Center for Information Services and High Performance Computing (ZIH)

Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods

ICSEA 2022, Lisbon (Portugal)

October 16 to October 20, 2022

Martin Zinner (martin.zinner1@tu-dresden.de)

Big Data

Definition & Considerations

➤ *Big Data* refers to enormous volumes of *data*
that cannot be processed effectively with the
traditional applications that exist.
Graphic: https://unsplash.com/s/photos/big-data

➤ Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing
Volumes of Data, http://www.gartner.com/newsroom/id/1731916

➤ An article by techjury states that Data is growing faster than ever before and in the
year 2020, about 1.7 megabytes of new information were created every second for
every human being on the planet, https://techjury.net/blog/how-much-data-is-
created-every-day/#gref

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of
Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International
Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,,
Portugal

Slide 2

TECHNISCHE
UNIVERSITÄT
DRESDEN

DRESDEN
concept

Small Data

How Small Data becomes bigger than Big Data

➢ In some cases processing small data becomes ineffective or impossible by using traditional applications.

➢ The impression is that we are dealing with inflated Small Data developing to Big Data.

➢ Some problems are difficult to solve even on small datasets.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 3

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Existing solutions to achieve early recognition and proactive handling of disruptions of services of computer centres

Industry-specific

➢ HPE InfoSight by Hewlett Packard (HP), uses predictive analytics to predict, prevent, and auto resolve problems from storage to VMs. Eliminates the pain of managing HP infrastucture.

➢ Ironstream integrates machine data from traditional legacy IBM systems into leading IT analytics platforms to work seamlessly with Splunk , ServiceNow , Micro Focus, Microsoft SCOM, Elastic, Apache Kafka.

➢ Splunk offers a central platform for analysing machine generated data by developing advanced analytics. It can collect data from various location and combines it into centralised indexes and aggregates the log files from many sources to make them centrally searchable.

➢ In order to support its in-memory database HANA, the software company SAP introduced onto the market a troubleshooting and performance analysis tool.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 4

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Existing solutions to achieve early recognition and proactive handling of disruptions of services of computer centres – cont.

Academic

➢ At the University of Technology Dresden/Germany, "jam–e jam", a prototype to analyse and predict system behaviour based on statistical analysis has been developed. It detects node-level failures on HPC systems, as early as possible, in order to employ appropriate protective measures in useful time.

TECHNISCHE
UNIVERSITÄT
DRESDEN

DRESDEN
concept

Motivation

Manufacturing

➢ 24/7/365 (round-the-clock) system availability: Continuous availability requirements for production-support IT systems, for example in the semiconductor industry, are the general rule.

➢ Real-time demands: Real-time information processing has become a significant requirement for the optimal functioning of the manufacturing plants.

➢ Rapidly increasing data amount: The total amount of data created, captured, and consumed globally is forecast to increase rapidly, reaching more than 180 Zettabytes in 2025, as opposed to 64.2 Zettabytes in 2020.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 6

Current state-of-the-art

Industry and academic

➢ The main focus of the existing industry-specific solutions is primarily system performance data on proprietary hardware like HPE InfoSight and Ironstream, and software like SAP HANA.

➢ Focus on gathering and interpreting event log data based on a proprietary language (SPL), which increases the learning curve, like Splunk.

➢ Intense academic research on event logs.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 7

Own strategy: Early Recognition and Proactive Handling of Disruptions (ERPHD)

Short description of our approach

➢ Our strategy comprises an overall solution, which is vendor independent, it does not need inside knowledge regarding the implementation and functioning of the applications and it is based on leading edge technologies, like Artificial Intelligence, Data Analytics, Trend Analysis, Machine Learning, Anomaly Detection, etc.

➢ It can be conceived as a central monitoring and troubleshooting environment.

➢ It proposes a uniform event log parsing and analysis strategy, which does not assume inside knowledge of the structure.

Challenge

➢ We are not aware of any commercially available tool which uses this approach.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 8

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Own strategy: Early Recognition and Proactive Handling of Disruptions (ERPHD) – cont.

In a nutshell: strategy to avoid the disruptions of the services of computer centres

➤ Monitoring the health of the IT systems.

➤ Predicting possible malfunctions, interruptions, and downtimes.

➤ Proactive actions in order to avoid possible degradation of the QoS delivered by the IT systems.

➤ Immediate alerts on failures.

➤ Appropriate actions in order to avoid subsequent degradation of the QoS delivered by the IT systems.

➤ Corrective actions in order to facilitate the resumption of the failed activity.

➤ Appropriate actions to avoid similar cases in the future.

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

The advantages of our strategy

A particular selection

➢ All existing applications can be monitored regarding early recognition and proactive handling of disruptions.

➢ A central monitoring system can be set up, versus many monitoring tools due to our uniform strategy.

➢ Automated evaluation of the event log files, long manual search is obsolete due to the harmonised parsing strategy for all event logs.

➢ It is based on reliable leading edge technology.

➢ Enables real-time computer centres and IT infrastructure due to our failure prevention strategy.

➢ Avoids or reduces "hot working phases" at night for the IT personnel.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 10

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

The difficulties of our strategy

A particular selection

➢ Difficult architectural set-up, i.e., new algorithms have to be designed and implemented.

➢ Heterogeneous development teams including mathematicians and data scientist should be built upon, i.e., the algorithmic part of the development may be sophisticated.

➢ Increased development costs due to the unconventional development strategies.

➢ IT staff has to be additionally trained due to unconventional architectural and maintenance strategies.

➢ Strong management commitment to overcome the difficulties due to the anticipated challenges.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 11

Lessons learnt

Regarding our strategy

➢ Assuring and/or improving the QoS of a data centre is a very complex endeavour, in which the human component also plays a very important role.

➢ There should be rules set up, such that the actions or decisions taken by the service team should not be based upon the individual skills of its members, but on generally accepted guidelines.

➢ The experience and know-how of an individual employee should not be lost if he leaves the company.

➢ If the computer centre or the IT infrastructure exceeds a certain size and importance, the advantages of the our strategy may prevail over its disadvantages.

.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 12

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Conclusion and Future Work

Main result

➢ Facilitates the paradigm shift from more or less subjectively designed individualistic conceptions in handling of disruptions regarding the service of computer centres and IT infrastructures towards objectively established optimal solutions.

Outlook

➢ The relevancy of the statistical approach applied on event logs versus the classical approach, where the event logs are content-wise analysed, is of utmost interest. Identifying the weak points of statistical approaches including the limits of applicability, may be of advantage.

➢ The question arises whether there are alternative approaches, which do not rely on proprietary information regarding the IT systems.

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 13

TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Thank you

Thank you for your attention

Questions?

Martin Zinner
Strategy for Early Recognition and Proactive Handling of Disruptions Regarding the Service of Computer Centres and IT Infrastructures Based on Statistical Methods; The Seventeenth International Conference on Software Engineering Advances ICSEA 2022 October 16 to October 20, 2022 – Lisbon,, Portugal

Slide 14

TECHNISCHE
UNIVERSITÄT
DRESDEN

DRESDEN
concept