



influx/days

Data Ingestion

Emanuele Della Valle

Prof. @ Politecnico di Milano & Partner
@ Quantia Consulting

Marco Balduini

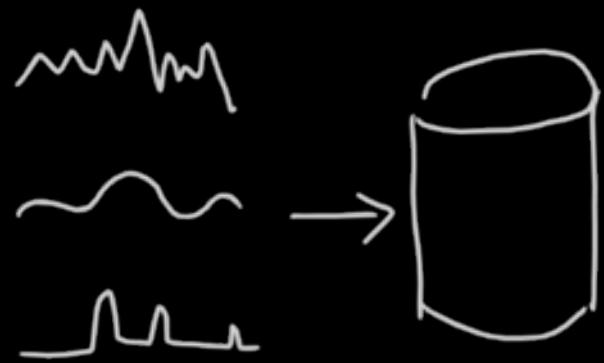
Founder & CEO @ Quantia Consulting



Introduction Data Ingestion

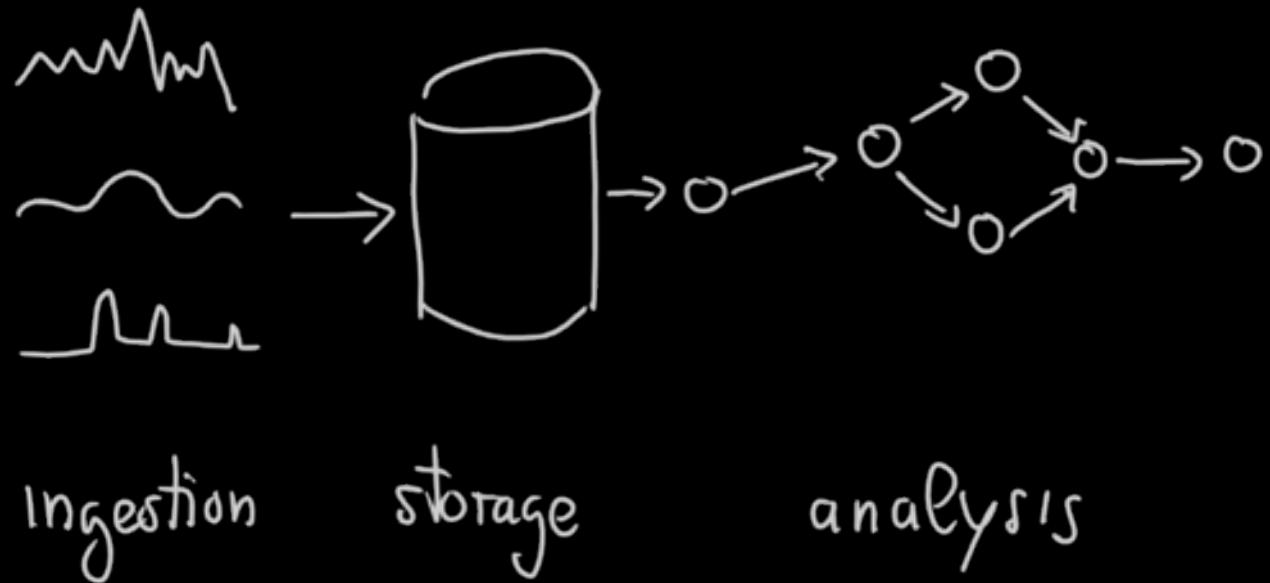


Data Lifecycle

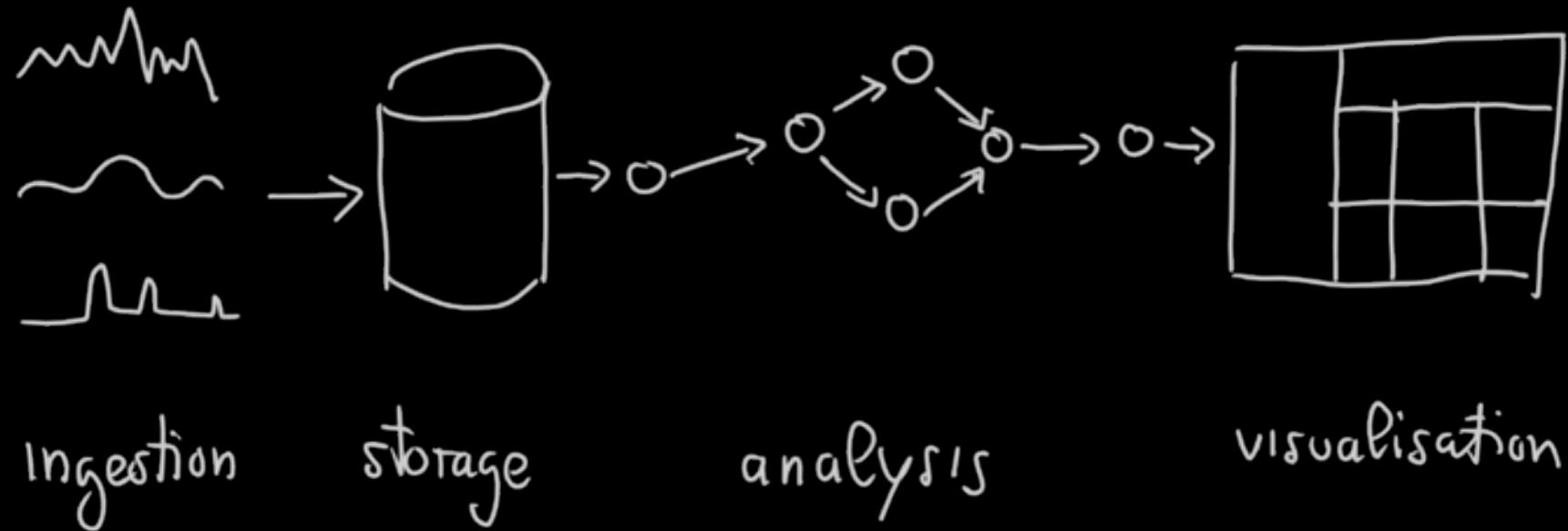


ingestion storage

Data Lifecycle



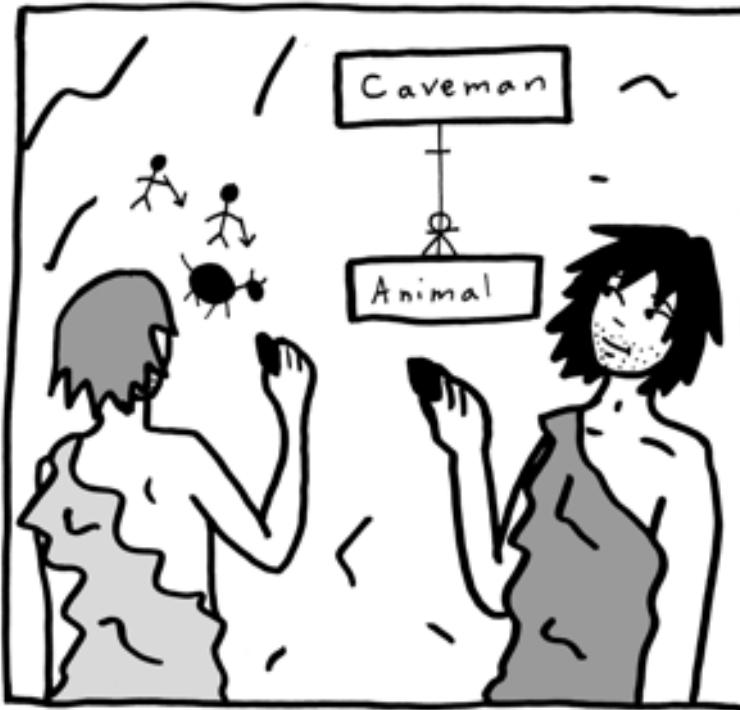
Data Lifecycle



How is data ingested?

Data Ingestion

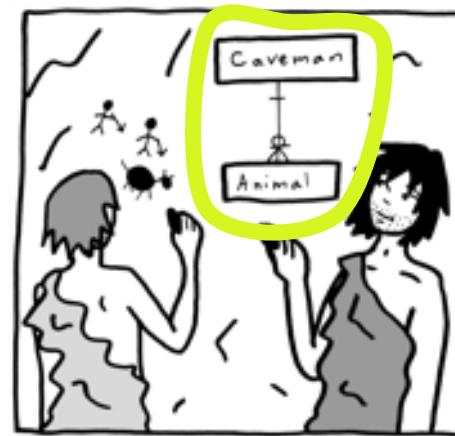
Conceptual vs. logical vs. physical views



[Src: <https://www.oreilly.com/library/view/data-modeling-made/9781935504481/>]

How is data ingested?

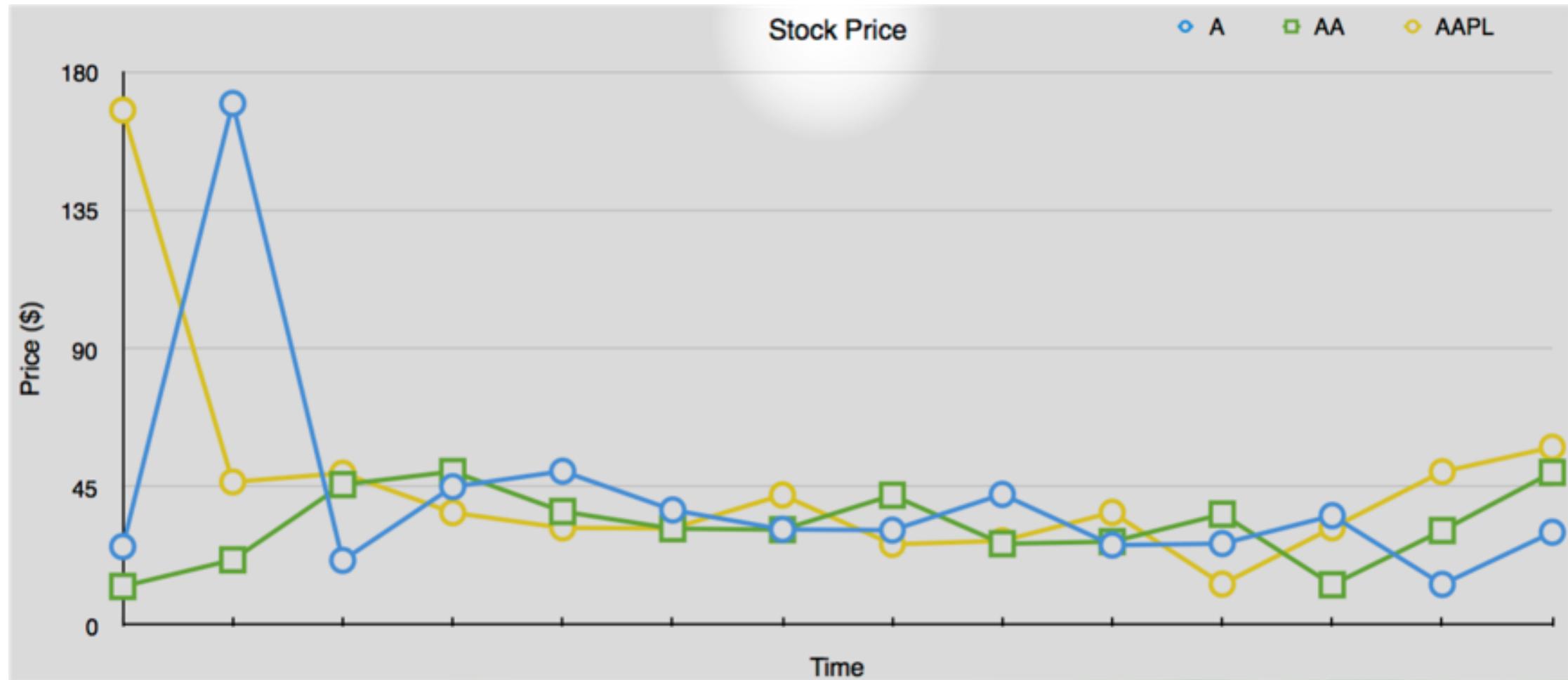
Conceptual View



Let's start from the anatomy of a Time-Series Line Graph



The type of measurement is the title of the Line Graph



Data Model

Measurement

- The name of the measurement used as high level grouping of data

Time stamps are on the X-Axis



Data Model

Measurement

- The name of the measurement used as high level grouping of data

Timestamp

- Time of the data

Data is on the Y-Axis



Data Model

Measurement

- The name of the measurement used as high level grouping of data

Field set

- Actual data

Timestamp

- Time of the data

The Legend distinguishes the three time series in the graph



Data Model

Measurement

- The name of the measurement used as high level grouping of data

Tag set

- Other lower level grouping criteria of data

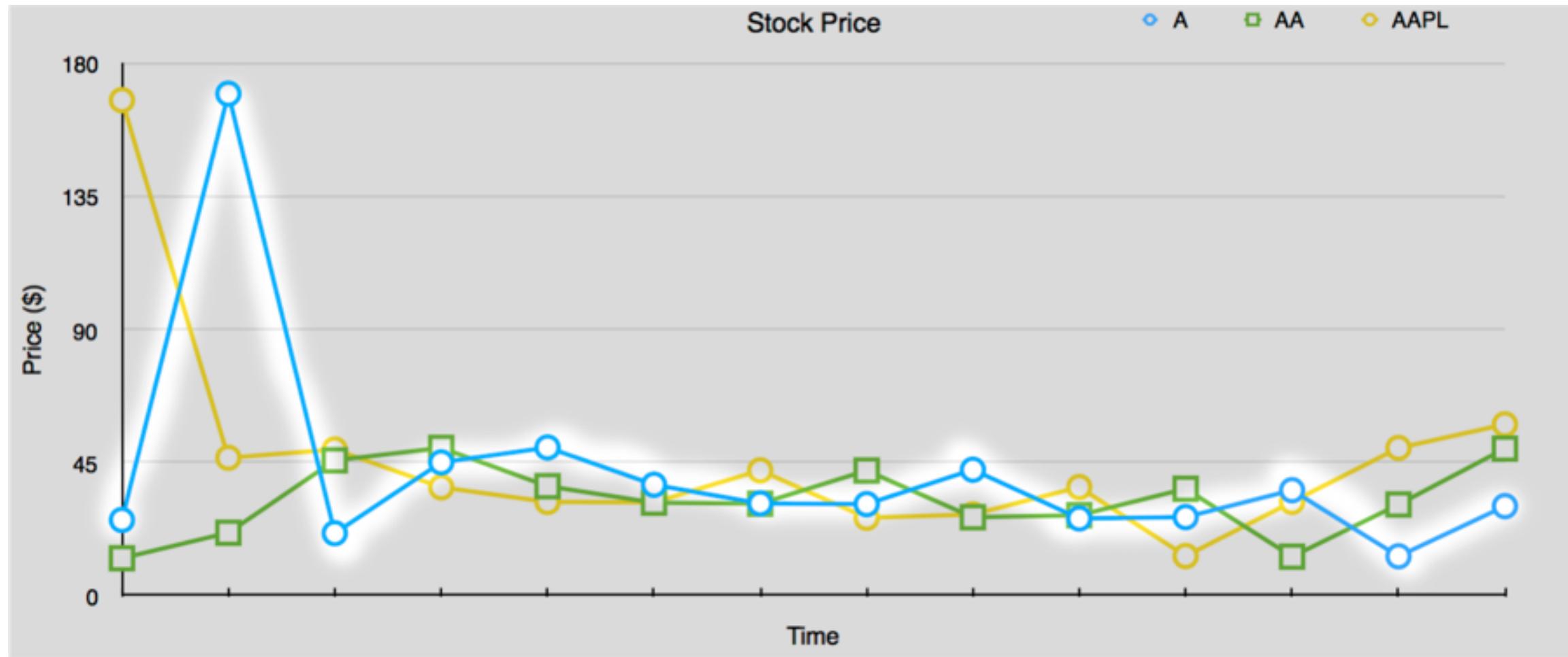
Field set

- Actual data

Timestamp

- Time of the data

A series in the graph thus is ...



Data Model

Measurement

- The name of the measurement used as high level grouping of data

Tag set

- Other lower level grouping criteria of data

Field set

- Actual data

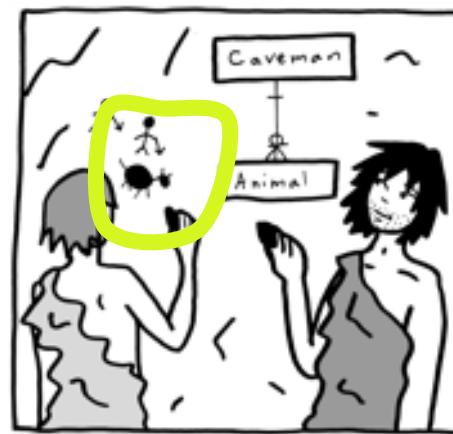
Timestamp

- Time of the data

Series

- Data points in time order grouped by measurements

How is data ingested? Logical View



Data Model

Measurement

- A **name** to group data at high level

Tag set

- A set of **key-value pairs** to group data at low level

Field set

- A set of **key-value pairs** to represent data

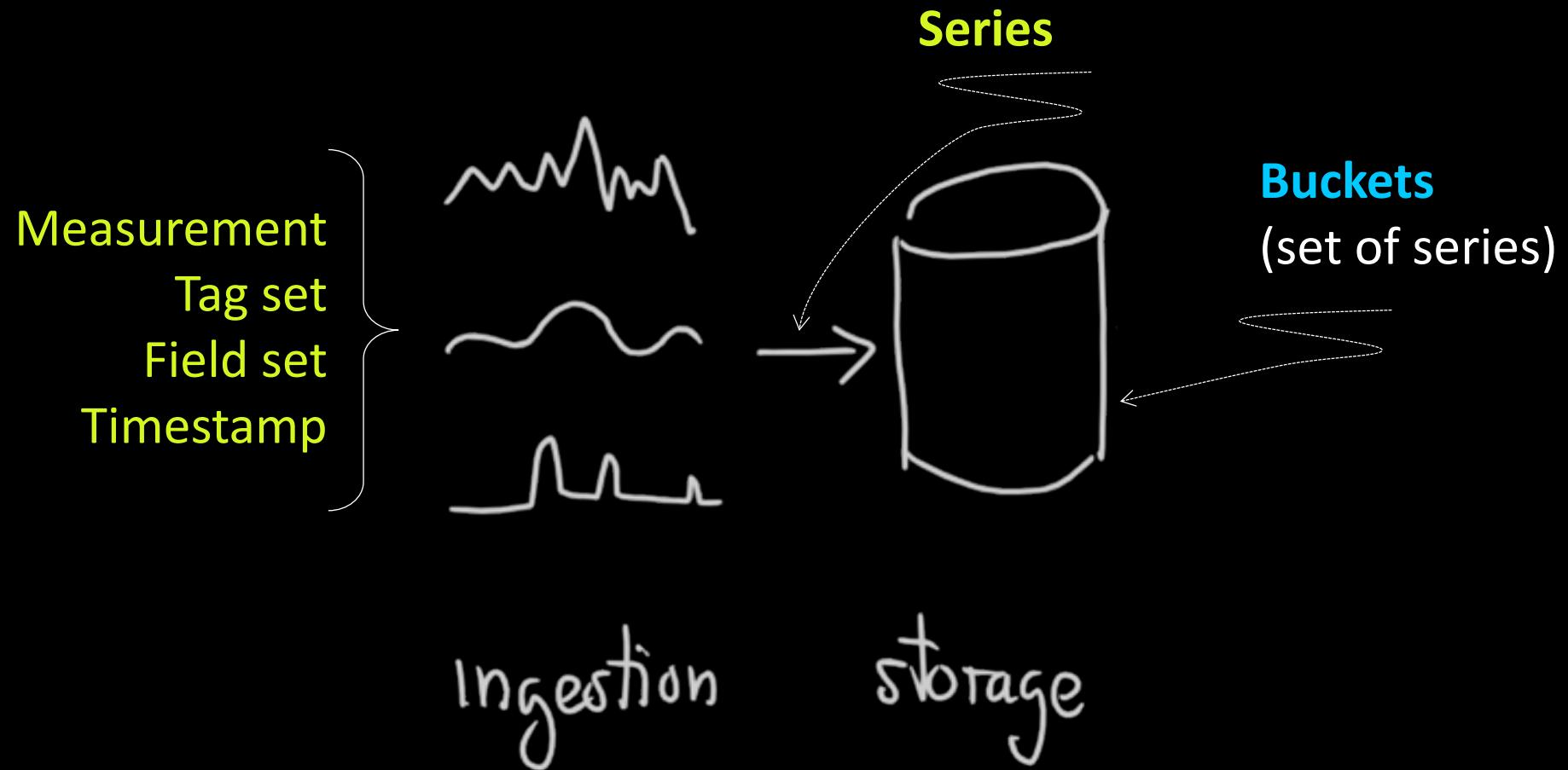
Timestamp

- **Time of the data with nanosecond precision**

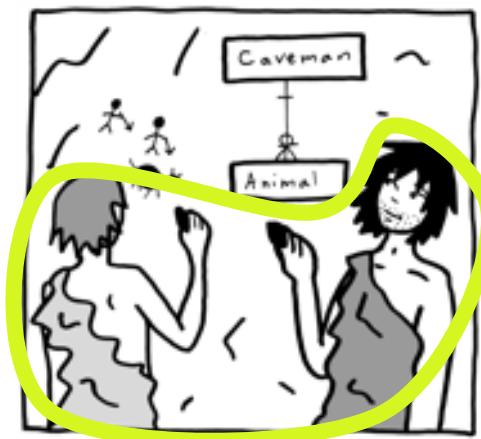
Series

- A unique combination of measurement+tags

Data model vs ingestion & storage



How is data ingested? Physical View



Line protocol

measurement,tags fields timestamp

whitespace I whitespace II

tags (a.k.a. tag set):

- syntax: key-value pair separated by a comma “,”
- e.g. ticker=A

Fields (a.k.a field set):

- syntax: key-value pair, separated by a comma “,”
- e.g. price=42

timestamp:

- up to nanosecond-precision Unix time

Reference: <https://v2.docs.influxdata.com/v2.0/reference/line-protocol/>

An example of Line Protocol

cpu,host=serverA,num=1,region=west idle=1.667,system=2342.2 1492214400000000000



Measurement

An example of Line Protocol

```
cpu,host=serverA,num=1,region=west idle=1.667,system=2342.2 1492214400000000000
```



An example of Line Protocol

```
cpu,host=serverA,num=1,region=west idle=1.667,system=2342.2 1492214400000000000
```



Fields

An example of Line Protocol

cpu,host=serverA,num=1,region=west idle=1.667,system=2342.2 1492214400000000000



timestamp

Bucket physical view

Columnar Data Stores

Best for column selection and aggregation thanks to

- Disk + Memory locality
- Cache locality

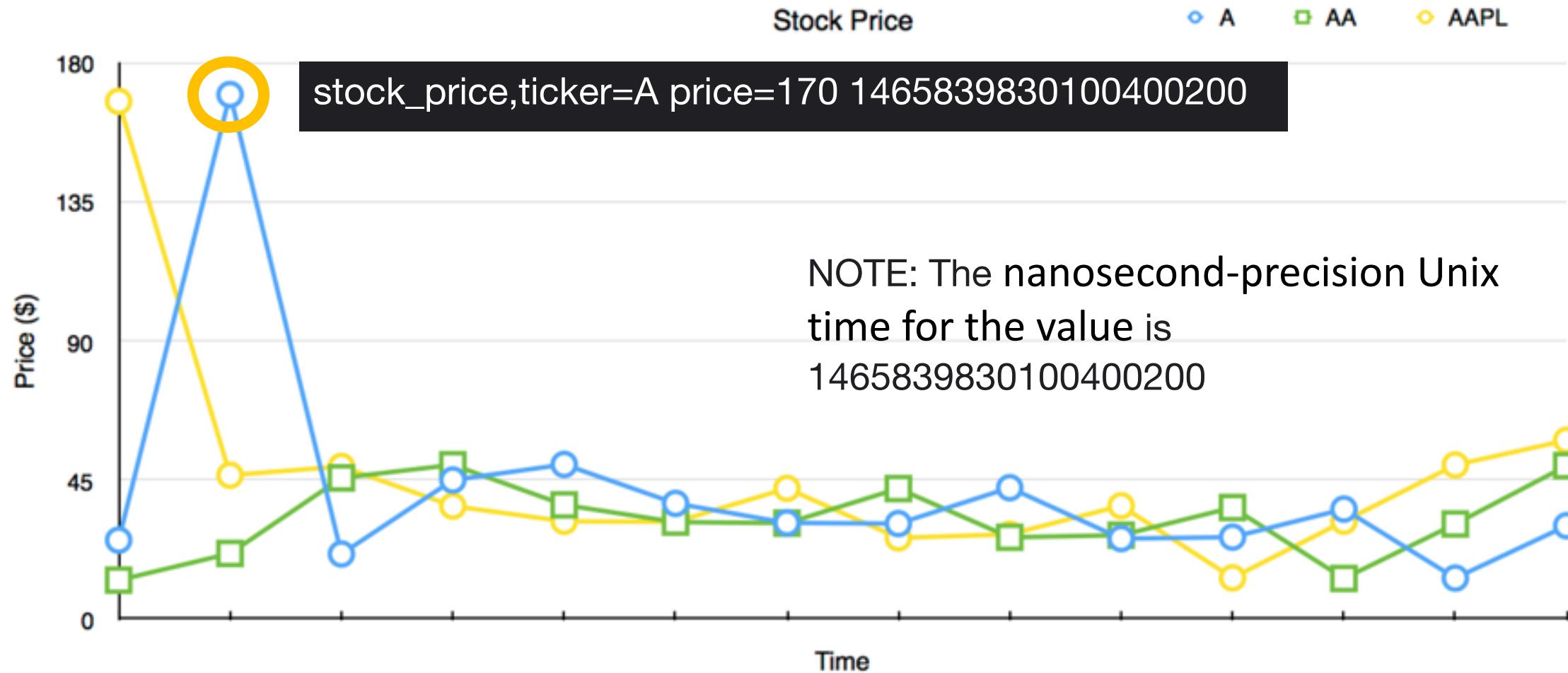
_time	host	num	region	idle	system
1492...0	serverA	1	west	1.667	2342.2
1492...1	serverB	1	west	1789.4	2.779
...

The diagram illustrates a columnar data store structure. It shows a table with six columns: _time, host, num, region, idle, and system. The first two rows of data are highlighted with red boxes. A bracket labeled "tags" spans across the first four columns, while a bracket labeled "fields" spans across the last two columns. Ellipses indicate more data rows below.

Let's get dirty!



What's the line protocol representation of this point?



Let's get **more** dirty!

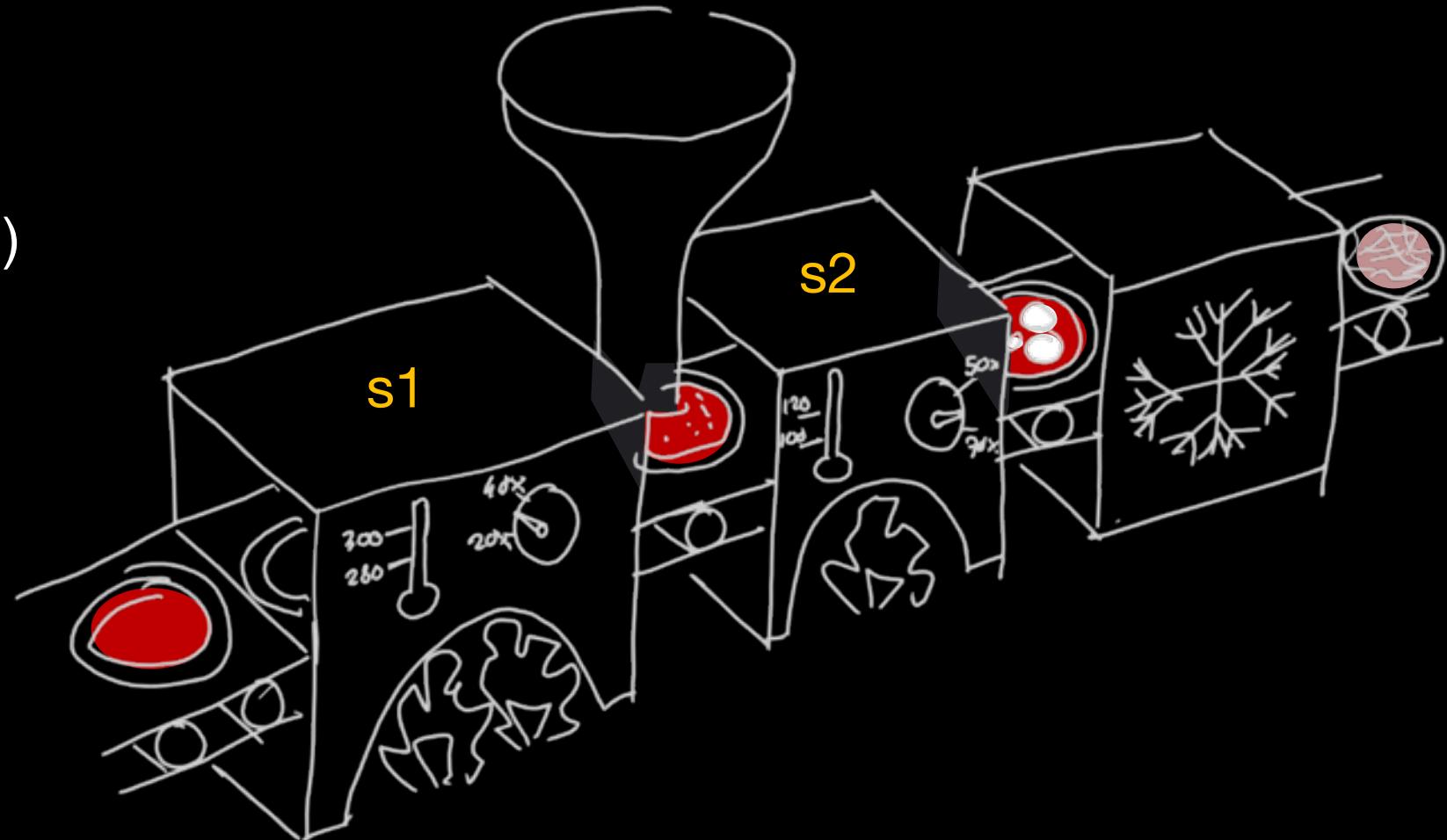


1

Use Case: Continuous Linear Pizza Oven

Sensors observe
temperature (C°)
relative humidity (%)
of the two ovens

Learning Goals
Basic
line protocol
usage



Task

Model the following data representing the temperatures, from both sensors, over time

measurement	sensor	value	ts
temperature	S1	290	1569888000000000000
temperature	S2	105	1569888015000000000
temperature	S1	305	1569888060000000000
temperature	S2	120	1569888075000000000

Let's continue on gitter

<https://gitter.im/quantia-cons/flux-training-sf-2019#>

Let's get even more dirty!



Use Case: Continuous Linear Pizza Oven

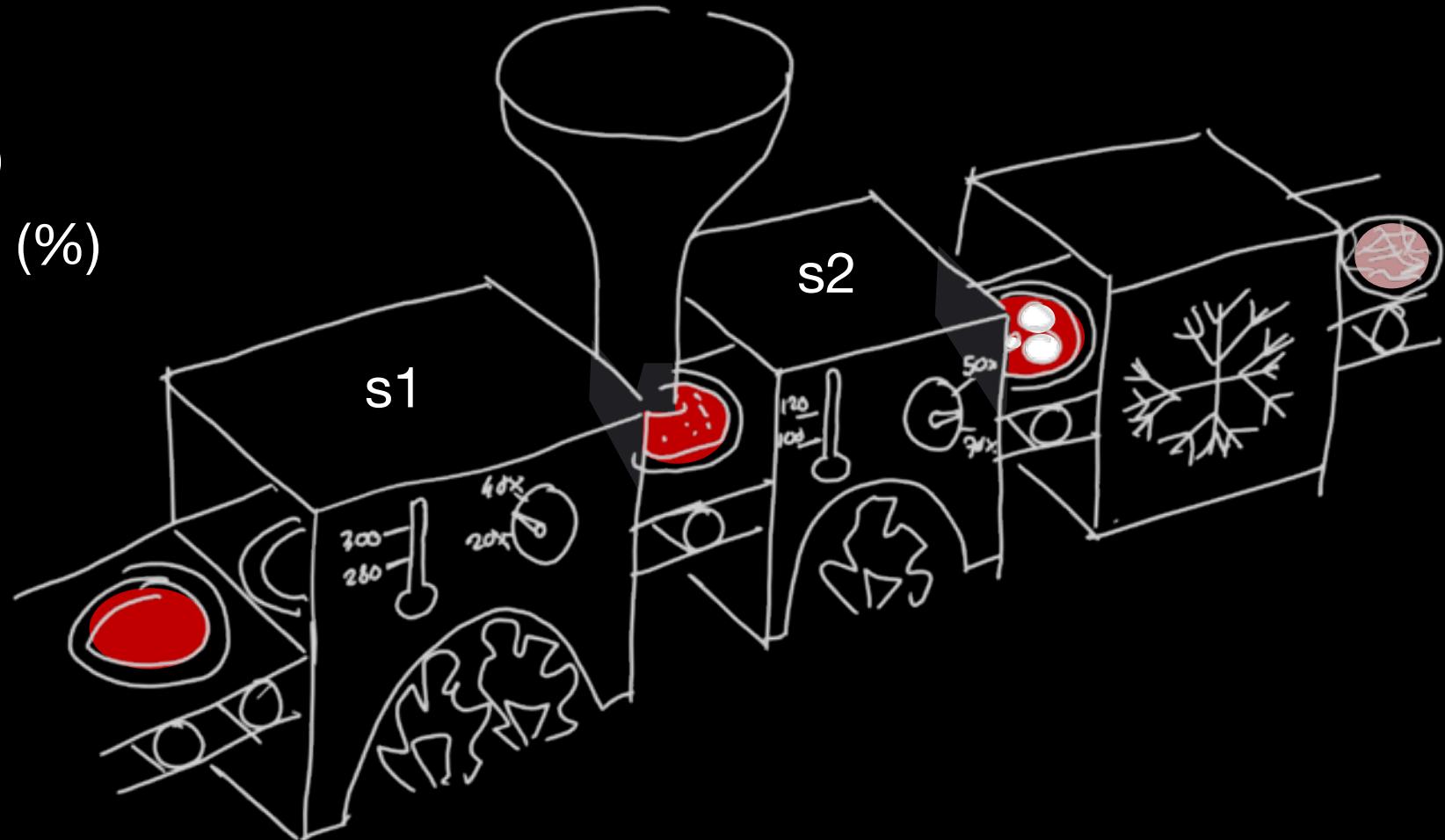
Sensors observe

- temperature (C°)
- relative humidity (%)

of the two ovens

Learning Goals

- Advanced line protocol usage



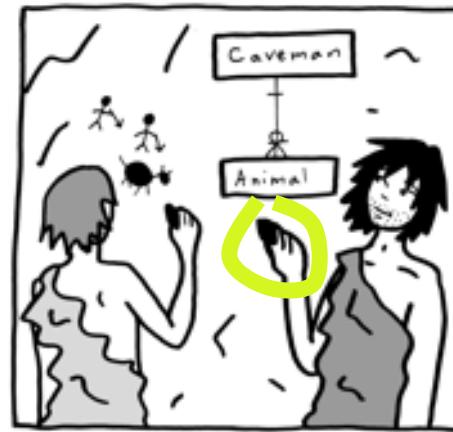
Task

Model the following data representing the temperatures and the humidities, from both sensors, over time.

measurement	sensor	temperature	humidity	ts
iot-oven	S1	290	30	1569888000000000000
iot-oven	S2	105	55	1569888015000000000
iot-oven	S1	305	38	1569888060000000000
iot-oven	S2	120	65	1569888075000000000

How is data **automatically** ingested?

Physical View



Telegraf

Telegraf is a data collection agent

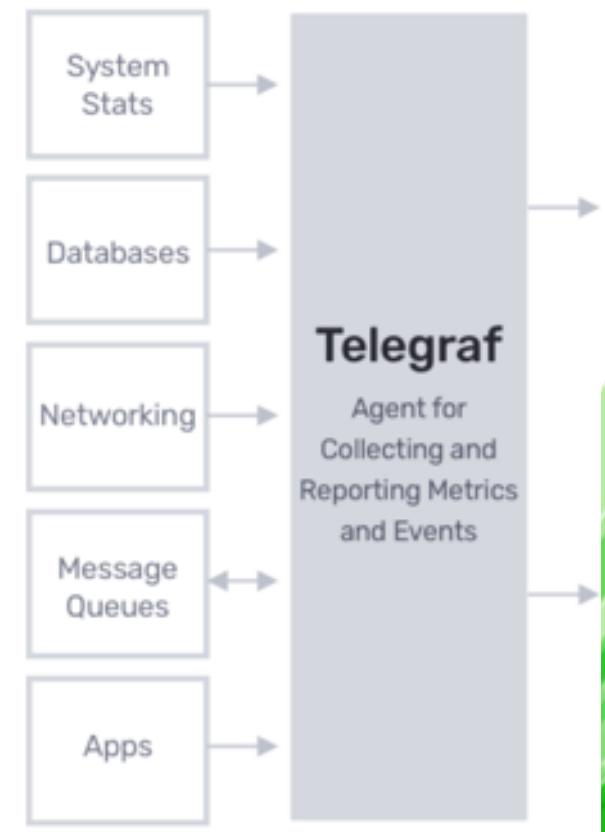
It is based on a Plug and Play architecture

It offers a variety of input plugins

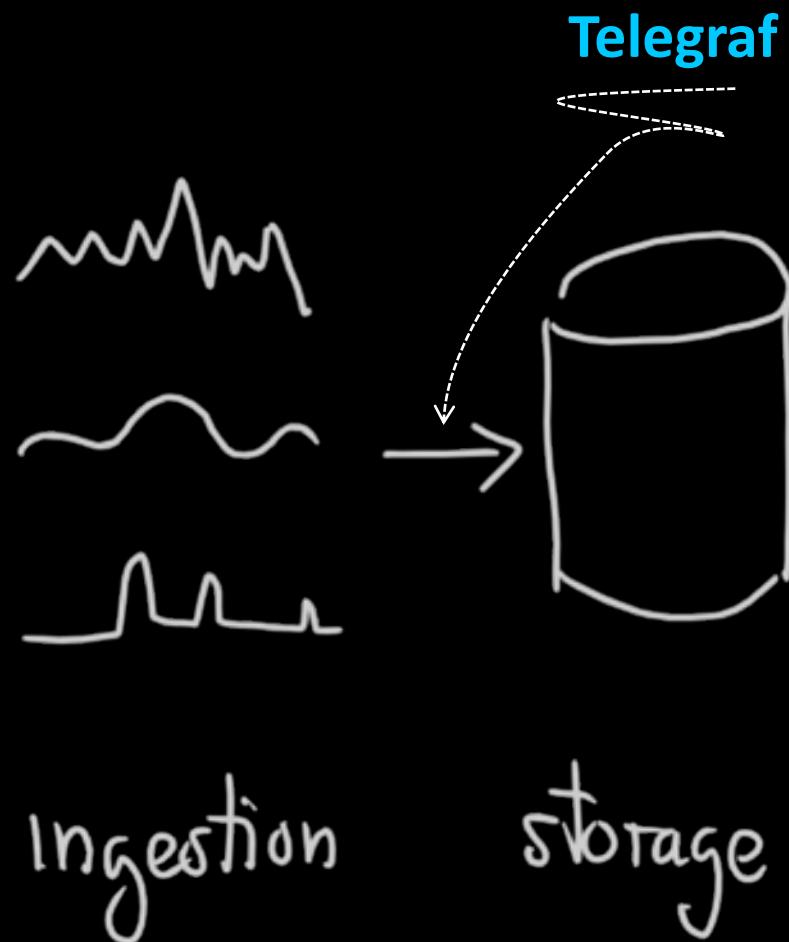
It can be configured from the influxDB cloud UI

Download and install:

<https://docs.influxdata.com/telegraf/v1.12/introduction/installation/>



Positioning Telegraf in the ingestion pipeline





influx/days