A peek into the future of the datacenter industry

Head of Lab
Tor Björn Minde
RISE ICE datacenter research

RISE ICE datacenter

2000 physical servers
250 kW
200 TB RAM
18 petabyte storage
50,000 cores
248 GPUs (with LTU)
1.1 M cuda cores
12.5 petaflops
HDFS clusters
OpenStack ECC
OCP servers

• 20 projects, from the ground to the cloud
• 25 employees
• 4 MEUR turnover
• Established 2016
• 50 MSEK invested

A full-scale research datacenter and test environment with the objective to increase knowledge, strengthen the AI & DC ecosystems and attract researchers.

Stakeholders: Ericsson, ABB, Vattenfall, Facebook, LTU, Region North

© RISE AB 2021
RISE ICE Datacenter test environment

Module 1 & 2 Compute clusters

Module 3 OCP Lab + Climate box + Heat box

Module 4 Facility hardware test lab

Edge + wind tunnels + liquid cooling test bed

FUTURE PREDICTIONS USING SCENARIOS
To predict the future with certainty is impossible!

Defining uncertainties

Low end-point  Uncertainty 1  High end-point

Low end-point  Uncertainty 2  High end-point

Low end-point  Uncertainty 3  High end-point
4 scenarios in a 2D map of 2 uncertainties

- Scenario 1: Very low - Very high
- Scenario 2: Very high - Very high
- Scenario 3: Very low - Very low
- Scenario 4: Very high - Very low

4 scenarios in a 1D map of 3 uncertainties

- Scenario 1
- Scenario 3
- Scenario 4
- Scenario 2
3 Uncertainties for datacenters

- Development of new processing technologies
  - No new
  - Many new

- Data processing growth
  - Very low growth
  - Very high growth

- Importance of Sustainability
  - Very low importance
  - Very high importance
Resource utilization (Uncertainty 1)

Sharing

100% → 2030

Energy efficiency (Uncertainty 1)

PUE

1.0 → 2030
End of Moore’s Law (Uncertainty 1)

Data growth – pace of change (Uncertainty 2)
Global power need

- Total need < 10-20 GW
- < 50-60 GW – 40 zettabyte
- 40 GW – 4 zettabyte
  = 1000 Facebook buildings

Source: Länsstyrelsen Norrbotten 2014

Market growth 2014-2020 of new mega datacenters

- A total 700 (>5 MW), based on sources from IDC, Gigaom and Datacenter Dynamics.

Source: Länsstyrelsen Norrbotten 2014
Growth of renewables (Uncertainty 3)

Growth of circular economy (Uncertainty 3)
**5G Requirements**

- **1000x** Mobile Data Volumes
- **10x-100x** Connected Devices
- **5x** Lower Latency
- **10x-100x** End-user Data Rates
- **10x** Battery Life for Low Power Devices

Source: Ericsson
Evolution Towards 6G

- 10000x Mobile Data Volumes
- 100x-1000x Connected Devices
- 50x Lower Latency
- 100x-1000x End-user Data Rates
- 100x Battery Life for Low Power Devices

5G/6G networks and edge datacenters

<table>
<thead>
<tr>
<th></th>
<th>3G</th>
<th>4G</th>
<th>5G</th>
<th>6G</th>
</tr>
</thead>
<tbody>
<tr>
<td>RTT&lt;sub&gt;UP&lt;/sub&gt;</td>
<td>20%</td>
<td>50%</td>
<td>88%</td>
<td>95%</td>
</tr>
<tr>
<td>RTT&lt;sub&gt;DOWN&lt;/sub&gt;</td>
<td>88%</td>
<td>50%</td>
<td>28%</td>
<td>5%</td>
</tr>
<tr>
<td>RTT&lt;sub&gt;TOTAL&lt;/sub&gt;</td>
<td>200 ms</td>
<td>40 ms</td>
<td>25 ms</td>
<td>21 ms</td>
</tr>
</tbody>
</table>

- Increasing part of the delay
- Approaching 20% for 5G
- Depending on many parameters

- 250 ms
- 10-40 ms

- < 5 ms
- 10-40 ms
What is Edge computing?

Edge computing is about pushing intelligence and processing capabilities closer to the end user or where the data originates or offer off-loading.

End user perspective

- Enhanced user experience
  - Response times, battery life, HW-extensions
- Edge service or edge plan
  - Improves applications marked with an “e” when edge capacity is available
- Virtual hardware (GPU) service
  - Improves e.g. gaming applications with HW-extensions
- Service & HW follows you
  - Edge compute containers moves
Size of Edge?

- Edge will not eat the cloud
- Edge will make the cloud grow

Use of Edge and Central cloud
Basic use case characteristics

<table>
<thead>
<tr>
<th>Type</th>
<th>Response times</th>
<th>Data amount</th>
<th>Traffic amount</th>
<th>Cache</th>
<th>DC location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cold storage</td>
<td>seconds</td>
<td>Gigabytes</td>
<td>Mb/s</td>
<td>Remote cloud</td>
<td></td>
</tr>
<tr>
<td>Off-line big data crunching</td>
<td>seconds</td>
<td>Gigabytes</td>
<td>Gb/s</td>
<td>Remote cloud</td>
<td></td>
</tr>
<tr>
<td>Chat/IoT/event type communication</td>
<td>100th milliseconds</td>
<td>kilobytes</td>
<td>kb/s</td>
<td>Remote cloud</td>
<td></td>
</tr>
<tr>
<td>Web/app rendering</td>
<td>100th milliseconds</td>
<td>Megabytes</td>
<td>Mb/s</td>
<td>Yes</td>
<td>Remote cloud</td>
</tr>
<tr>
<td>Up/Down-link streaming</td>
<td>10^6 milliseconds</td>
<td>Gigabytes</td>
<td>Mb/s</td>
<td>Yes</td>
<td>mix</td>
</tr>
<tr>
<td>Remote operation/ Real-time AV conferencing</td>
<td>10^6 milliseconds</td>
<td>Megabytes</td>
<td>Mb/s</td>
<td>Yes</td>
<td>mix</td>
</tr>
<tr>
<td>Real-time aggregation/ analytics/ VR/ AR</td>
<td>milliseconds</td>
<td>Megabytes</td>
<td>Gb/s</td>
<td>Yes</td>
<td>edge</td>
</tr>
<tr>
<td>Transaction/Control loops</td>
<td>milliseconds</td>
<td>kilobytes</td>
<td>kb/s</td>
<td>edge</td>
<td></td>
</tr>
</tbody>
</table>
**SCENARIOS**

- **Scenario 1** “Bold”
- **Scenario 2** “Disrupt”
- **Scenario 3** “Status Quo”
- **Scenario 4** “Reduction”

Data processing growth:
- Very high growth
- Very low growth

Development of new processing technologies:
- No new
- Many new
Uncertainties 3 dimensions

Scenario 1 “Bold”
- No new
- Very low growth
- Very low importance

Scenario 2 “Disrupt”
- Many new
- Very high growth
- Very high importance

Scenario 3 “Status Quo”
- No new
- Very low growth
- Very low importance

Scenario 4 “Reduction”
- Many new
- Very high growth
- Very high importance

Importance of Sustainability
- Very high importance
- Very low importance

Data processing growth
- Very high growth
- Very low growth

Development of new processing technologies
- Many new
- No new
GROWTH OF NEW MEGA DATACENTERS/YEAR WILL EXCEED 100

EDGE WILL BE UBIQUITOUS AND AN INTEGRAL PART OF 6G
EDGE WILL MOVE NON LATENCY DEPENDENT COMPUTE OUT OF CITIES

NORDICS WILL HOST 50% OF ALL EUROPEAN NON LATENCY-DEPENDENT CLOUD
EDGE WILL BE POWERED BY HYDROGEN OR BIOGAS IN CITIES

LIQUID COOLING WILL BE EVERYWHERE
CLOUD SITES WILL BE SELECTED FOR INDUSTRIAL SYMBIOSIS

ALL MATERIAL IN THE DATACENTER WILL BE IN CIRCULAR ECONOMY
ALL PARTS OF THE DATACENTER WILL BE SOFTWARE DEFINED

ALL DATACENTERS WILL SELF-LEARN AND OPTIMIZE OUTCOME/W
NEW COOLING METHODS WITH NON MOVING PARTS NEEDED

SECURED CONTROL AND INFERENCE NEEDED IN EDGE NODES
CODE OPTIMIZATION WILL GROW POPULAR AGAIN FOR APPLICATION SOFTWARE

Thank you!

- Tor Björn Minde
  - tor.bjorn.minde@ri.se
  - +46 70 624 2959
  - @torshammer, @ICEbyRISE