ECMWF data centre migration from UK to Italy

Stanislav Burlakov
ECMWF
European Centre for Medium-Range Weather Forecasts

Established in 1975
- 23 member states
- 12 cooperating states
- About 430 staff from over 35 countries

Produces global Numerical Weather Predictions (NWP)
- A 24/7 operational service and a research institute
- Headquarters in Reading (UK)
- Sites in Bologna (Italy) and Bonn (Germany)

Has one of the largest supercomputing facilities and meteorological archives in the world
- Multiple Top 500 list supercomputers
- Data archive of approximately 600PB
- Disseminate tens of Terabytes of data daily
Migration of Data Centre to Italy

ECMWF’s data centre has been in Reading for over 40 years.

ECMWF’s next supercomputers should provide a tenfold increase in computational capacity.

The current facility does not offer the required flexibility for future growth.

New data centre location decided based on an international bidding process.
New Network and Security Infrastructure Design

- IP Fabric Architecture (Leaf-Spine)
- Multi-site Topology (Hall Blueprint)
- Branch Office Architecture
- Security Layer Approach

An opportunity to make a change
Technical Migration challenges

- Long-distance transfers – effects of packet loss and latency on throughput
  - WAN emulator testing
  - 100Gbit bandwidth tests and kernel tuning to establish a performance baseline

- Complex data flows
  - Application owners are not network engineers – data flows not always clear
  - Complex inter-site routing topology

- New set of technologies
  - Change to different network equipment vendor and need for upskilling
  - Change from “classic” routing to EVPN/VXLAN IP fabrics
WAN: Internet Edge design

Wide Area Network

ECMWF (AS 42081)

- JISC (AS 786)
  - JISC-PE2
  - JISC-PE1
- GARR (AS 137)
  - GARR-PE1
  - GARR-PE2
- LEPIDA (AS 31638)
  - LEPIDA-PE1
  - LEPIDA-PE2

RMDCN

- GTT (AS 3257)
  - RMDCN-CE1
  - RMDCN-CE2

GARR (AS 137)

- JISC (AS 786)
  - JISC-PE2
  - JISC-PE1

LEPIDA (AS 31638)

- JISC (AS 786)
  - JISC-PE2
  - JISC-PE1

ECMWF Bologna

GARR

- JISC (AS 786)
- LEPIDA (AS 31638)
- GTT (AS 3257)

eBGP

1G 10G 100G 2x100G

ECMWF
Stanislav Burlakov
Master-Network-Diagrams.vdx
Page-1
26/10/2022
Data Centre: Data Hall Design

- VDI
- Linux Clusters
- Storage
- HPC Peripherals

- Virtual Hosts Enterprise Test/Dev/DMZ/Web/DB

- Firewall Cluster
- CDS

- Internet Edge Router
- ADC

- DHS Spines

- L3 Fabric
- EVPN-VXLAN
- Multiple Network Segments

- Spine
- Leaf
- 25G
- 100G
- 2x 100G

- Multiple Pairs

- Data Hall

- To WAN
- To other DC

Stanislav Burlakov - ECMWF

26/10/2022
Data Centre design

Offices Firewall

FW-DH1 MGMT
FW-DH2 MGMT

Internet Edge 1

Internet Edge 2

Bologna Offices

Data Hall 1 (DH1)

Data Hall 2 (DH2)

Anycast GW-10

VRF A, VXLAN 10, Stretched

VRF B, VXLAN 21, DH1

VRF A, VXLAN 10, Stretched

VRF B, VXLAN 22 DH2

Data Handling

Lepida CE1
RMDCN CE1
GARR CE1
GARR CE2
RMDCN CE2
Lepida CE2

Internet

Firewall learns all routes from all VRFs and advertises a Default back to all VRFs* in both DH.

*RMDCN VRF learns all routes

Both FW clusters use AS PATH PREPEND to direct all traffic to the active cluster. FW-DH1 is generally preferred

VRF A, VXLAN 10, Stretched
Target share their routes between DH1 and DH2

12 EVPN stretch

DKI

13 EVPN routing

Network Diagrams.vsdx

Page 1
26/10/2022

Stanislav Burlakov - ECMWF
WAN: Inter-site connectivity

Firewall learns all routes from all VRFs and advertises a Default back to all VRFs* in both DH.
*RMDCN VRF learns all routes

Both FW clusters use AS PATH PREPEND to direct all traffic to the active cluster. FW-DH1 is generally preferred.

FW-DH1 MGMT

FW-DH2 MGMT

VRF A, VXLAN 10, Stretched

VRF B, VXLAN 22, DH2

Anycast GW-22

Anycast GW-10

L2 EVPN stretch

L3 EVPN routing

DCI

VRF A, VXLAN 10, Stretched

VRFA, VXLAN 21, DH1

VRFB, VXLAN 21, DH1

Anycast GW-21

Anycast GW-10

L2 EVPN stretch

EVPN all VRFs with the same Route Target share their routes between DH1 and DH2

Both FW clusters use AS PATH PREPEND to direct all traffic to the active cluster. FW-DH1 is generally preferred.
WAN: EUMETCAST with PIM-SSM

Stanislav Burlakov - ECMWF

GEANT

EUMETCAST

GARR (AS 137)

GARR-PE1

GARR-PE2

Telecommunications room 1

Telecommunications room 2

Firewall Cluster

Data Hall 1

Data Hall 2

EUMETCAST Server

EUMETCAST Server

IGMPv3 [S,G] Join

EV PN Type 3 / Type 6-8 Routes

PIM Register

PIM Register

PIM Register

PIM Register

BGP multicast

IGMP

EVPN

PIM
eBGP

Terrestrial

SLS

BLS

Master-Network-Diagrams.vsdx

Page-1

27/10/2022

ECMWF

Stanislav Burlakov
Thank you! Questions?