

Advanced Data Analysis Joining Time Series

Emanuele Della Valle

Prof. @ Politecnico di Milano

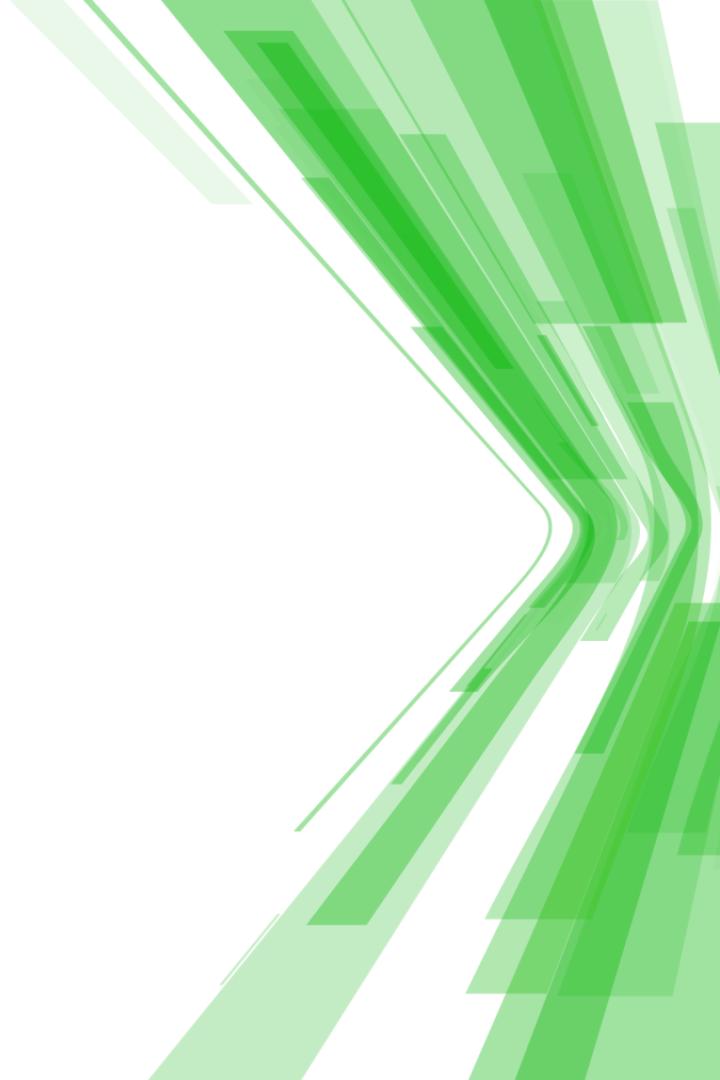
Founder & Partner @ Quantia Consulting

Marco Balduini

Founder & CEO @ Quantia Consultin



Exploring flux Joining Time Series



Joining two time series



The ideal case



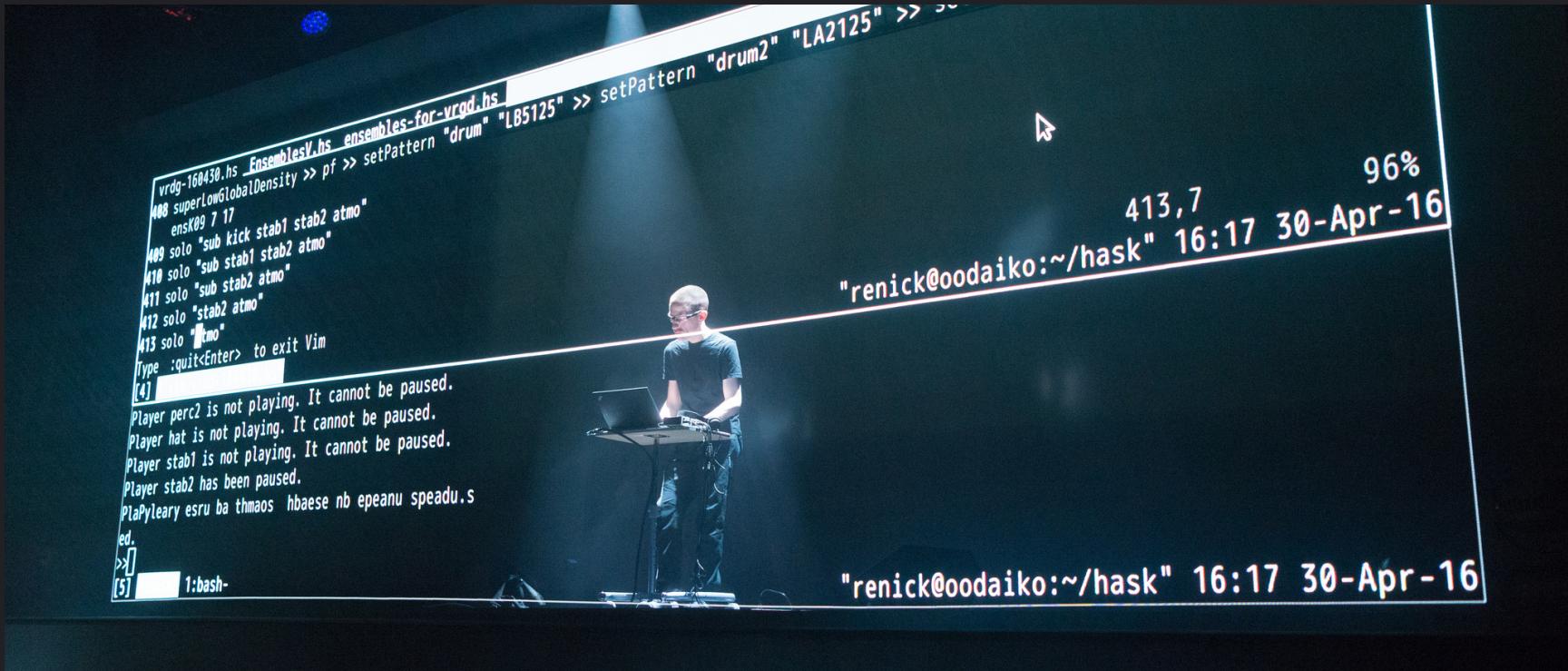
Running example on pizza quality issues

- If the temperature difference between the base cooking area and the mozzarella melting area is lower than 180° or greater than 190° , the mozzarella cheese will not melt properly or it will burn.

Ok ... sounds easy



Let's do some live coding



Here we go

```
ts12 = from(bucket: "training")
|> range(start: 2020-06-07T12:00:00Z, stop: 2020-06-07T12:05:00Z)
|> filter(fn: (r) => r._measurement == "iot-oven")
|> filter(fn: (r) => r._field == "temperature")

ts1 = ts12 |> filter(fn: (r) => r.sensor == "S1")
ts2 = ts12 |> filter(fn: (r) => r.sensor == "S2")

join(tables: {key1: ts1, key2: ts2}, on: ["_time"], method: "inner")

join |> map(fn: (r) => ({ _time: r._time,
                           _value: r._value_s1 - r._value_s2 }))
```

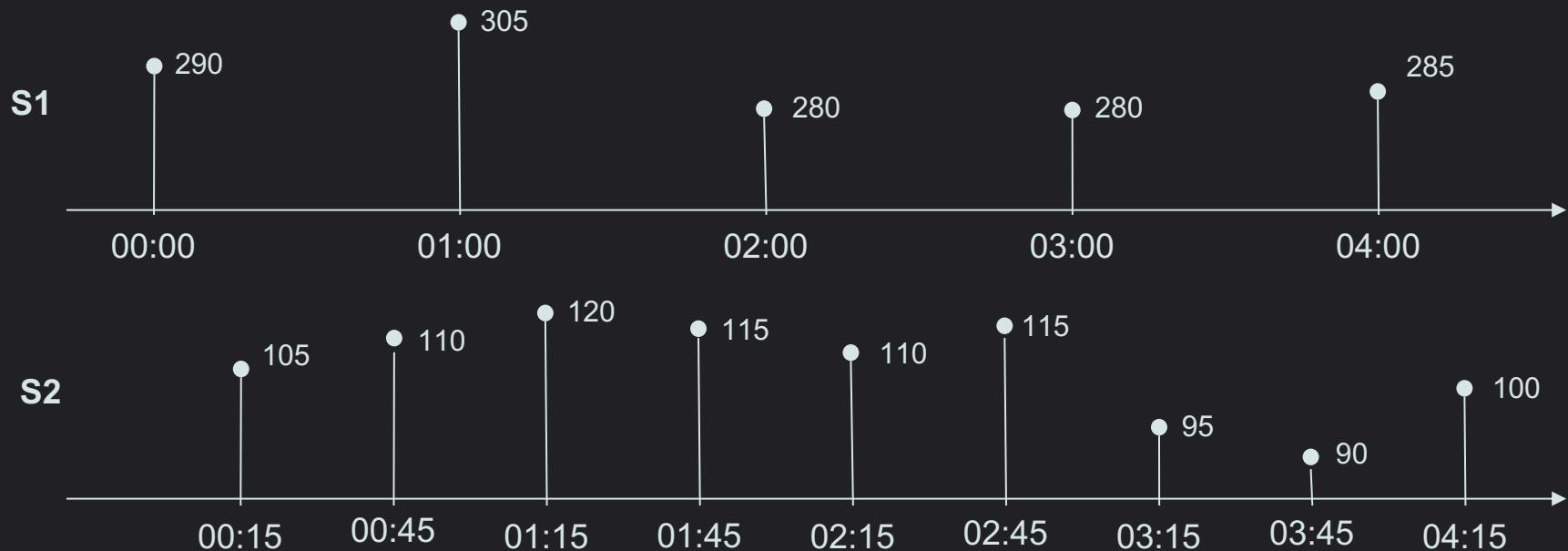
It gives no results ... why?



The reality is always more difficult than expected

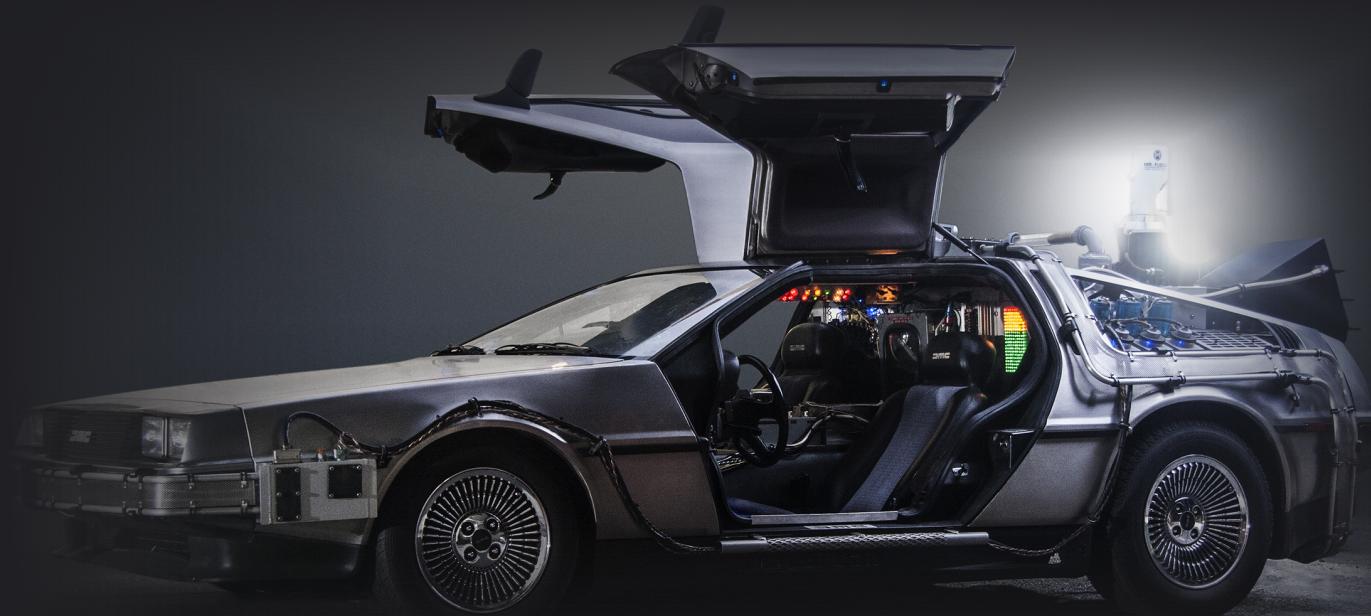


A close look to the distribution of our data over time

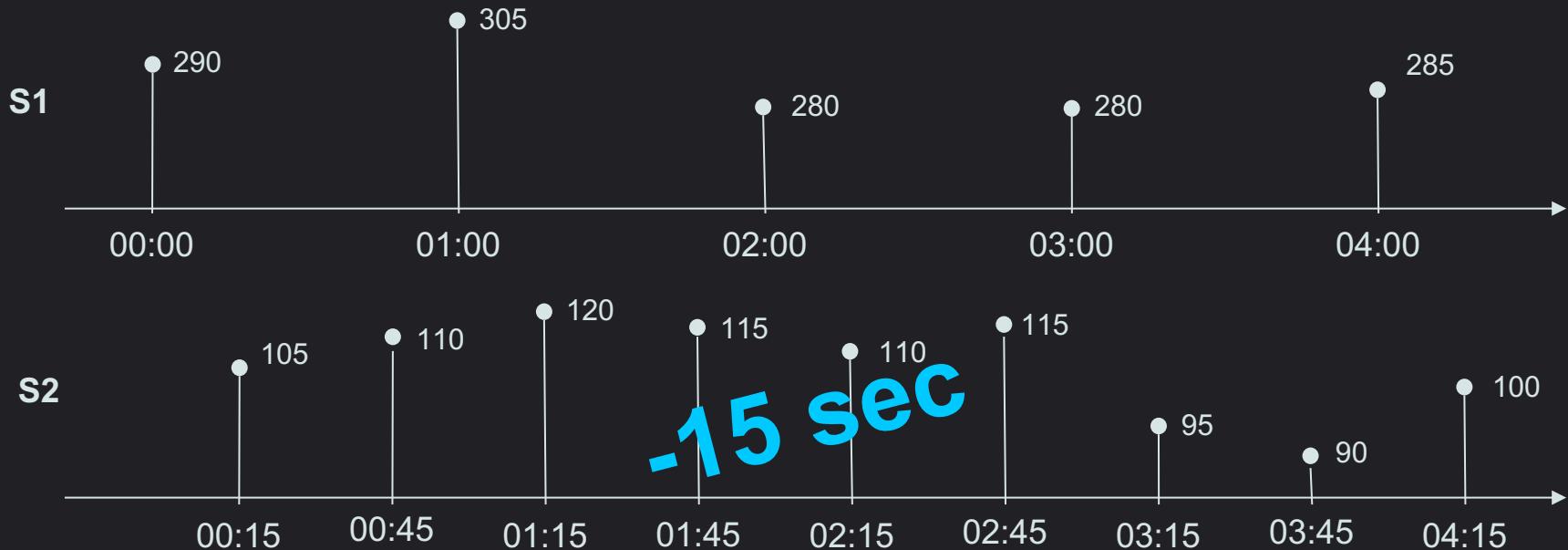


NOTE: S1 and S2 are not synchronized: metrics don't occur at the same time and rate

Can we synchronize them? Let's do a time shift!



Can we synchronize them with a time shift?

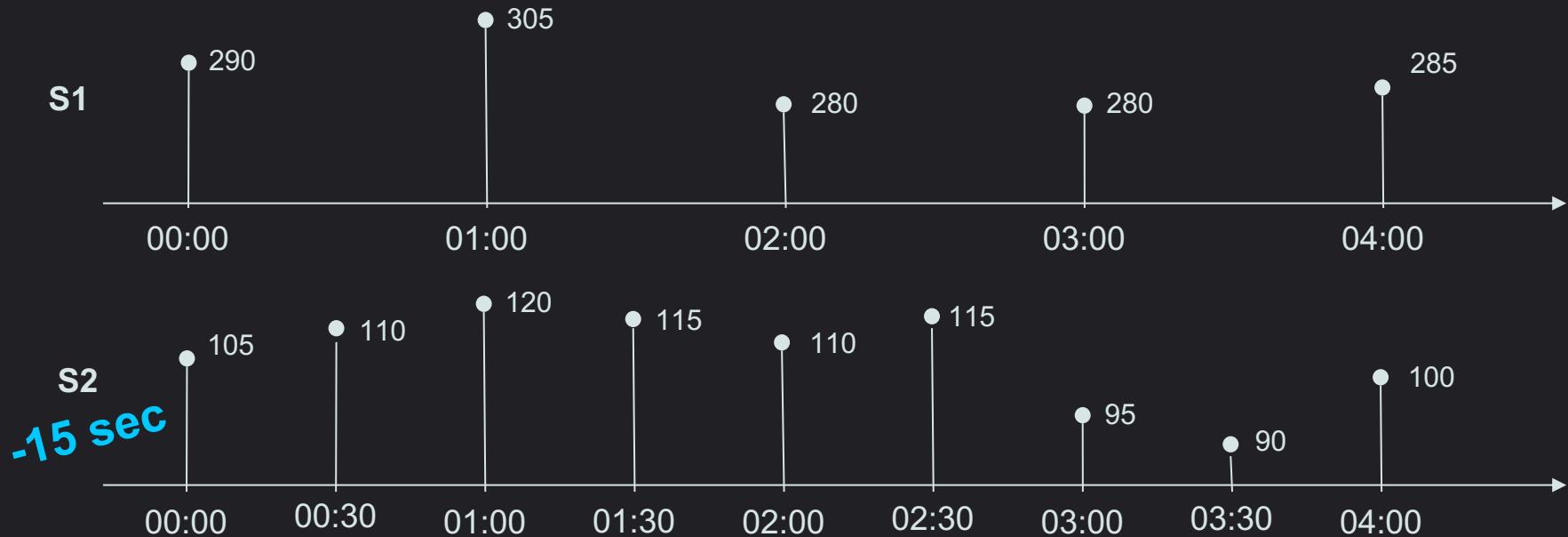


Here we go

```
temp = from(bucket: "training")
|> range(start: 2020-01-01T00:00:00Z, stop: 2020-01-01T00:05:00Z)
|> filter(fn: (r) => r._measurement == "iot-oven")
|> filter(fn: (r) => r._field == "temperature")
tempS1 = temp |> filter(fn: (r) => r.sensor == "S1")
tempS2 = temp |> filter(fn: (r) => r.sensor == "S2")
|> timeShift(duration: 15s, columns: ["_time"] )

join = join(  tables: {s1:tempS1, s2: tempS2},  on: ["_time"] )
join |> map(fn: (r) => ({ _time: r._time,
                           _value: r._value_s1 - r._value_s2 }))
```

Can we synchronize them with a time shift?



NOTE: some metrics now occur at the same time, but the rate is still different

Moreover, is -15 sec correct? Shall I do -30 sec, instead?



Shall we tumble?

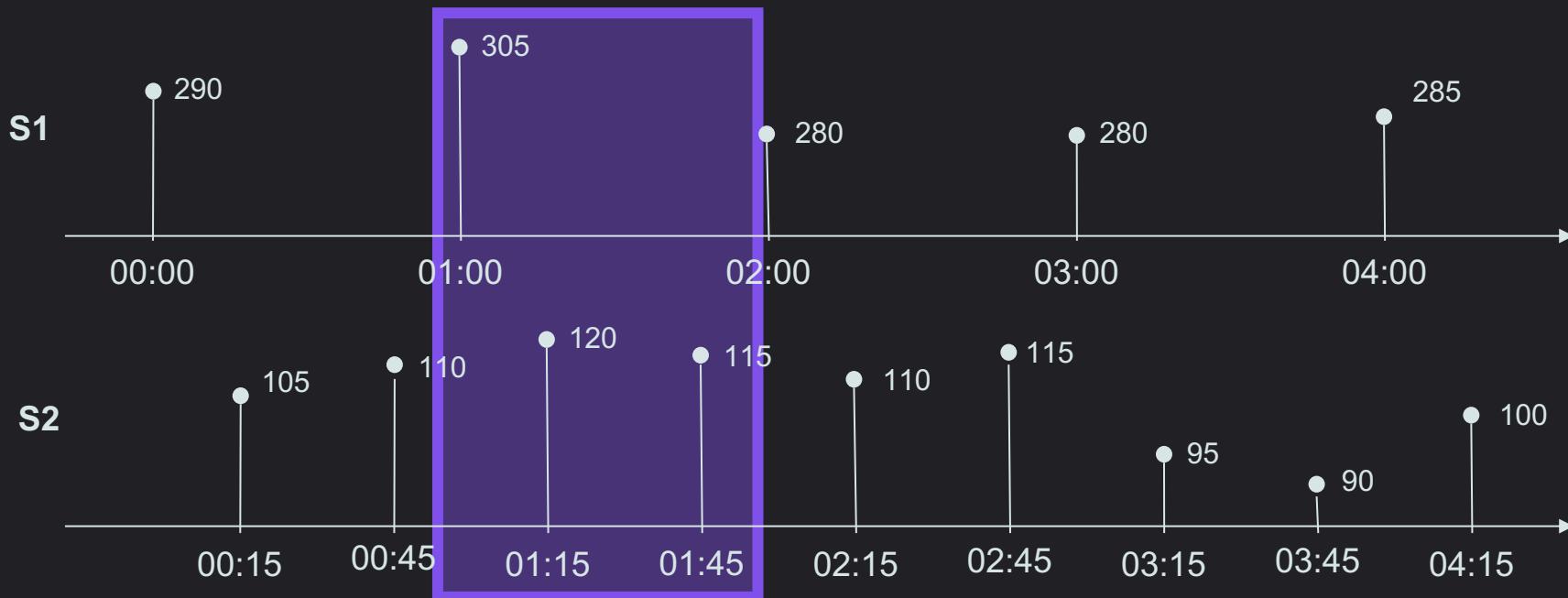


Can we synchronize them with a tumbling window?



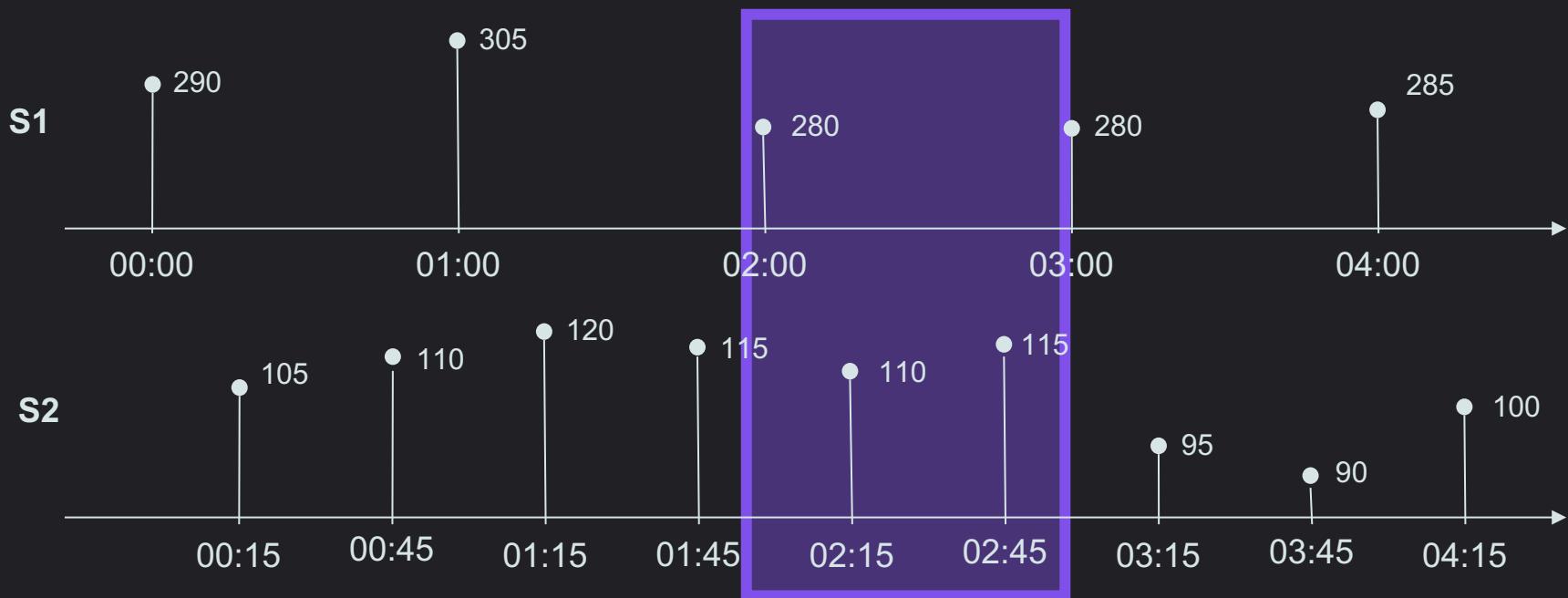
NOTE: S1 and S2 are not synchronized: metrics do not occur nor at the same time nor at the same rate

Can we synchronize them with a tumbling window?



NOTE: S1 and S2 are not synchronized: metrics do not occur nor at the same time nor at the same rate

Can we synchronize them with a tumbling window?



NOTE: S1 and S2 are not synchronized: metrics do not occur nor at the same time nor at the same rate

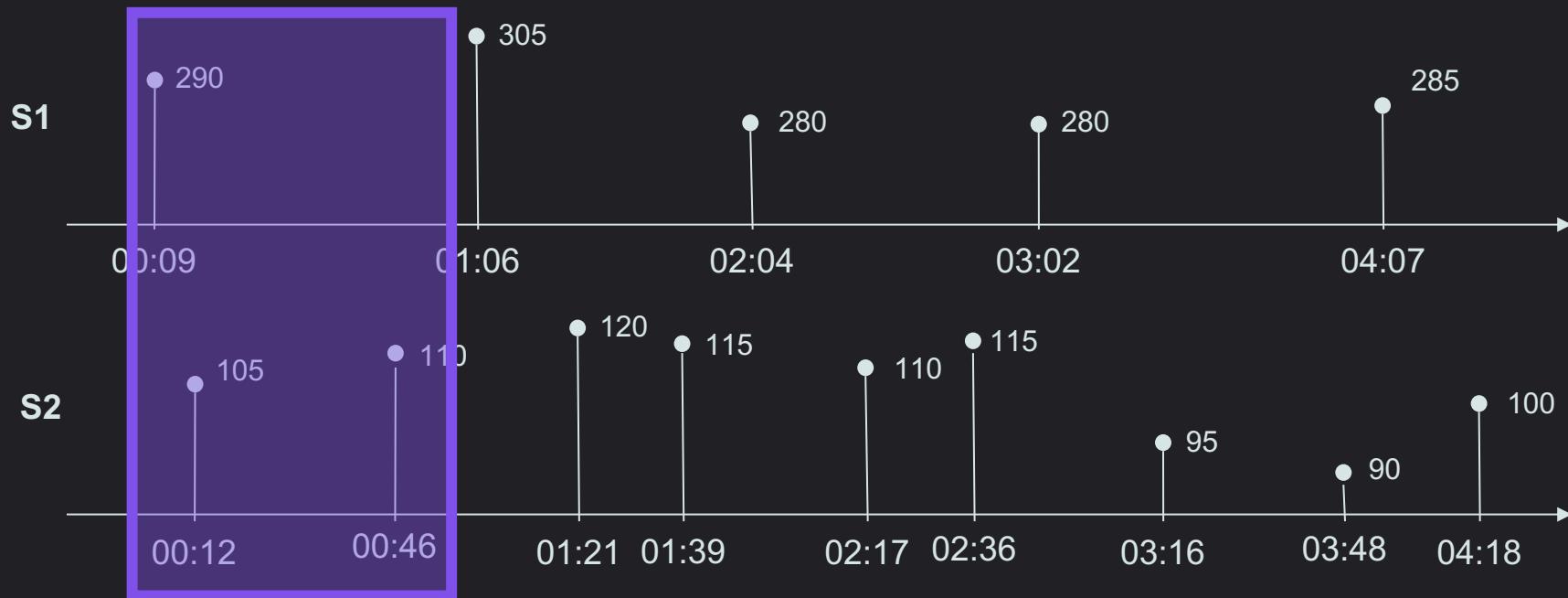
Here we go

```
temp = from(bucket: "training")
|> range(start: 2020-04-30T10:00:00Z, stop: 2020-04-30T10:05:00Z)
|> filter(fn: (r) => r._measurement == "iot-oven")
|> filter(fn: (r) => r._field == "temperature")
tempS1 = temp |> filter(fn: (r) => r.sensor == "S1")
                |> aggregateWindow(every: 60s, fn: mean)
tempS2 = temp |> filter(fn: (r) => r.sensor == "S2")
                |> aggregateWindow(every: 60s, fn: mean)
join = join(  tables: {s1:tempS1, s2: tempS2},  on: ["_time"] )
join |> map(fn: (r) => ({ _time: r._time,
                            _value: r._value_s1 - r._value_s2 }))
```

It works!

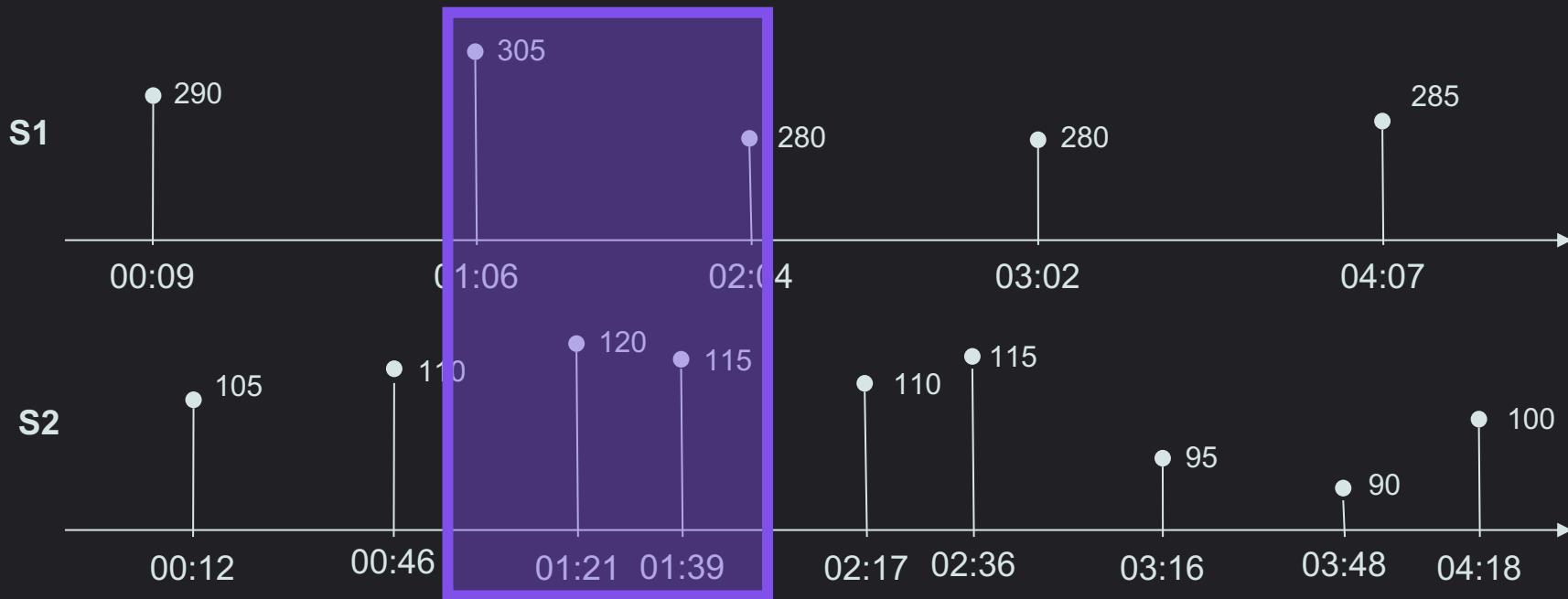


Bonus point: tumbling windows also synchronize events!



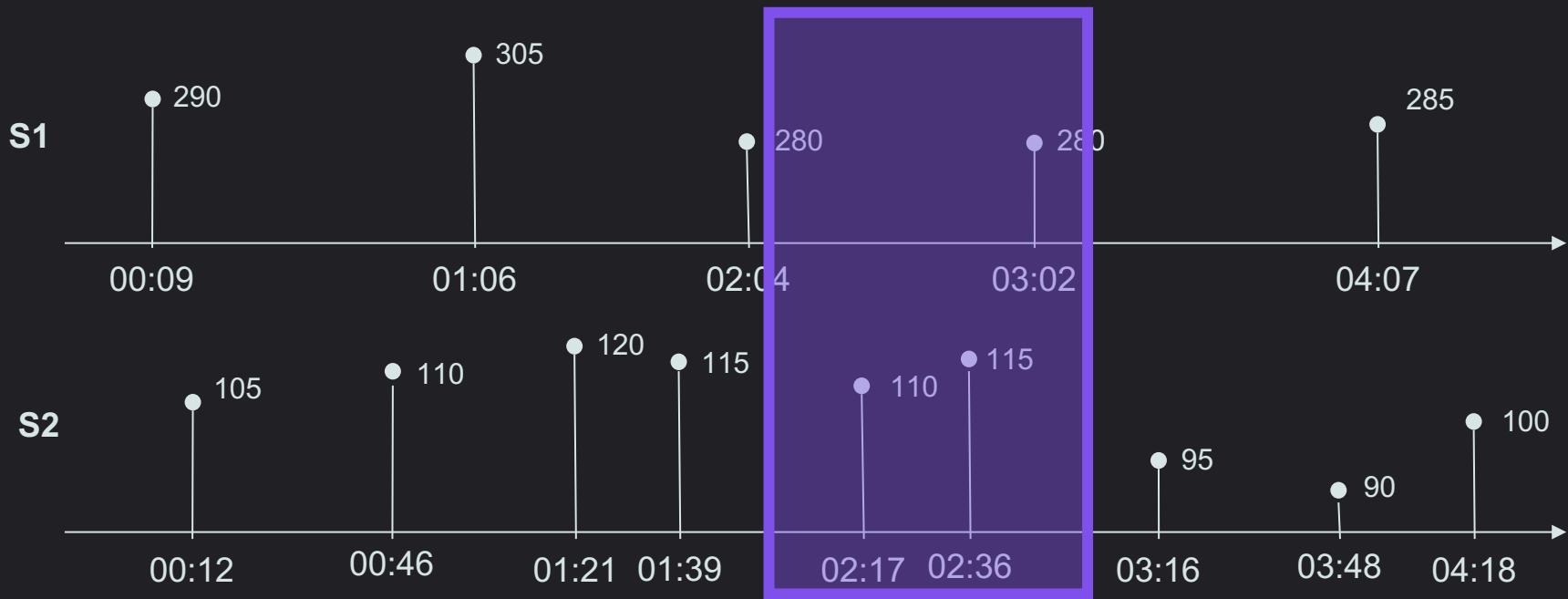
MENO: events are irregular time series

Bonus point: tumbling windows also synchronize events!



MENO: events are irregular time series

Bonus point: tumbling windows also synchronize events!



MENO: events are irregular time series

Quiz

- Which synchronization method shall you use before joining?
none | timeShift | aggregateWindow
 - one timeseries is regular, but the other is not - ?
 - two timeseries are regular, have the same rate, but metrics occur at a different time - ?
 - two timeseries are regular, have the same rate, and metrics occur at the same time - ?
 - two timeseries are irregular - ?
 - two timeseries are regular, but metrics don't occur at the same time and rate - ?

Quiz answers

- Which synchronization method shall you use before joining?
none | timeShift | aggregateWindow
 - one timeseries is regular, but the other is not - aggregateWindow
 - two timeseries are regular, have the same rate, but metrics occur at a different time - timeShift
 - two timeseries are regular, have the same rate, and metrics occur at the same time - none
 - two timeseries are irregular - aggregateWindow
 - two timeseries are regular, but metrics don't occur at the same time and rate - aggregateWindow

Let's get dirty!

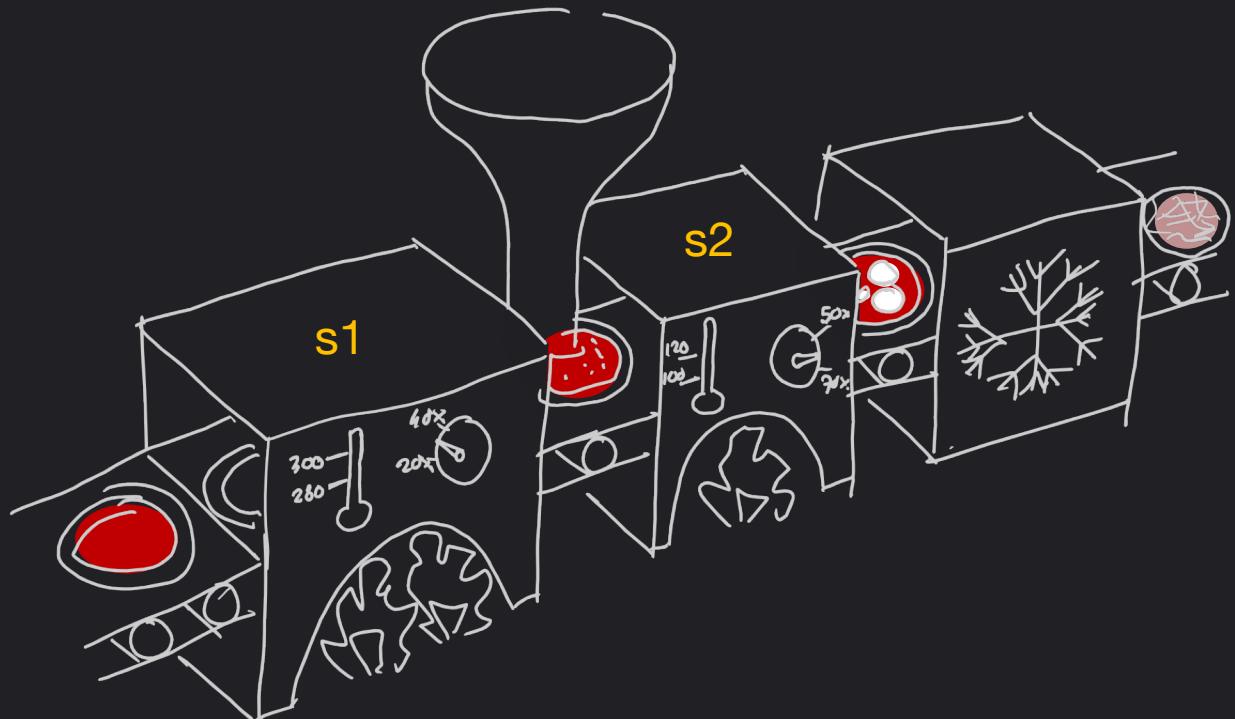


8

Continuous Linear Pizza Oven

Learning goals:

- Join data



Task

- Extract the difference between the humidity levels of the base cooking area and the mozzarella melting area. Find if the differences are lower than 20% or greater than 30%



Advanced Data Analysis Joining Time Series

Emanuele Della Valle Prof. @ Politecnico di Milano & Partner @ Quantia Consulting
Marco Balduini Founder & CEO @ Quantia Consulting