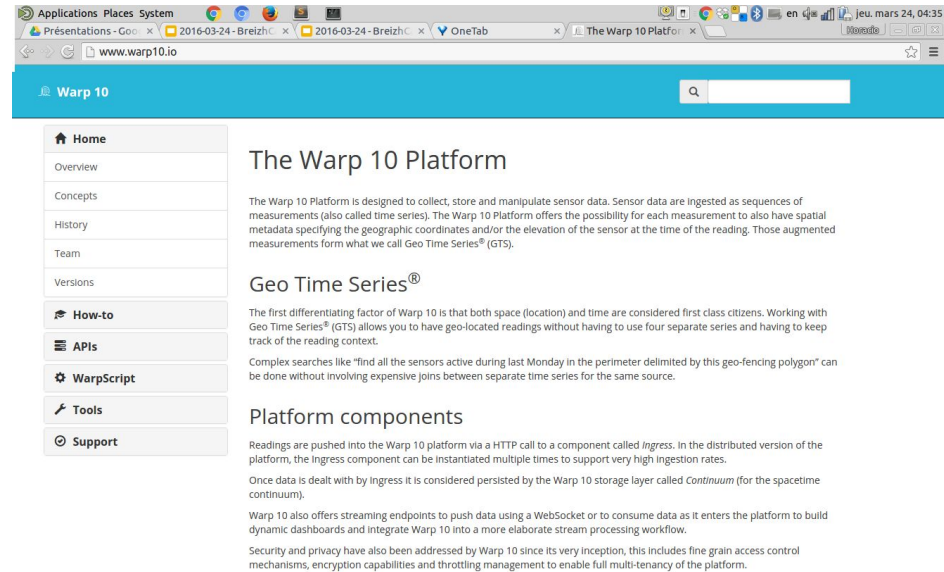- Peaks at ~10M datapoints/s

- 500M unique series

Warp 10

Open-source Time Series Database

More than a Time Series DB

Warp 10 is a software platform that

● Ingests and stores time series

● Manipulates and analyzes time series

Many time-series solutions

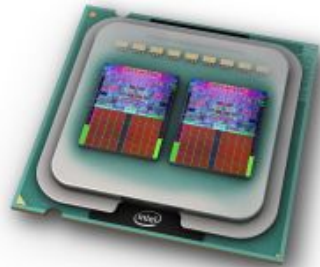# But they are only stores...

Fetching data is only the tip of the iceberg

Analysing the data

High level analysis must be done elsewhere

Algorithms are resource hungry

Your computer is not a datacenter

Manipulating Time Series

To be scalable, analysis must be done in the TSDB engine, not in user's computer

A true Time Series analysis toolbox

- ○ Hundreds of functions
- ○ Manipulation frameworks
- ○ Analysis workflow

Why not a simple REST API?

- One endpoint by function?
- How to chain a workflow analysis?

REST API not suitable for
complex manipulations

Why not a SQL dialect?

- How do you do a simple moving average in SQL?
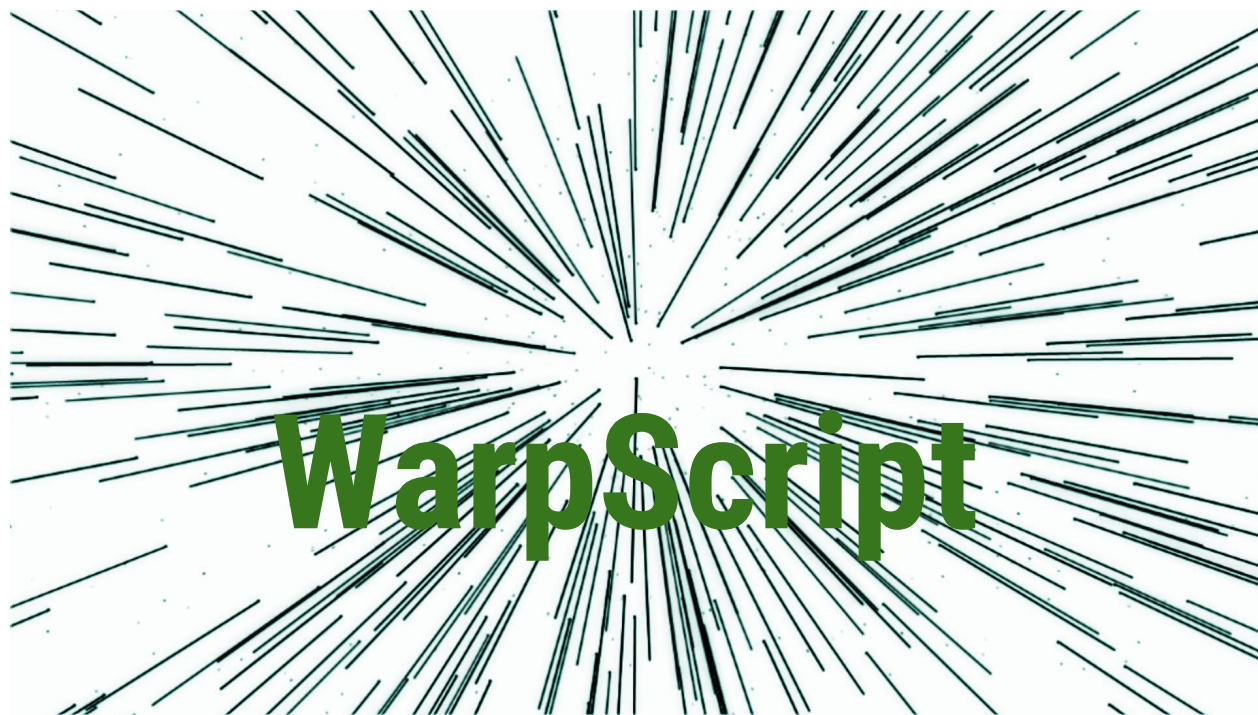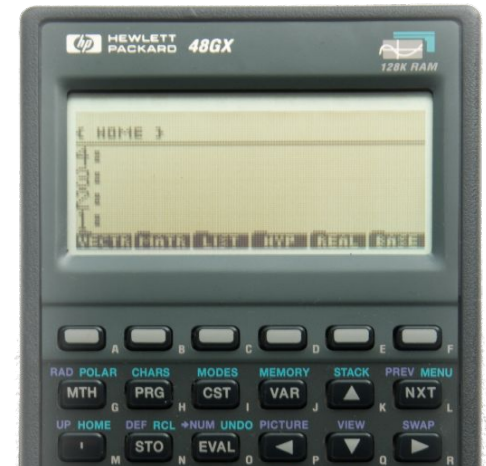- How do you geo-time fencing in SQL?

SQL is not adapted to TS analysis!
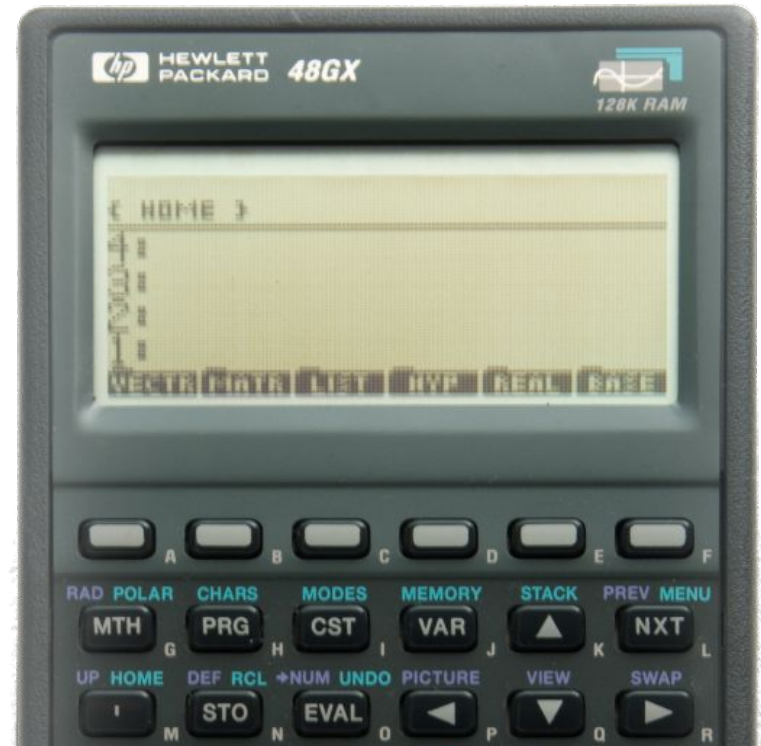
Solution: a Time Series manipulation language

WarpScript

WarpScript

Time Series manipulation language

A stack based language

Input	2	3	add	11	mul	1	add
Stack	2	3 2	5	11 5	55	1 55	56

Basic operations

```
// This is a commentary
'a'    // string value
true  // boolean value
42    // long value
3.14159 // double value
+     // operations

20 22 + // several items in one line
```

Five frameworks

- BUCKETIZE
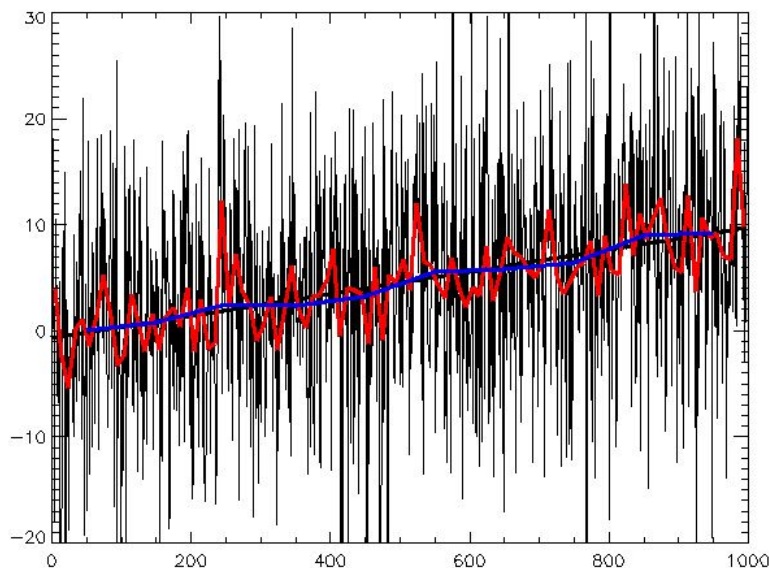- MAP
- REDUCE
- FILTER
- APPLY

More than 800 functions

Trigonometry

Macros

Data & Time

List & Maps

Logic Structures

Strings

Maths & statistics

Loop Structures

Time series functions

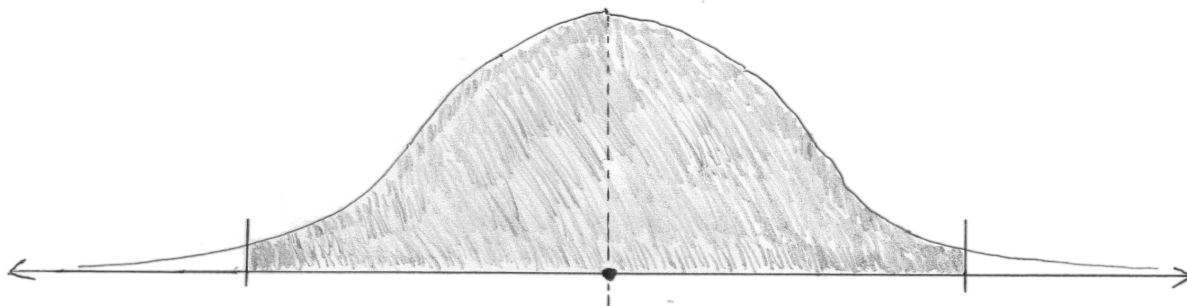Pauline & compagnie

TIMECLIP

TIMESPLIT

SHRINK

MERGE

...

Time series functions

MUSIGMA

NORMALIZE

NSUMSUMQ

STANDARDISE

ZSCORE

Geo-Time Series functions

Geo mapping (WKT)

Horizontal & vertical speed

Horizontal & vertical distance

Haversine

...

Quantum IDE

Warp 10 platform and tools

Warp 10

And this is only the surface

Possibilities are endless

"Once you stop learning, you start dying" ~ Albert Einstein ~
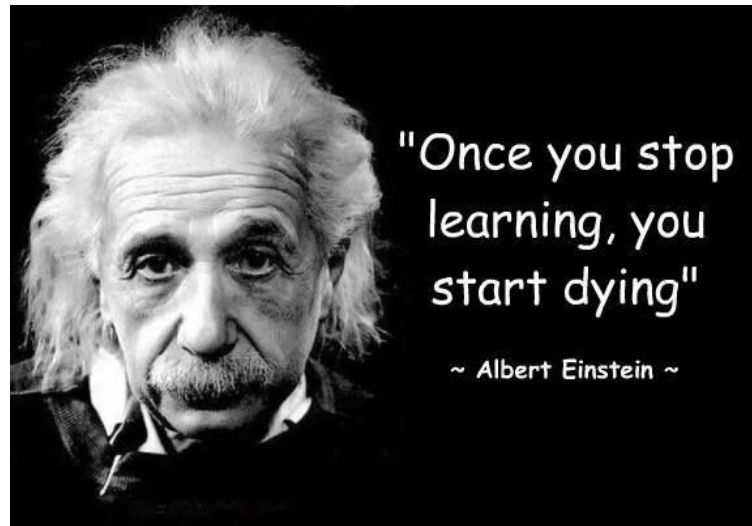
Think differently
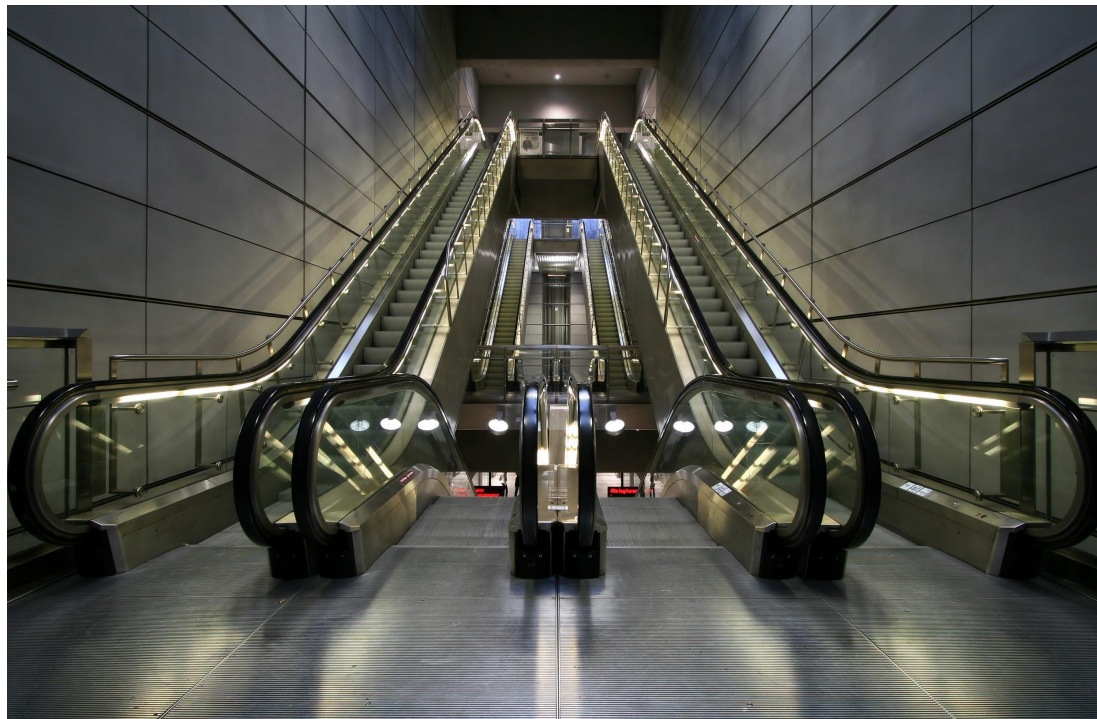
Time Series are everywhere

Thank you!

Rediscover the known Universe with NASA datasets
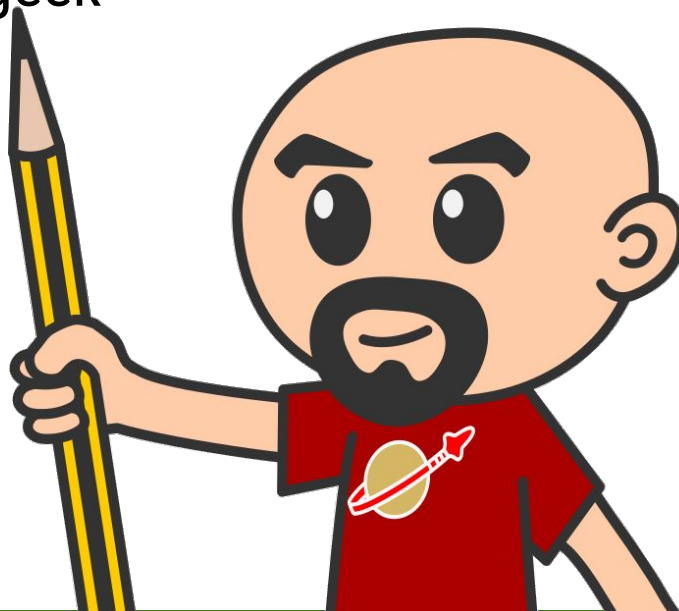
Horacio Gonzalez

@LostInBrittany

Horacio Gonzalez

@LostInBrittany

Spaniard lost in Brittany, developer,
dreamer and all-around geek

OVH
Team DevRel

HelloExoWorld

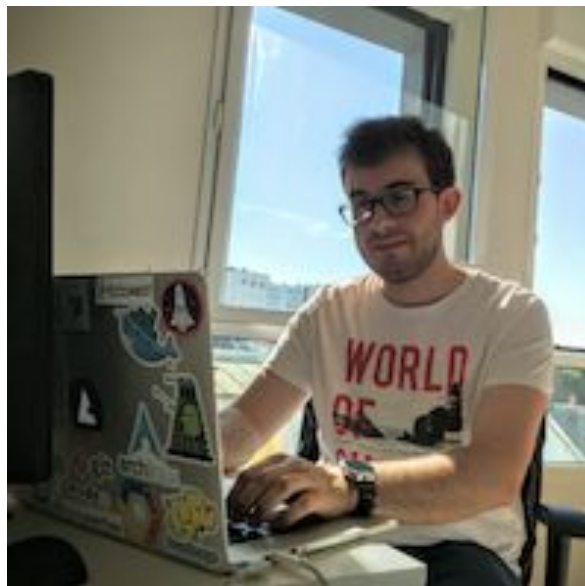## Looking for exoplanets in NASA datasets

HelloExoWorld

Once upon a time…
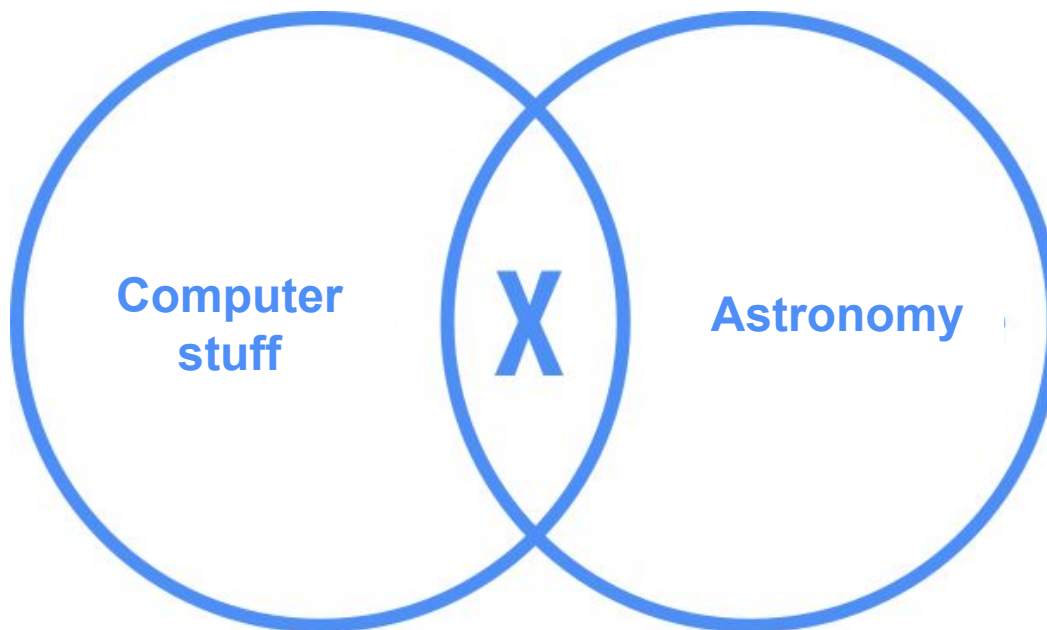
An amateur astronomer

Pierre Zemb, DevOps OVH

What not to do if you love astronomy

Live in Brest

Looking for solutions

Computer stuff   X   Astronomy

Mixing passions

Google is your friend...

Let's find a project

Exoplanets?
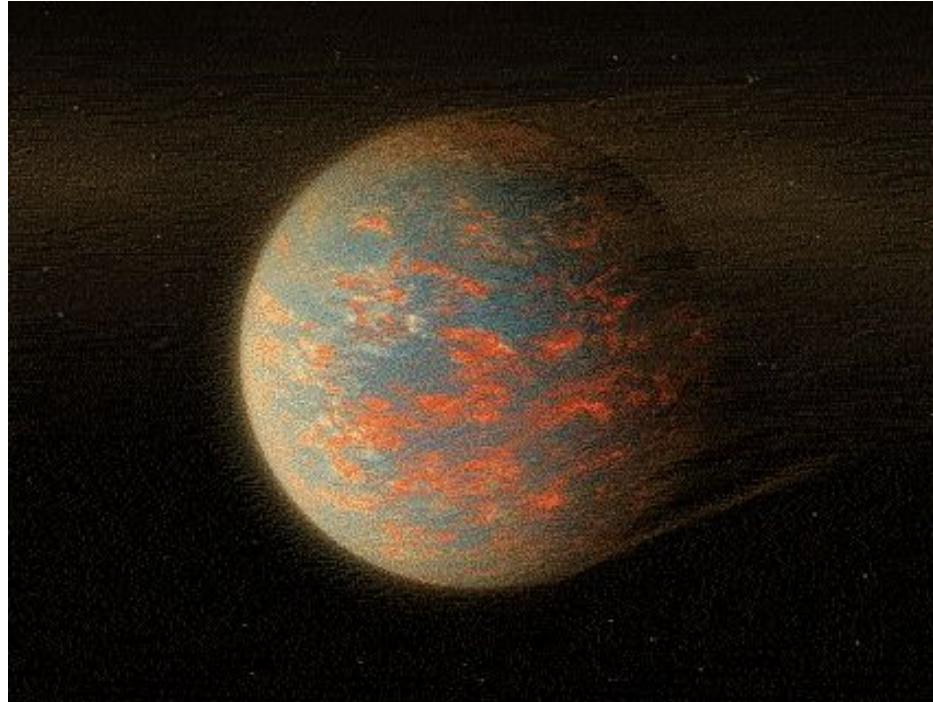
Planets orbiting stars far away

How do we find them?

Cumulative Detections Per Year

25 Jan 2018
exoplanetarchive.ipac.caltech.edu

Radial Velocity
Transits
Microlensing
Imaging
Timing Variations
Orbital Brightness
Modulation
Astrometry
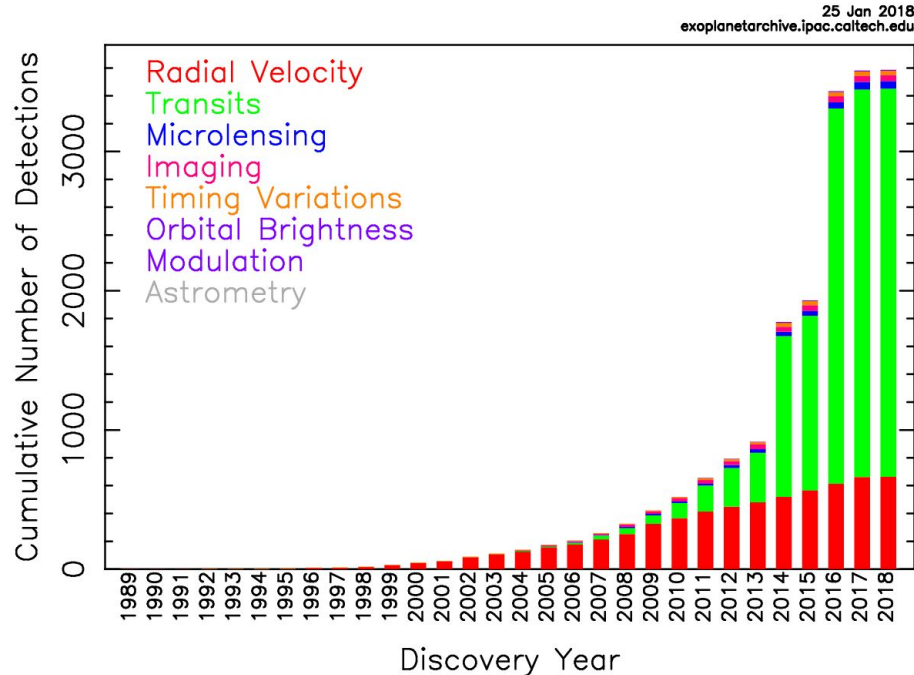
The transit method seems the best

The transit method

Apparent Brightness of Star

How do we look for transits?
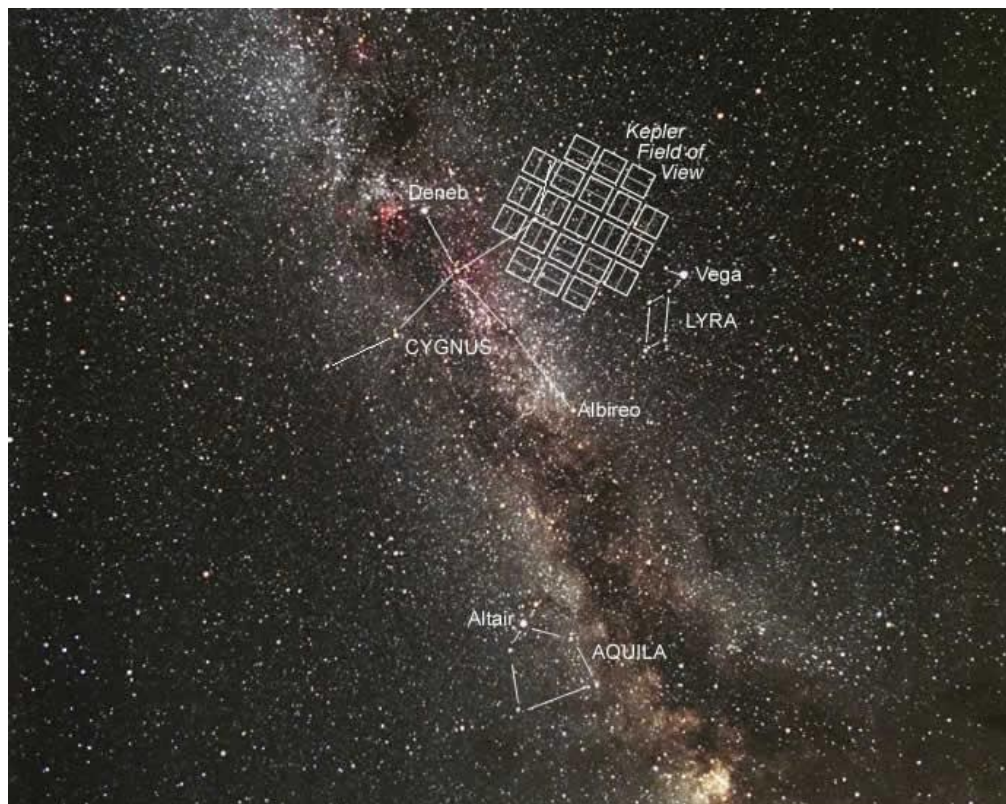

Image credits : NASA
Kepler

Watching the sky

And what kind of data we get?

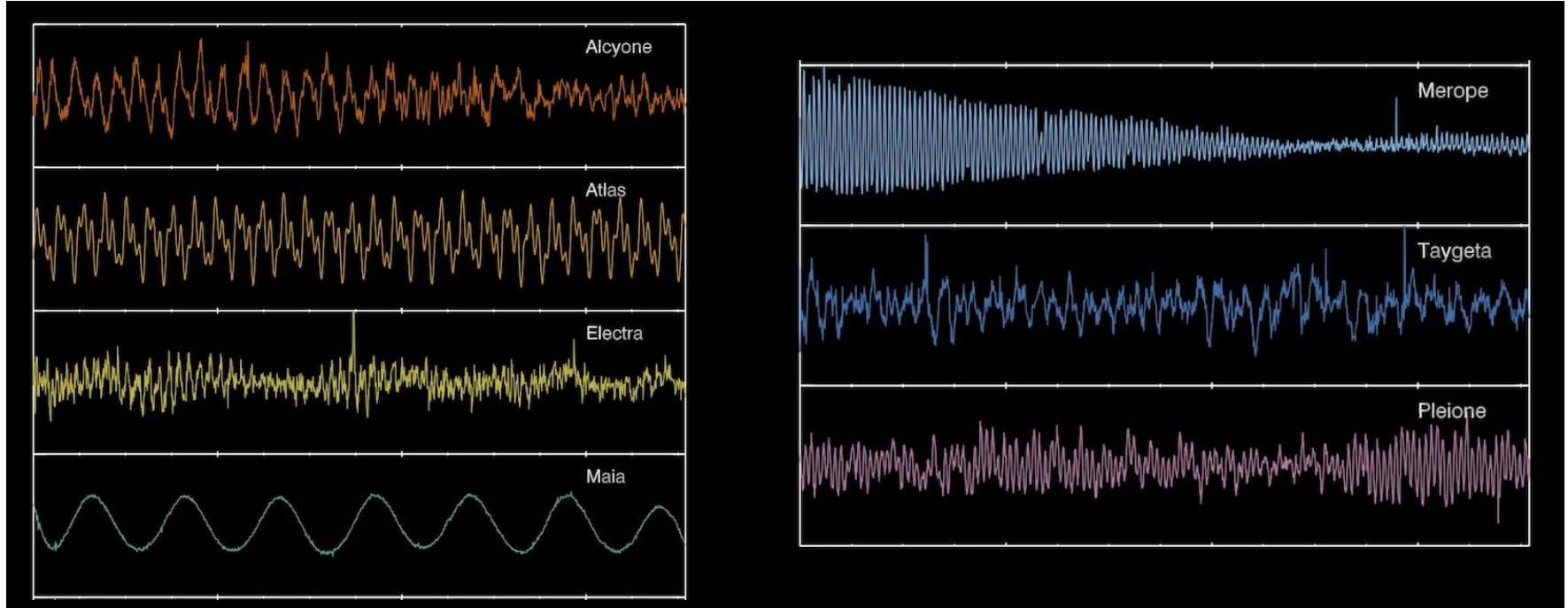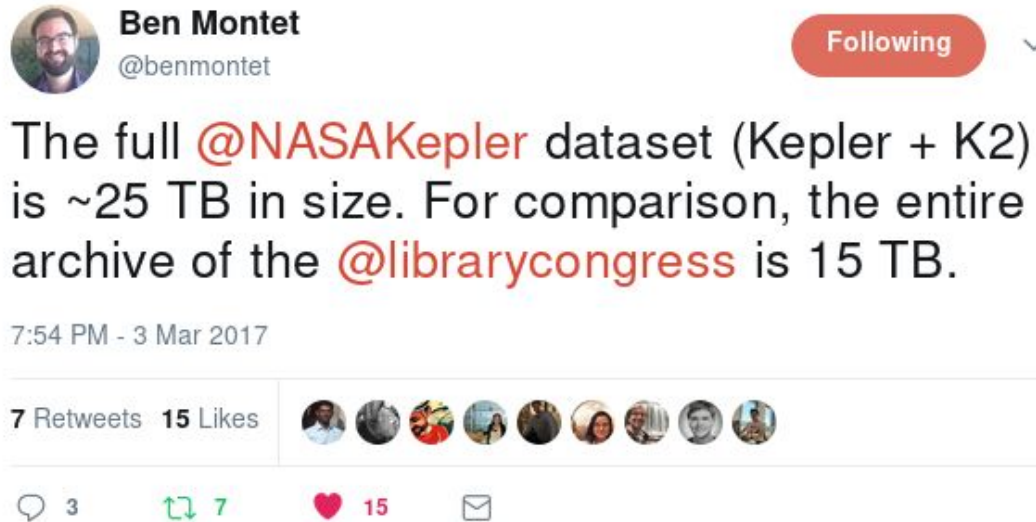Pleia...   ...mmon

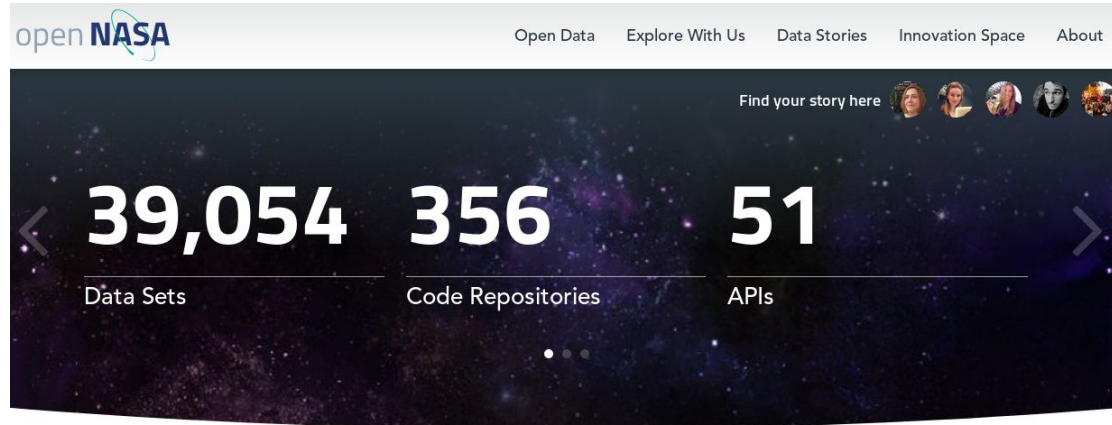# Well, that's the problem

Seven stars, seven different profiles

Kinda big data

Ben Montet
@benmontet
Following

The full @NASAKepler dataset (Kepler + K2) is ~25 TB in size. For comparison, the entire archive of the @librarycongress is 15 TB.

7:54 PM - 3 Mar 2017

7 Retweets 15 Likes

3   7   15

Over 40 million light curves

Big AND open data

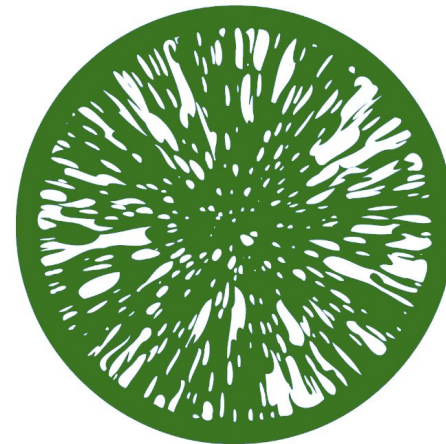
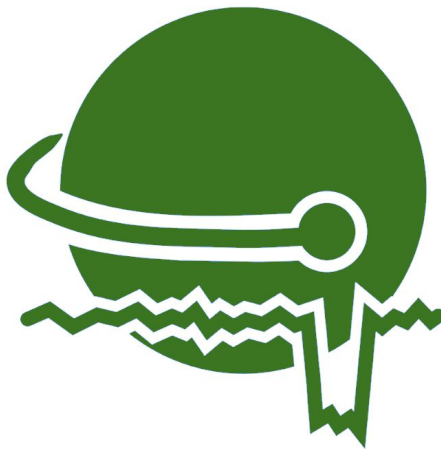## Lots of datasets in #opendata

And we can help with that!

Let's use our tools to analyse the data

A match made in heaven

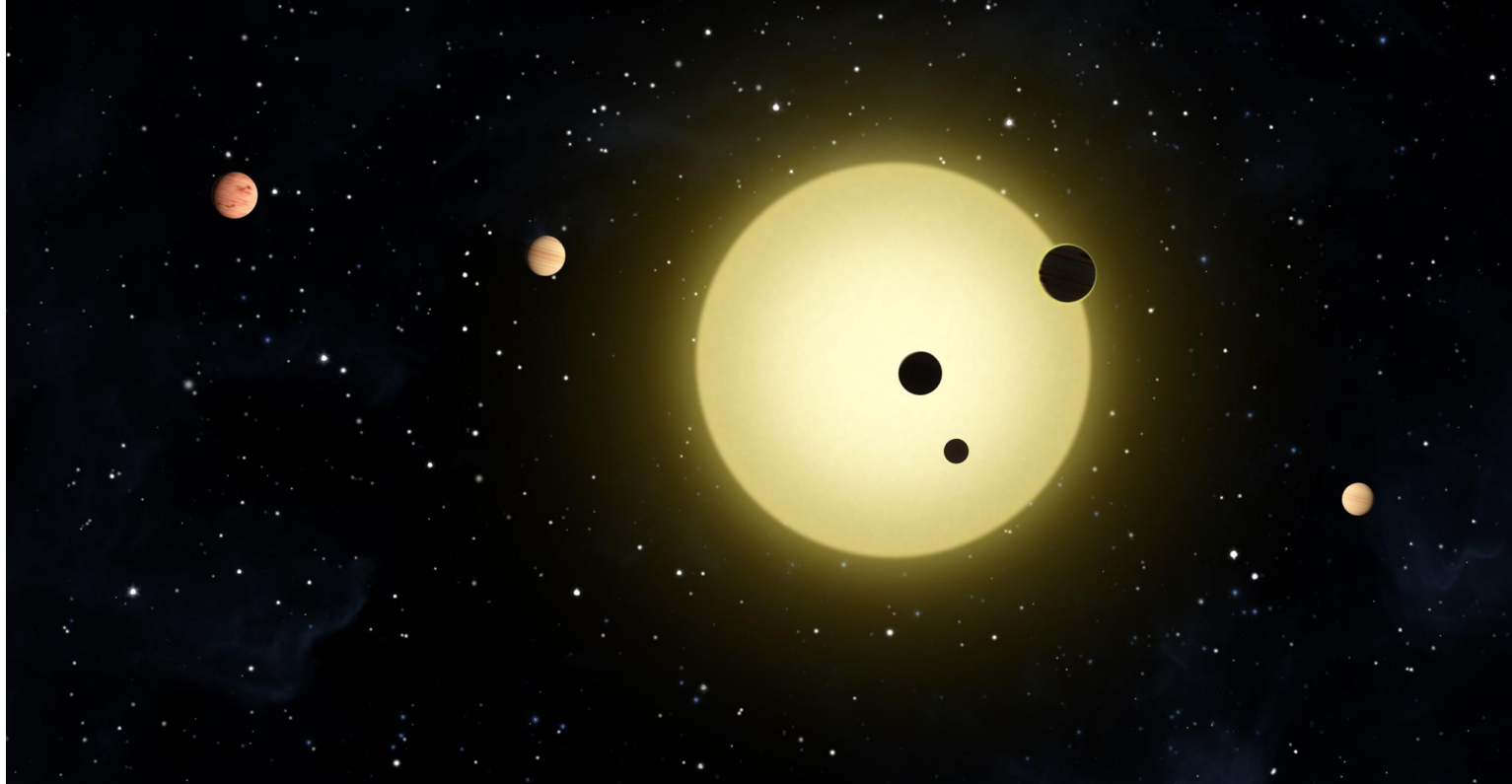## Warp 10, OVH Metrics and HelloExoWorld
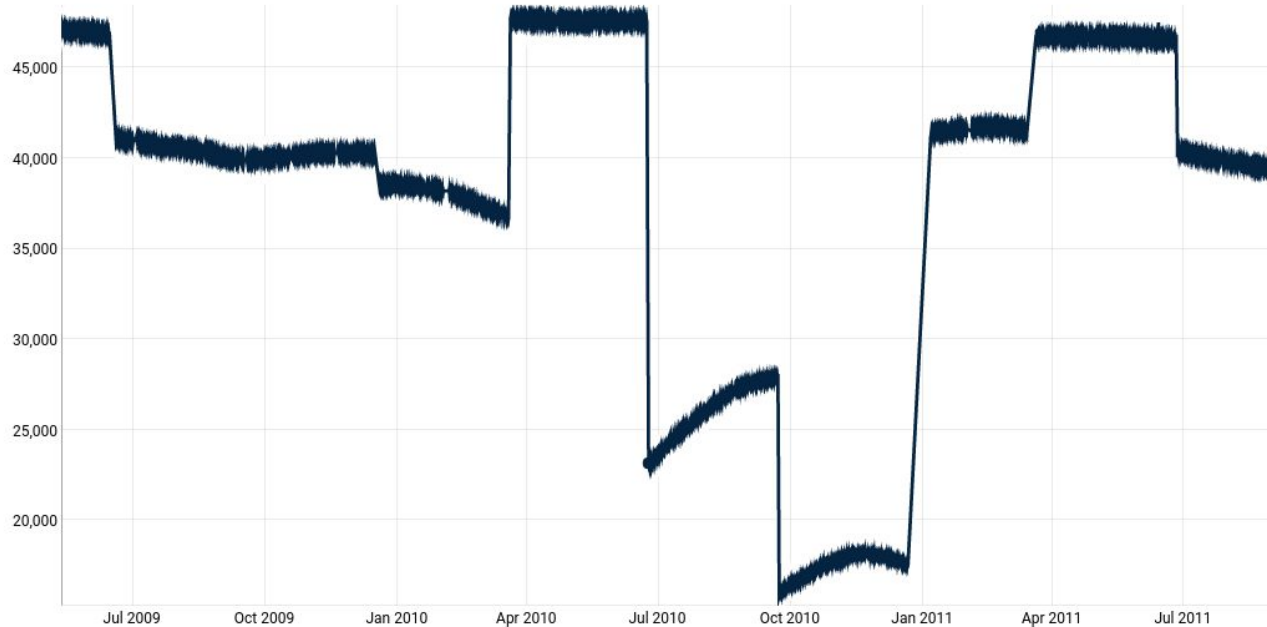
What we have done

- Downloaded and parsed 40 millions of FITS files
- Pushed it to OVH Metrics
- Select a cool subset as training set
- Verified we could find the same planets as NASA

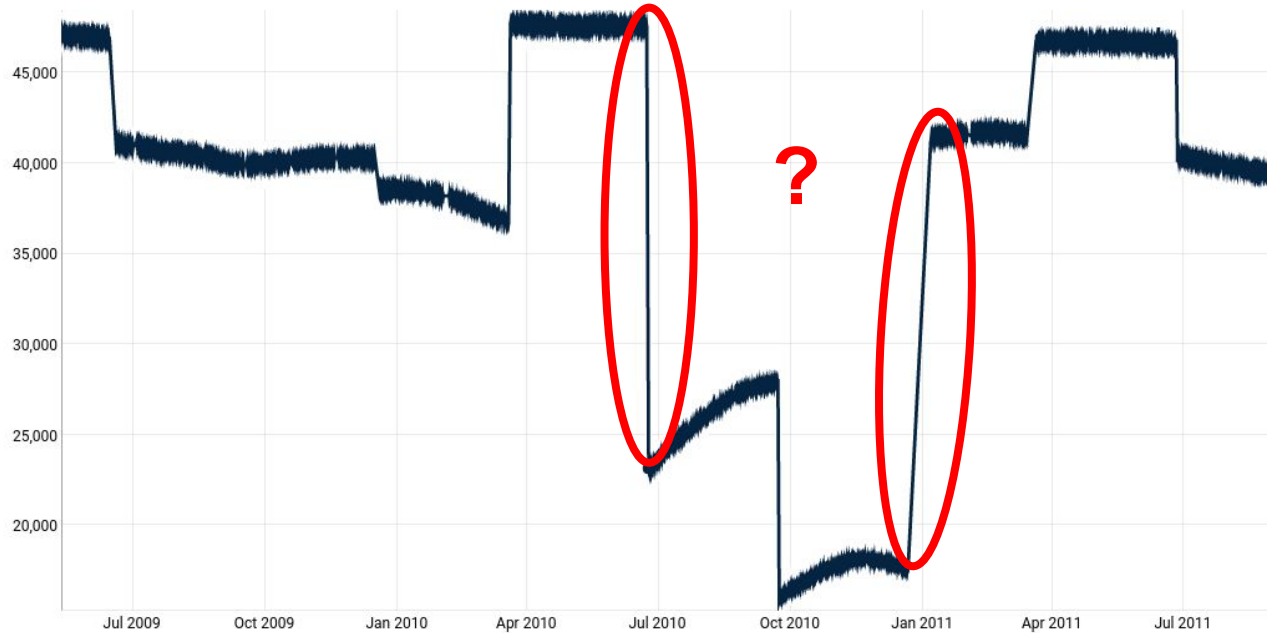# Choosing a star: Kepler 11

Looking at the raw signal…

SAP_FLUX:
The flux in units of electrons per second contained in the optimal aperture pixels collected by the spacecraft.
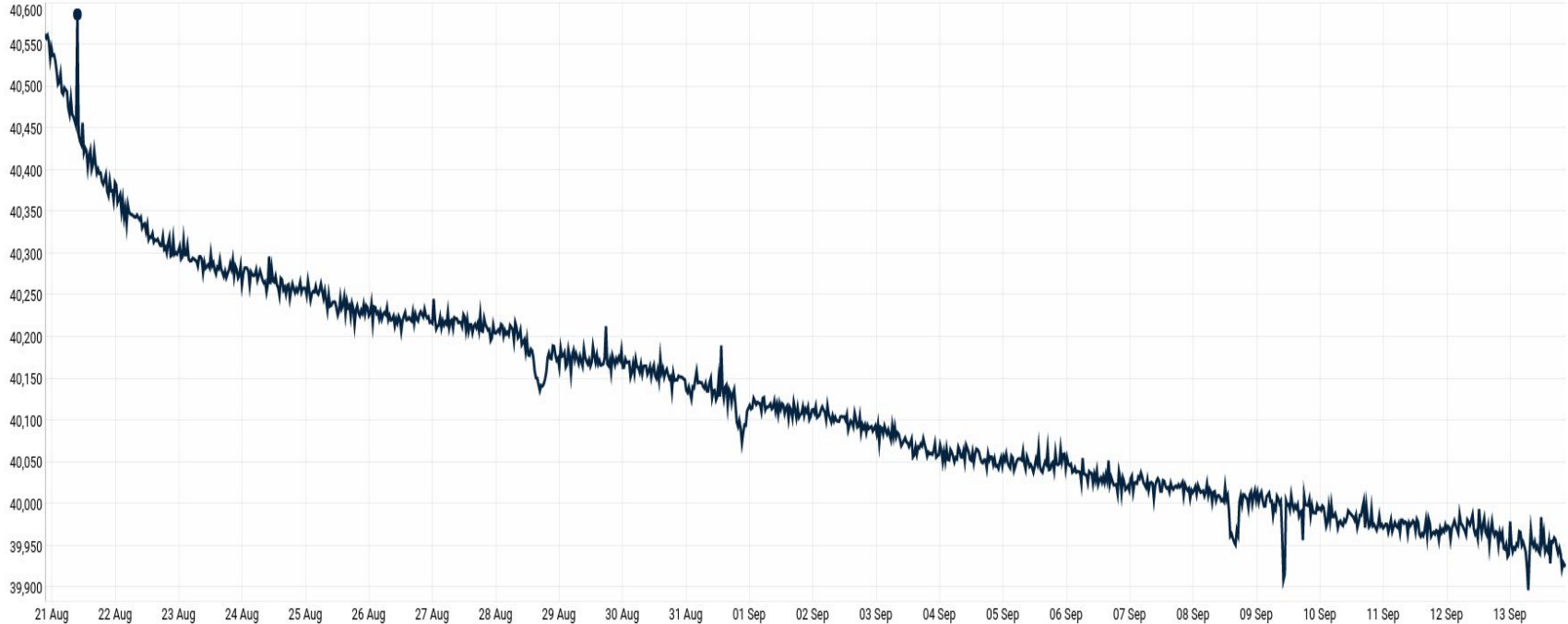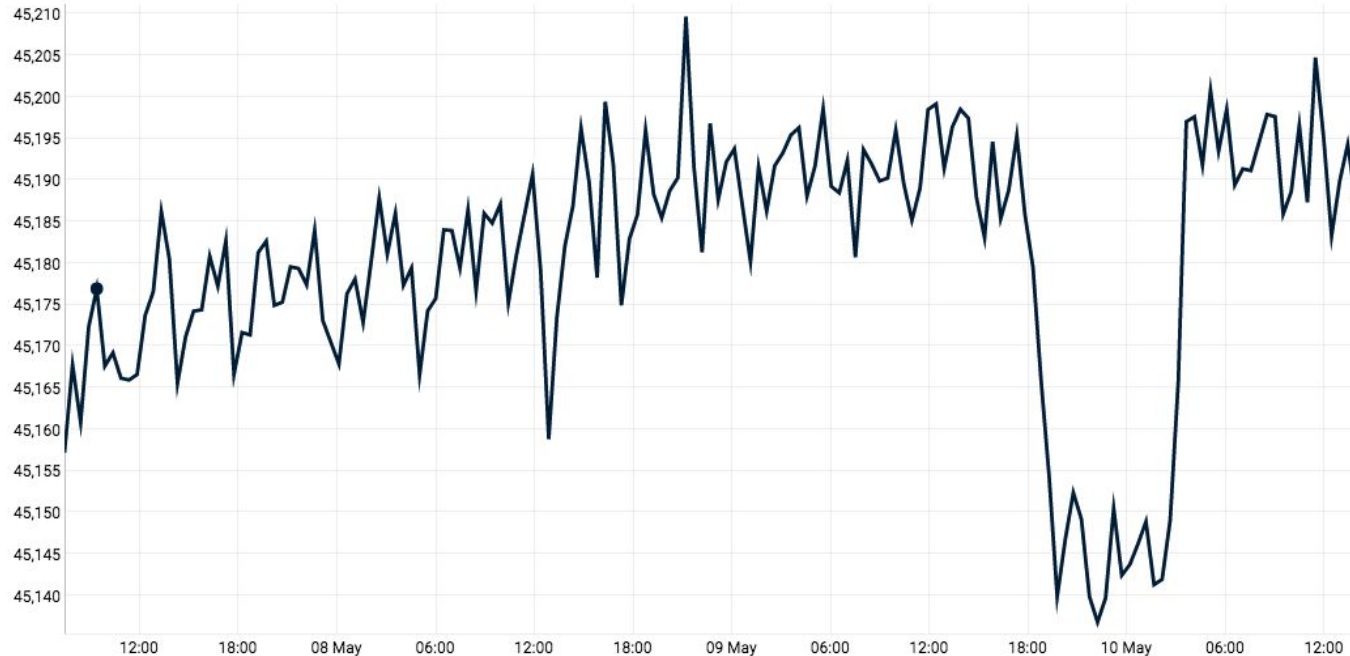
Looking at the raw signal…

SAP_FLUX:
The flux in units of electrons per second contained in the optimal aperture pixels collected by the spacecraft.

Looking at one record
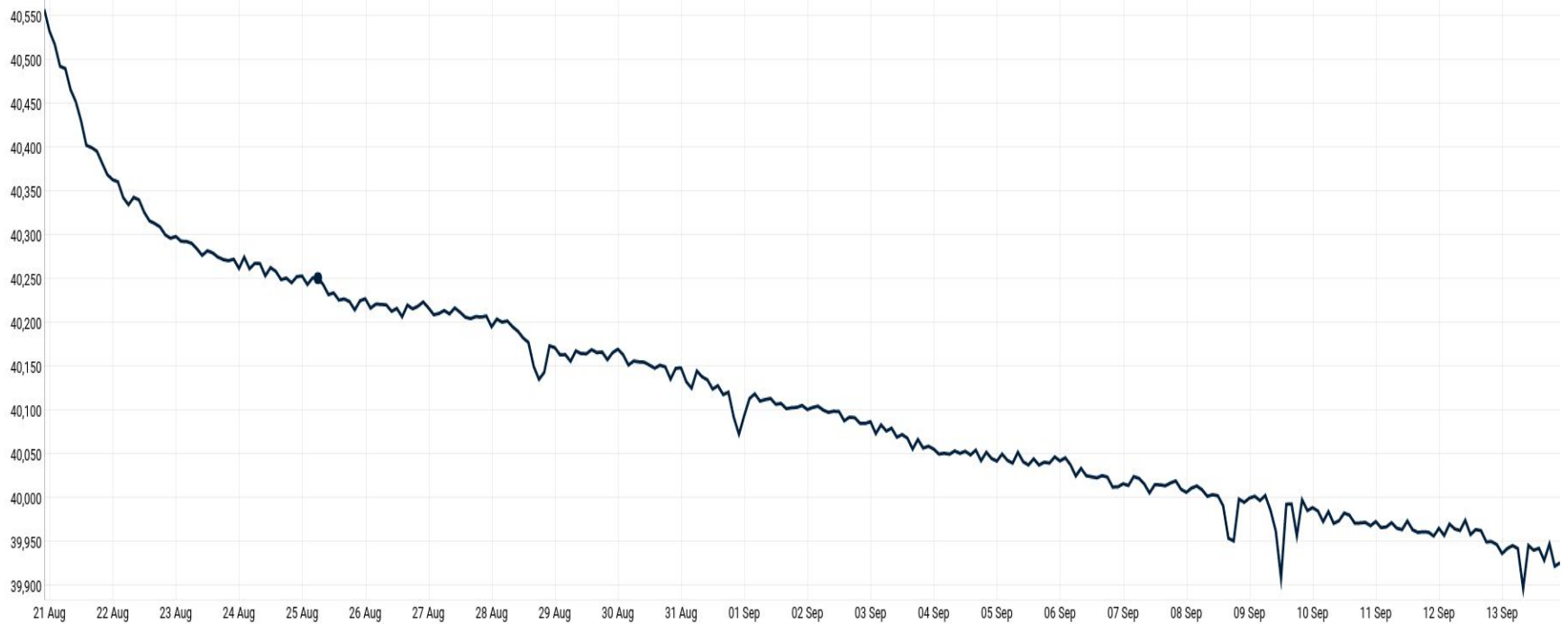

Perturbations in dirty signals
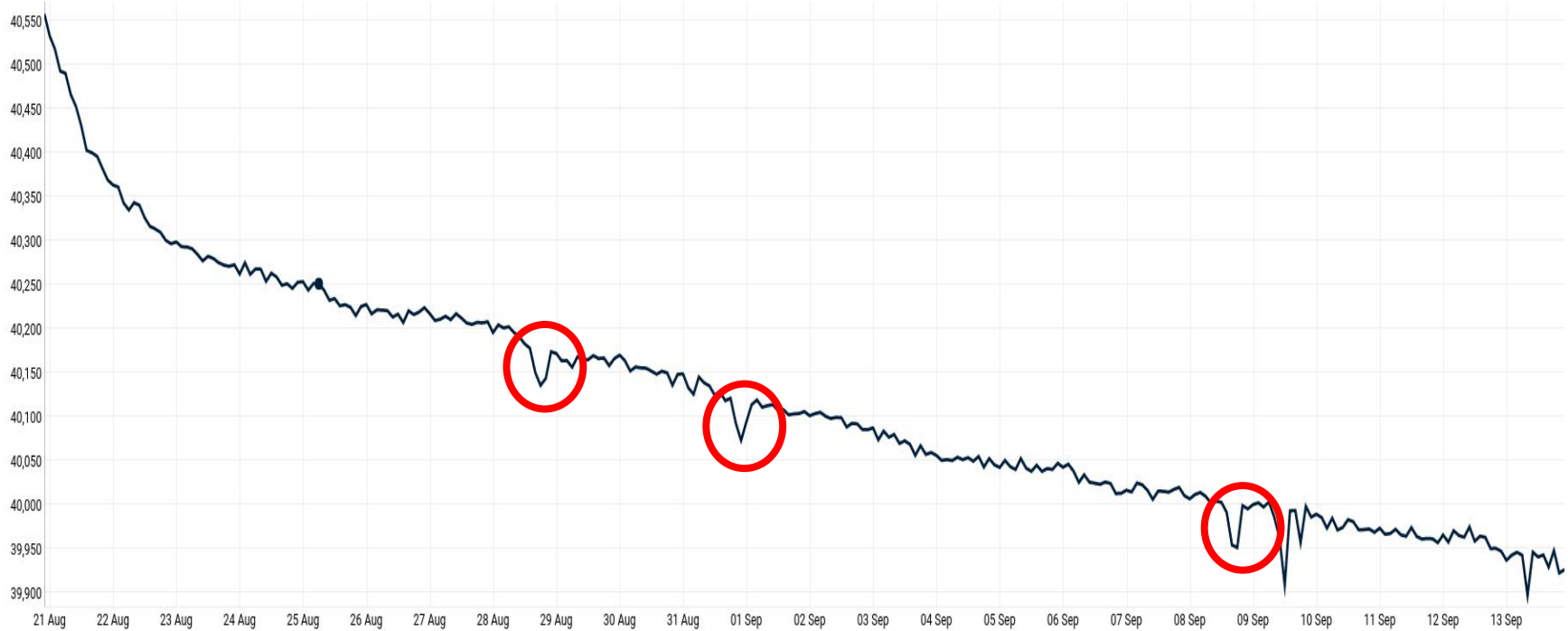
Transits are tiny

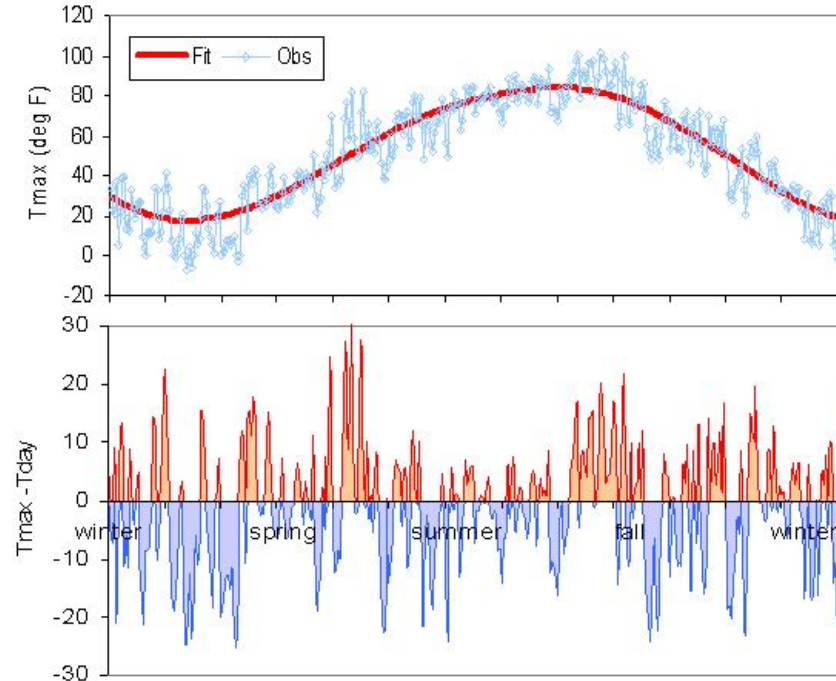## ~40 electrons per second

First step: downsampling

First step: downsampling

You can see the transit candidates...
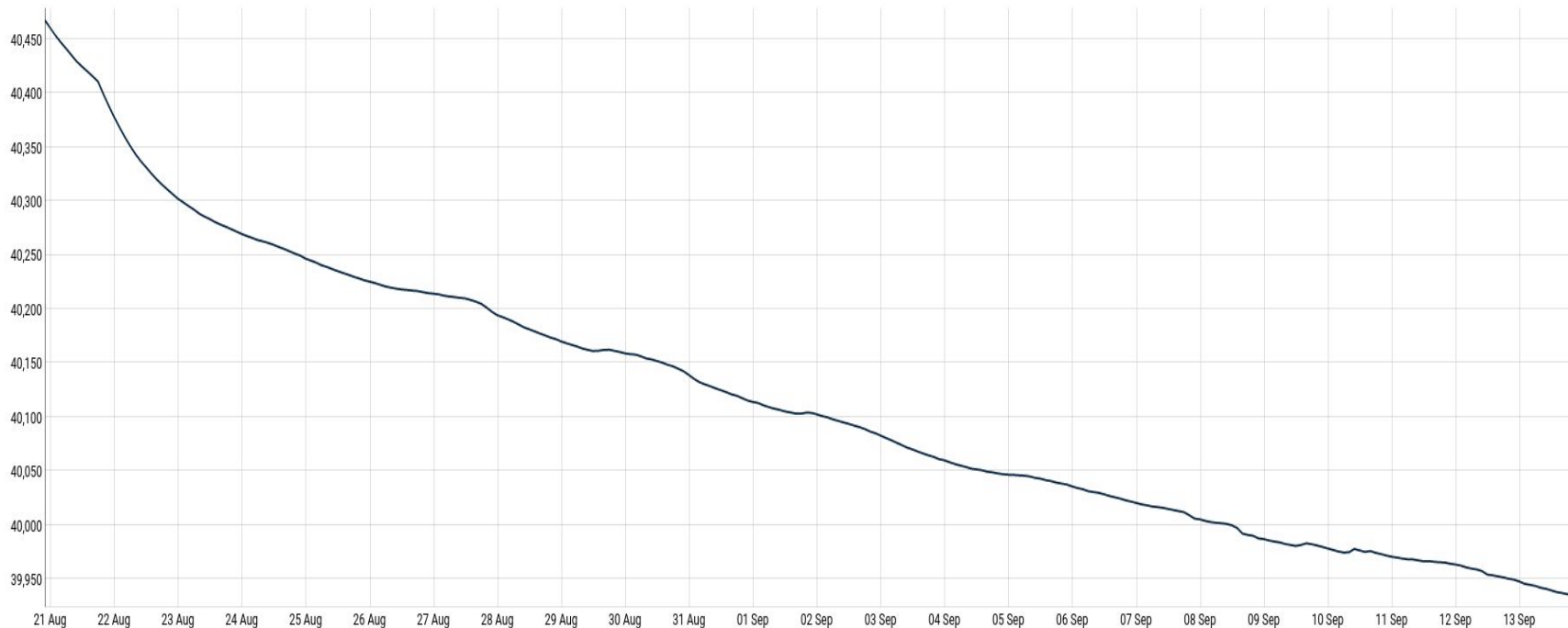but how can we teach the computer to see them?
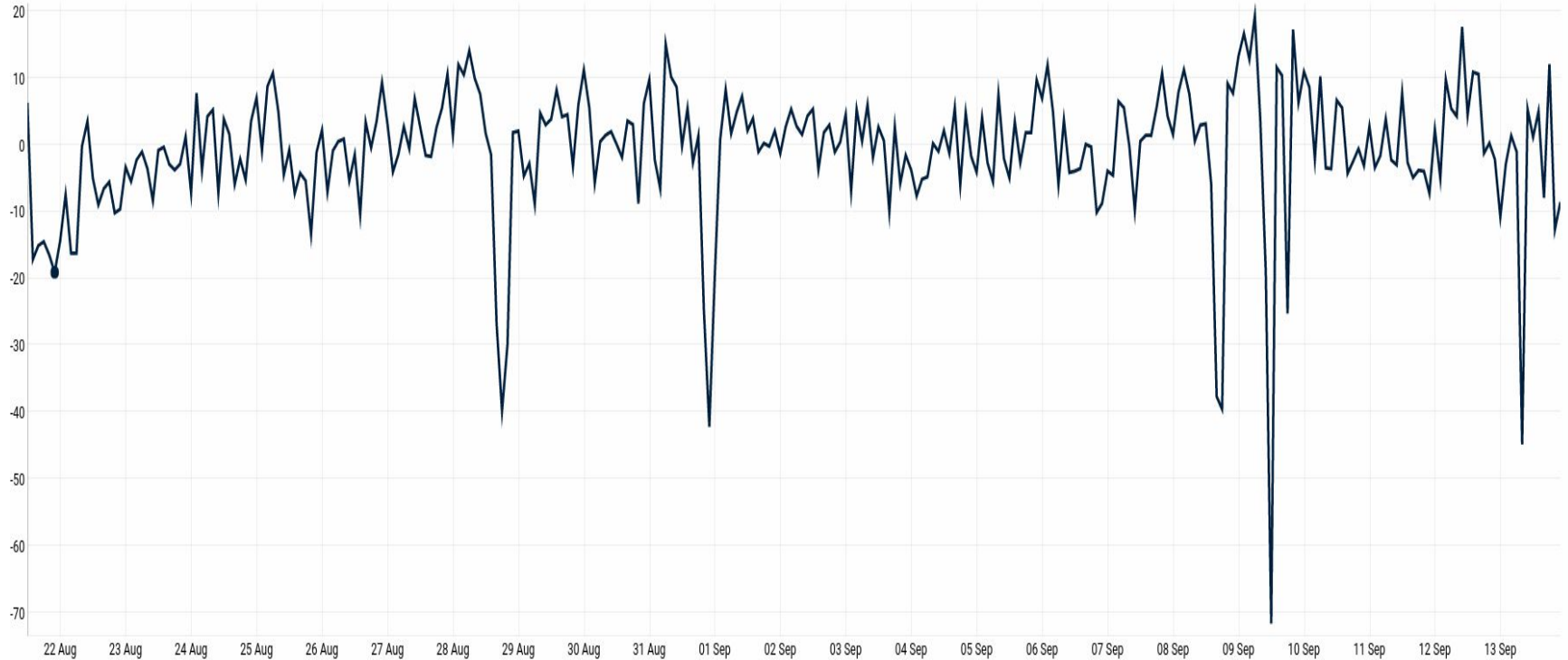
If you ♥ signal processing

High pass filter

Poor person's high pass filter

Using the trend
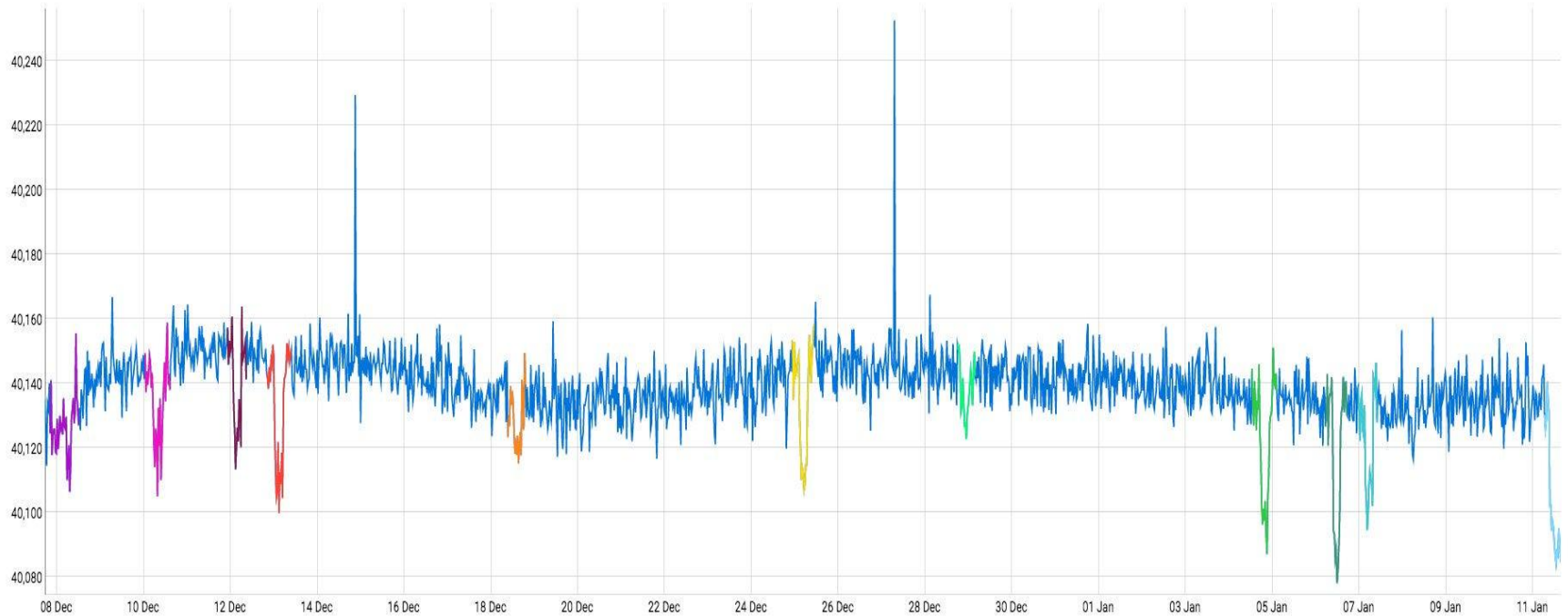
Signal - Trend

Now you can see them well

After some tuning

We have our transit candidates

What's next?

Where do we go from here?

Only the beginning

New import method

Better detection

Deep learning

Explorer

satellite/star location

Yours?

A growing team

And you!

https://xkcd.com/1371/

Join us!
https://helloexo.world

Thank you!