

# Mobility Insights at Swisscom : Understanding Collective Mobility in Switzerland

Spark Summit, October 2016

francois.garillot@swisscom.com @huitseeker

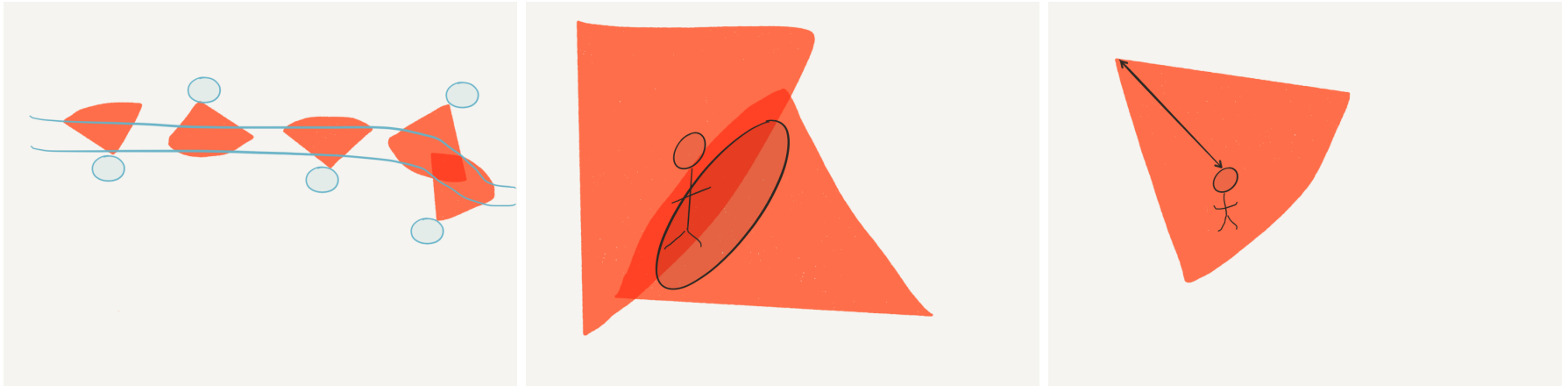
mohamed.kafsi@swisscom.com @mou7

# Agenda

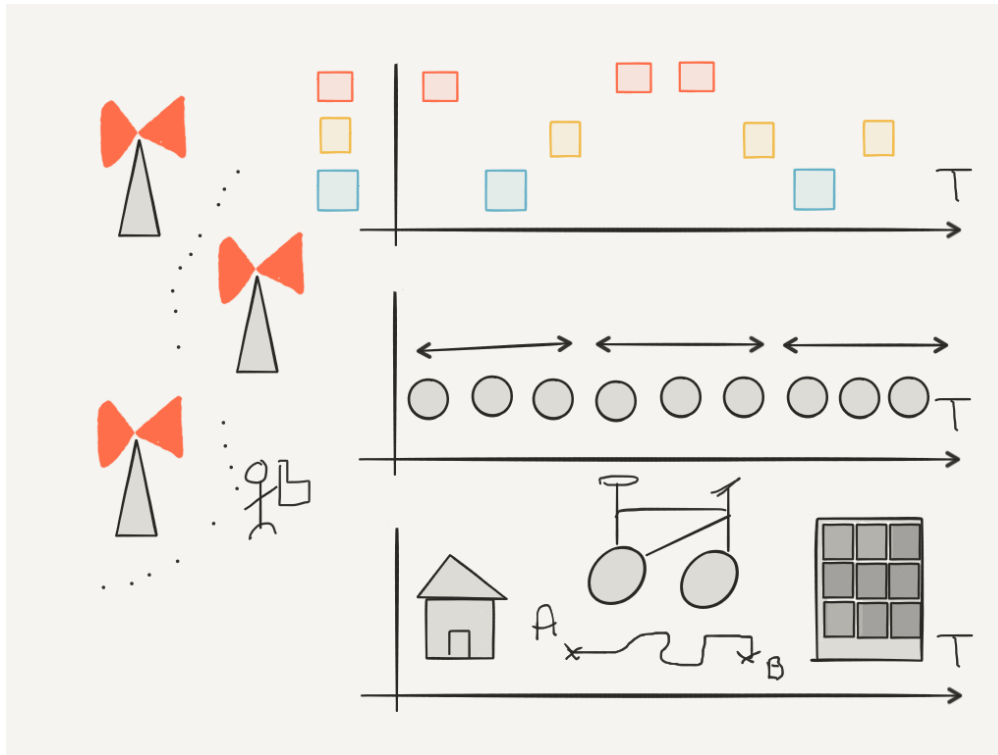
- Intro
- Smart-Data
- Big Data Architecture
- Trajectory Classification
- Streaming
- Data challenges

# Introduction : Positioning

# Positioning users in a modern network



- no triangulation at scale
- positioning based on cell attachment history, prec ~200m
- cell-to-cell handover, prec ~50m around limit
- Timing Advance (roundtrip) : better results on good data sources



# Trajectory data mining

- time series reconstruction
- trajectory segmentation
- map matching, clustering
- mode of transport detection
- ...

# How to create value with positioning at Swisscom ?

- with competitive analytics & data sources,
- and by making sure it embodies the right values.

# Smart Data

# On (not) tracking (any users)

*"Swisscom strictly complies with all applicable legislations, in particular with the telecommunications law and the data protection initiative."*

Jürg Studerus, Swisscom Senior Manager, Corporate Responsibility



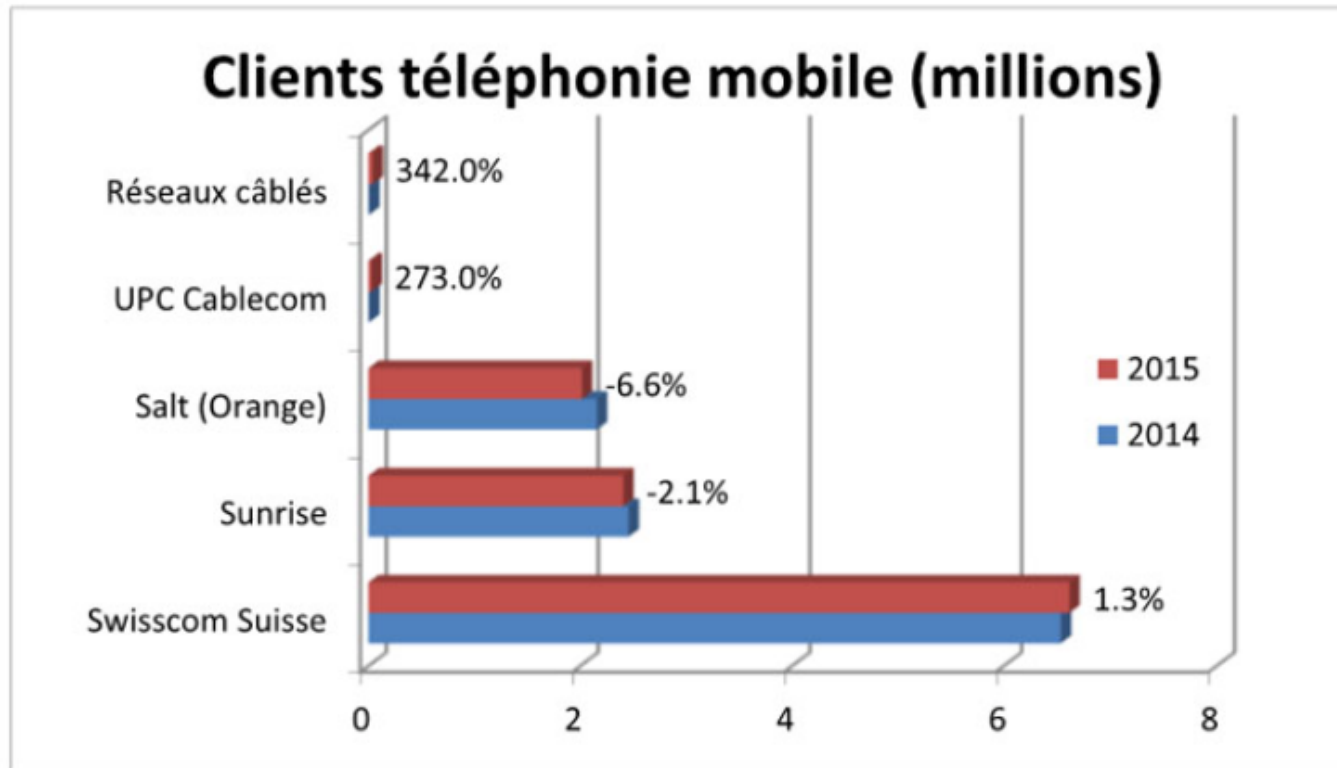
# Smart Data : Big Data without Big Brother

- Privacy preservation is an asset
- It makes sense to care as much about your customer as they do about you.

We technically enforce this

- answering only synoptic questions, no individual ones,
- with data flow control : we neutralize quasi-identifiers at every stage

# Swisscom mobile subscribers



source: [xavierstuder.com](http://xavierstuder.com), MD&A reports

# Our choices

- public good applications: making Switzerland run better,
- understanding places, not individuals,
- anonymized aggregations

# A first product : City



*"It's a dream for civil engineers" -- Alexandre Machu, Urban systems engineer, Pully*

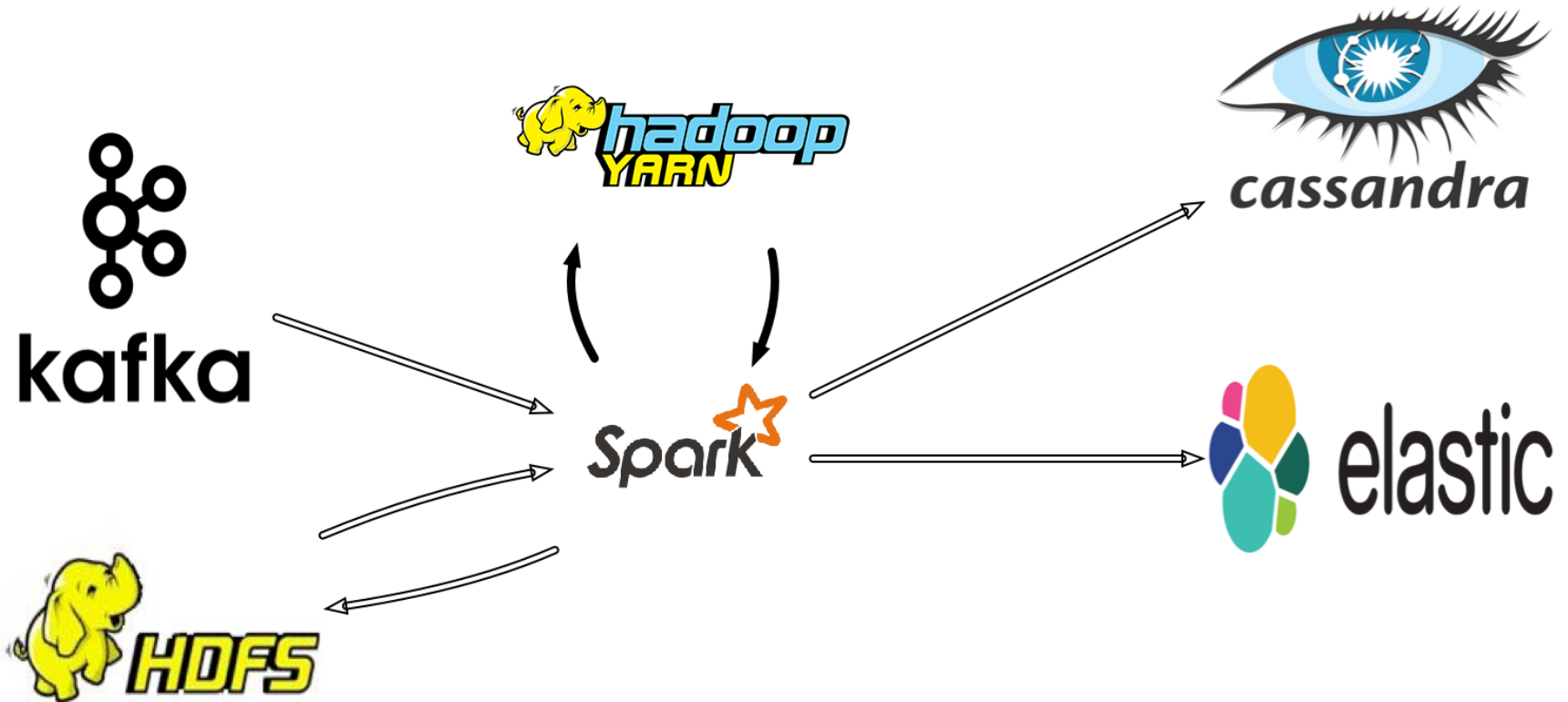
Demo time

# Usages

- New roads to divert transit traffic out of downtown (informs a 50M\$ project)
- Parking lot expansion and transformation (informs a 10M\$ project)
- Electric car charging station deployment

# Big Data architecture

# In the backend





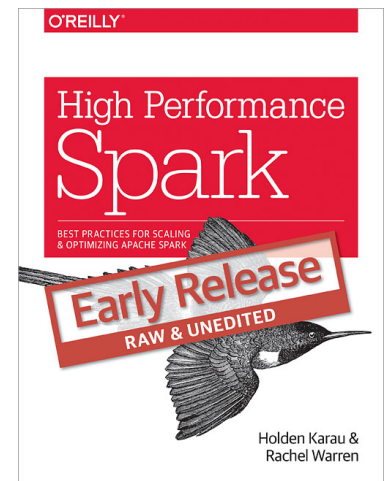
# Spark configuration essentials for enterprise jobs

```
spark.executor.memory="not the default 1g"  
spark.kryo.registrator="something custom" // among others
```

```
spark.shuffle.service.enabled="true"  
spark.dynamicAllocation.enabled="true"
```

```
spark.deploy.recoveryMode="ZOOKEEPER"  
spark.deploy.recoveryDirectory="/path/to/state"  
spark.deploy.zookeeper.url="quorumMachine1:2181, ..."
```

NOT the only valuable settings, see  
<https://techsuppdiva.github.io>



# Scala (1/2)

```
type ChronoHistory = List[UEupdate] @@ Chronological
type AnteChronoHistory = List[UEupdate] @@ AnteChronological

implicit class Chrono(l: List[UEupdate]) {
  def asChrono: ChronoHistory = {
    chronoCheck(l)
    l.asInstanceOf[ChronoHistory]
  }
  def asAnteChrono: AnteChronoHistory = {
    anteChronoCheck(l)
    l.asInstanceOf[AnteChronoHistory]
  }
}
```

## Scala (2/2)

```
implicit def reverseChrono(l: ChronoHistory): AnteChronoHistory  
implicit def reverseAnteChrono(l: AnteChronoHistory): ChronoHis
```

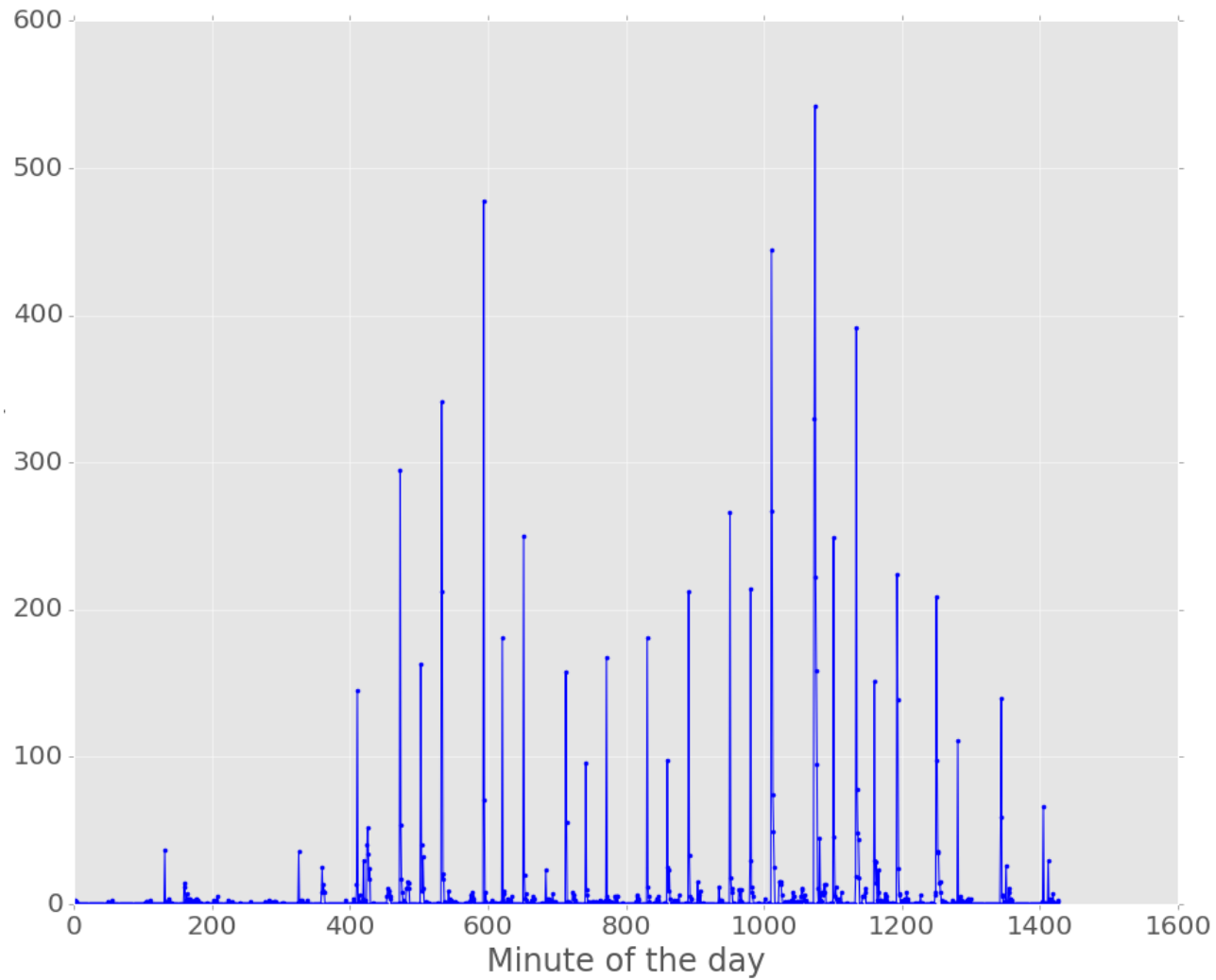
# Trajectory Classification

What is the proportion of trips  
associated with trains?

# Mode of Transport Detection

- **Input:** Sequence of network events
- **Output:** Mode of transport (train vs. other)
- Network events associated with cells
- Create fingerprints of cells
- **Intuition:** cells with intermittent increases in the number of connections are associated with collective mode of transports

# Bursty Cell



Number of devices vs. minute of day

# Burstiness

Random process with mean  $\mu$  and variance  $\sigma^2$ , the relative variance is

$$D = \frac{\sigma^2}{\mu}.$$



# Machine Learning with Spark

- Periodic Spark job to compute cell features
- Supervised training on labeled data (train vs. others)
- Training and test with Spark ML

# Spark (1/2)

```
val labeledPoints: RDD[LabeledPoint] = data.map {  
  case (transportMode, tripFeatures) =>  
    LabeledPoint(  
      labelOf(transportMode).toDouble,  
      featuresToFeatureVector(tripFeatures)  
    )  
} // generate labeled data  
labeledPoints.cache()  
  
def trainNewModel = // Fix the used model  
  new LogisticRegressionWithLBFGS()  
    .setIntercept(true)  
    .setNumClasses(numberOfClasses)  
    .run(_: RDD[LabeledPoint])
```

## Spark (2/2)

*// train a model for performance evaluation*

```
val model = trainNewModel(trainingData)
```

*// Evaluate model on test instances and compute test error*

```
val labelAndPreds = testData.map { point =>  
  val prediction = model.predict(point.features)  
  (point.label, prediction)  
}
```

```
val testErr =  
  (labelAndPreds  
    .filter(r => r._1 != r._2)  
    .count().toDouble) / testData.count()
```

*// train final model on the whole dataset*

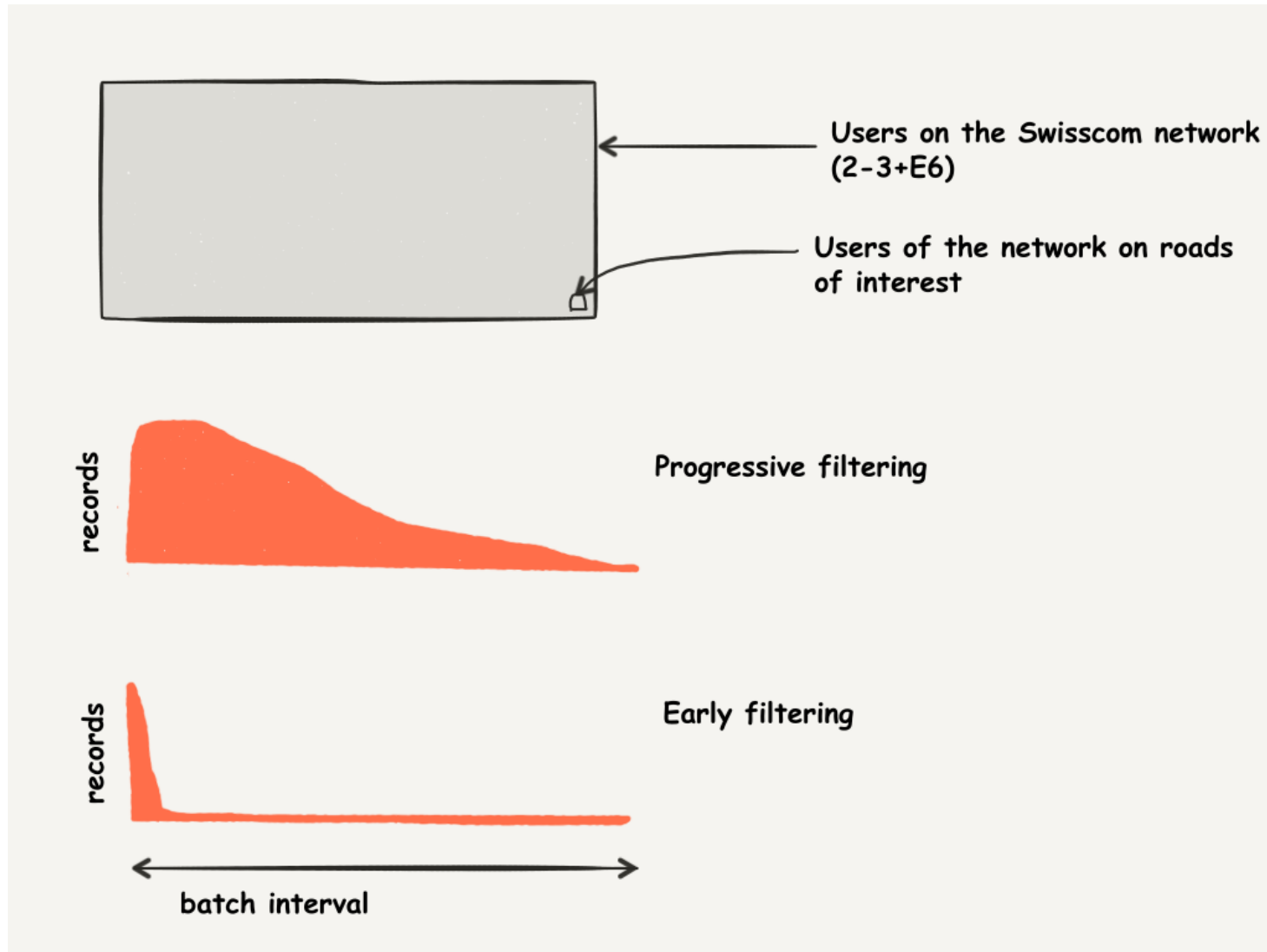
```
val finalModel = trainNewModel(labeledPoints)
```

# Streaming Analytics

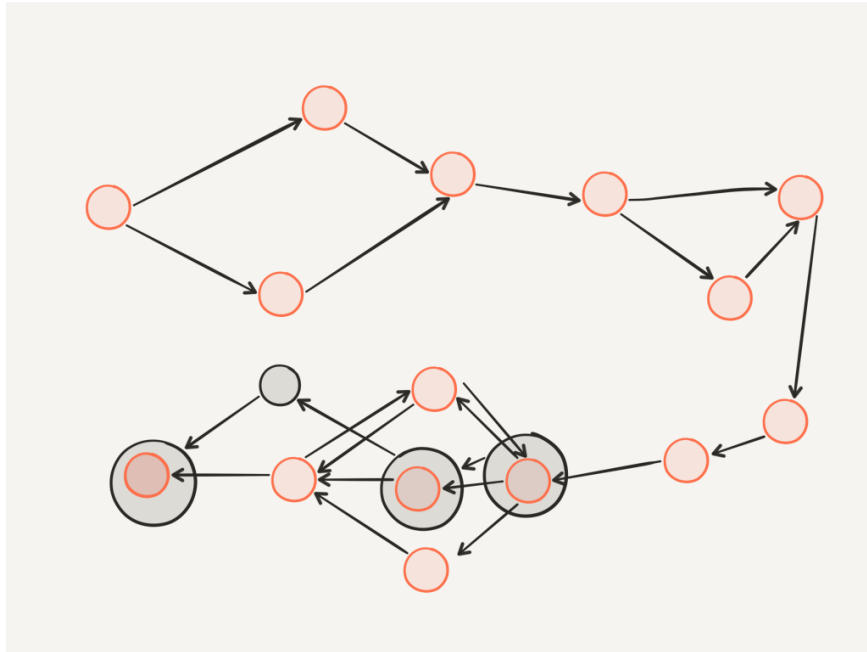
# Road conditions on highways



# Selecting users on a path of Interest



# Graph matching



Locality-sensitive hashing :

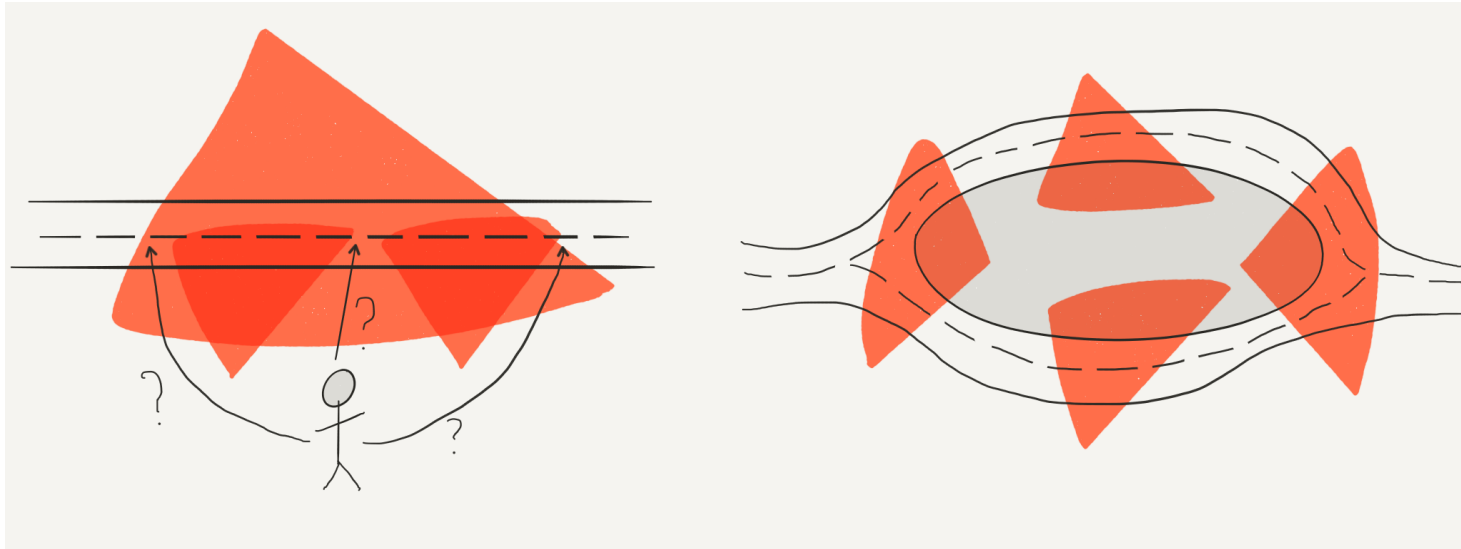
A **family**  $H$  of hashing functions is  $(r, cr, p_1, p_2)$ -sensitive if:

- if  $p - q \leq r$  then  $Pr_H[h(q) = h(p)] \geq p_1$
- if  $p - q \geq cr$  then  $Pr_H[h(q) = h(p)] \leq p_2$

More:

- *Locality Sensitive Hashing By Spark*, Uber, Spark Summit 2016
- *A Gentle Introduction to Locality-Sensitive Hashing with Apache Spark*, Scala By The Bay 2015

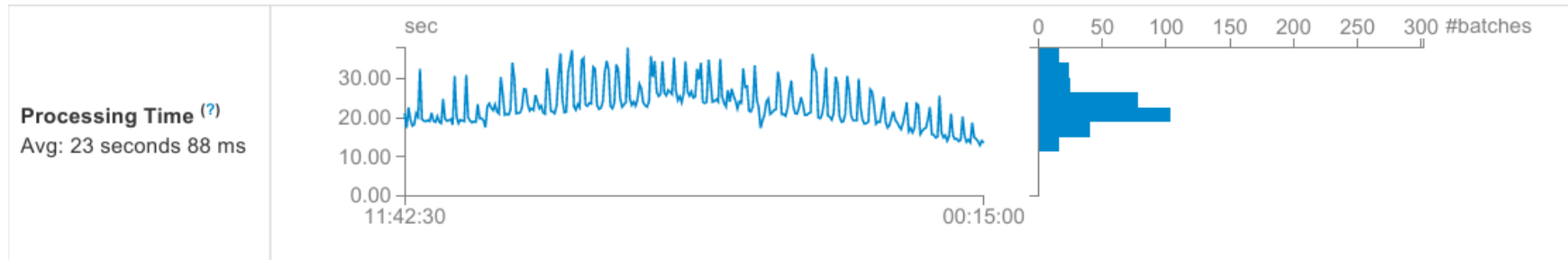
# Computing speeds: Solving graph constraints



- given a history of cells, where was the user, exactly ? (twice)
- what's the path between 2 positions ?
- linear query **per user**



# Checkpointing: Set the checkpoint interval



- are you checkpointing too often ?
- every  $k$  batches, you'll need  $p$  batches to recover from checkpointing time loss
- make sure  $k \geq p$

# Data Challenges

# Crucial elements

- Quality, reliability of data sources
- Automated ground truth checking
  - sensors
  - TEMS fleet
- What's the ground truth for mode of transport, domicile, etc ?
- Colleagues and friends volunteers

Questions ?