# UTILIZING MONITORING AND REPORTING TECHNIQUES IN DATA PROCESSING SYSTEMS TO IMPROVE THE VALUE ADDING OF DATA

**Johanna Senft, Henrike Barkmann, Sven Stönner, Max Wegner**

DLR

# Introduction

- **DLR‘s Remote Sensing Data Center (DFD)** processes daily
  → more than 20 TB of remote sensing data
  → using more than 70 processing chains.

- Processing chains **Input data** of is of varying quality
  → it may be degraded by external and internal influences.

- **DFD’s Monitoring & Reporting System** is needed to ensure:
  → reliable data processing,
  → adequate performance,
  → effective operation,
  → and stable resource usage.

DLR Remote Sensing Data Center, Oberpfaffenhofen

DLR Remote Sensing Data Center, Neustrelitz

# Agenda

- Section I : IMPC and IMPC data process
- Section II: DFD’s Monitoring & Reporting System (DFD M&R)
- Section III: Use cases
- Section IV: Observation Discussion; Conclusion

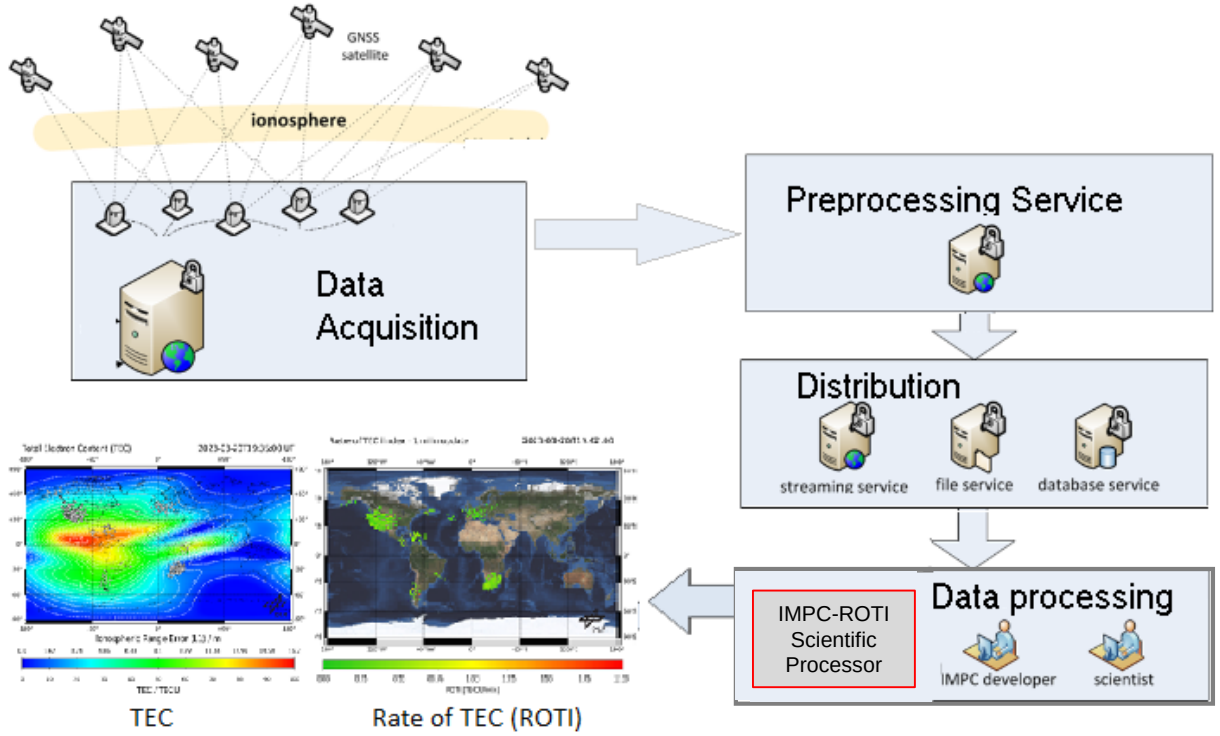
**UTILIZING MONITORING AND REPORTING TECHNIQUES IN DATA PROCESSING SYSTEMS TO IMPROVE THE VALUE ADDING OF DATA**

# Section I: The IMPC / IMPC data process

*Ionospheric disturbances affect the performance of space-based communication, navigation and remote sensing. They degrade the accuracy of Global Navigation Satellite Systems (GNSS) i.e. GPS or Galileo.*

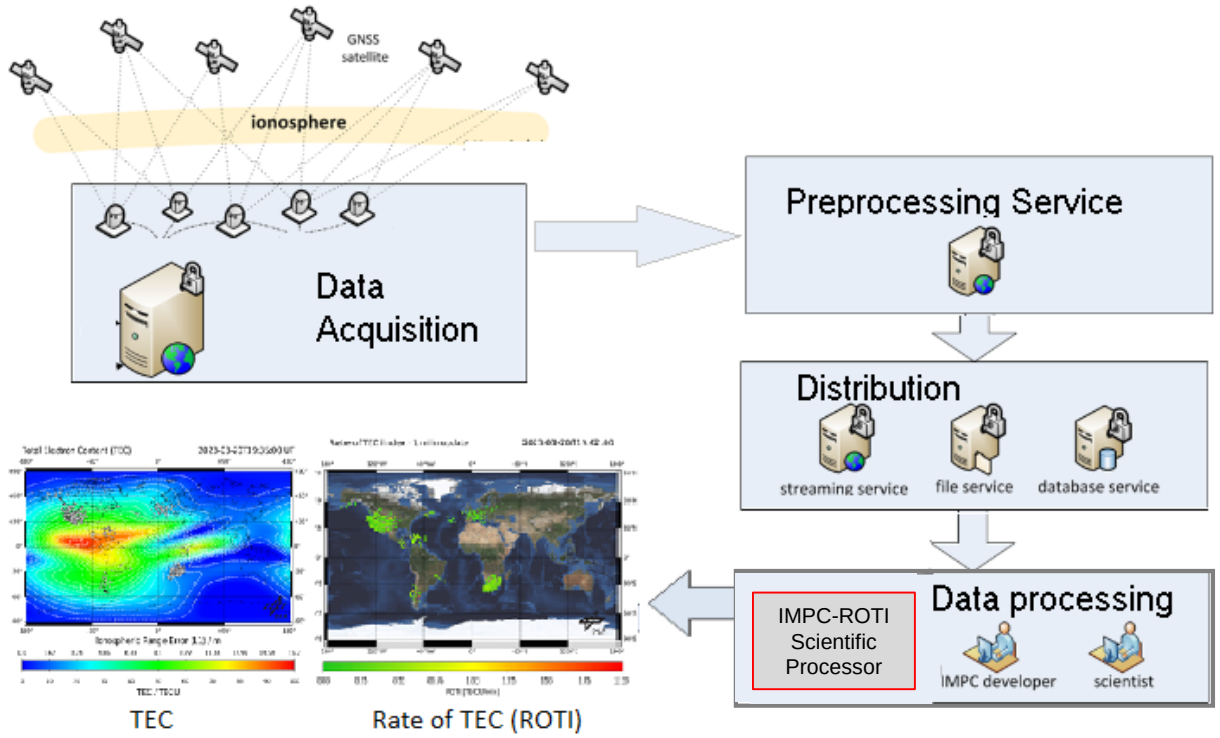
# Section I: The IMPC / IMPC data process

*Ionospheric disturbances affect the performance of space-based communication, navigation and remote sensing. They degrade the accuracy of Global Navigation Satellite Systems (GNSS) i.e. GPS or Galileo.*



# Section I: The IMPC / IMPC data process

*Ionospheric disturbances affect the performance of space-based communication, navigation and remote sensing. They degrade the accuracy of Global Navigation Satellite Systems (GNSS) i.e. GPS or Galileo.*

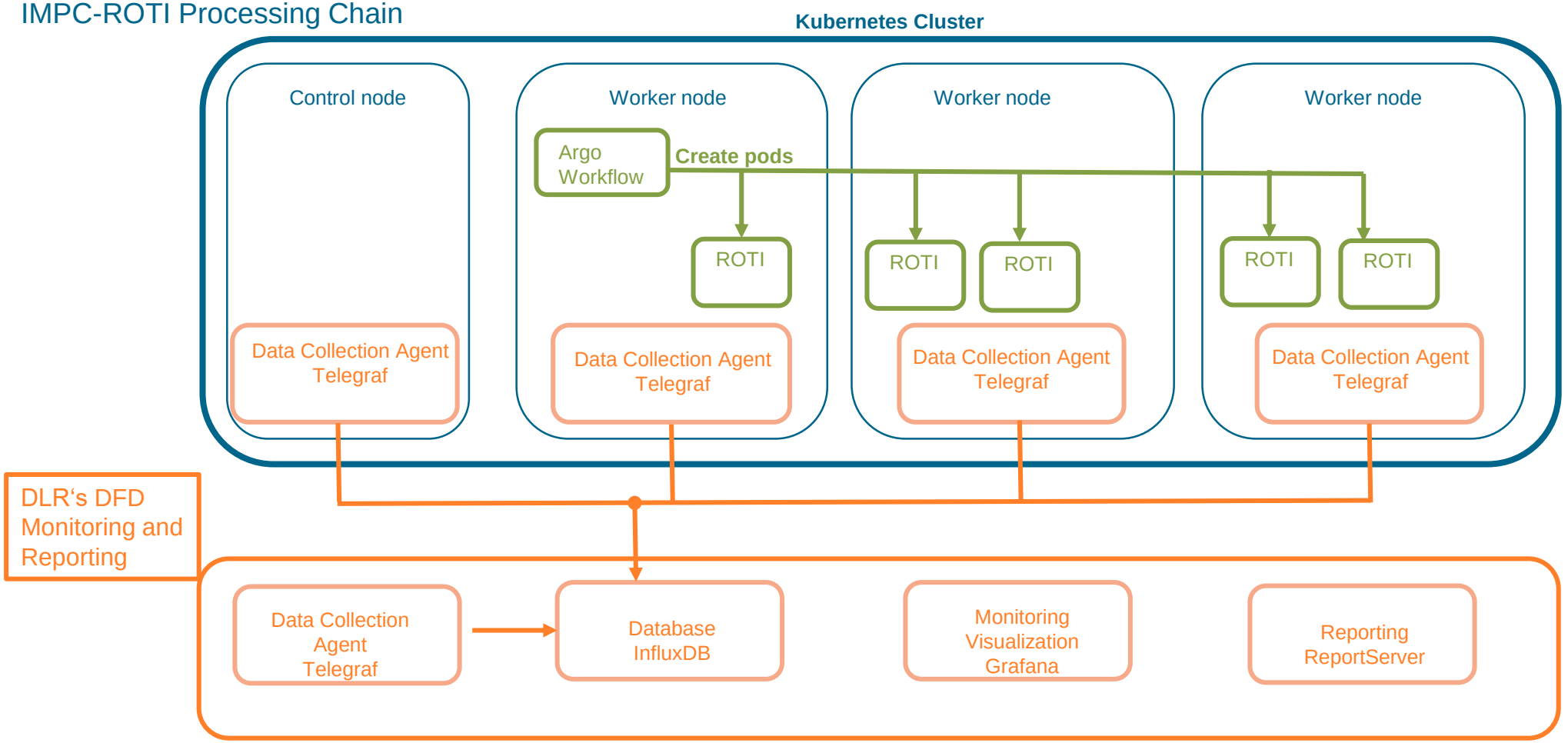



- Daily more than 50 GiB ionosphere data are processed and archived
- Including 2888 ROTI, 288 TEC products
- 1588 products for GNSS users

# Section I: The IMPC / IMPC data process

## System architecture

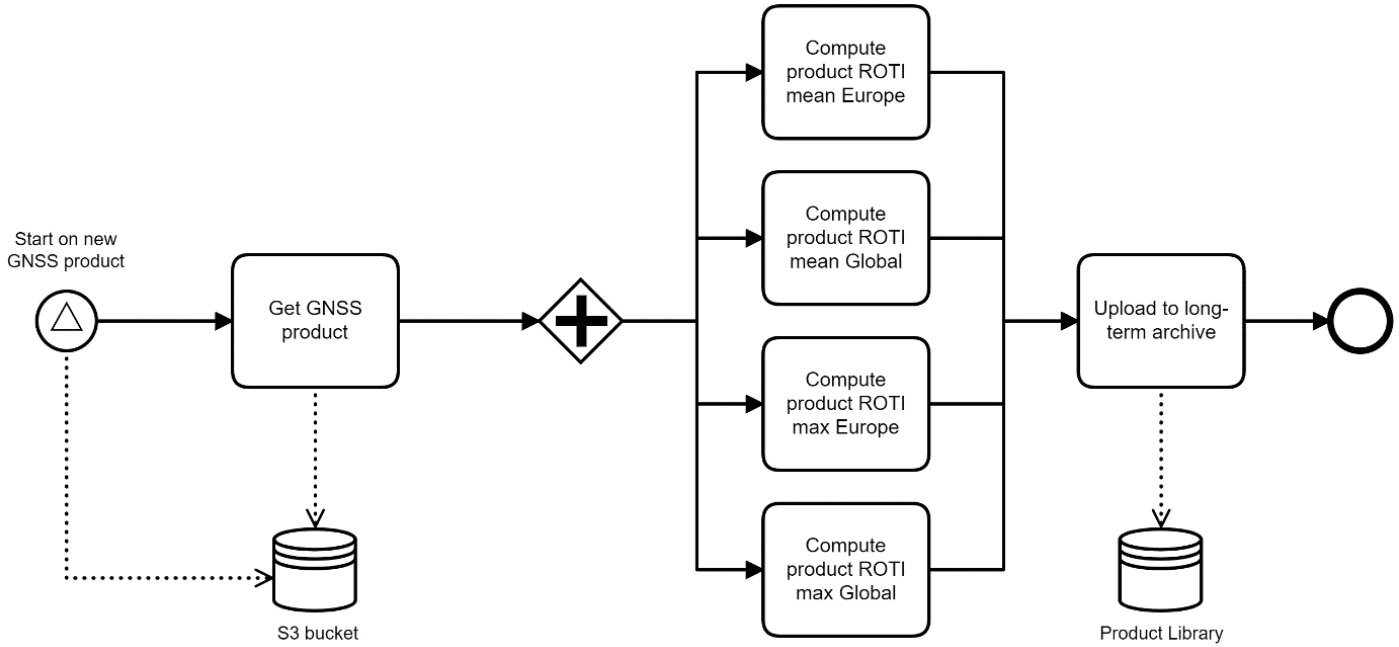
IMPC-ROTI Processing Chain



# Section I: The IMPC / IMPC data process

## IMPC-ROTI Scientific processor - workflow chain

Consists of five key steps:

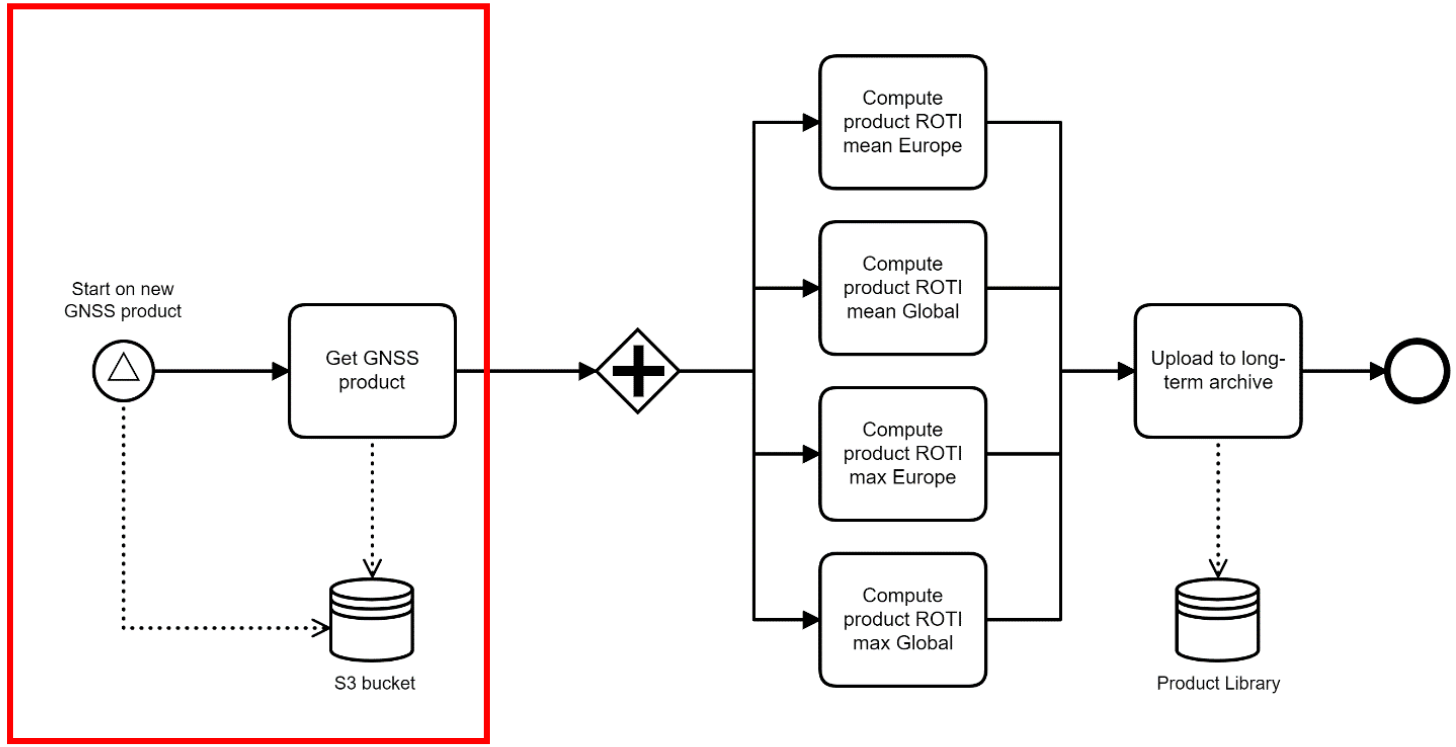


# Section I: The IMPC / IMPC data process

## IMPC-ROTI Scientific processor - workflow chain

Consists of five key steps:

- Read data from datasource

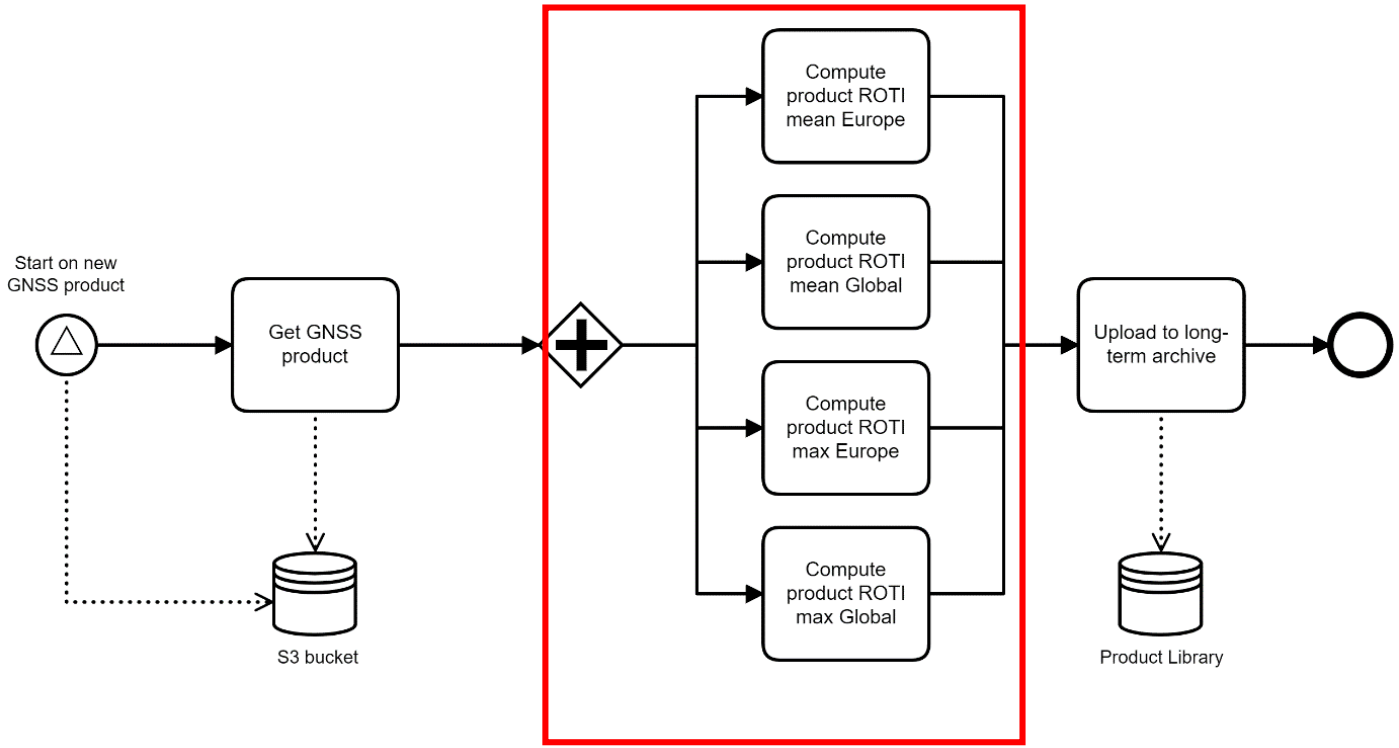


# Section I: The IMPC / IMPC data process

## IMPC-ROTI Scientific processor - workflow chain

Consists of five key steps:

- Read data from datasource
- Transform data
- Generate additional information

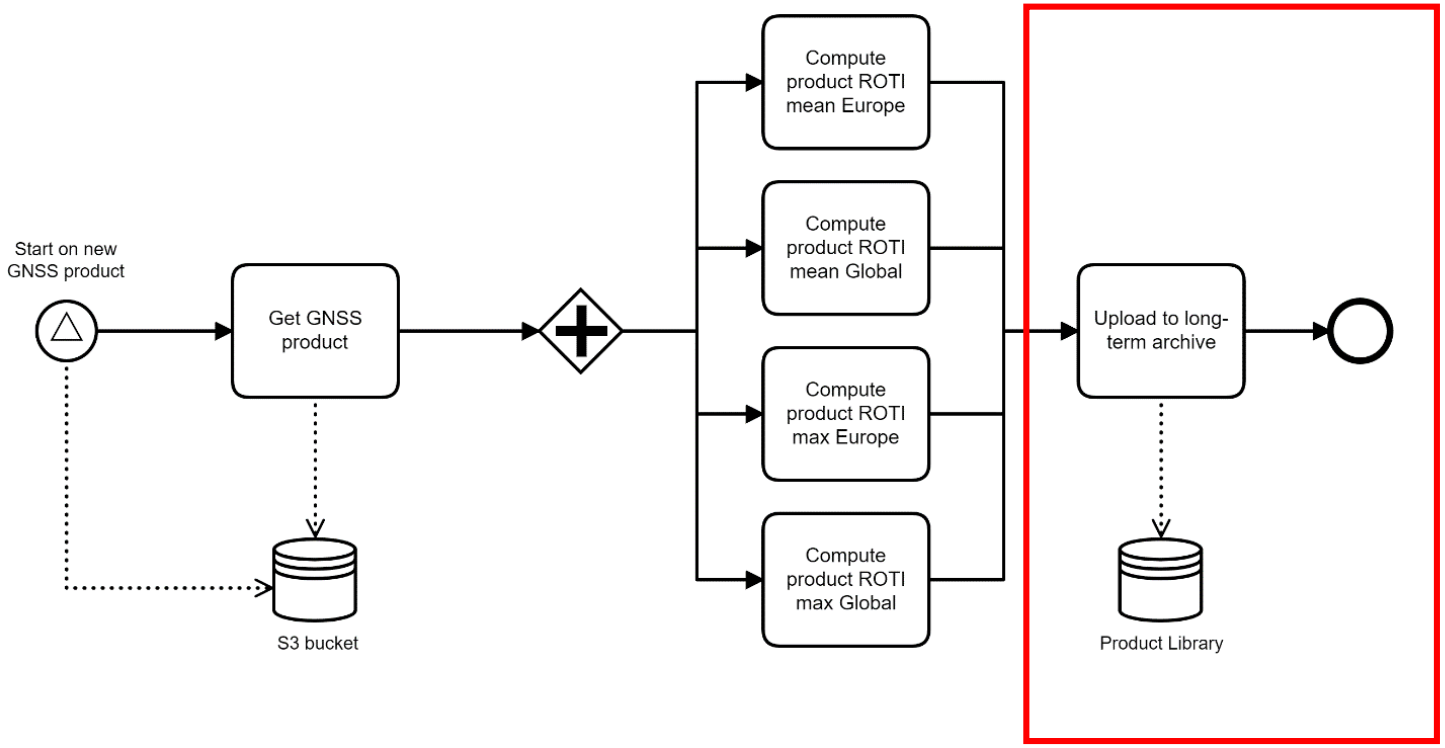


# Section I: The IMPC / IMPC data process

## IMPC-ROTI Scientific processor - workflow chain

Consists of five key steps:

- Read data from datasource
- Transform data
- Generate additional information
- Transfer data into a long-term archive
- Upload data to a remote data sink.



# Section I: The IMPC / IMPC data process

Processing chains are vulnerable to errors

Core problems

Extensive usage of processing capacity → System overload

# Section I: The IMPC / IMPC data process

Processing chains are vulnerable to errors

| | | |
|---|---|---|
| | | Core problems |
| Extensive usage of processing capacity | → | System overload |
| Introducing new processing products | → | Performance decrease |

# Section I: The IMPC / IMPC data process

Processing chains are vulnerable to errors

| | Core problems |
|---|---|
| Extensive usage of processing capacity → | System overload |
| Introducing new processing products → | Performance decrease |
| Malfunction of processing steps → | Bottlenecks - processing workflow |

# Section I: The IMPC / IMPC data process

Processing chains are vulnerable to errors

| | Core problems |
|---|---|
| Extensive usage of processing capacity → | System overload |
| Introducing new processing products → | Performance decrease |
| Malfunction of processing steps → | Bottlenecks - processing workflow |
| Disfunction between system components → | Instable system inter-operation |

# Section I: The IMPC / IMPC data process

Processing chains are vulnerable to errors

| | Core problems |
|---|---|
| Extensive usage of processing capacity → | System overload |
| Introducing new processing products → | Performance decrease |
| Malfunction of processing steps → | Bottlenecks - processing workflow |
| Disfunction between system components → | Instable system inter-operation |

Requires Monitoring and Reporting to ensure:

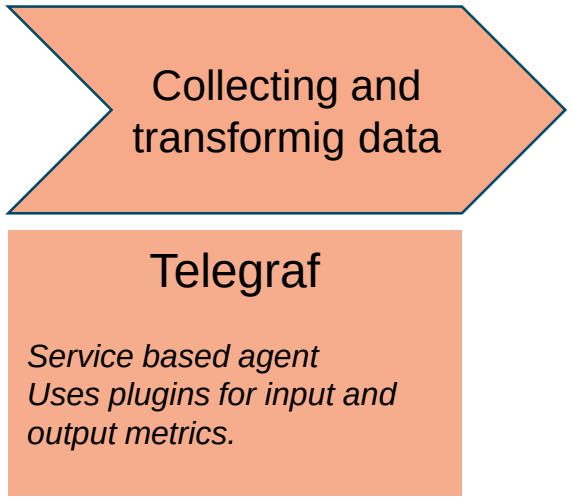
- well-functioning
- performance
- overall data quality

# Agenda

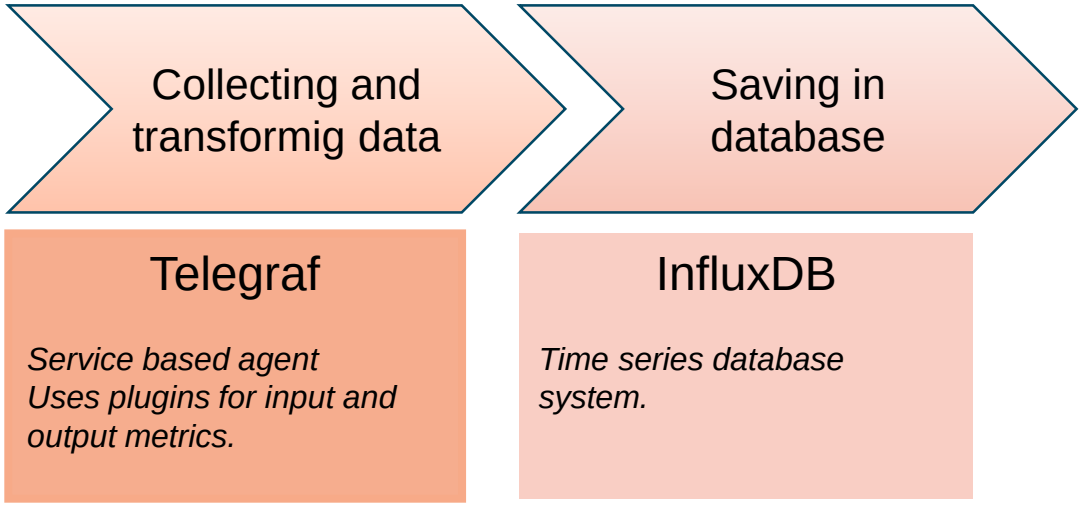
- Section I : IMPC and IMPC data process
- Section II: DFD’s Monitoring & Reporting System (DFD M&R)
- Section III: Use cases
- Section IV: Observation Discussion; Conclusion

**UTILIZING MONITORING AND REPORTING TECHNIQUES IN DATA PROCESSING SYSTEMS TO IMPROVE THE VALUE ADDING OF DATA**

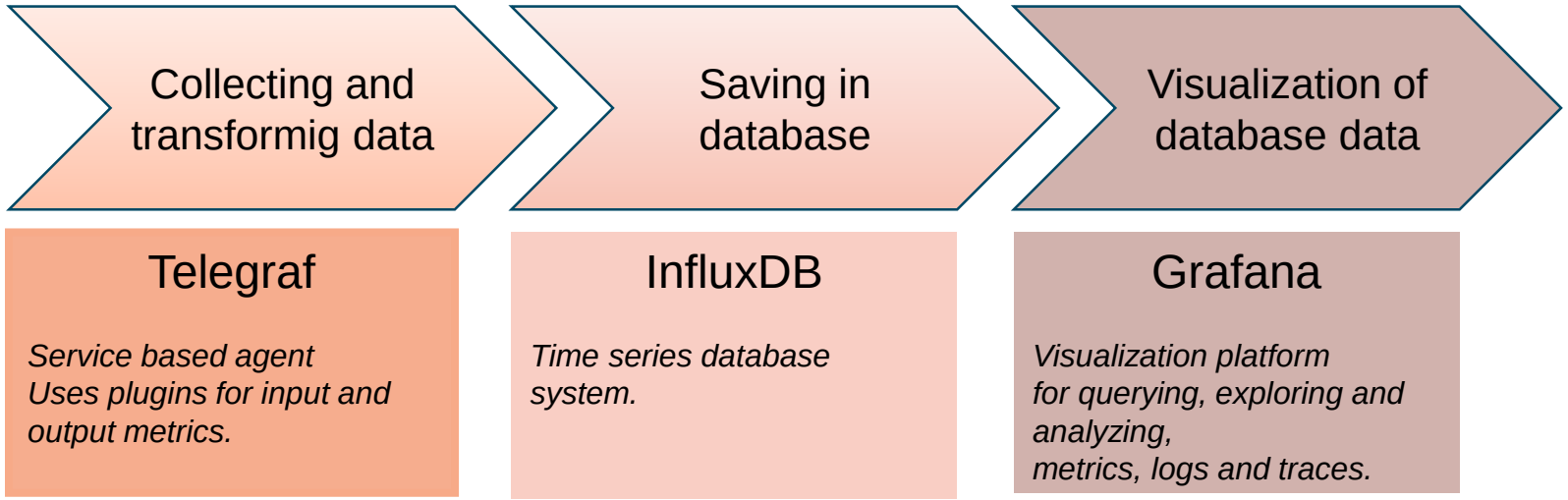
# Section II: DFD’s Monitoring & Reporting System (DFD M&R)



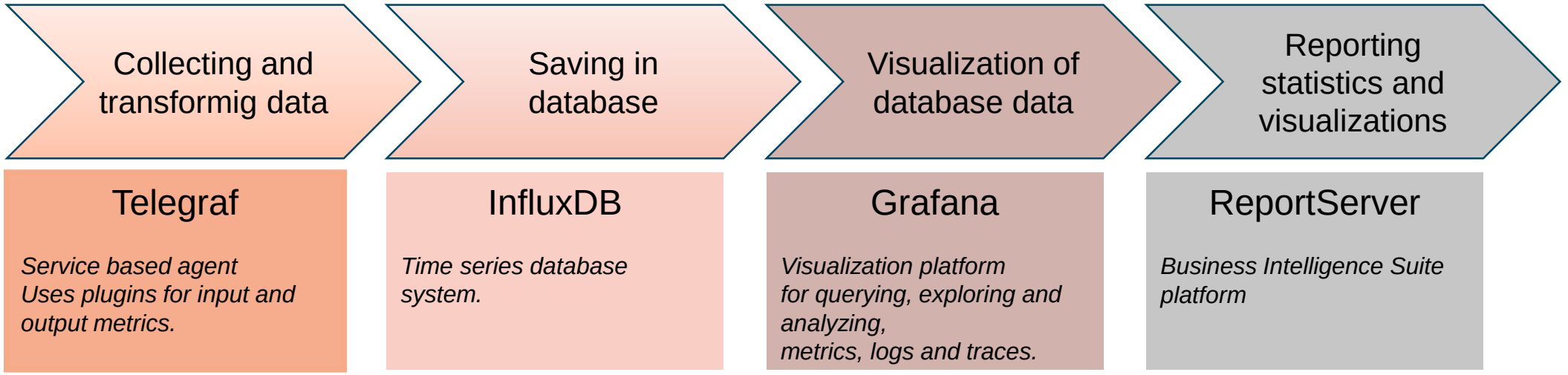
# Section II: DFD’s Monitoring & Reporting System (DFD M&R)



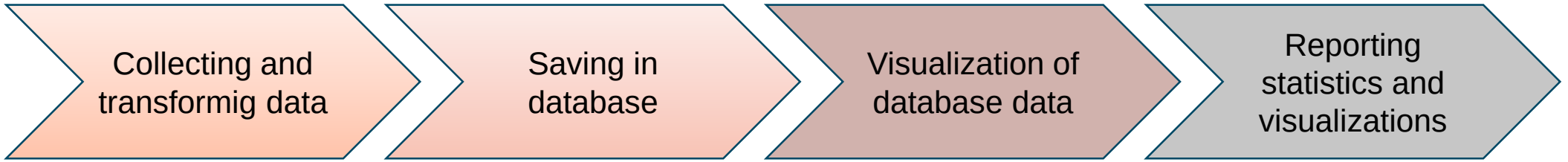
# Section II: DFD's Monitoring & Reporting System (DFD M&R)



# Section II: DFD's Monitoring & Reporting System (DFD M&R)



# Section II: DFD's Monitoring & Reporting System (DFD M&R)




## Telegraf

*Service based agent*
*Uses plugins for input and output metrics.*

**Application IMPC-ROTI**

*Collects:*

- *system metrics,*
- *Kubernetes metrics*
- *Argo-WF metrics*
- *Saves metrics directly in Database*

## InfluxDB

*Time series database system.*

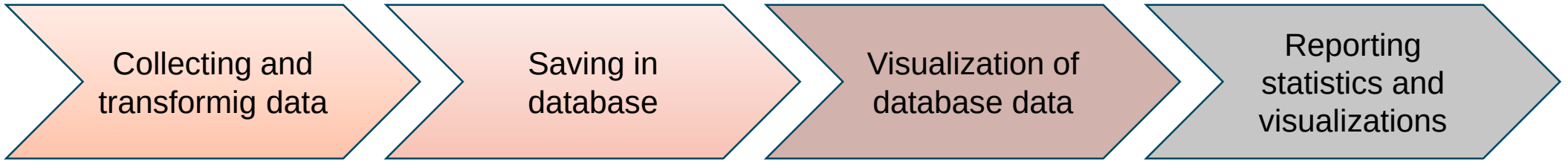
## Grafana

*Visualization platform for querying, exploring and analyzing, metrics, logs and traces.*

## ReportServer

*Business Intelligence Suite platform*

# Section II: DFD's Monitoring & Reporting System (DFD M&R)




## Telegraf

*Service based agent*
*Uses plugins for input and output metrics.*

**Application IMPC-ROTI**

*Collects:*
- *system metrics,*
- *Kubernetes metrics*
- *Argo-WF metrics*
- *Saves metrics directly in Database*

## InfluxDB

*Time series database system.*

**Application IMPC-ROTI**

*Configured with one bucket containing all received data*

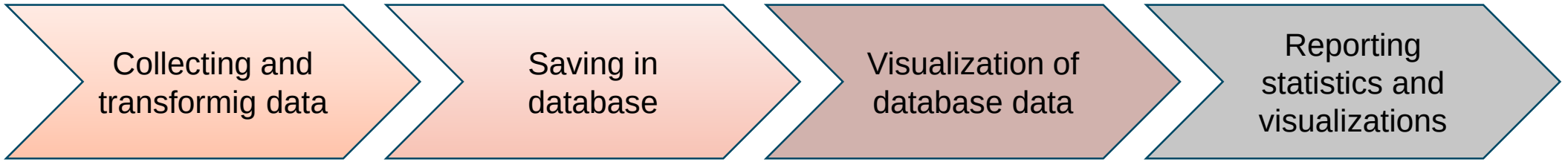
## Grafana

*Visualization platform for querying, exploring and analyzing, metrics, logs and traces.*

## ReportServer

*Business Intelligence Suite platform*

# Section II: DFD's Monitoring & Reporting System (DFD M&R)




## Telegraf

*Service based agent*
*Uses plugins for input and output metrics.*

**Application IMPC-ROTI**

*Collects:*
- *system metrics,*
- *Kubernetes metrics*
- *Argo-WF metrics*
- *Saves metrics directly in Database*

## InfluxDB

*Time series database system.*

**Application IMPC-ROTI**

*Configured with one bucket containing all received data*

## Grafana

*Visualization platform for querying, exploring and analyzing, metrics, logs and traces.*

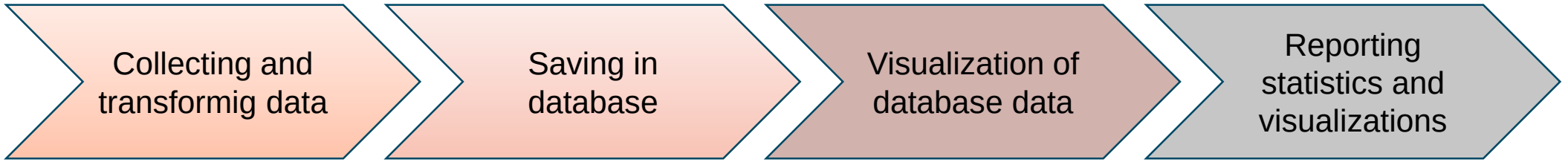
**Application IMPC-ROTI**

*Displays collected data in an operator friendly manner*
- *Quick Overview Dashboard*
- *WF Status Dashboard*
- *Used Components Dashboard*
- *Kubernetes Dashboard*

## ReportServer

*Business Intelligence Suite platform*

# Section II: DFD's Monitoring & Reporting System (DFD M&R)




## Telegraf

*Service based agent*
*Uses plugins for input and output metrics.*

**Application IMPC-ROTI**

*Collects:*
- *system metrics,*
- *Kubernetes metrics*
- *Argo-WF metrics*
- *Saves metrics directly in Database*

## InfluxDB

*Time series database system.*

**Application IMPC-ROTI**

*Configured with one bucket containing all received data*

## Grafana

*Visualization platform for querying, exploring and analyzing, metrics, logs and traces.*

**Application IMPC-ROTI**

*Displays collected data in an operator friendly manner*
- *Quick Overview Dashboard*
- *WF Status Dashboard*
- *Used Components Dashboard*
- *Kubernetes Dashboard*

## ReportServer

*Business Intelligence Suite platform*

**Application IMPC-ROTI**

*Compiles reports by querying Influx-DB:*
- *System Usage Report*
- *System Error Report*
- *Resources Report*

# Agenda

- Section I : IMPC and IMPC data process
- Section II: DFD’s Monitoring & Reporting System (DFD M&R)
- Section III: Use cases
- Section IV: Observation Discussion; Conclusion

**UTILIZING MONITORING AND REPORTING TECHNIQUES IN DATA PROCESSING SYSTEMS TO IMPROVE THE VALUE ADDING OF DATA**

# Section III: Use cases: System overload

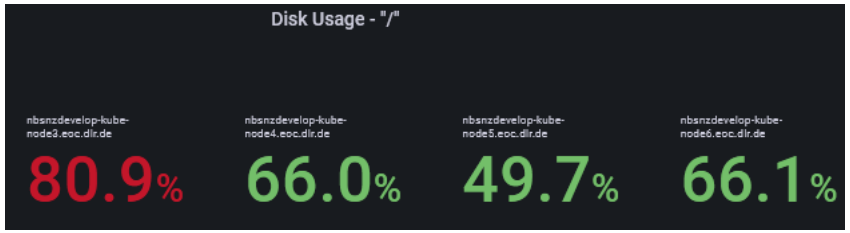
# Section III: Use cases: System overload

CPU Usage

# Section III: Use cases: System overload

CPU Usage

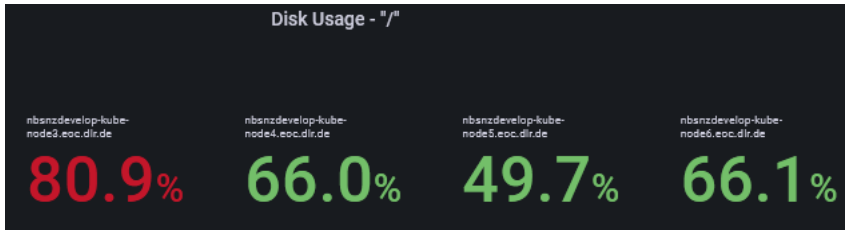
Disk Usage

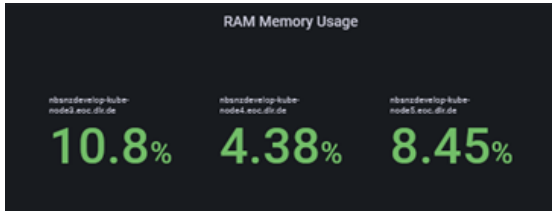
# Section III: Use cases: System overload

CPU Usage


Disk Usage



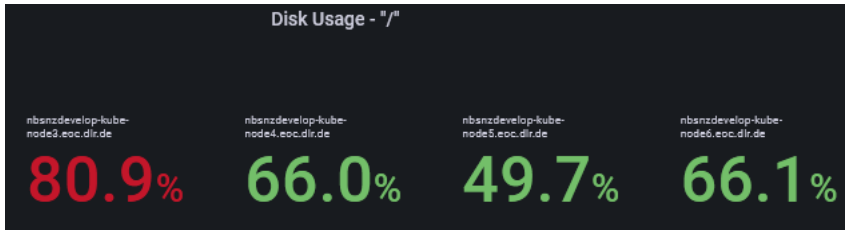

RAM Memory Usage



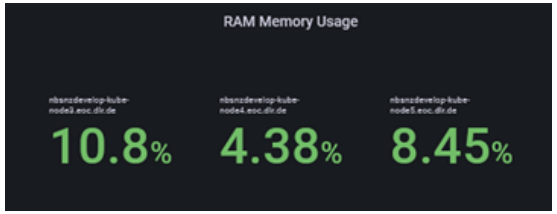
# Section III: Use cases: System overload

CPU Usage


Disk Usage




RAM Memory Usage




Operator alarms

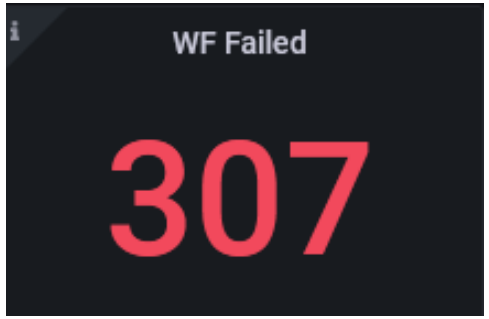
System crash can be averted

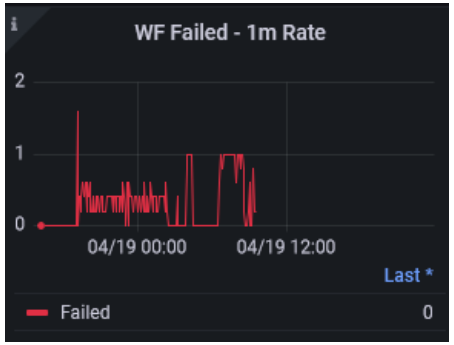
Reporting: system usage, ressources report

# Section III: Use Cases: Performance - decrease

# Section III: Use Cases: Performance - decrease

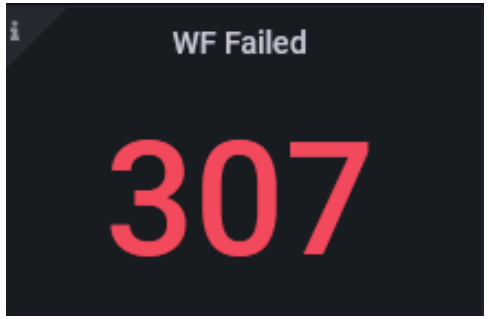
Workflow Errors

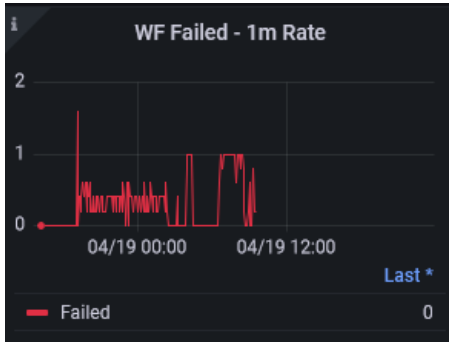


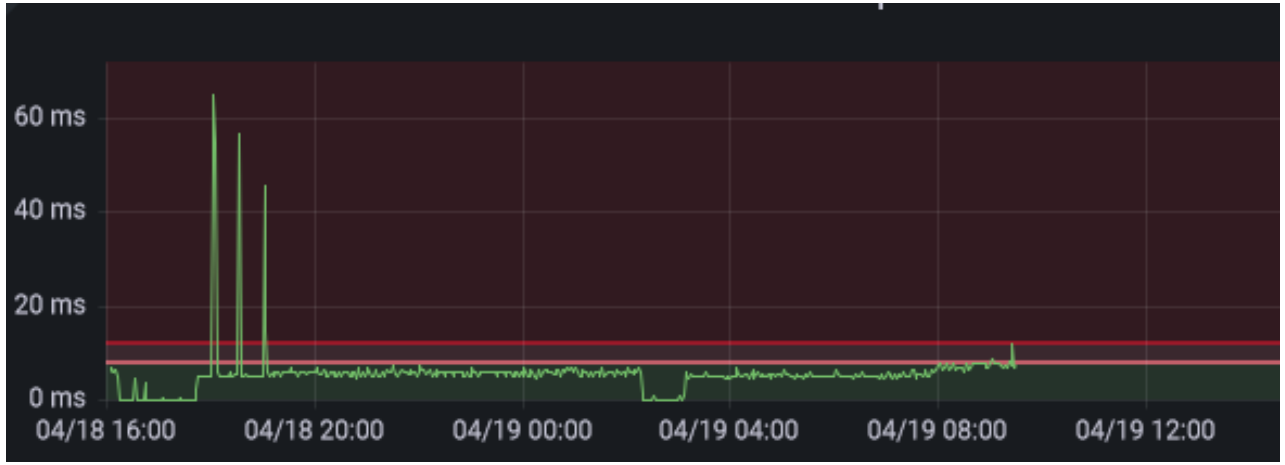
# Section III: Use Cases: Performance - decrease

## Workflow Errors



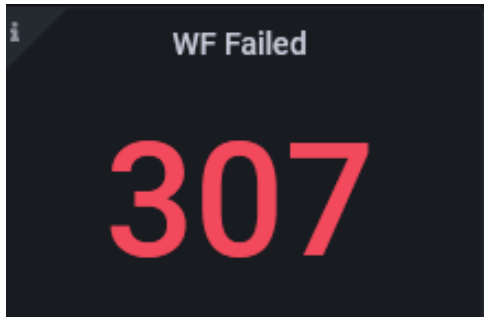



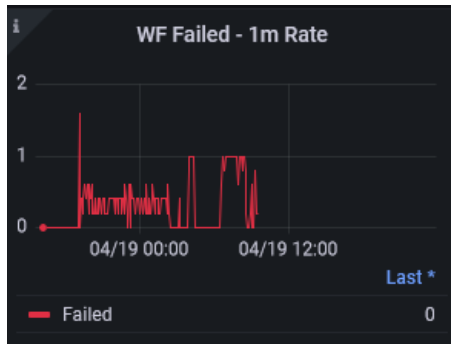

## Workflow Duration of operation



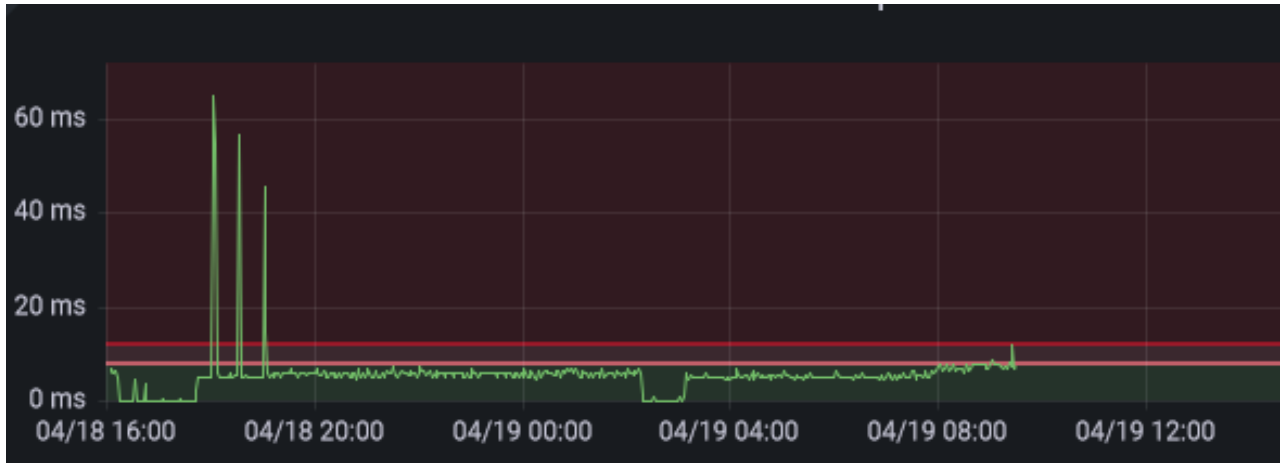
# Section III: Use Cases: Performance - decrease

Workflow Errors




Workflow Duration of operation




Operator alarms

Recognize performance problems caused by new processing products

Reporting: system error, ressources report

# Section III: Use cases: Bottlenecks - processing workflow

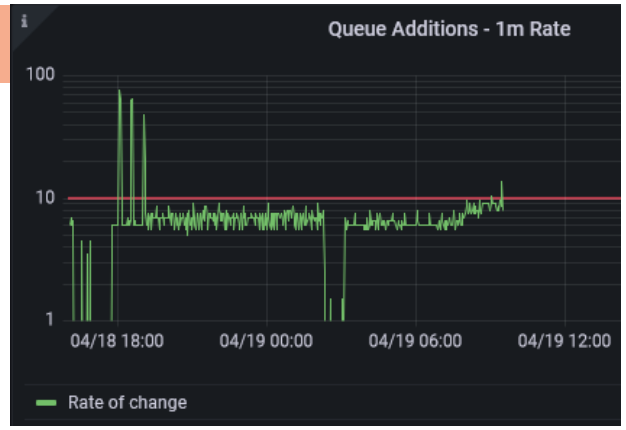
# Section III: Use cases: Bottlenecks - processing workflow

## Workflow - pods state

# Section III: Use cases: Bottlenecks - processing workflow

## Workflow - pods state

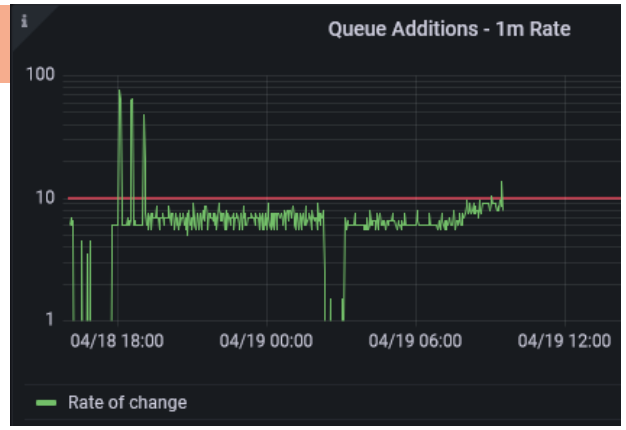

## Addition to WF - Queue



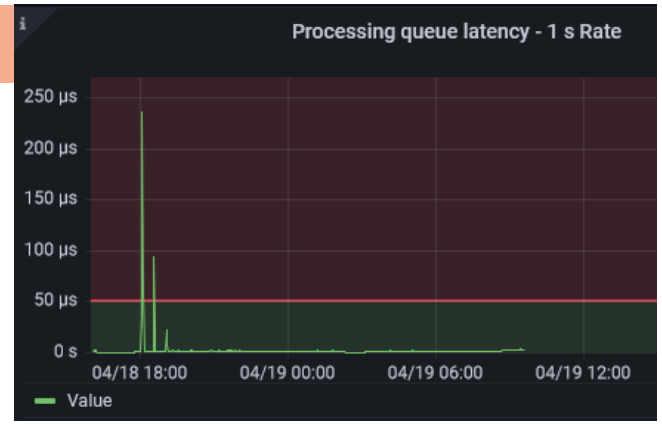
# Section III: Use cases: Bottlenecks - processing workflow

## Workflow - pods state


## Addition to WF - Queue



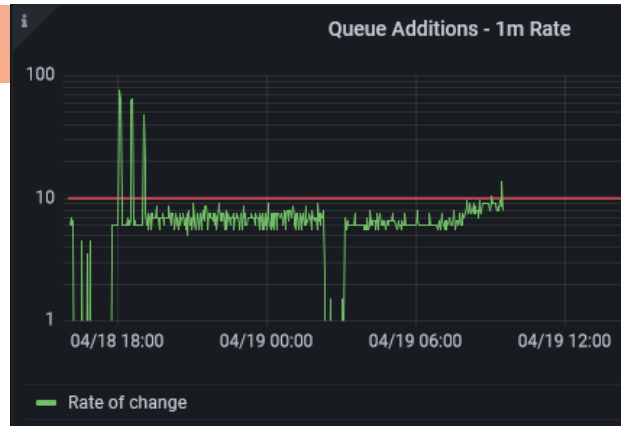

## WF-Queue Latency



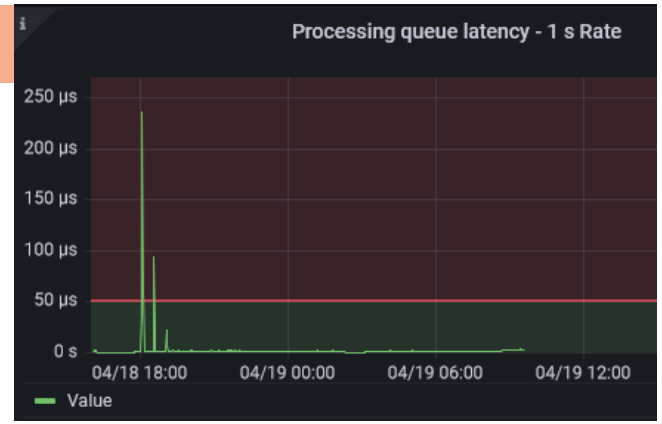
# Section III: Use cases: Bottlenecks - processing workflow

## Workflow - pods state


## Addition to WF - Queue




## WF-Queue Latency




## Operator alarms

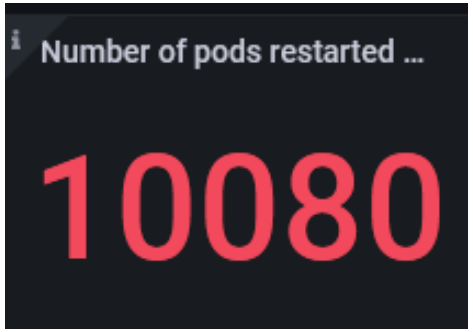
Bottlenecks are identified quickly

Reporting: system error, system usage, ressources report

# Section III: Use cases: Instable system inter-operation

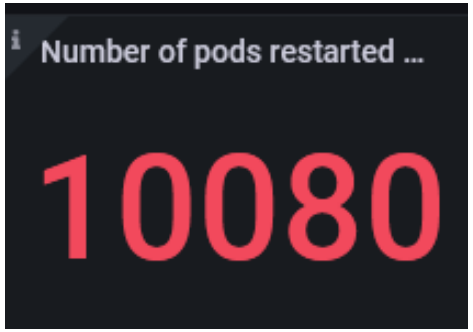
# Section III: Use cases: Instable system inter-operation

Number of pods restarted by Kubernetes

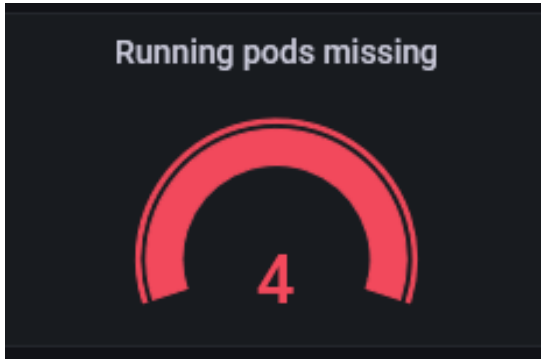


# Section III: Use cases: Instable system inter-operation

Number of pods restarted by Kubernetes

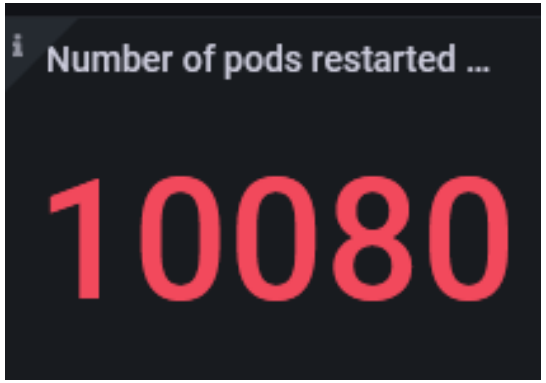



Missing pods in Argo-WF terminated by Kubernetes

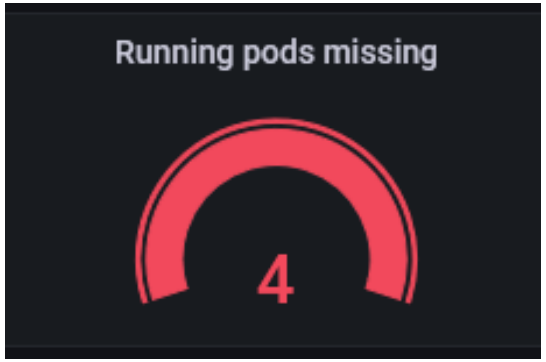


# Section III: Use cases: Instable system inter-operation

Number of pods restarted by Kubernetes




Missing pods in Argo-WF terminated by Kubernetes




Operator alarms

Provide a stable operation between system components

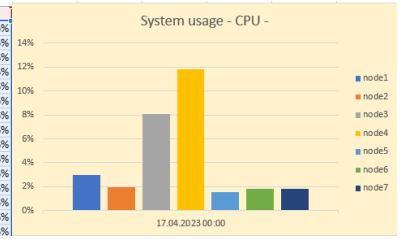
Reporting: system error, system usage, ressources report

# Section III: Use cases: Reports

## System Usage Report

| CPU- , Memory-, Disk-Usage | Designed for operators |
|---|---|

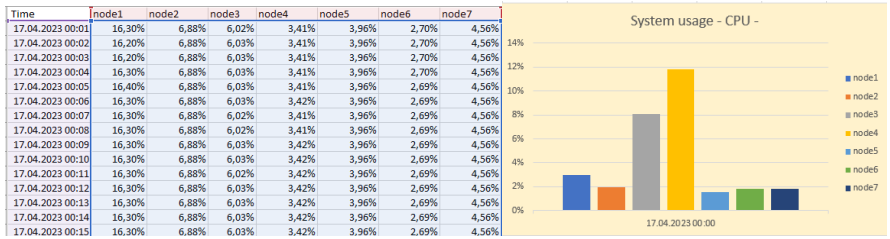
| Time | node1 | node2 | node3 | node4 | node5 | node6 | node7 |
|---|---|---|---|---|---|---|---|
| 17.04.2023 00:01 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:02 | 16,20% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:03 | 16,20% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:04 | 16,30% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:05 | 16,40% | 6,88% | 6,03% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:06 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:07 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:08 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:09 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:10 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:11 | 16,30% | 6,88% | 6,02% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:12 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:13 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:14 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:15 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |



# Section III: Use cases: Reports

## System Usage Report

| | |
|---|---|
| CPU- , Memory-, Disk-Usage | Designed for operators |

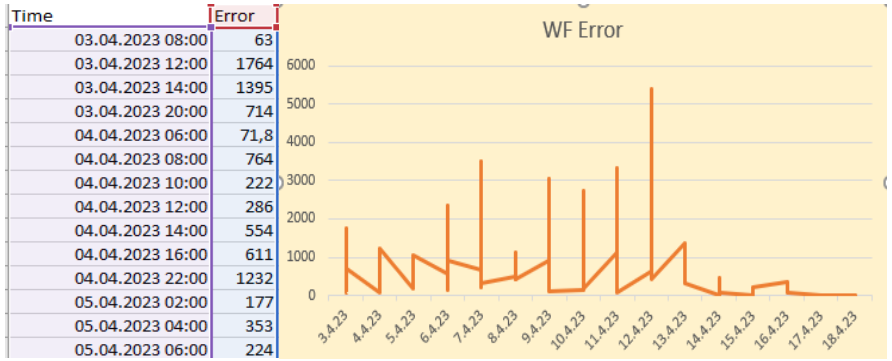


| Time | node1 | node2 | node3 | node4 | node5 | node6 | node7 |
|---|---|---|---|---|---|---|---|
| 17.04.2023 00:01 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:02 | 16,20% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:03 | 16,20% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:04 | 16,30% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:05 | 16,40% | 6,88% | 6,03% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:06 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:07 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:08 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:09 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:10 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:11 | 16,30% | 6,88% | 6,02% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:12 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:13 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:14 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:15 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |



## System Error Report

| | |
|---|---|
| Occured failures and errors | Designed for developers |



| Time | Error |
|---|---|
| 03.04.2023 08:00 | 63 |
| 03.04.2023 12:00 | 1764 |
| 03.04.2023 14:00 | 1395 |
| 03.04.2023 20:00 | 714 |
| 04.04.2023 06:00 | 71,8 |
| 04.04.2023 08:00 | 764 |
| 04.04.2023 10:00 | 222 |
| 04.04.2023 12:00 | 286 |
| 04.04.2023 14:00 | 554 |
| 04.04.2023 16:00 | 611 |
| 04.04.2023 22:00 | 1232 |
| 05.04.2023 02:00 | 177 |
| 05.04.2023 04:00 | 353 |
| 05.04.2023 06:00 | 224 |

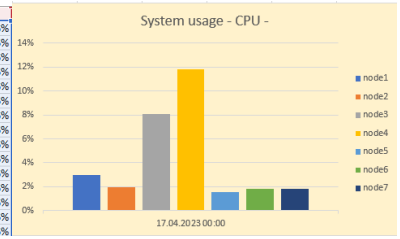
# Section III: Use cases: Reports

## System Usage Report

CPU- , Memory-, Disk-Usage

Designed for operators

| Time | node1 | node2 | node3 | node4 | node5 | node6 | node7 |
|---|---|---|---|---|---|---|---|
| 17.04.2023 00:01 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:02 | 16,20% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:03 | 16,20% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:04 | 16,30% | 6,88% | 6,03% | 3,41% | 3,96% | 2,70% | 4,56% |
| 17.04.2023 00:05 | 16,40% | 6,88% | 6,03% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:06 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:07 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:08 | 16,30% | 6,88% | 6,02% | 3,41% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:09 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:10 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:11 | 16,30% | 6,88% | 6,02% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:12 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:13 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:14 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |
| 17.04.2023 00:15 | 16,30% | 6,88% | 6,03% | 3,42% | 3,96% | 2,69% | 4,56% |

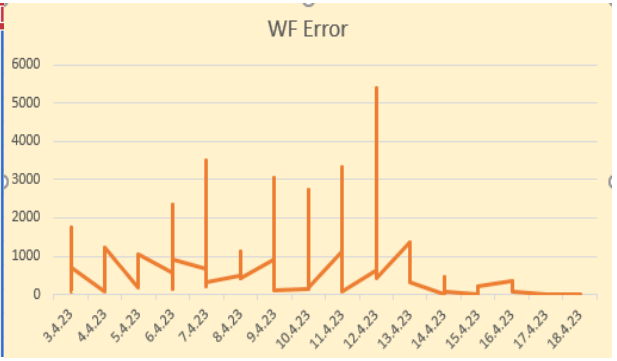



## System Error Report

Occured failures and errors

Designed for developers

| Time | Error |
|---|---|
| 03.04.2023 08:00 | 63 |
| 03.04.2023 12:00 | 1764 |
| 03.04.2023 14:00 | 1395 |
| 03.04.2023 20:00 | 714 |
| 04.04.2023 06:00 | 71,8 |
| 04.04.2023 08:00 | 764 |
| 04.04.2023 10:00 | 222 |
| 04.04.2023 12:00 | 286 |
| 04.04.2023 14:00 | 554 |
| 04.04.2023 16:00 | 611 |
| 04.04.2023 22:00 | 1232 |
| 05.04.2023 02:00 | 177 |
| 05.04.2023 04:00 | 353 |
| 05.04.2023 06:00 | 224 |




## Ressources Report

Used ressources
Good future planning instrument for introduction of new products

Designed for project leaders, system and software developer

# Agenda

- Section I : IMPC and IMPC data process
- Section II: DFD’s Monitoring & Reporting System (DFD M&R)
- Section III: Use cases
- Section IV: Observation Discussion; Conclusion

**UTILIZING MONITORING AND REPORTING TECHNIQUES IN DATA PROCESSING SYSTEMS TO IMPROVE THE VALUE ADDING OF DATA**

# Section IV: Observation, Discussion

## Benefits

- By impeding system overload
- By increasing system stability
- By improving system performance
- By detecting bottleneck

# Section IV: Observation, Discussion

## Benefits

- By impeding system overload
- By increasing system stability
- By improving system performance
- By detecting bottleneck

- Increased system reliability
- Faster troubleshooting
- Data throughput optimization
- Reduced operation efforts
- Improved data consistency
- Improved data quality
- Reliable planning information

# Section IV: Observation, Discussion

## Drawbacks

**monitoring complex systems**

- creates a large amount of data
- strain the system load,
- strain the application capacity,
- strain the data traffic,
- are memory consuming

# Section IV: Observation, Discussion

## Drawbacks

**monitoring complex systems**

- creates a large amount of data
- strain the system load,
- strain the application capacity,
- strain the data traffic,
- are memory consuming

- Slow down the entire system
- Reducing data quality
- Worst-case lead to system crash

# Section IV: Observation, Discussion

**Carefully consider**

- The amount of collected monitoring metrics
- The retention policy of stored data
- The data cardinality in databases

## Drawbacks

**monitoring complex systems**

- creates a large amount of data
- strain the system load,
- strain the application capacity,
- strain the data traffic,
- are memory consuming

- Slow down the entire system
- Reducing data quality
- Worst-case lead to system crash

# Section IV: Observation, Discussion

## Outlook - future development

- Introduction of user defined *Custom Metrics;* specific for each application.
- Monitor Log-Errors by introduction of log-shipment and log-files analyzing
- Extension of DFD M&R procedures to components outside the Kubernetes Cluster

# Section IV: Conclusion

Application of DFD M&R to a complex data processing system

can improve essentially all data processing chains.

Careful configuration and implementation is paramount to the successful use of the DLR DFD Monitoring and Reporting System.