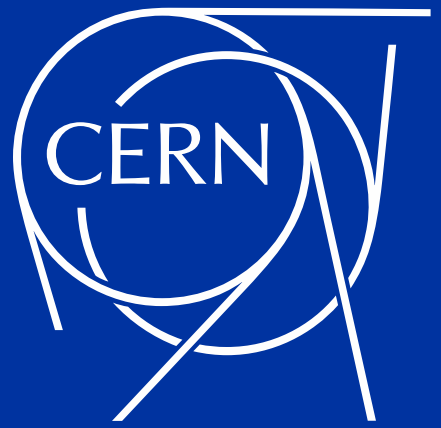


Pre-processing for Anomaly Detection on Linear Accelerator

Martin Molan
Dr. Yann Donon
Dr. Alberto Di Meglio

Introduction

Anomaly detection for linear accelerators (LINACs, LINAC4 H- source for LHC)

Approaches to anomaly detection

Semi-supervised approaches:

- Anomalies are rare events

Autoencoders

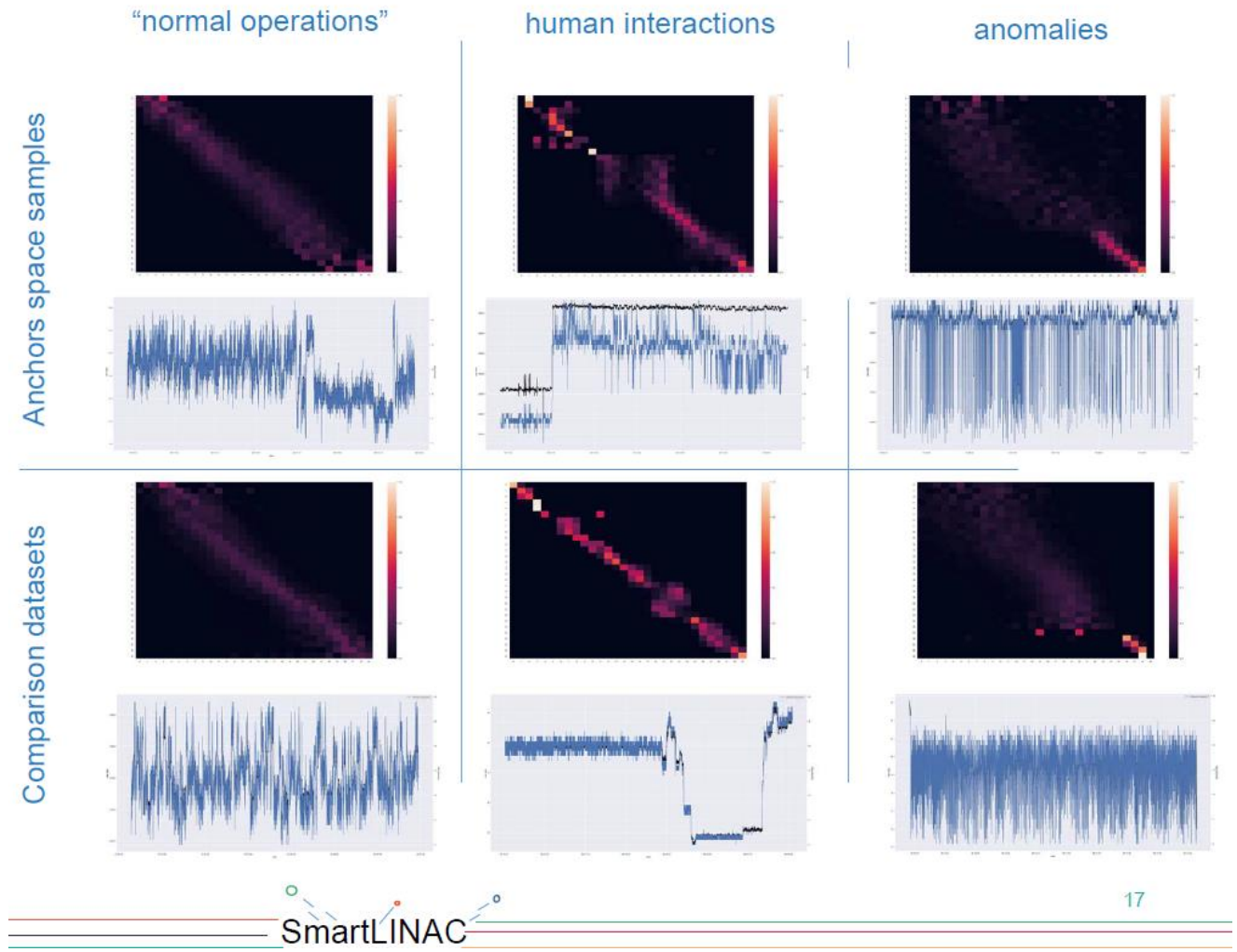
- Reproduction error

Hidden Markov models

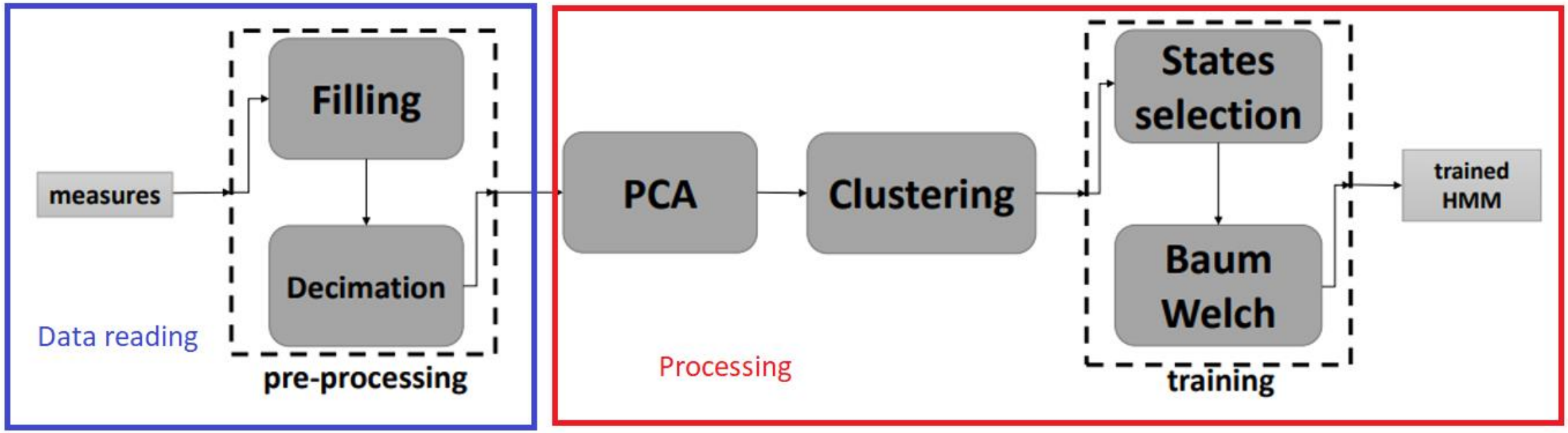
- Probability of sequence

Proposed method

- Stochastic matrix as input for CNN



Estimating confusion matrix



Parallelization

Dask (running on Swan notebook)

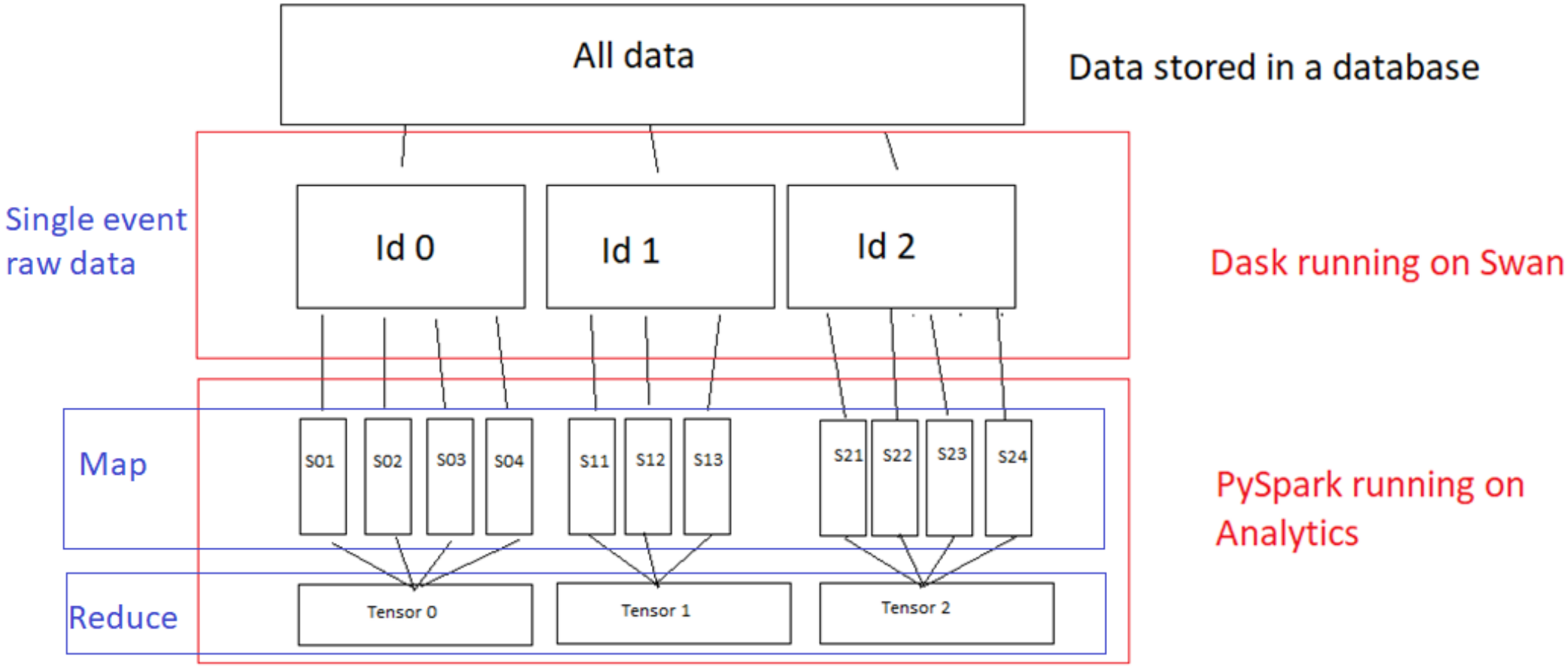
- Data reading
- Input: Id
- Output: (relevant) raw data

Spark (running on analytics)

- Map:
  - Gaussian filter
  - Binning
  - Stochastic matrix
- Reduce
  - Combine tensors

```
def submit_spark_job(context, local_bind_address = 3306, idSource = 27, increment = 5000, test = False):
    try:
        results_l = get_raw_data_list(local_bind_address, idSource, increment, test)
    except Exception as e:
        print(e)
        return []
    rdd = context.parallelize(results_l)
    rdd_calc = rdd.map(statisticalAnalysis_raw)
    fr = rdd_calc.reduce(combine_tensor)
    return fr
```

Approach to parallelization



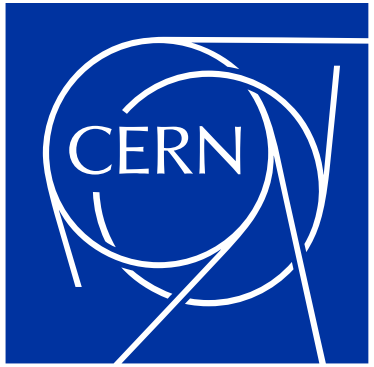
Implementation and performance

Apache Spark: 2 EXECUTORS 4 CORES Jobs: 21 COMPLETED

Job ID	Job Name	Status	Stages	Tasks	Submission Time	Duration
59	reduce	COMPLETED	1/1	4/4	an hour ago	7s
60	reduce	COMPLETED	1/1	4/4	43 minutes ago	2s
61	reduce	COMPLETED	1/1	4/4	43 minutes ago	0s
62	reduce	COMPLETED	1/1	4/4	38 minutes ago	4s
63	collect	COMPLETED	1/1	4/4	37 minutes ago	3s
64	reduce	COMPLETED	1/1	4/4	36 minutes ago	2s
65	reduce	COMPLETED	1/1	4/4	34 minutes ago	2s
66	reduce	COMPLETED	1/1	4/4	32 minutes ago	2s
67	reduce	COMPLETED	1/1	4/4	31 minutes ago	4s
68	reduce	COMPLETED	1/1	4/4	28 minutes ago	4s
69	reduce	COMPLETED	1/1	4/4	25 minutes ago	4s
70	reduce	COMPLETED	1/1	4/4	24 minutes ago	0s
71	reduce	COMPLETED	1/1	4/4	23 minutes ago	4s
72	collect	COMPLETED	1/1	4/4	23 minutes ago	6s
73	reduce	COMPLETED	1/1	8/8	22 minutes ago	15s
74	reduce	COMPLETED	1/1	4/4	17 minutes ago	4s
75	reduce	COMPLETED	1/1	4/4	7 minutes ago	4s
76	reduce	COMPLETED	1/1	4/4	5 minutes ago	2s
77	reduce	COMPLETED	1/1	4/4	4 minutes ago	3s
78	reduce	COMPLETED	1/1	4/4	a few seconds ago	7s
79	reduce	COMPLETED	1/1	4/4	a few seconds ago	1s

Total read time: 1.6 h

Total map reduce time: 80s



home.cern