HPC2N's high density compute room

NDGF Manager
Mattias Wadenstein
<maswan@ndgf.org>
Niklas Edmundsson
<nikke@hpc2n.umu.se>

2022-11-02
HEPiX Fall 2022
Umeå, Sweden

NORDIC E-INFRASTRUCTURE COLLABORATION

Overview

- A trip down memory lane
- Changes and experiences
- Future thoughts

As seen at HEPiX!

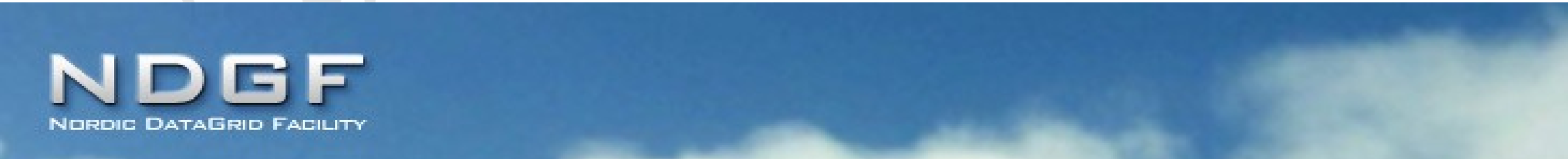
- I have given a few updates on this as we built and started using it
- Some lucky fellows got a tour in the 2009 Spring HEPiX
- But first, a recap:

NDGF Site Report

Mattias Wadenstein
System Integrator, NDGF

HEPiX Spring 2008 at CERN
Geneva, 2008-05-05

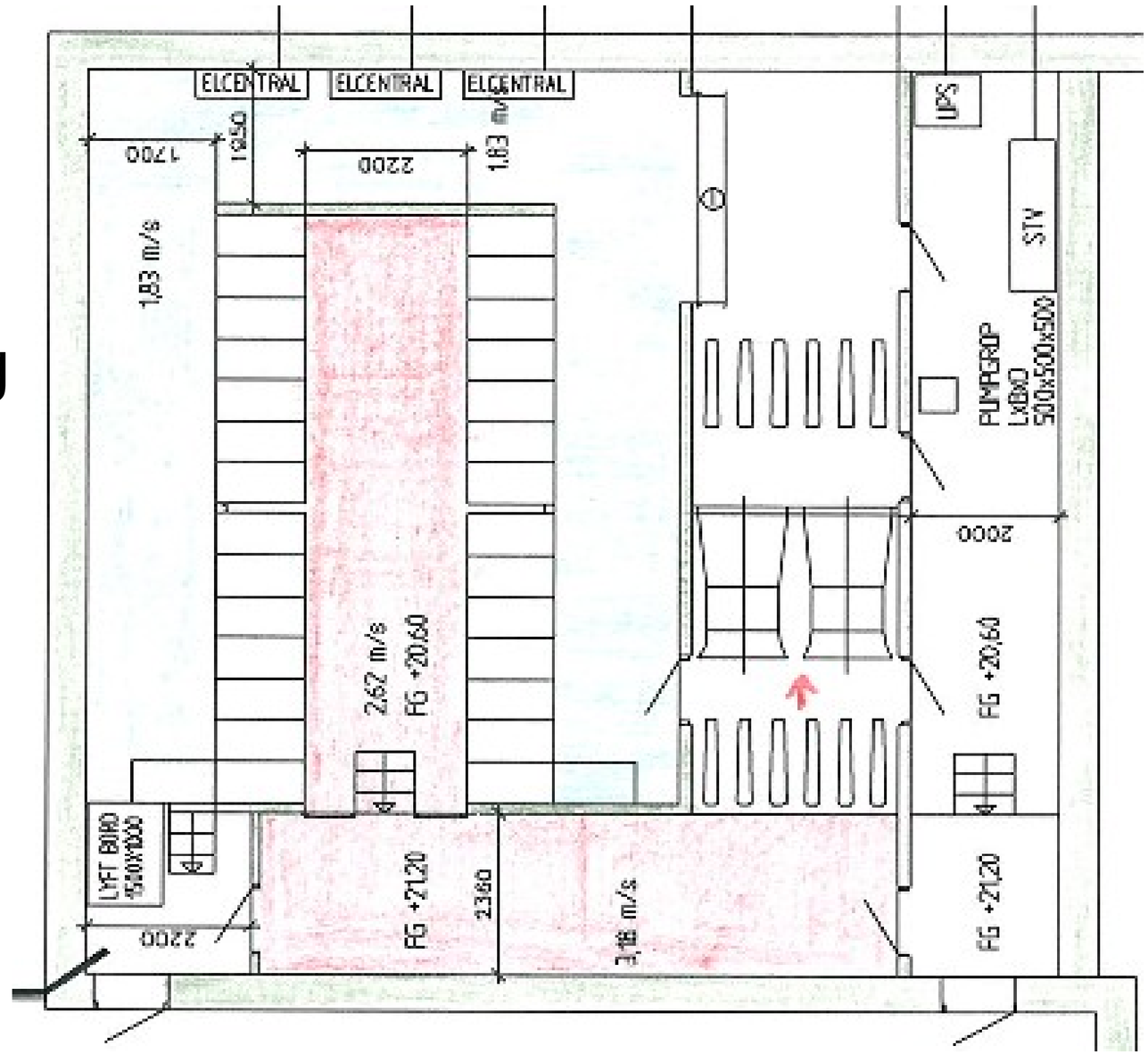
NDGF
NORDIC DATAGRID FACILITY



Example site: HPC2N

New machine room

- strict hot/cold aisle
- 28 racks
- 25 kW front to back aircooling per rack
- total ~400kW
- no redundant power
- Only for computing, storage&servers in old machine room
- no raised floor
- 3.5m high



NDGF
Nordic DataGrid Facility

Example site: HPC2N

Working almost as well as imagined

- Slight issue with back-draft where the air does a right angle turn
- Some minor regulating issues (ventilation fans running faster than needed)

Cools 20kW at 0.5kW overhead

Cools 250kW at 3-4kW overhead

NDGF
NORDIC DATAGRID FACILITY

Example site: HPC2N

Per rack:

1 x 32A 400V 3-phase

Or 3 x 32A 230V 1-phase

Separate cabling ladders for:

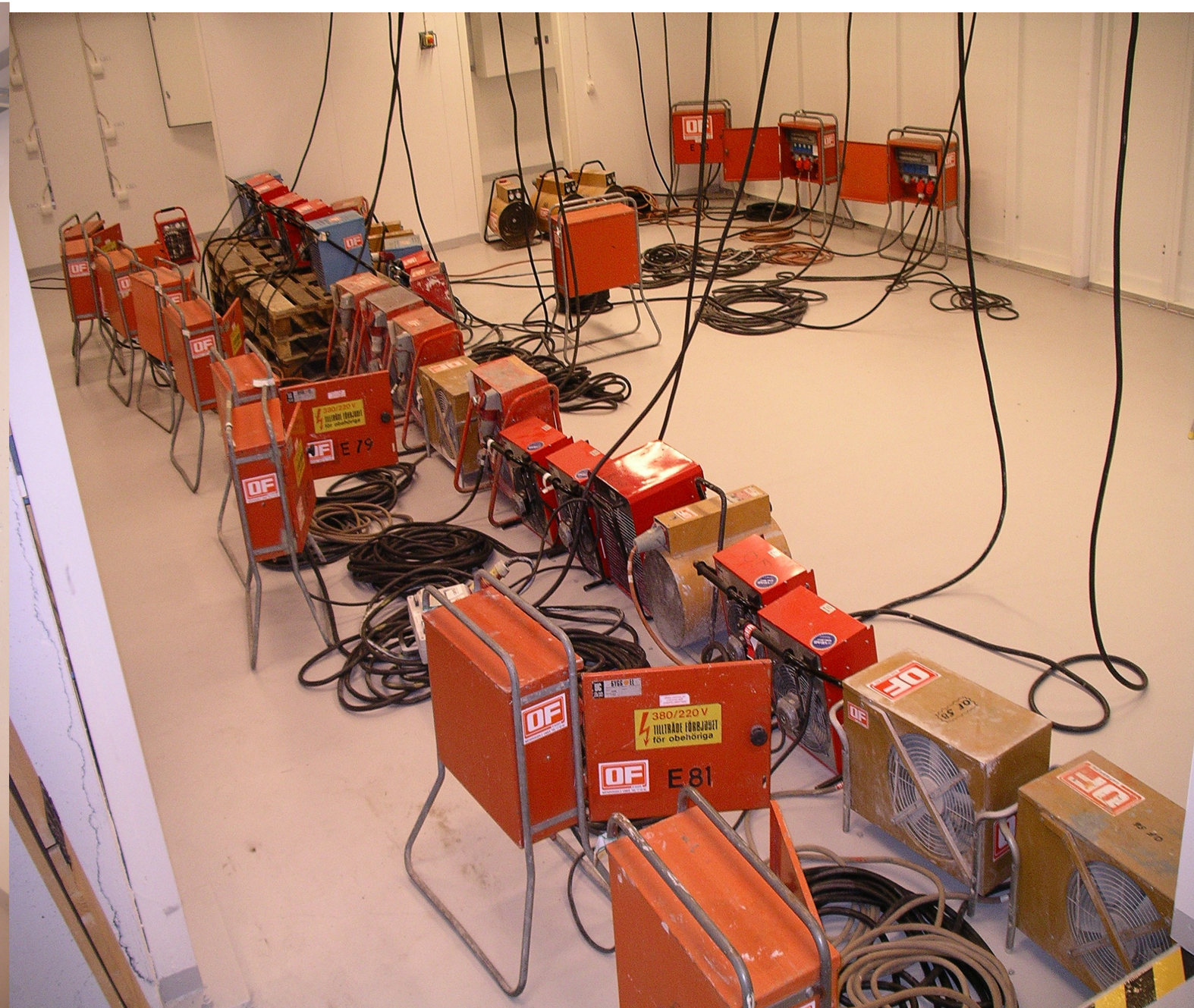
HPC interconnect (infiniband, myrinet, etc)

Ethernet

Power

NDGF
NORDIC DATAGRID FACILITY

Example site: HPC2N



NDGF
NORDIC DATAGRID FACILITY

Example site: HPC2N

Things fixed along the way

- Ventilation fan speed regulation, base on differential pressure instead of return temperature.
  - We want as high return temperature as possible, counterproductive and a waste of energy to increase ventilation fan speed based on temperature.
  - Differential pressure between cold/hot aisle is an accurate indicator of airflow need.
  - We aim for a slight overpressure on the cold side, computer fans just need to drag the air they're served through the chassis.

Things fixed along the way

- Work environment, add some sound dampening to walls to reduce noise that's hard on human ears.
  - Concrete-only floor/wall/ceiling
- Fire alarm detection improved
  - Initial detection had regular room sensors and assumed air was not moving...

Going from 400 kW to 650 kW capacity (sustained)

- Adding another 300 kW class cluster would exceed initial 400 kW total
  - Infrastructure upgrade in 2011
  - Increased power feed capacity by upgrading transformer (800 kVA)
  - Increased cooling capacity by upgrading circulation pump
  - Added more power feed to racks, standardize on 2x 32A 3 phase outlets (started out with mixed bag 1 or 3 phase).
  - 44 kW power peak per rack

Going from 400 kW to 650 kW capacity (sustained)

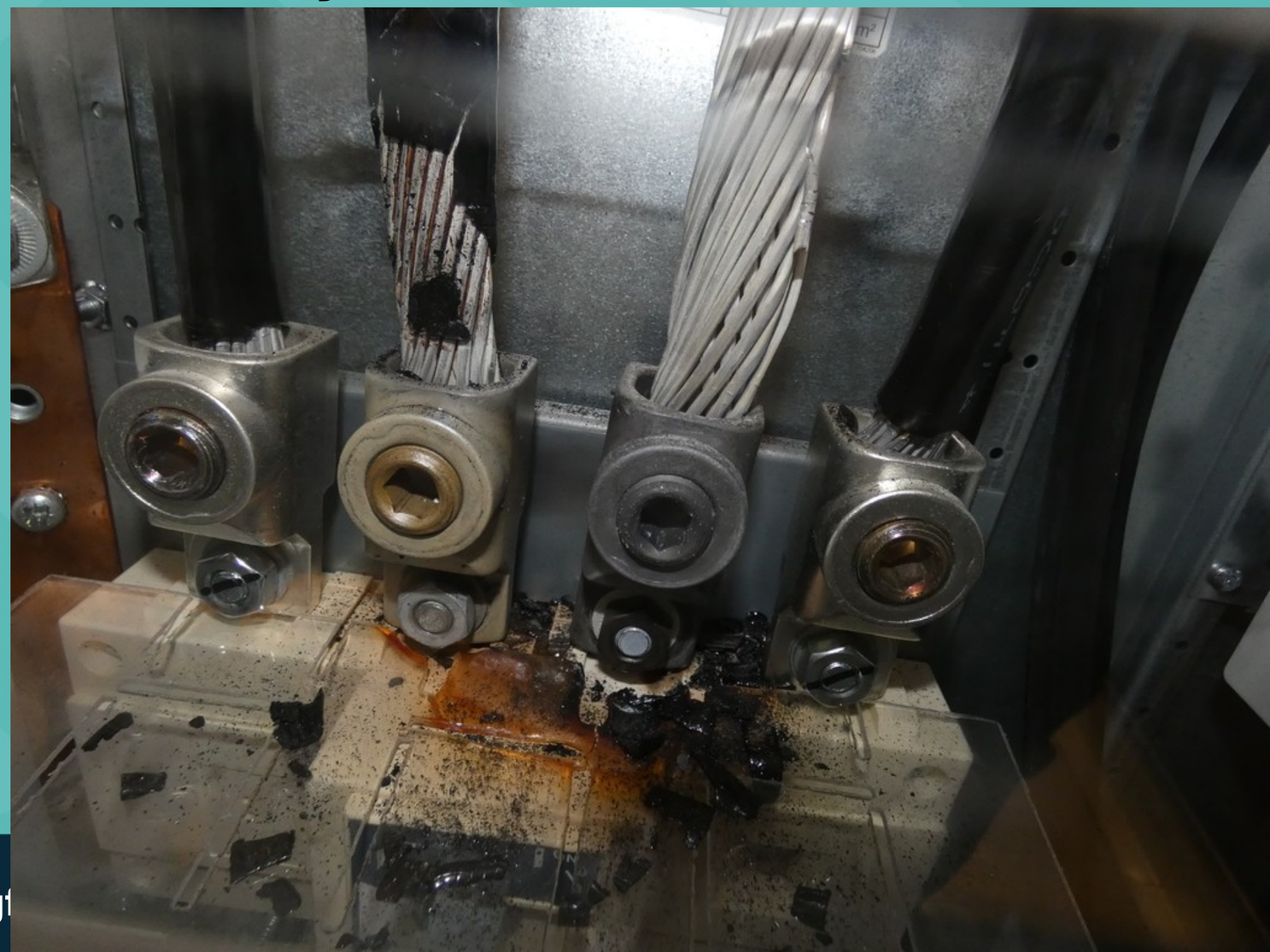
- Temperature in cold aisle raised from 18C to 22C, 50%Rh.
  - Needed to handle more power in datacenter without increasing airflow (more humidity to carry the heat)
  - Also resulted in drastically improved comfort for humans.
- Uneven airflow/pressure
  - Assymetric layout between the two aisles
  - Different equipment/load/airflow
  - Solved by an adjustable "wing"

Challenges

- Min 20 degrees delta-t required to have 650 kW cooling capacity with margins.
- Procuring systems with 20 degrees delta-t
  - Individual compute nodes OK, but switches etc drags the delta-t down.
- Getting vendors to do (not only promise):
  - reasonably sealed front-to-back racks.
  - Power draw evenly balanced on all phases both on rack and cluster level
    - We have 6 power distribution cabinets, not good enough to only care on a cluster level.
    - Very common with slight imbalance, repeated for all racks, skewing the entire system balance. Or a switch rack that just blows cool air through.

Mishaps

- Breaker meltdown in one of the power distribution cabinets, happened after 11 years.
  - Root cause was probably cable torqueing.
  - Added periodic inspection with heat camera to discover issues in time to avoid future smoke-emitting incidents.



Experience with UPS-less compute-only datacenter

- Working as expected wrt researchers/compute job experience
  - Job crash is the same regardless of cause (power loss vs node crash)
  - Remedy: Resubmit job
- No increase in equipment/PSU failure
  - PSU:s that did fail did so regardless of power outages etc.
- Requires good-enough power/cooling stability
  - Was a bit taxing when on-campus excavators homed in in power infrastructure...
  - Outages due to cooling loss same order of magnitude as power

Thoughts about future

- Highly efficient and flexible design
- Good return temperature on cooling water possible
  - Able to heat buildings without heat pump from aircooled servers
- No practical limit on cooling capacity per rack
  - as long as the equipment can cool by pulling in enough 22C air from the front
- Never want raised floors again!

Questions?

NORDIC E-INFRASTRUCTURE COLLABORATION