

Using Multi-System Monitoring Time Series to Predict Performance Events

Andreas Schörgenhumer

Mario Kahlhofer

Peter Chalupar

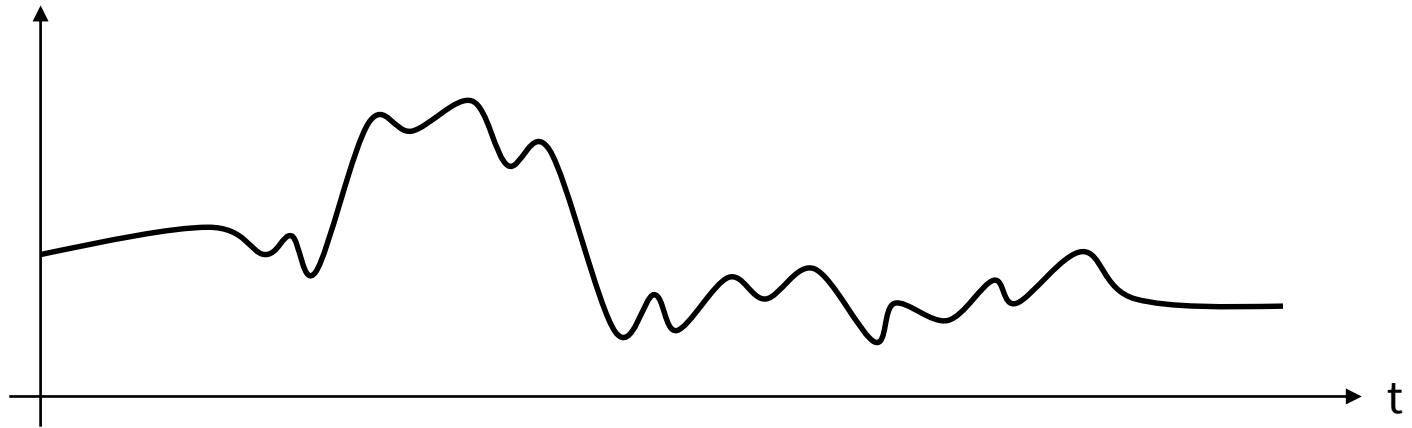
Hanspeter Mössenböck

Paul Grünbacher

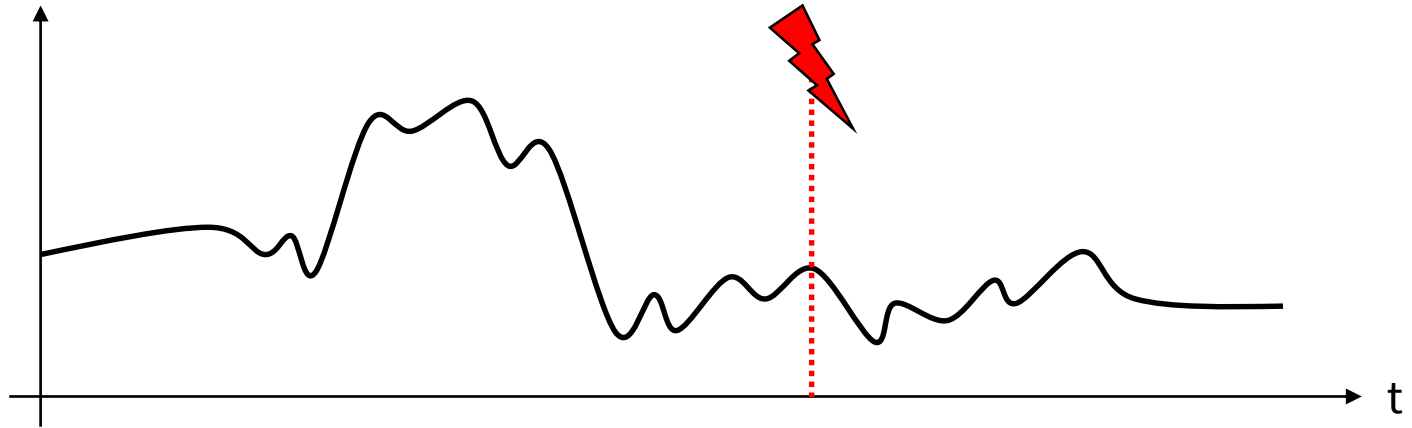
09.11.2018



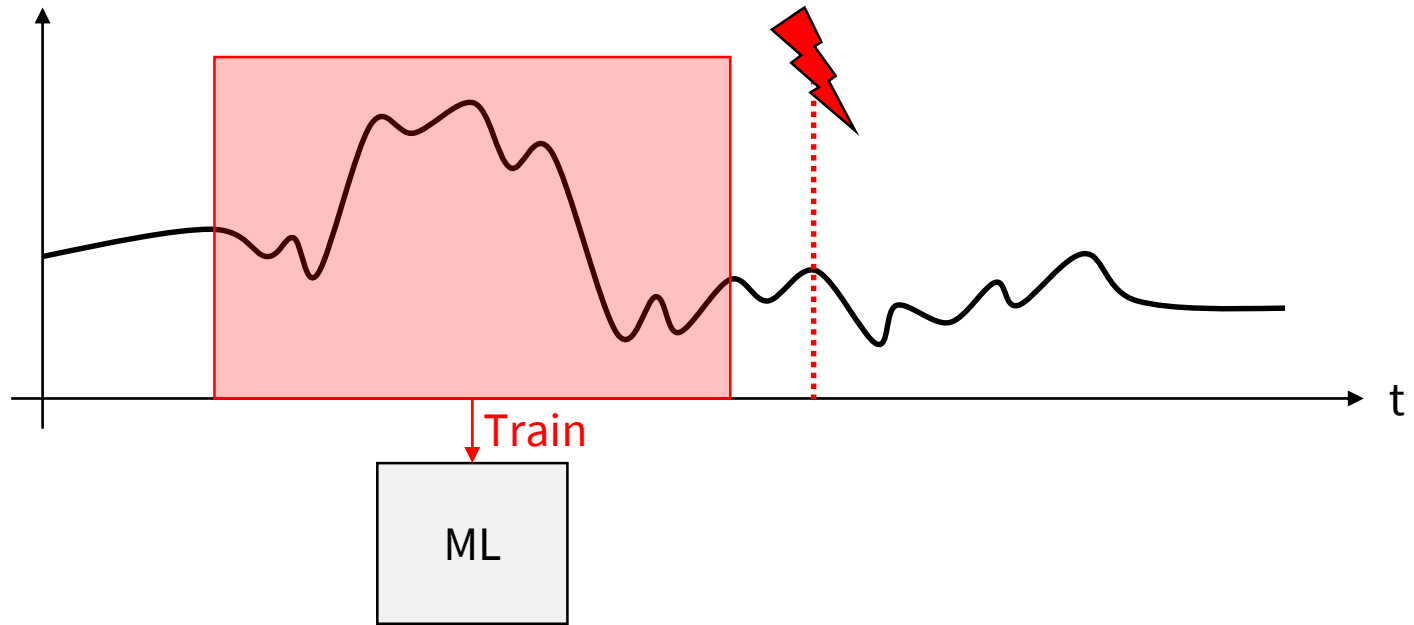
Motivation



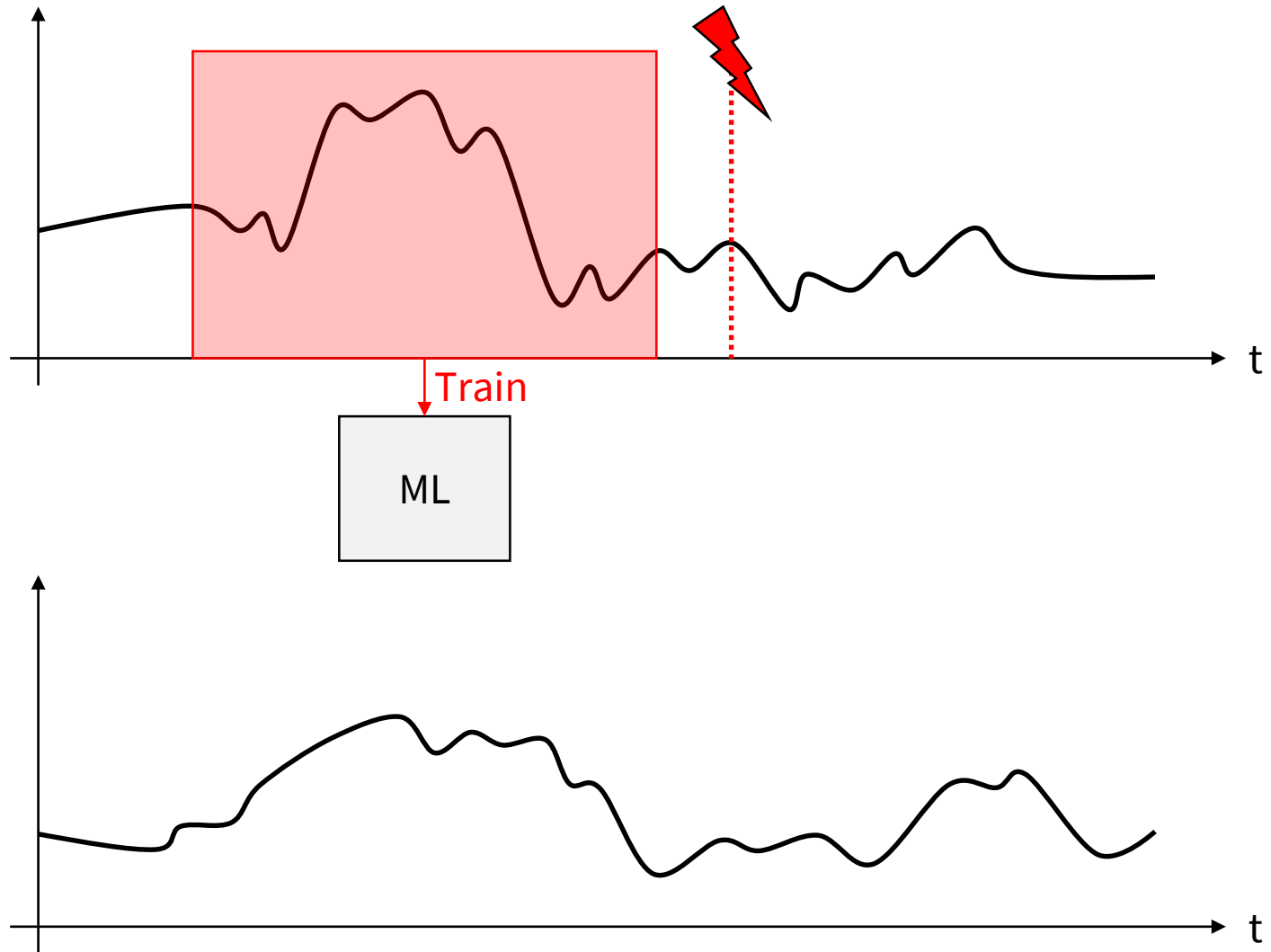
Motivation



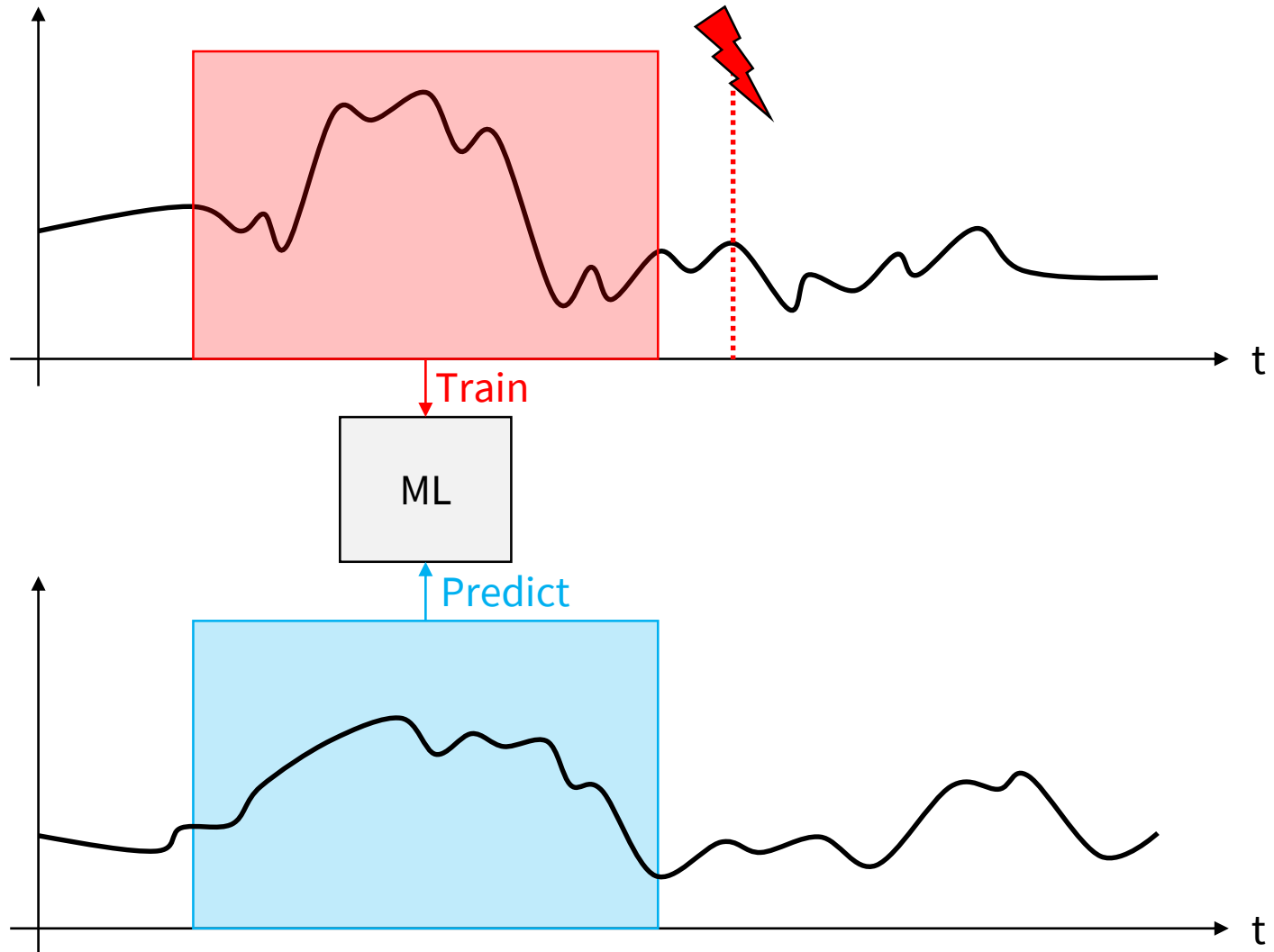
Motivation



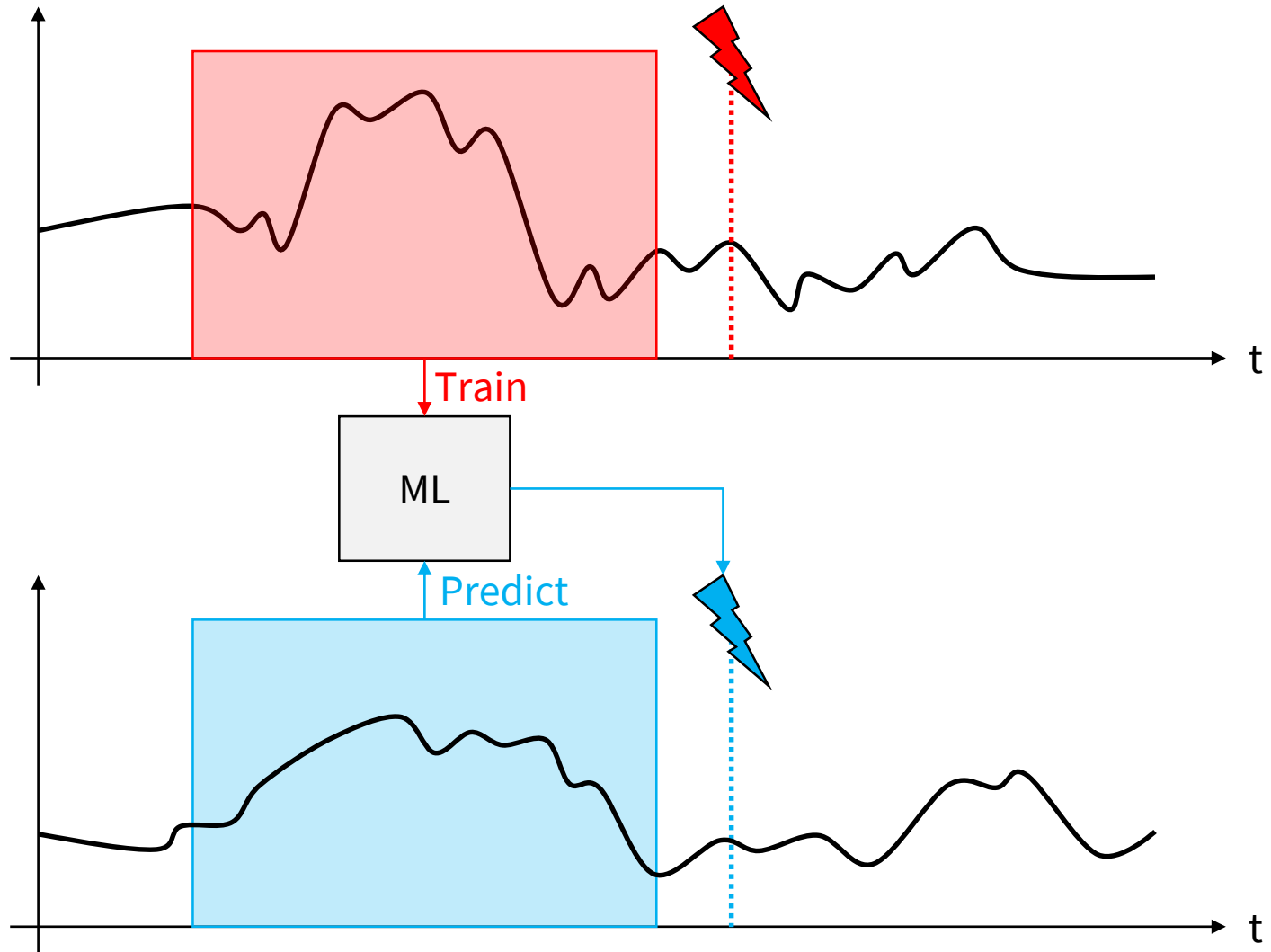
Motivation



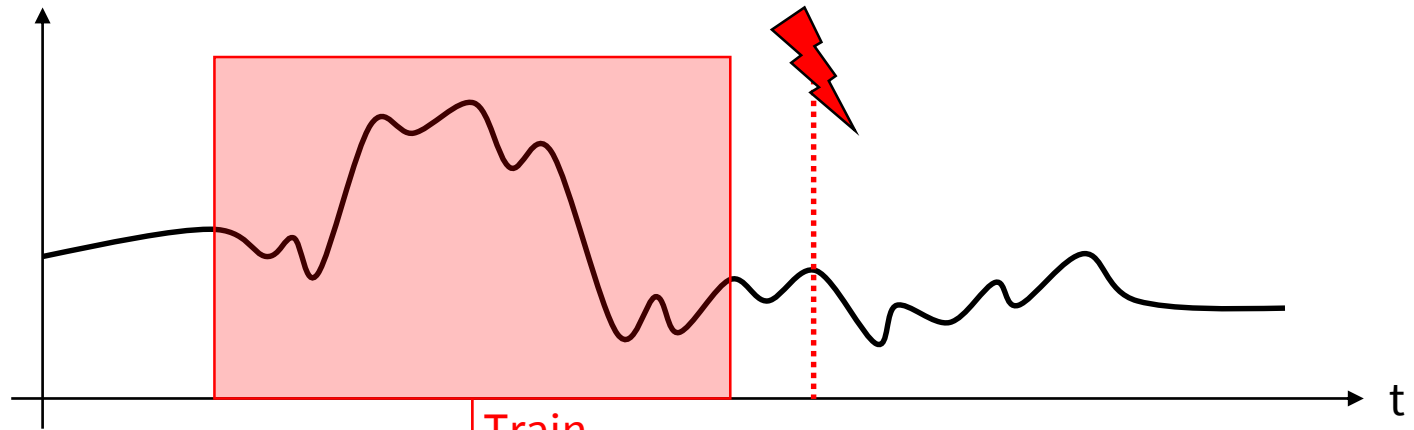
Motivation



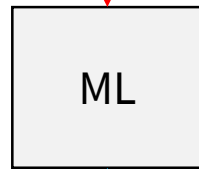
Motivation



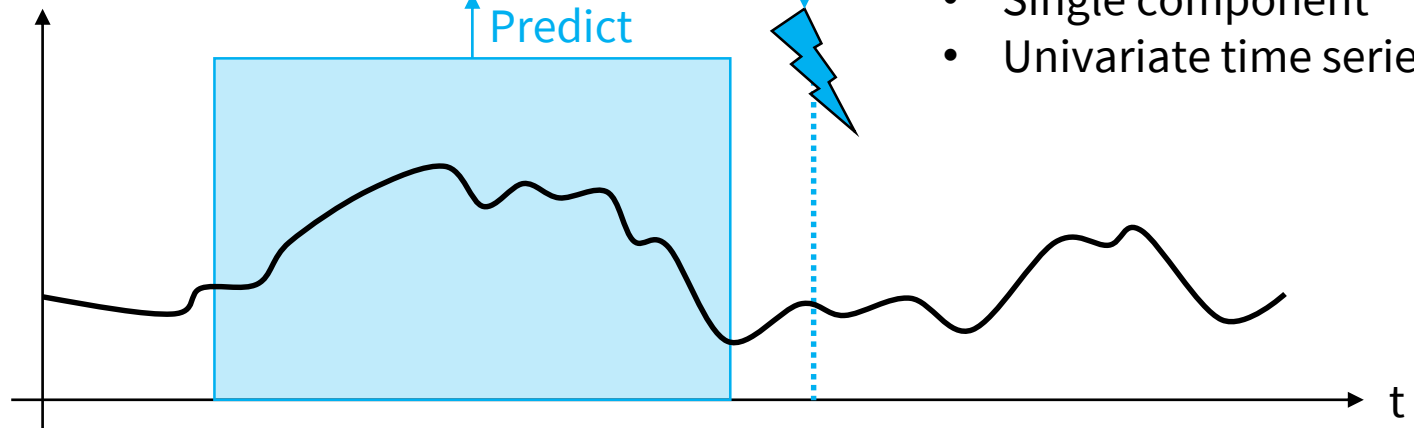
Motivation



Train



Predict

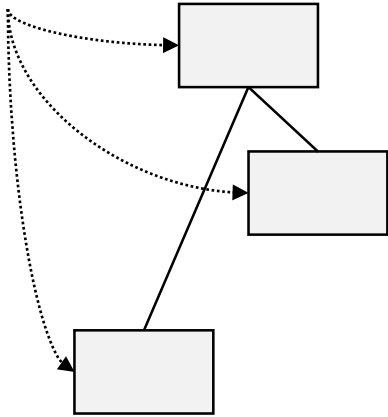


Straightforward:

- Single system
- Single component
- Univariate time series

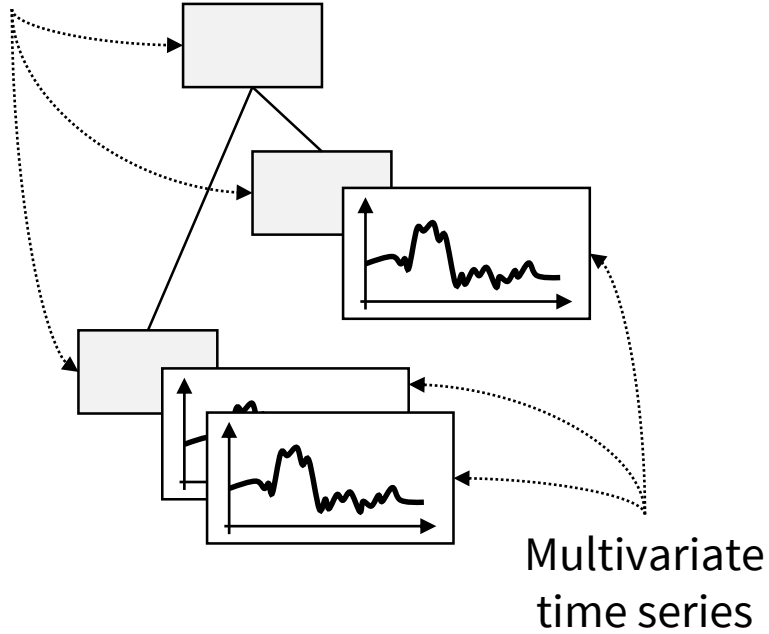
Motivation

Multiple,
interlinked
components

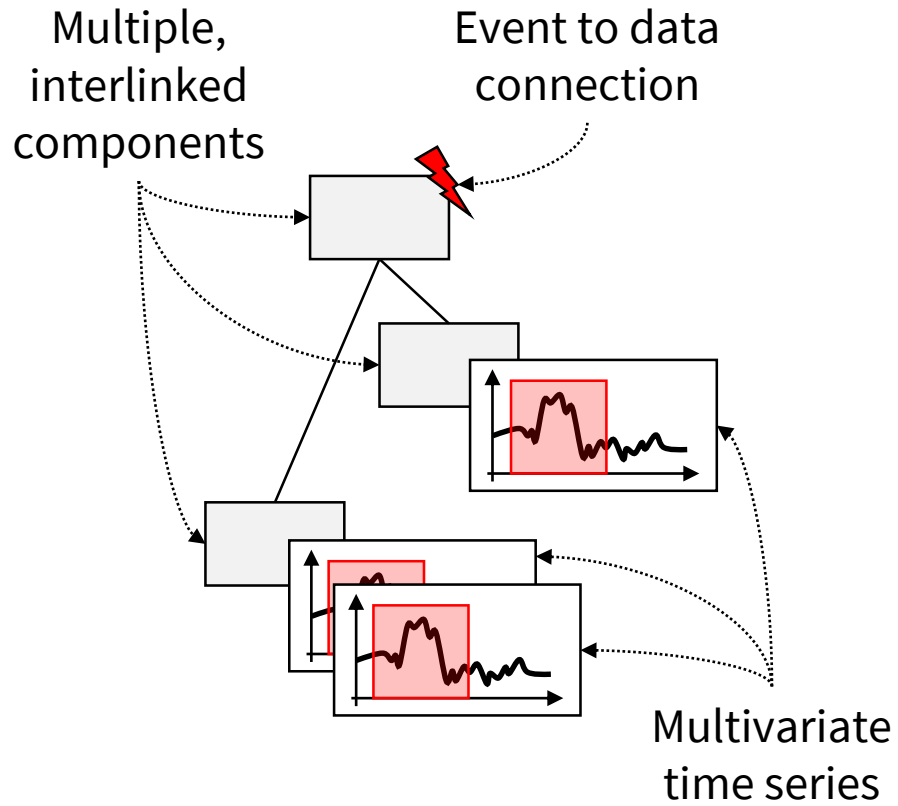


Motivation

Multiple,
interlinked
components

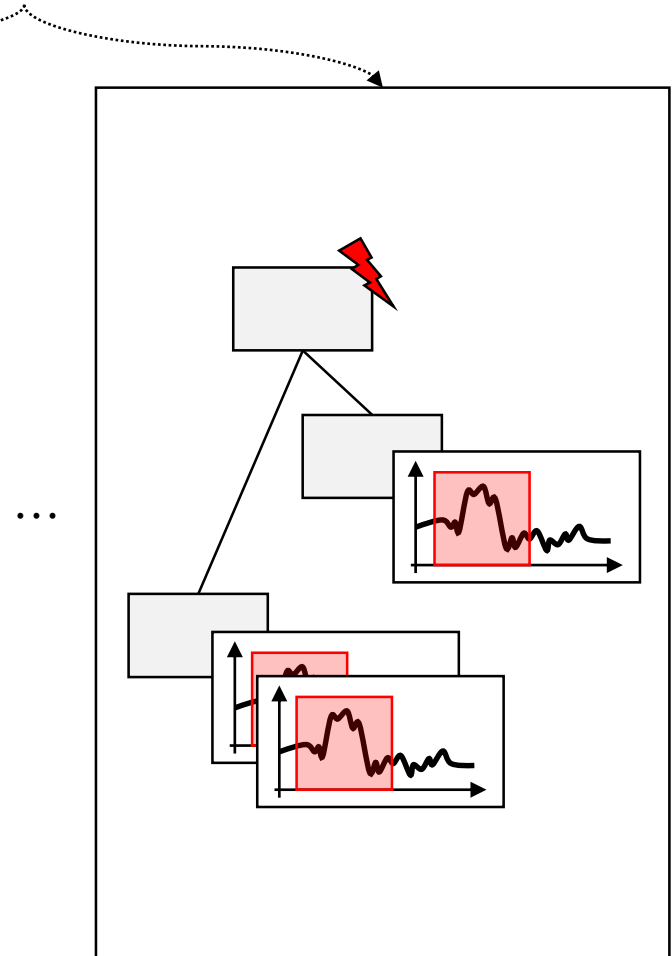
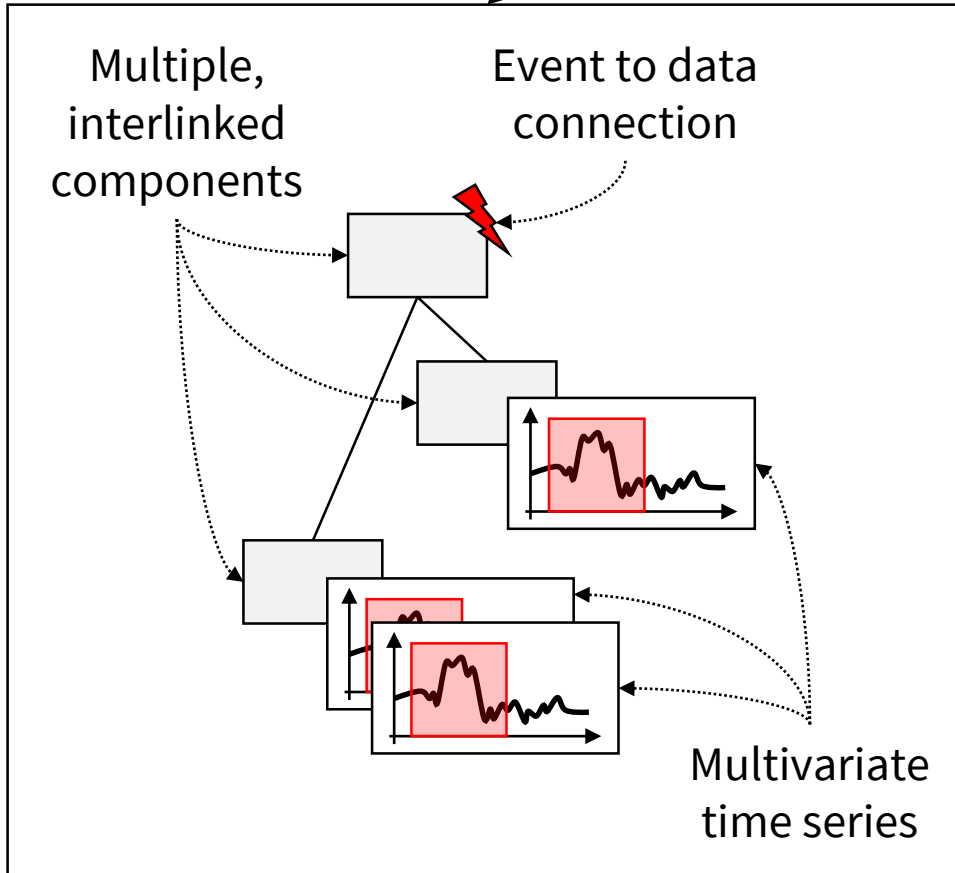


Motivation

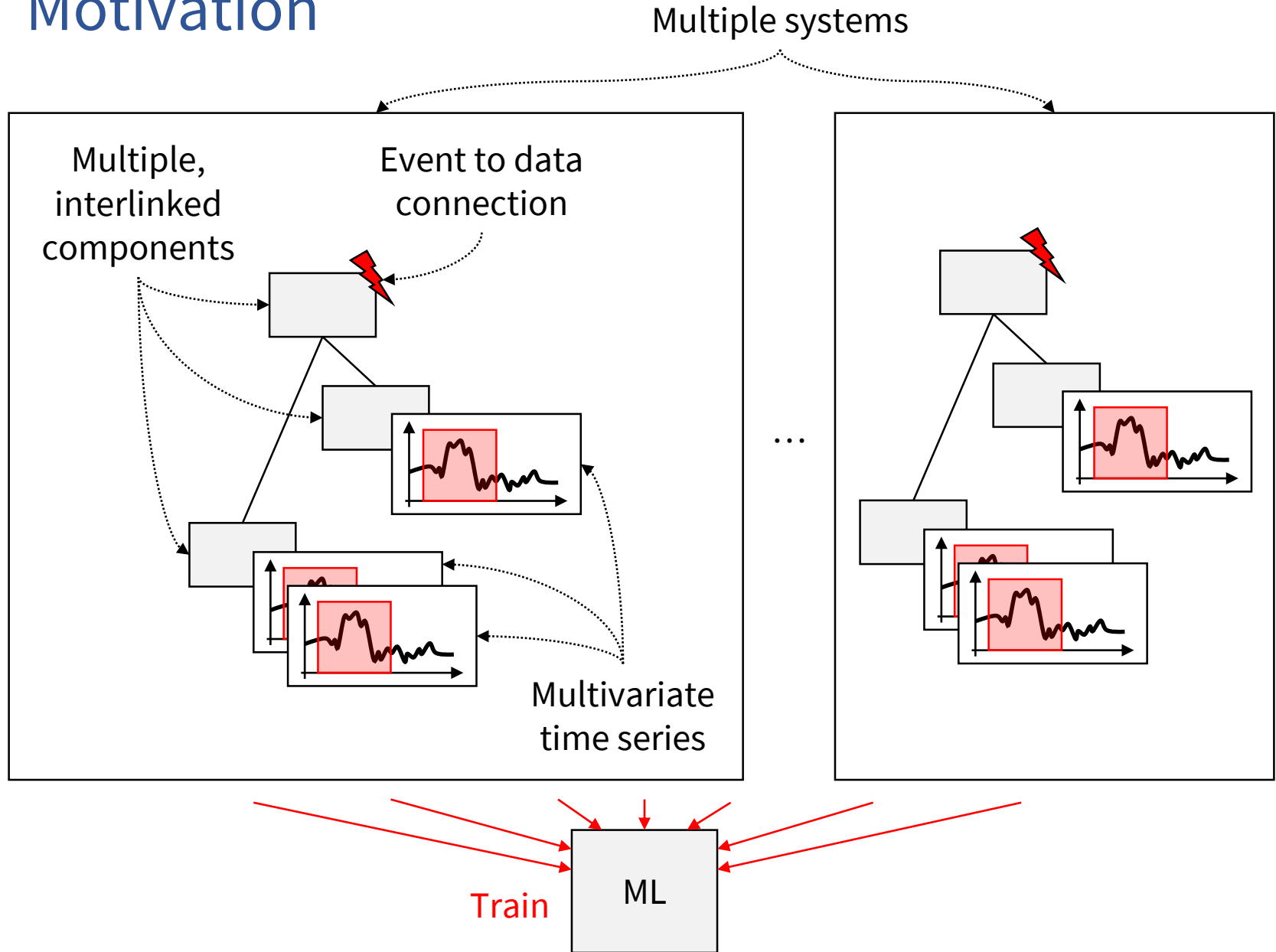


Motivation

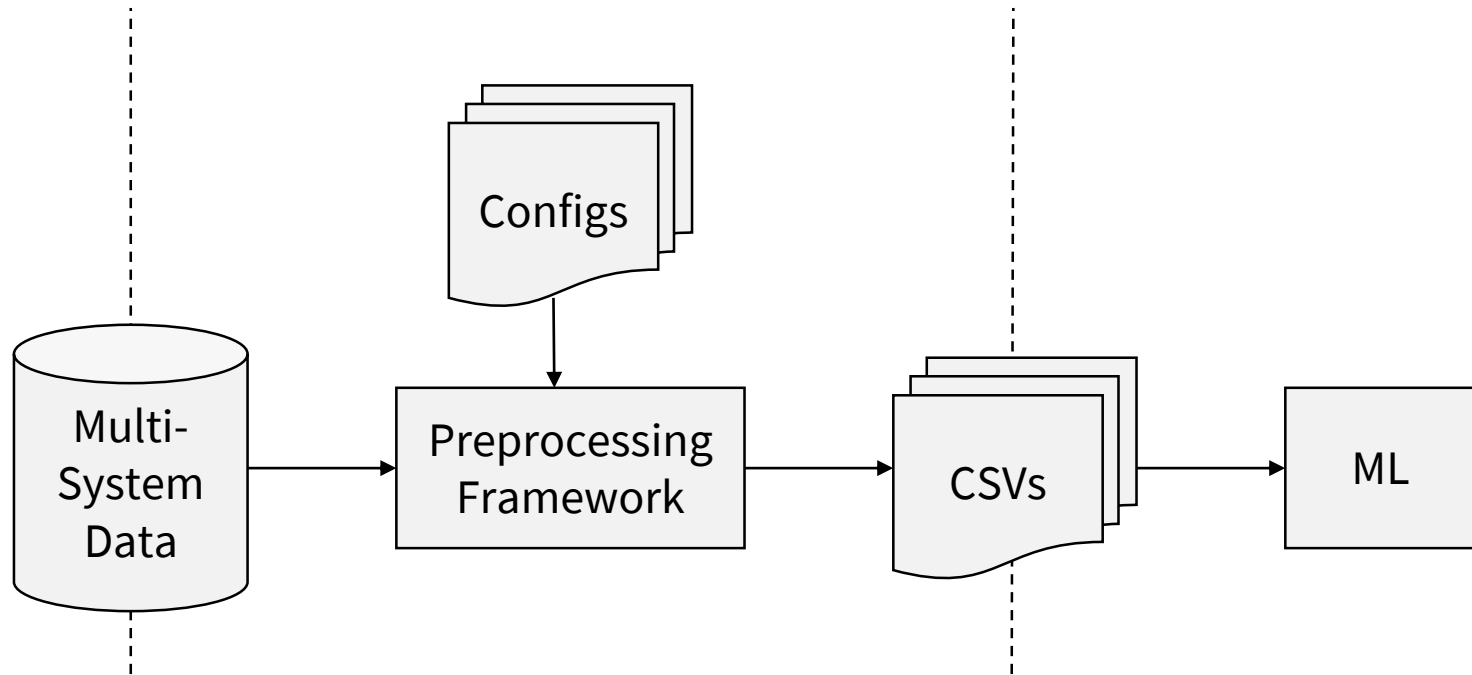
Multiple systems



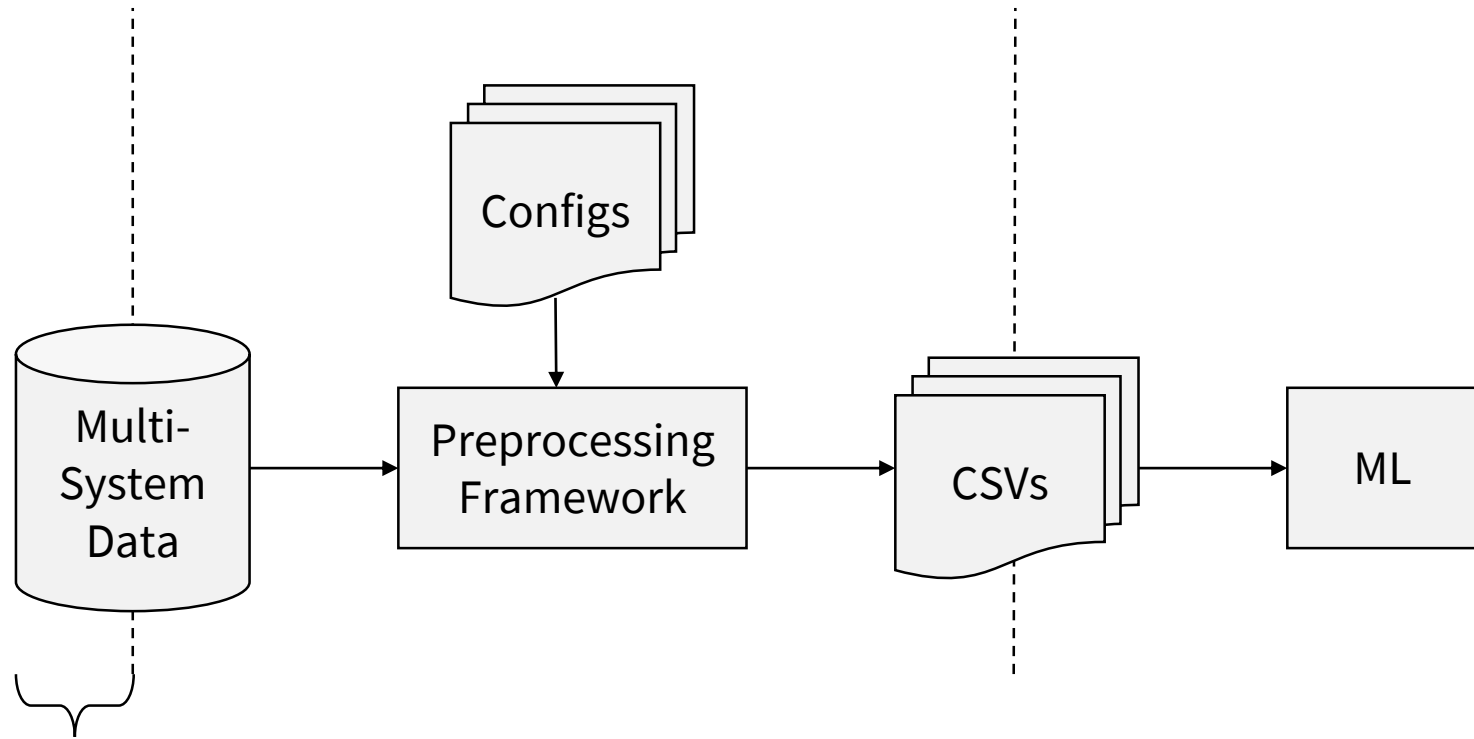
Motivation



Approach

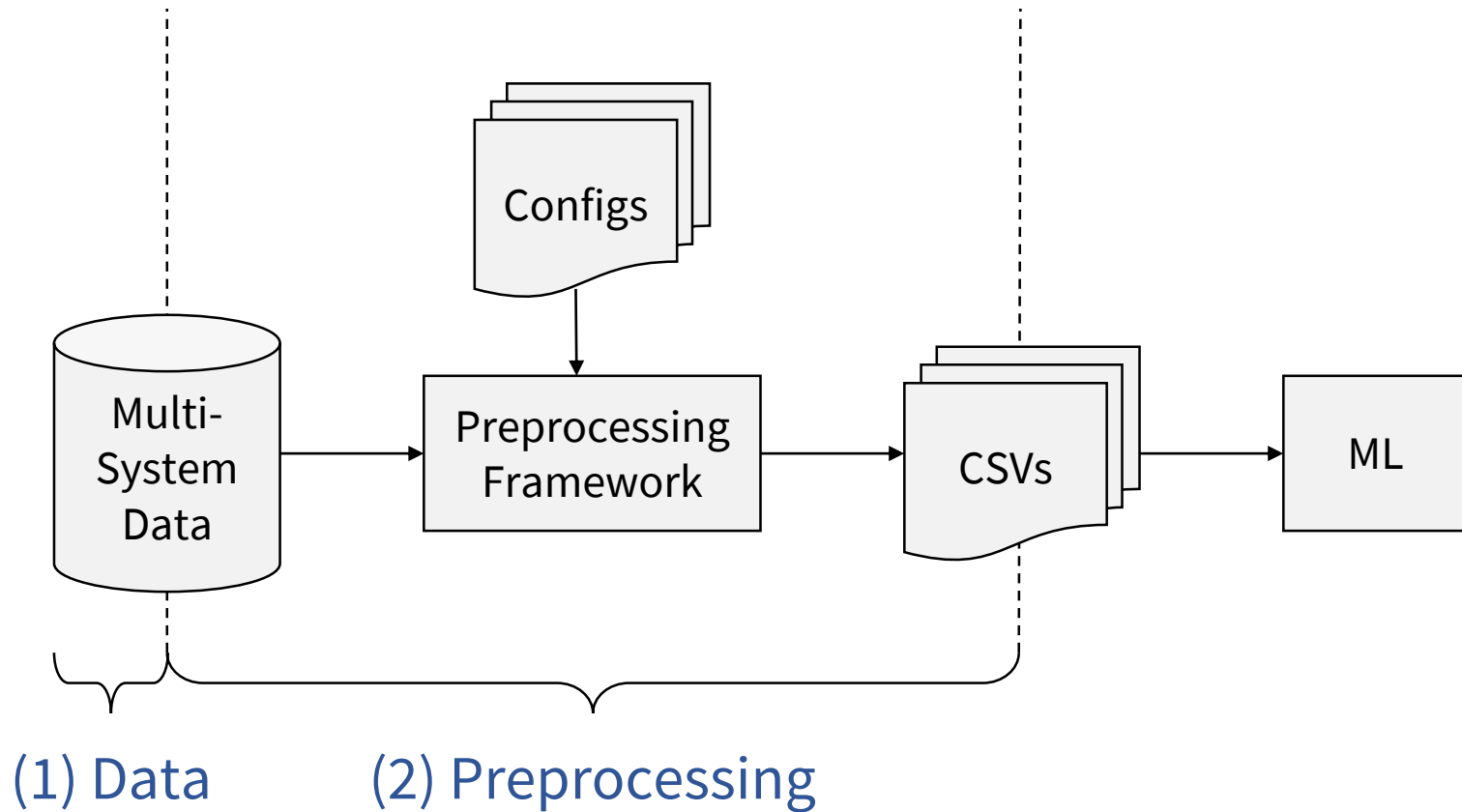


Approach

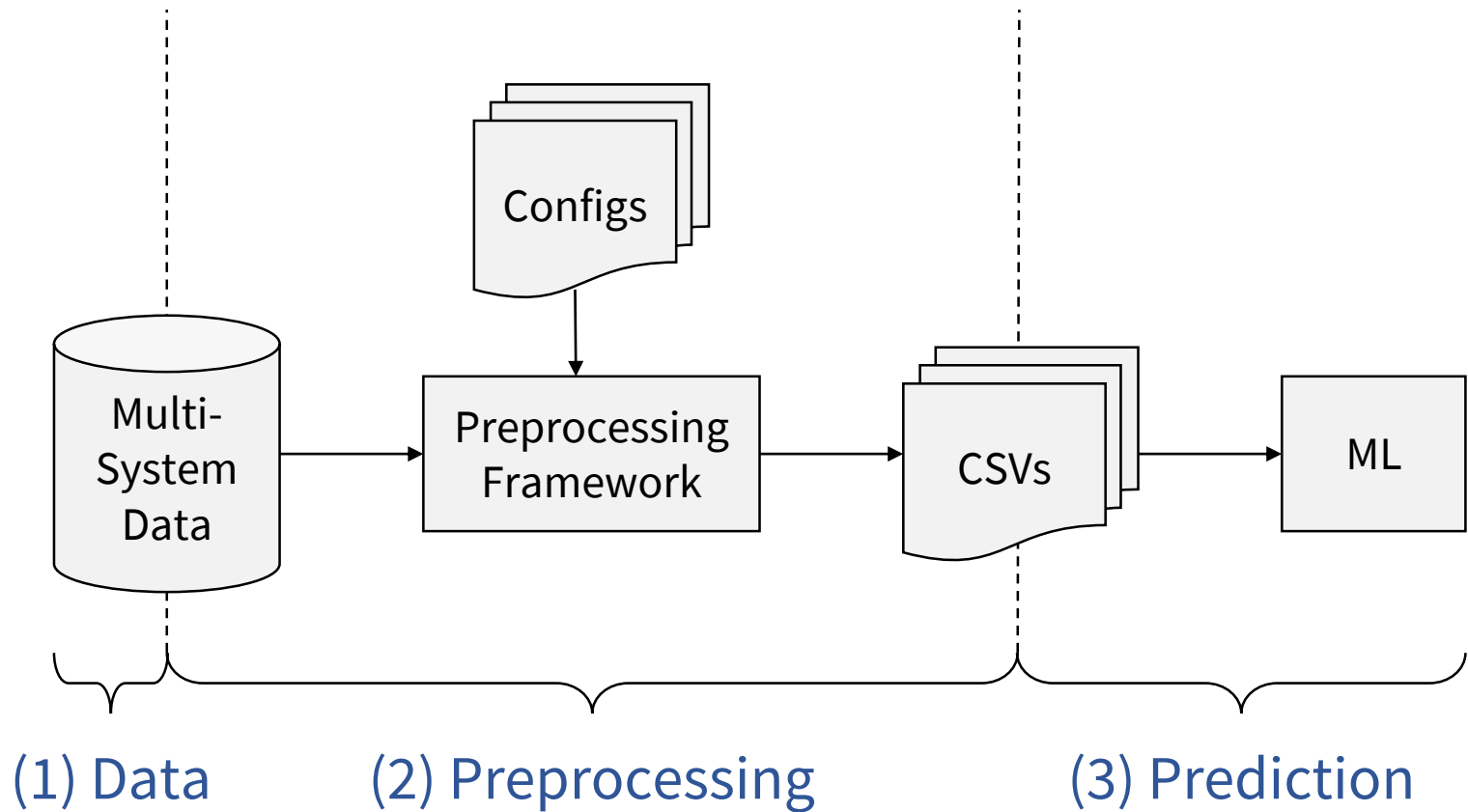


(1) Data

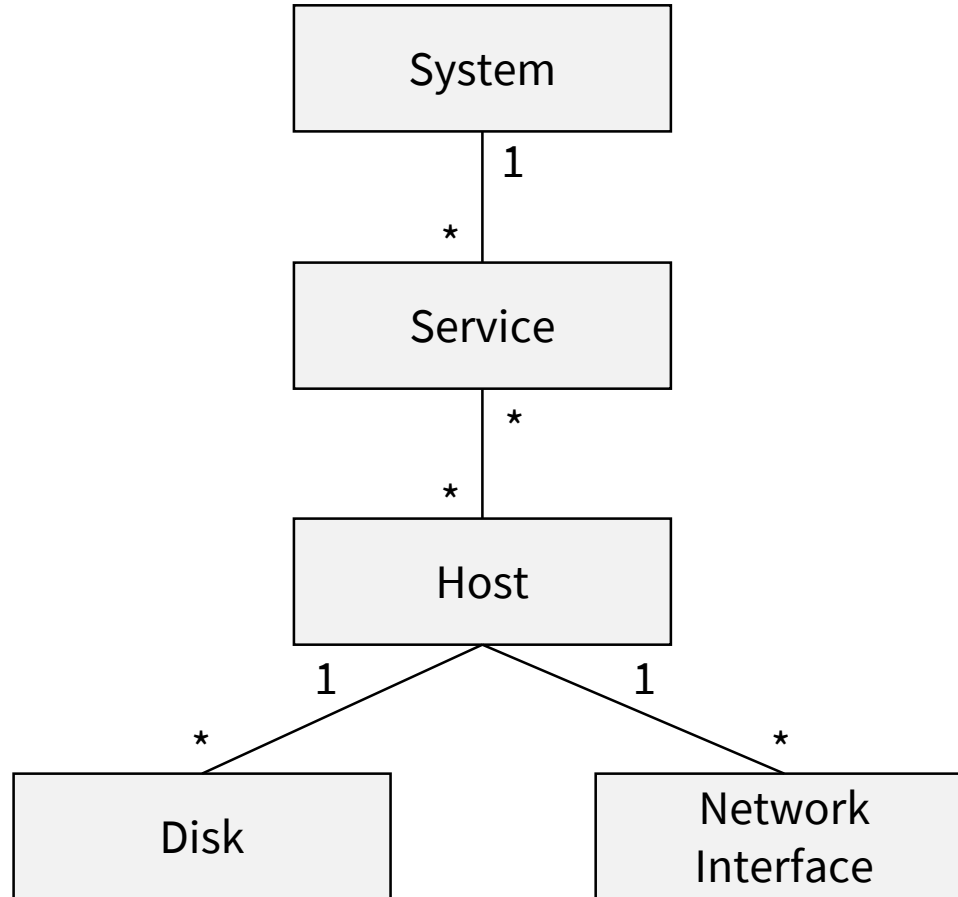
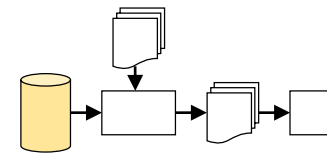
Approach



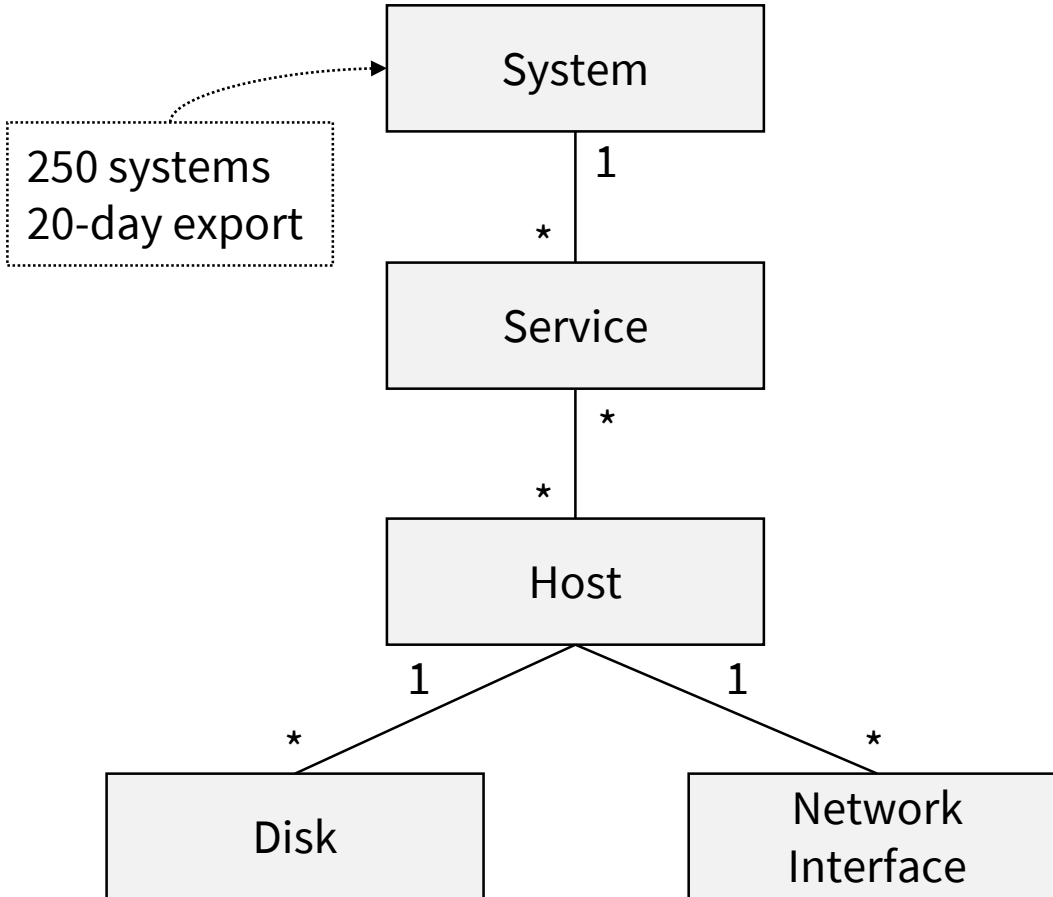
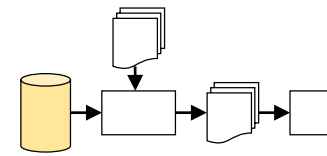
Approach



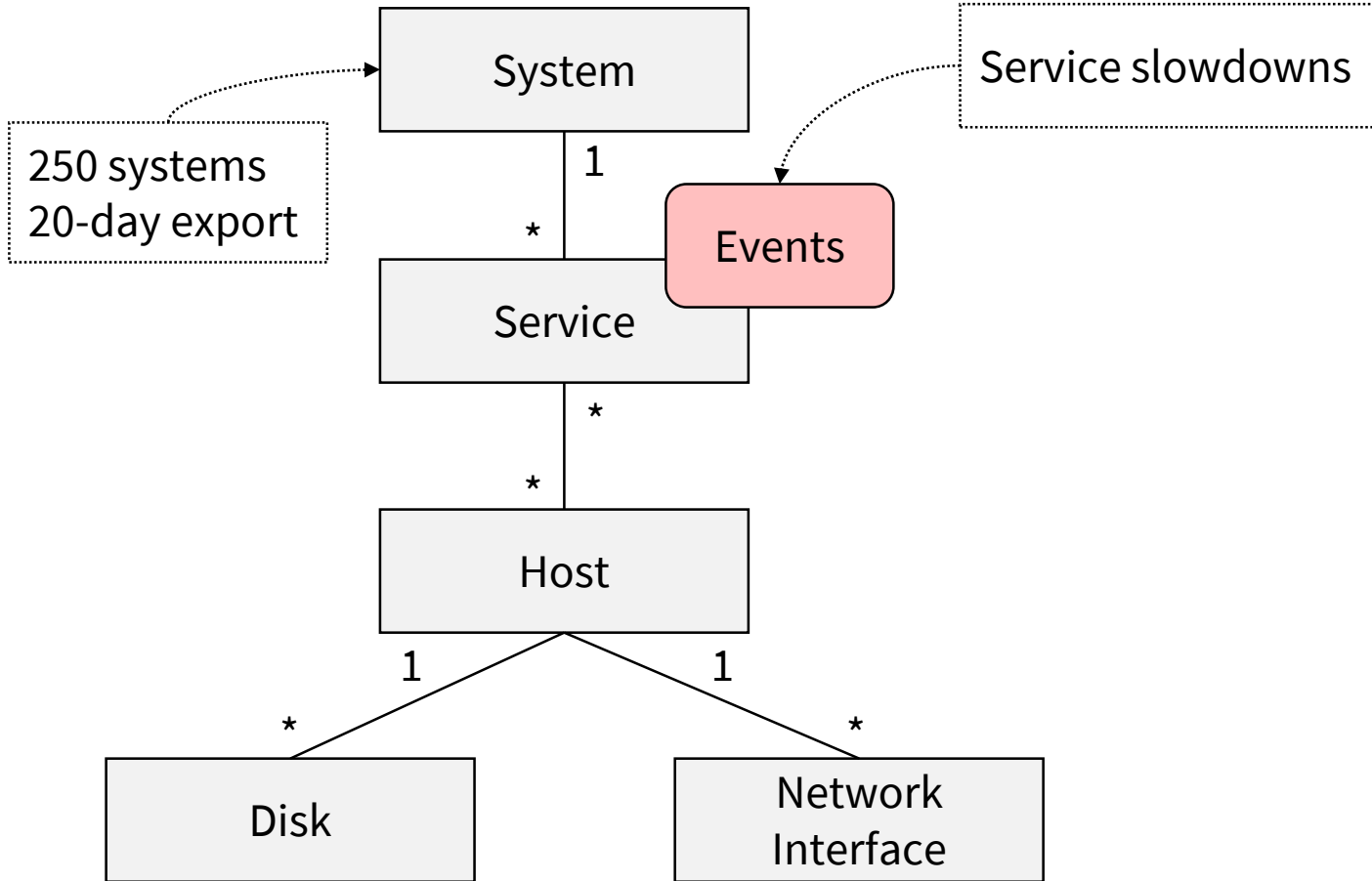
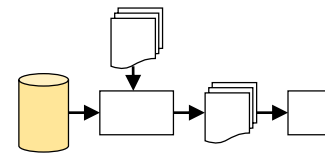
(1) Data



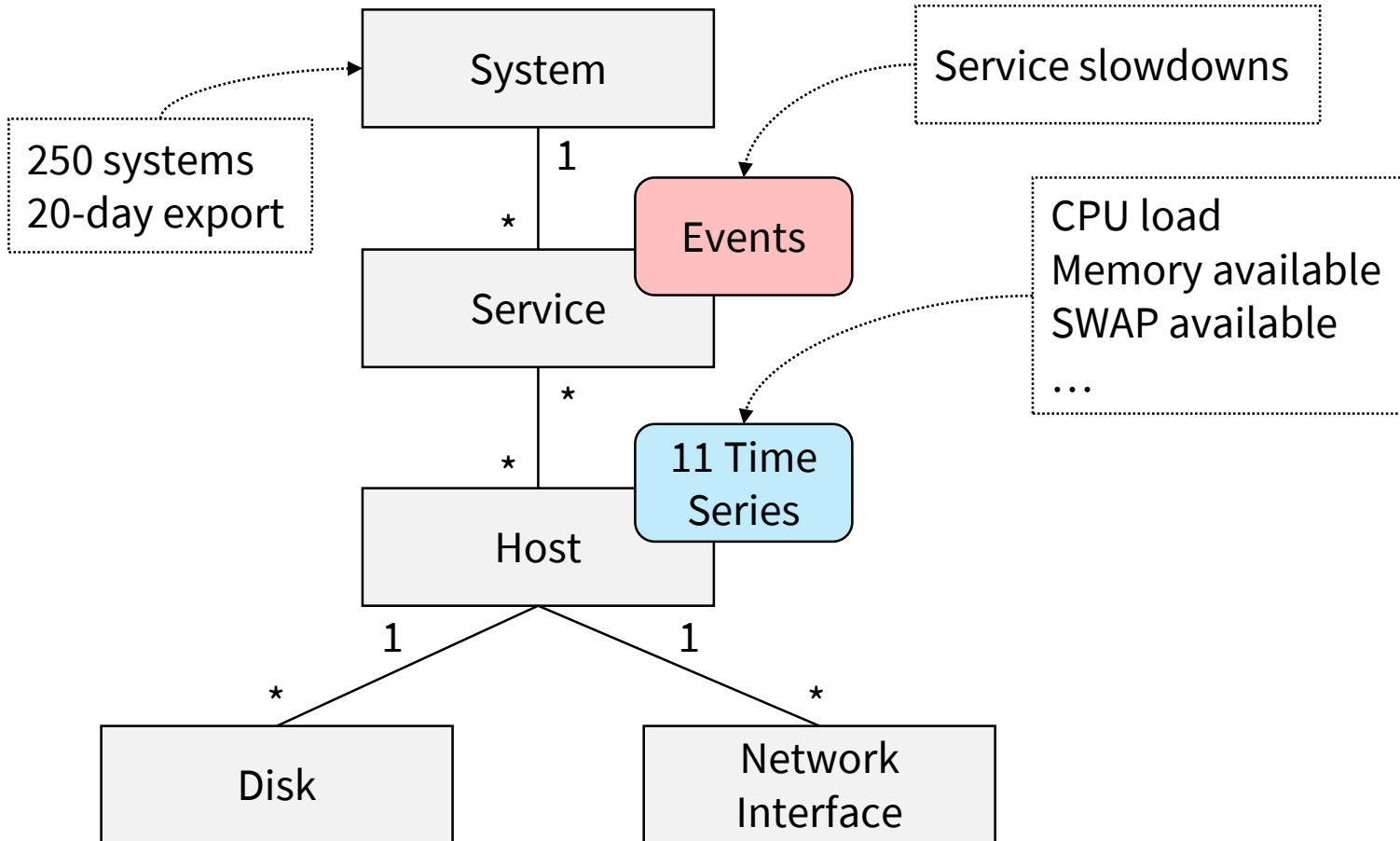
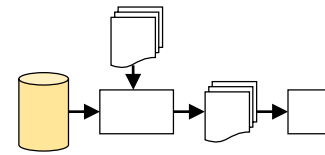
(1) Data



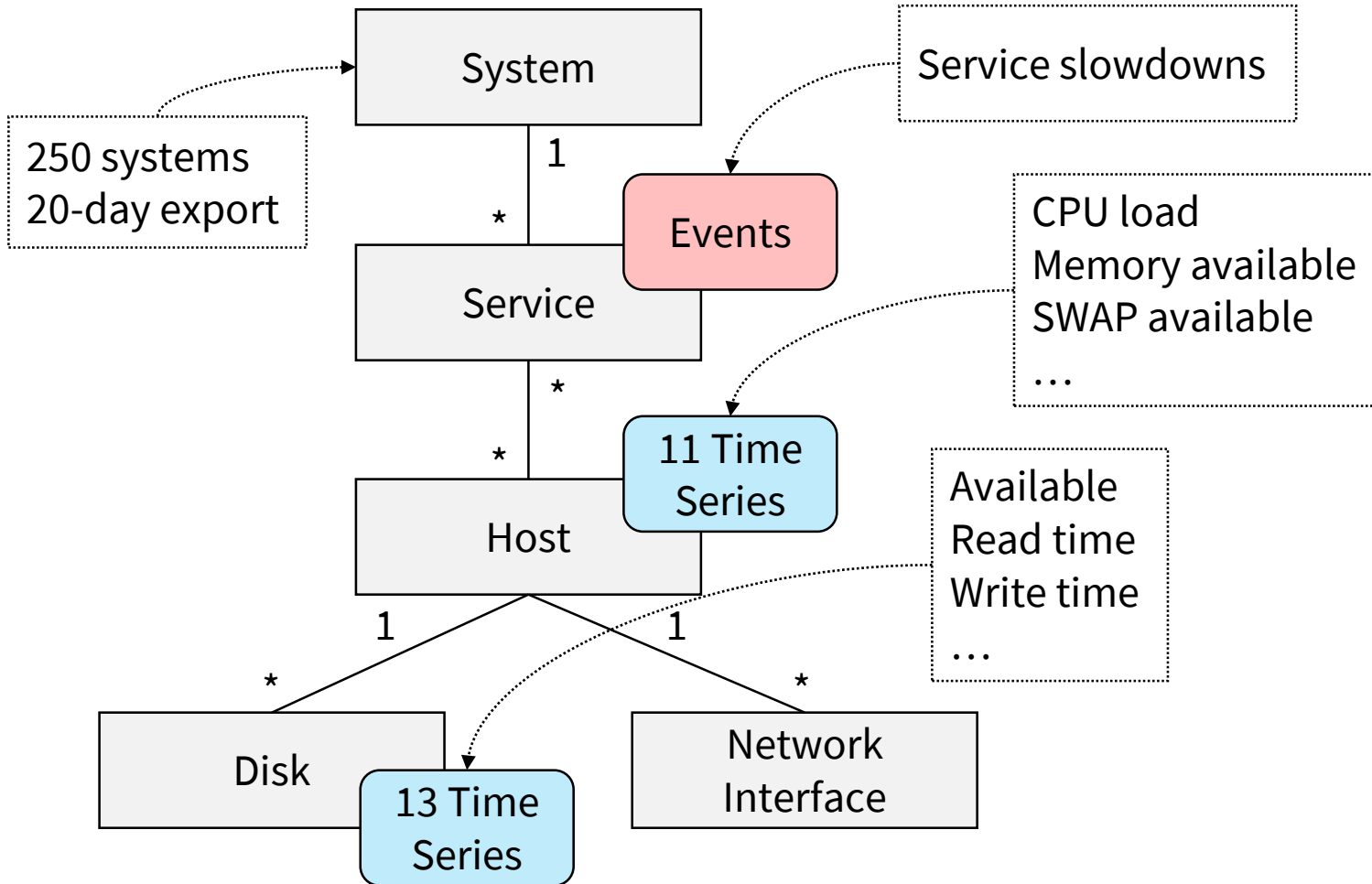
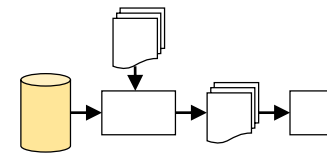
(1) Data



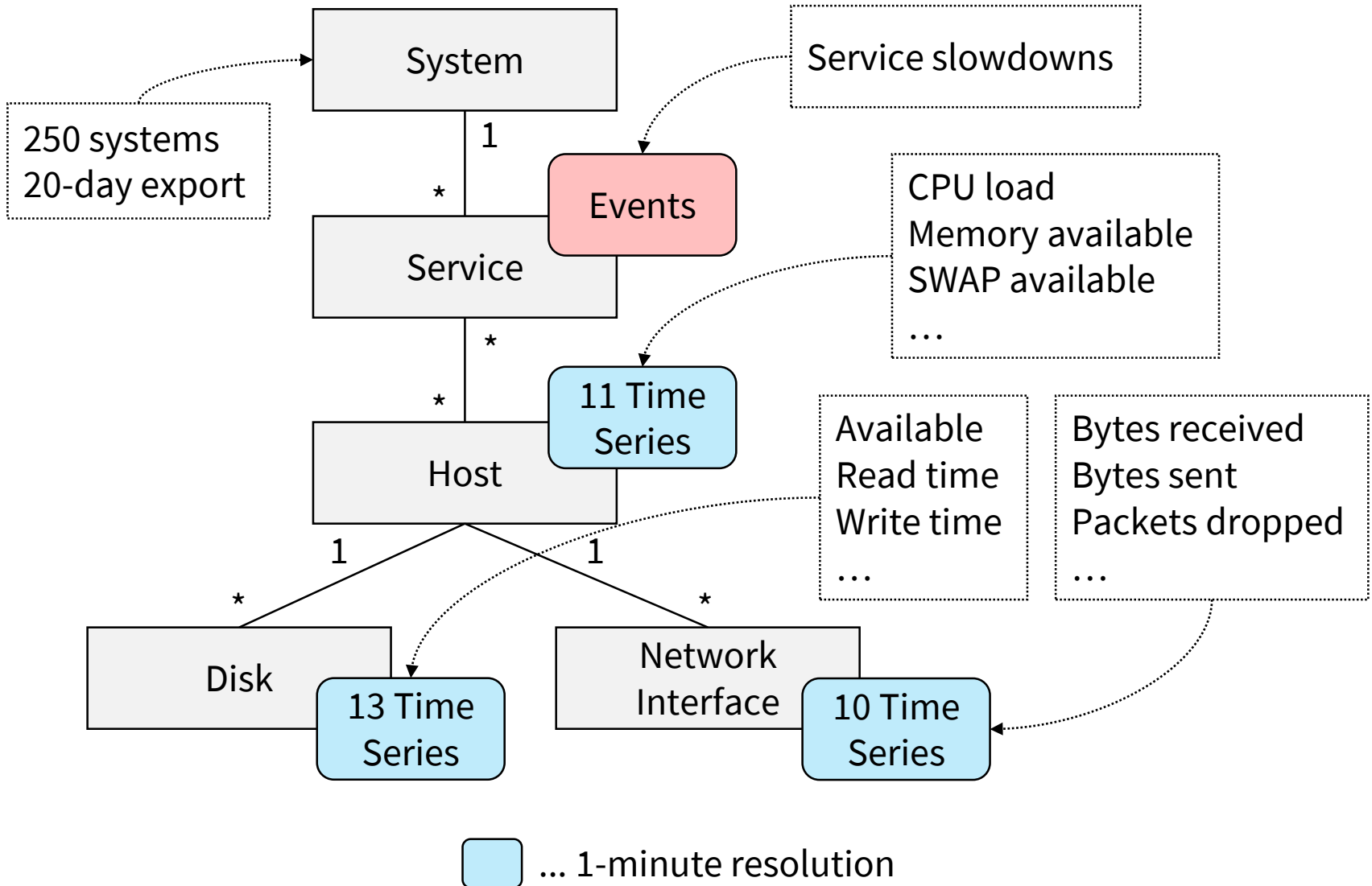
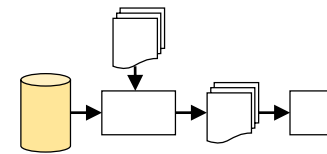
(1) Data



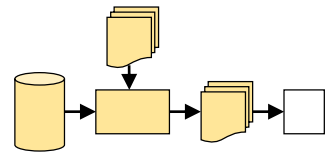
(1) Data



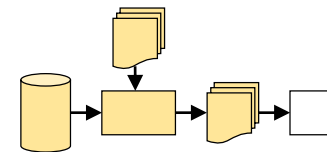
(1) Data



(2) Preprocessing – Framework

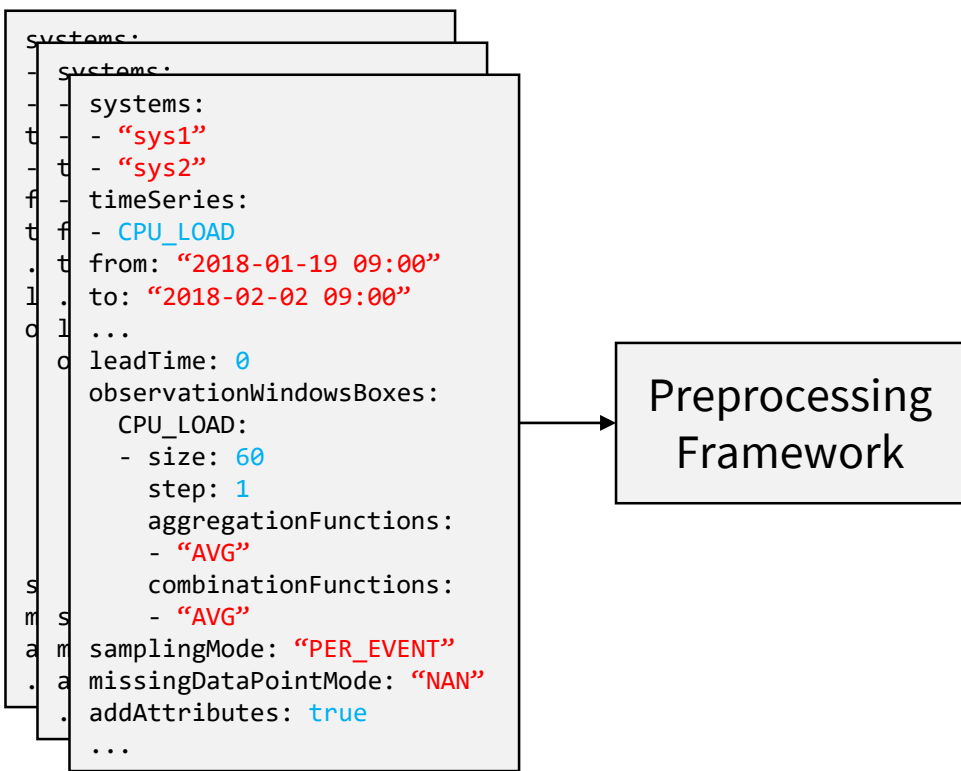


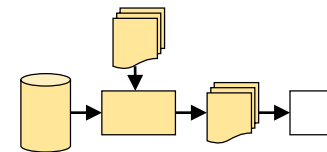
Preprocessing
Framework



(2) Preprocessing – Framework

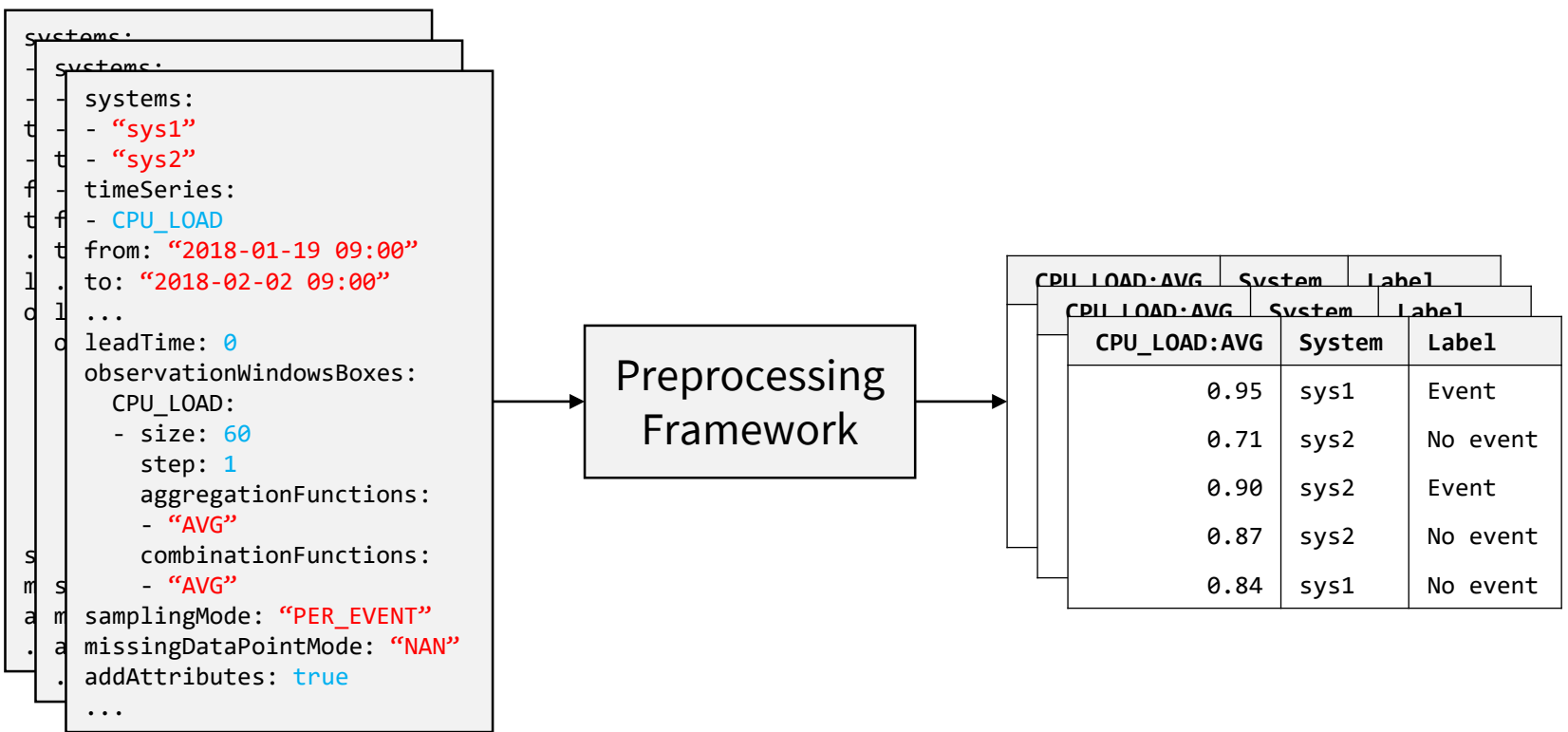
- **Input: YAMLS (configurations/configs)**
 - Contains all necessary data processing settings
 - Easily changeable due to YAML format



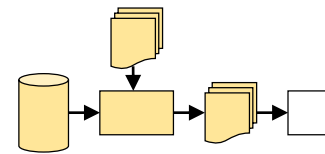


(2) Preprocessing – Framework

- **Input: YAMLS (configurations/configs)**
 - Contains all necessary data processing settings
 - Easily changeable due to YAML format
- **Output: CSVs (feature vectors)**
 - Portable format, directly useable for ML

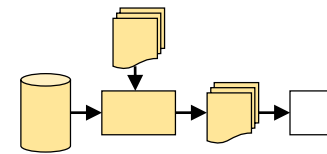


(2) Preprocessing – Config Settings



Setting	Example
Systems	<code>[sys1, sys2, ...]</code>
Time series	<code>[Host: CPU_LOAD, Disk: AVAILABLE, ...]</code>
Time frame	From: <code>19-01-2018 09:00</code> To: <code>02-02-2018 09:00</code>
Sampling mode	<code>PER_EVENT, SLIDE_THROUGH</code>
Negative sampling source	<code>NON_EVENT_SERVICES, EVENT_SERVICES, ...</code>
Lead time	<code>10 min</code>
Observation windows	<code>60 min, AVG aggregation, AVG combination</code>
Missing data mode	<code>DROP, NAN, LAST_VALUE, ...</code>
Metadata	<code>System, special attributes, ...</code>
...	...

(2) Preprocessing – Example

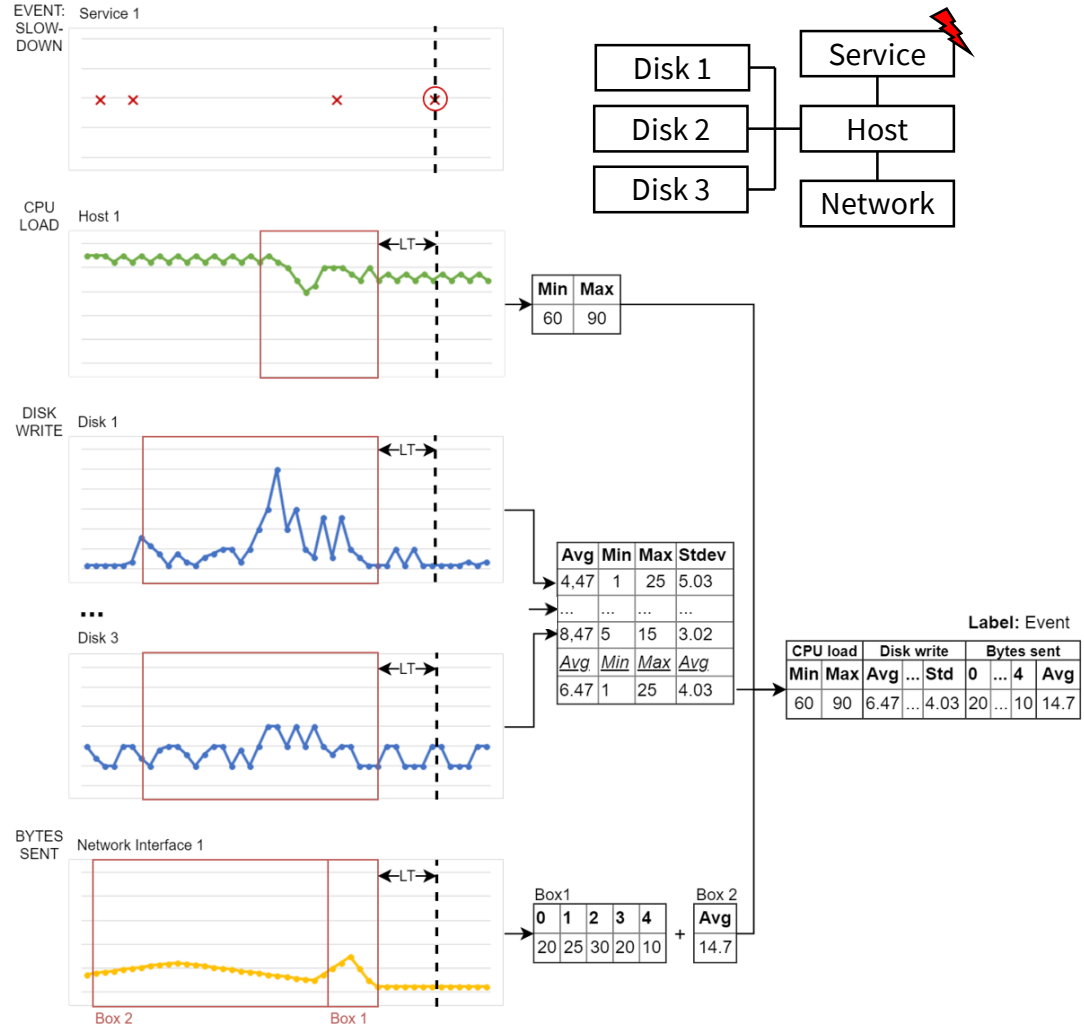


```

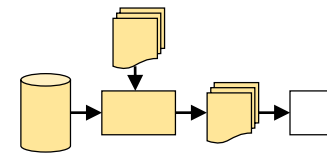
...
samplingMode: "PER_EVENT"
leadTime: 5
observationWindowsBoxes:
  CPU_LOAD:
    - size: 15
      aggregationFunctions:
        - "MIN"
        - "MAX"
    ...
  DISK_WRITE:
    - size: 30
      aggregationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "STD_DEV"
      combinationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "AVG"
    ...
  BYTES_SENT:
    - size: 5
      aggregationFunctions:
        - "NONE"
    ...
    - size: 30
      aggregationFunctions:
        - "AVG"
    ...
...

```

Sample: Customer System X, Service 1, Timestamp: 2018-02-01 17:30



(2) Preprocessing – Example

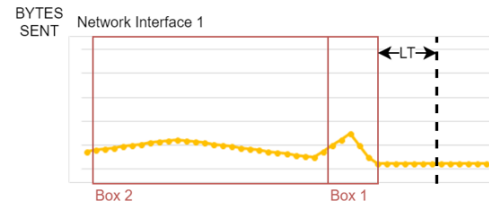
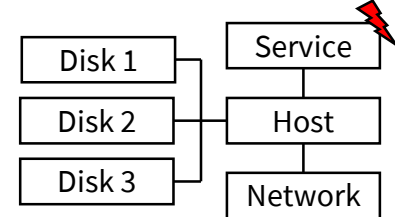
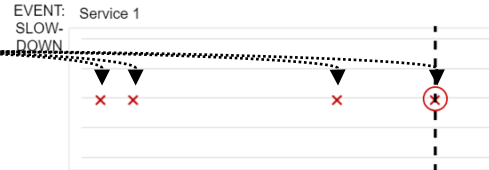


Sample: Customer System X, Service 1, Timestamp: 2018-02-01 17:30

```

...
samplingMode: "PER_EVENT"
leadTime: 5
observationWindowsBoxes:
  CPU_LOAD:
    - size: 15
      aggregationFunctions:
        - "MIN"
        - "MAX"
    ...
  DISK_WRITE:
    - size: 30
      aggregationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "STD_DEV"
      combinationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "AVG"
    ...
  BYTES_SENT:
    - size: 5
      aggregationFunctions:
        - "NONE"
    ...
    - size: 30
      aggregationFunctions:
        - "AVG"
    ...
...

```



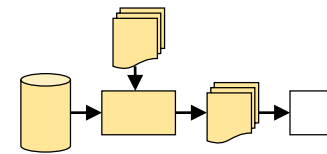
Min	Max
60	90

Avg	Min	Max	Stdev
4.47	1	25	5.03
...
8.47	5	15	3.02
Avg	Min	Max	Avg
6.47	1	25	4.03

Label: Event						
CPU load		Disk write		Bytes sent		
Min	Max	Avg	Std	0	4	Avg
60	90	6.47	4.03	20	10	14.7



(2) Preprocessing – Example

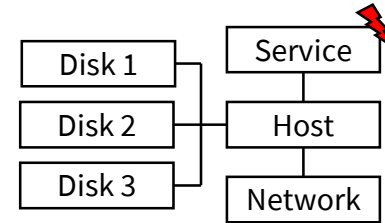
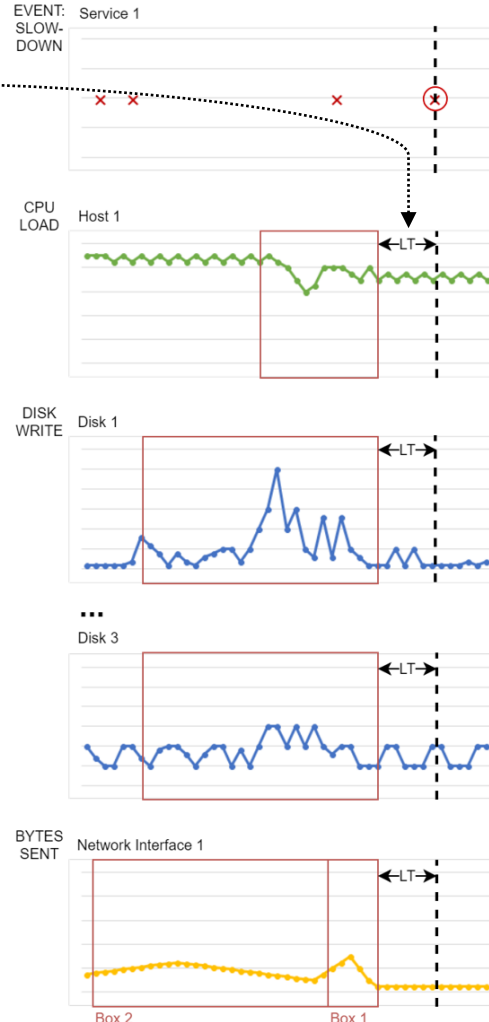


Sample: Customer System X, Service 1, Timestamp: 2018-02-01 17:30

```

...
samplingMode: "PER_EVENT"
leadTime: 5
observationWindowsBoxes:
  CPU_LOAD:
    - size: 15
      aggregationFunctions:
        - "MIN"
        - "MAX"
    ...
  DISK_WRITE:
    - size: 30
      aggregationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "STD_DEV"
      combinationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "AVG"
    ...
  BYTES_SENT:
    - size: 5
      aggregationFunctions:
        - "NONE"
    ...
    - size: 30
      aggregationFunctions:
        - "AVG"
    ...
...

```



Min	Max
60	90

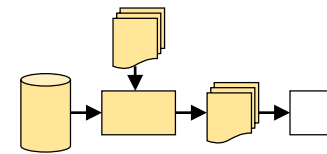
Avg	Min	Max	Stdev
4.47	1	25	5.03
...
8.47	5	15	3.02
...
6.47	1	25	4.03

CPU load		Disk write			Bytes sent			Label: Event
Min	Max	Avg	...	Std	0	...	4	Avg
60	90	6.47	...	4.03	20	...	10	14.7

Box 1					Box 2
0	1	2	3	4	Avg
20	25	30	20	10	14.7



(2) Preprocessing – Example

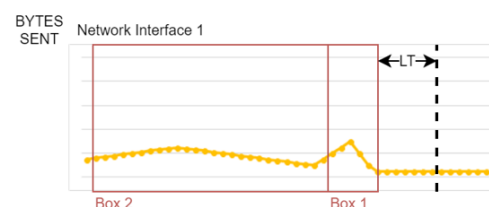
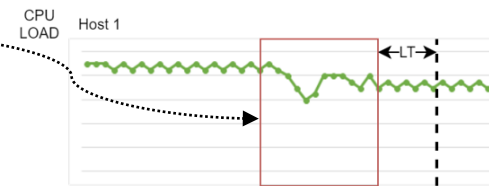
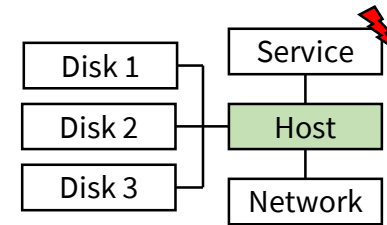


Sample: Customer System X, Service 1, Timestamp: 2018-02-01 17:30

```

...
samplingMode: "PER_EVENT"
leadTime: 5
observationWindowsBoxes:
  CPU_LOAD:
    - size: 15
    aggregationFunctions:
      - "MIN"
      - "MAX"
    ...
  DISK_WRITE:
    - size: 30
    aggregationFunctions:
      - "AVG"
      - "MIN"
      - "MAX"
      - "STD_DEV"
    combinationFunctions:
      - "AVG"
      - "MIN"
      - "MAX"
      - "AVG"
    ...
  BYTES_SENT:
    - size: 5
    aggregationFunctions:
      - "NONE"
    ...
    - size: 30
    aggregationFunctions:
      - "AVG"
    ...
...

```



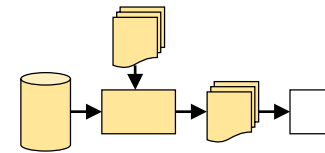
Min	Max
60	90

Avg	Min	Max	Stdev
4.47	1	25	5.03
...
8.47	5	15	3.02
6.47	1	25	4.03

CPU load		Disk write			Bytes sent		Label: Event		
Min	Max	Avg	...	Std	0	...	4	Avg	
60	90	6.47	...	4.03	20	...	10	14.7	



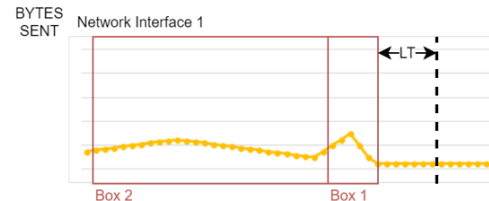
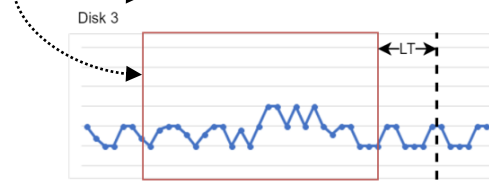
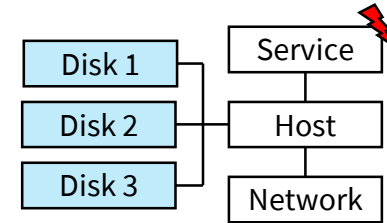
(2) Preprocessing – Example



Sample: Customer System X, Service 1, Timestamp: 2018-02-01 17:30

```

...
samplingMode: "PER_EVENT"
leadTime: 5
observationWindowsBoxes:
  CPU_LOAD:
    - size: 15
      aggregationFunctions:
        - "MIN"
        - "MAX"
    ...
  DISK_WRITE:
    - size: 30
      aggregationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "STD_DEV"
      combinationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "AVG"
    ...
  BYTES_SENT:
    - size: 5
      aggregationFunctions:
        - "NONE"
    ...
    - size: 30
      aggregationFunctions:
        - "AVG"
    ...
...
  
```



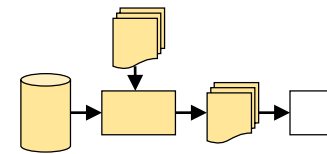
Min	Max
60	90

Avg	Min	Max	Stdev
4.47	1	25	5.03
...
8.47	5	15	3.02
Avg	Min	Max	Avg
6.47	1	25	4.03

CPU load		Disk write			Bytes sent			Label: Event
Min	Max	Avg	...	Std	0	...	4	Avg
60	90	6.47	...	4.03	20	...	10	14.7



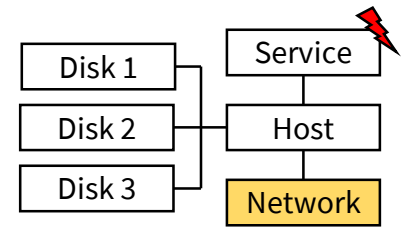
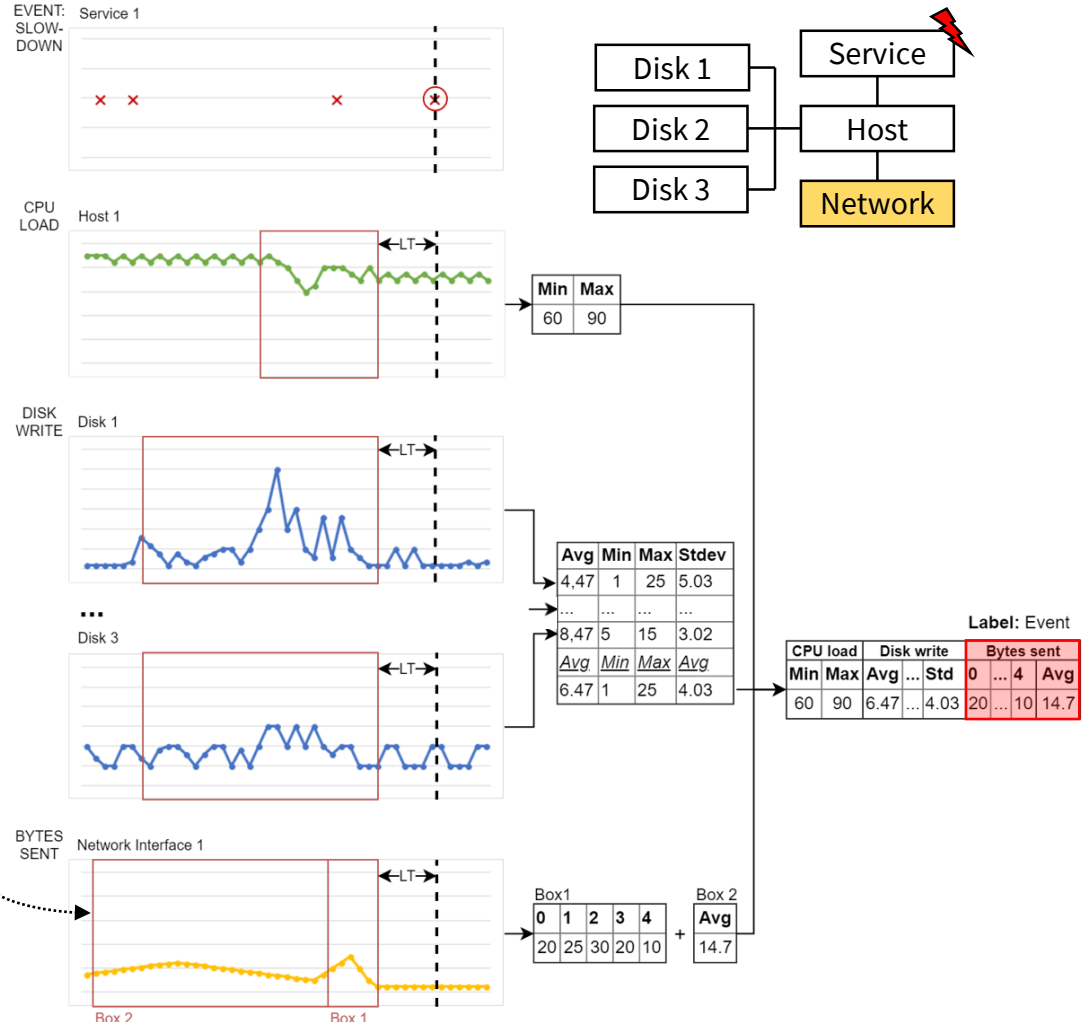
(2) Preprocessing – Example



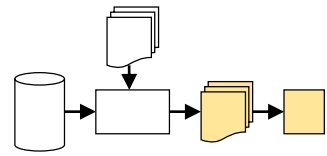
Sample: Customer System X, Service 1, Timestamp: 2018-02-01 17:30

```

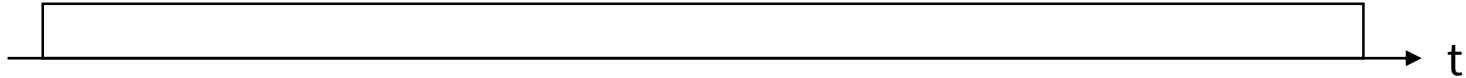
...
samplingMode: "PER_EVENT"
leadTime: 5
observationWindowsBoxes:
  CPU_LOAD:
    - size: 15
      aggregationFunctions:
        - "MIN"
        - "MAX"
    ...
  DISK_WRITE:
    - size: 30
      aggregationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "STD_DEV"
      combinationFunctions:
        - "AVG"
        - "MIN"
        - "MAX"
        - "AVG"
    ...
  BYTES_SENT:
    - size: 5
      aggregationFunctions:
        - "NONE"
    ...
    - size: 30
      aggregationFunctions:
        - "AVG"
    ...
...
  
```



(3) Prediction

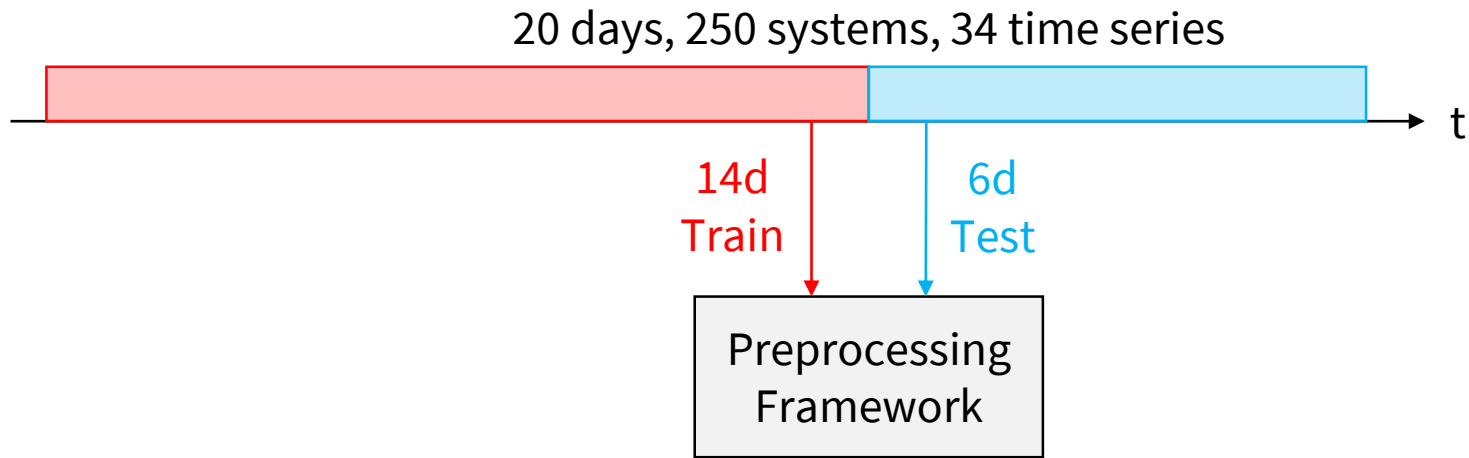
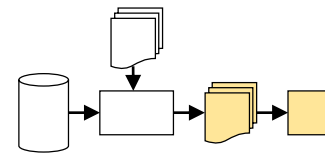


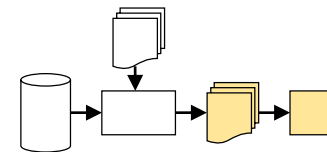
20 days, 250 systems, 34 time series



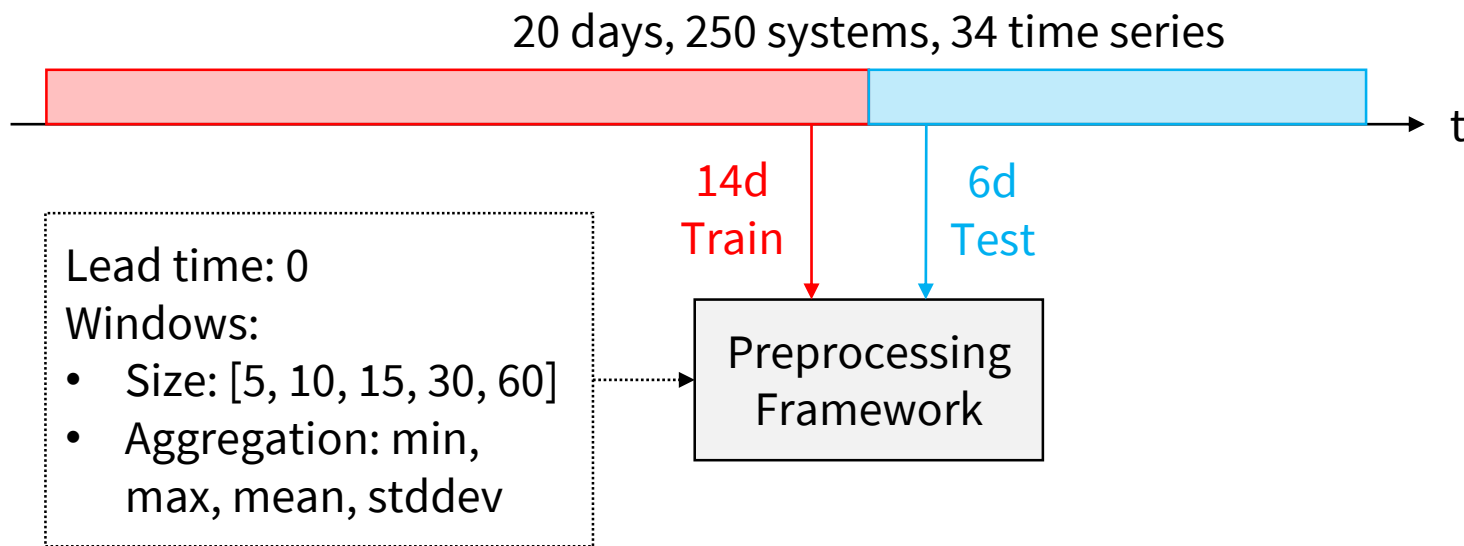
Preprocessing
Framework

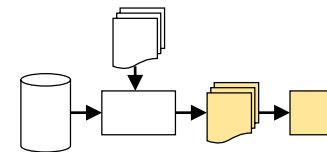
(3) Prediction



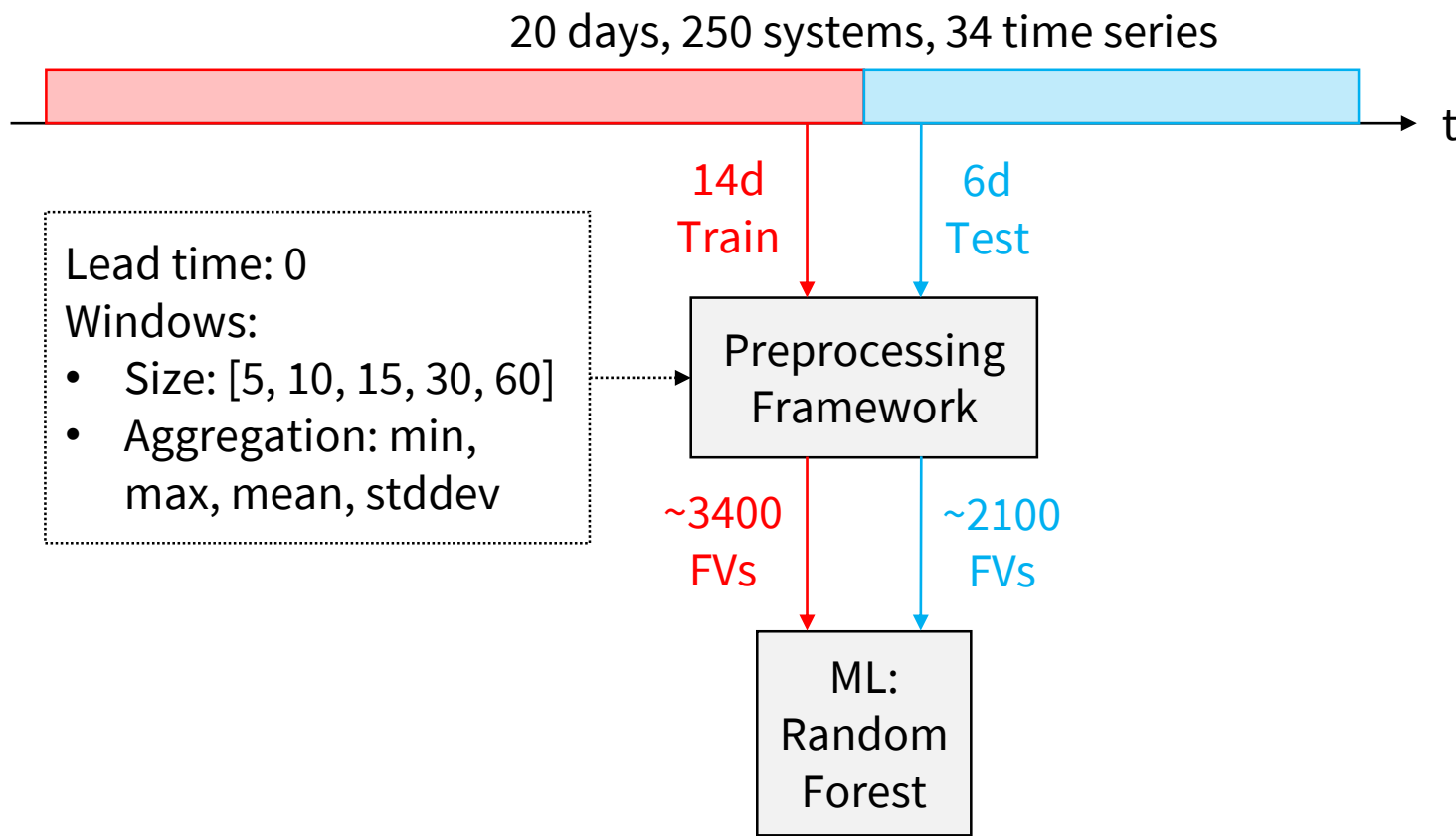


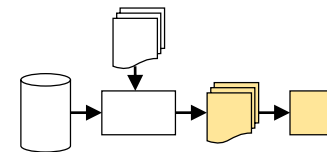
(3) Prediction





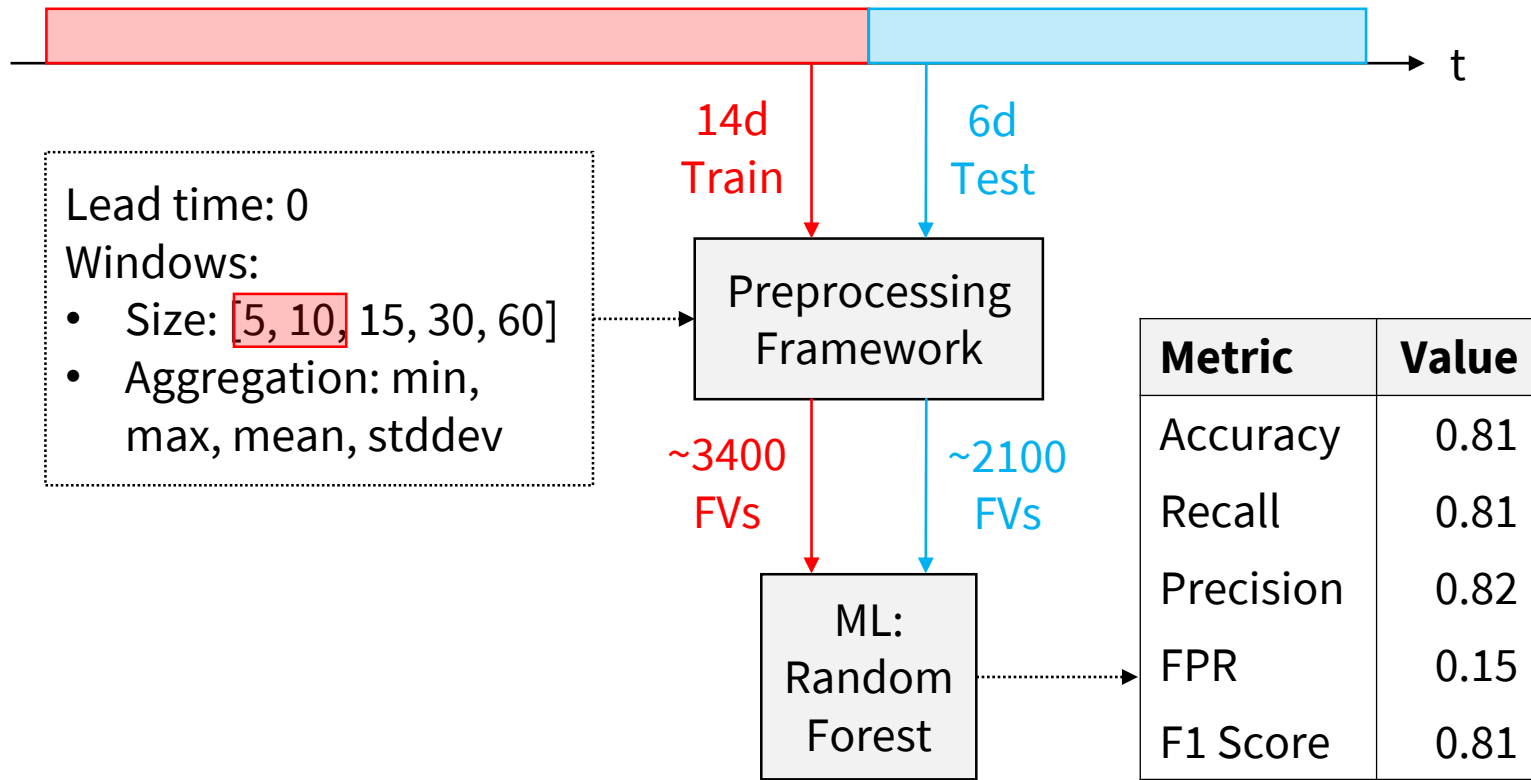
(3) Prediction





(3) Prediction

20 days, 250 systems, 34 time series



Future Work

- More training and testing:
 - Splits
 - Config settings
 - System generalization

Future Work

- More training and testing:
 - Splits
 - Config settings
 - System generalization
- Other ML models

Future Work

- More training and testing:
 - Splits
 - Config settings
 - System generalization
- Other ML models
- Other events

Using Multi-System Monitoring Time Series to Predict Performance Events

Andreas Schörgenhumer

Mario Kahlhofer

Peter Chalupar

Hanspeter Mössenböck

Paul Grünbacher

09.11.2018

