RISE and Open Source Monitoring of legacy OCP Systems funded by H2020

Tor Bjorn Minde, Chief Executive Officer
Dr Jon Summers, Research Leader in Data Centers
RISE SICS North, Lulea, Sweden
Agenda

- Introducing RISE ICE
- Introduce BTDC project
- Detailing the open source setup
- Showing results from the monitoring of the infrastructure
- Show results of the OCP systems operations
  - Heat sinks
  - Fans
  - Wind Tunnel and results
  - Workloads
- Summarise
SICS - ICE
A LARGE SCALE DATACENTER RESEARCH & TEST ENVIRONMENT
Mission and Vision

**Mission:** Turn great ideas into proven innovations in the data center space

- To build competence within the area of sustainable efficient data centers, as well as cloud applications and data analysis by enabling the use of testing in large-scale data centers
- To attract and thereby contribute to an increase of the number of scientists within the area of sustainable efficient data centers, as well as cloud applications and data analysis

**Vision:** To become the leading data center research institute in Europe

- The leading data center research institute in Europe within the area of sustainable efficient data centers, as well as cloud applications and data center centric data analysis
SICS ICE. A full-scale research datacenter and test environment with the purpose to increase knowledge, strengthen the DC ecosystem, and attract researchers.

- 1000 physical servers
- 250 kW
- 200 TB RAM
- Upto 10 petabyte storage
- 20,000 cores
- 80 GPUs
- HDFS clusters
- OpenStack ECC
- OCP servers
Business model

- Big data
- Machine learning
- IT Cloud
- Facility HW
- Utility

Delivering experiments and testing as-a-Service

Universities
R&I projects
Companies
The leading datacenter research facility in the Nordics

Testing in a flexible full-scale datacenter – without having to invest. With access to massive amounts of research data. Add a stand-by team of world-leading scientists at your disposal, contributing to your innovation. That is SICS ICE.

See what we could offer

https://ice.sics.se
"Ordinary" web-scale data center with optimized and stable environmental conditions

Mainly for datacenter monitoring and modeling and testing of data analytics and IT/cloud-related applications

- 5000 cores, Dell R730
- 40TB RAM
- Up to 7PB storage
- 80kW
- 10/40 Gb/s network
- Separate management network
- GPU accelerated Hadoop clusters
- 80 GPUs
- Openstack racks for ECC
- One OCP rack with 40 OCP windmill servers
A flexible data center facility lab, with unique possibilities to vary environmental conditions

Mainly for testing of data center facility installations and equipment

- Dell servers R530/430
- SICS East HPE cluster, HOPS staging cluster, Nyx Hortonworks cluster
- Prepared for both liquid cooling and free air cooling solutions
- Quick couplings for water, power and fiber
- Flexible raised floor solution to be able to compare with slab floors – reconfigurable.
ICE – Module 3
Open Compute Project Lab

A flexible data center facility lab, focused on OCP equipment

Mainly for testing of data center facility installations and OCP equipment

- OCP windmill servers
- 480 servers / 144 kW peak
- Prepared for free air cooling
- AC/DC power distribution using Trilogy HPDUs
ICE – Experiment area
Micro grid – Edge – Rack Wind tunnel

An experimental area for edge datacenter integration with a smart electrical micro grid and cooling network

Mainly for testing of data center facility integration, peak shaving, operations

- 10 kW solar panels
- 30 kWh electricity storage
- 22 kW cooling power
- 2 m³ cold water storage
- 10 kW IT-load

Immersion based liquid cooling experimental set-up
The Boden Type Data Center (BTDC) project is funded by the EU.

To build the most cost and energy efficient sub 1MW prototype data center in the world!

Visit https://bodentypeDC.eu

- Three Data Center PODs
- Designed for 500kW
- OCP Test POD will take contents of RISE North Module 3
- POD2 and 3 are designed for HPC/GPU donors for a target 350kW.
Rapid growth in data center energy consumption in Europe from 86TWh in 2013 to an anticipated 104TWh in 2020 (reference P. Bertoldi from the EU JRC).

Aim of BodenType DC project to build a prototype data center that is energy efficient – needs a comprehensive monitoring and measurement tool.

Efficiency is focussed on reducing power losses – no UPS, reducing cooling power consumption – using direct air (measured comparison with other methods) and better utilization of the IT systems – workload deployment and management.

BodenType DC H2020 Project

The Project partners are:

Visit https://bodentypeDC.eu
Construction of the Boden Type Data Center

Visit https://bodentypeDC.eu
Conceptual design of the Boden Type Data Center

Visit https://bodentypeDC.eu
Open source monitoring of BTDC One

Visit https://bodentypeDC.eu
Montoring data center Modules 1 and 2 (RISE ICE)

- **IT**
  - Server load
  - Server temps and fan metrics
  - Network load
  - Other IT information

- **Facility**
  - Temperature
  - Humidity
  - Power/per server/per rack
  - Cooling power
  - Cooling fan power
  - Etc. etc.
Analytics – What is the data be used for?

- The data has been used for
  - Energy statistics and operational metrics
    - PUE and other metrics. Continuous or ”batch”.
    - Is the data center operating in the best way possible?
  - Condition monitoring
    - Server behaviour
    - Automatic fault identification and root cause analysis
  - Optimization and control (closing loop)
    - Load balancing
    - Operational schemes, temperatures
  - Experimental evidence and verification
    - Validating simulations results (e.g. CFD)
    - Fine grained energy savings from strategies and tools
Data collection

- Open Source Tool
- Protocols to be read include SNMP, IPMI, Modbus, M-bus
- Useful in data center operation
  - Track what is happening
- Hard to export data for post-analysis
- Difficult to do efficient searches.
Long term storage – Need a data structure

- **Metric.** The metric should represent a specific “thing” or measurement, like Ethernet packets or temperature. The definition of what the metric represents is defined with the `<tag>`s.

- **Tag(s).** The `<tag>` field applicable in the SICS ICE installation and can have the following keys:
  - **dc** Defines what specific data center in the case of several data centers being monitored.
  - **pod** Data center module (POD) number, 1, 2, 3 etc.
  - **rack** The rack number in the pod. (This is left empty for equipment not mounted in rack.)
  - **host** Defines which host the values are read from, servers, weather station, CRAC, etc. (This data will in some cases be redundant.)
  - **source** What is the source of the measurement on the current host. For temperature metrics this could e.g. be cpu, ram, exhaust, top-front, etc.
  - **unit** The unit the metric is represented in.
  - **id** If one host has multiple sources of the same type id is used to distinguish between them, e.g. there can be multiple cpus on the same host.
  - **opt** Optional key, very rarely used.
Data access

- HDFS

- Divided in hourly files, in date folders.
- All data stored in plain text files.
- Human readable JSON format.
Data access - KairosDB

- Same information stored as in the HDFS
- Web-interface
- Possible to search, plot and export.

```plaintext
temperature <time> 23.1 dc=lulea pod=1 rack=3 host=p01r02emx888 unit=C source=front-top
temperature <time> 45.8 dc=lulea pod=1 rack=5 host=p01r05srv18 unit=C source=cpu id=1
temperature <time> -1.2 dc=lulea host=weather_station unit=C source=air
current <time> 1.1 dc=lulea pod=1 rack=5 host=p01r05hpdu18 unit=A source=outlet id=1
voltage <time> 231.1 dc=lulea pod=1 rack=5 host=p01r05hpdu18 unit=V source=outlet id=1
speed_rpm <time> 2280.0000 dc=lulea pod=2 rack=8 host=p02r08srv05 unit=RPM source=system id=3
```
Data "transformation"

- How to get the data from Zabbix to the long term storage?

- Kafka
  - Distributed streaming platform
  - Reliably get data between systems or applications
  - Publish and subscribe (useful for smaller experiments)
  - Zabbix input module
    - Read data from Zabbix DB and get it into the Kafka "stream"
  - KairosDB output module
    - Extract from Kafka and write to KairosDB
  - HDFS output module
    - Extract from Kafka and write to HDFS
Example data, POD2 maintenance work
Cooling innovation with the fan controllers

Traditional
Set supply
Temp

Experimental
Set supply
Pressure

Pod 1 – **Cooling Units set** $T_{\text{supply}}$

**IT Equipment**

$T_{\text{return}}$

$\Delta T$ fixed for all IT units

Wind tunnel to determine $\Delta T$ – **IT equipment sets** $T_{\text{return}}$
Replaced aluminium heat sinks containing embedded heat pipes with copper heat sinks.

Aim to achieve lower flow rates and higher delta temperature across the Windmill servers.

Flowrate control is also achieved by the use of some developed fan controller cards.
Heat pipe aluminium versus copper heat sinks.

The above are initial results from a desktop experiment. Ongoing work in the wind tunnel for better control and hopefully more conclusive results.
FAN Controller results – taking control over the BIOS

Fan signal from OCP motherboard
- Set temperature with the HEX
- Flow rate with the FAN – can pressurize the cold aisle.
- Humidity?
- Upstream and downstream temperature measurements.
- Can support between 1U and 4U.
Results with the OCP Windmill Server

- Wind tunnel results are captured using the same data center monitoring tools.

- Uses the same monitoring system as the data center modules

- Enables remote access and control.

- Temperatures are in °C
Thanks to UK startup Edgetic for the results as a collaborator making use of the wind tunnel with its remote control capability and open source monitoring system.

- Blue points are measured and yellow are actually predicted from careful profiling of the OCP system behaviour.
Types of workloads that will operate in the BTDC POD

Workload Classification

Based on Processing Model
- Batch
- Interactive

Based on Resource
- Memory
- CPU
- IO
- Database

Based on Computer Environment
- Physical
- Virtual
- Cloud

Based on Generation
- Synthetic
- Real
- Hybrid
- Application based Load Patterns

Based on Application
- Web
- Social Network
- Video Service
- Smart Cities
- Industrial Internet of Things
- etc
Building realistic workloads for BTDC project

User Story 1
UC 2  UC 3
SUC 2  SUC 3  SUC 4  SUC 5
AP1 2  AP1 3  AP1 4  AP1 5
AP0 2  AP0 3  AP0 4

domain user stories
use cases
(sub) use cases
Application Primitives level 1
Application Primitives level 0

Server Type A
Server Type B
Server Type C
Server Type D

Docker Host
Docker Host
Docker Host
Docker Host
Docker Host
Docker Host
Docker Host
Docker Host
Docker Host

Docker Image DB

Workload Definition

Data Center HW

Building IOT equivalent workloads
SUMMARY

- Operate and monitor experimental test data centers.
- Platform based on open source software to monitor data center facilities, IT utilisation and lab based experiments.
- Created a 150kW IT footprint using legacy OCP Windmill servers, with power and network.
- IT footprint will go into a prototype data center being built in Boden and cooled by direct air with an ambitious linking of facility and server fans.
- Characterized the thermal and power of the OCP windmill server using a combination of heat sinks and locally programmed fan controllers, together with a uniquely developed server wind tunnel.
- The EU funded project has developed a strategy for deploying synthetic and near to real Industrial IOT workloads.

Visit https://bodentypeDC.eu
OPEN. FOR BUSINESS.