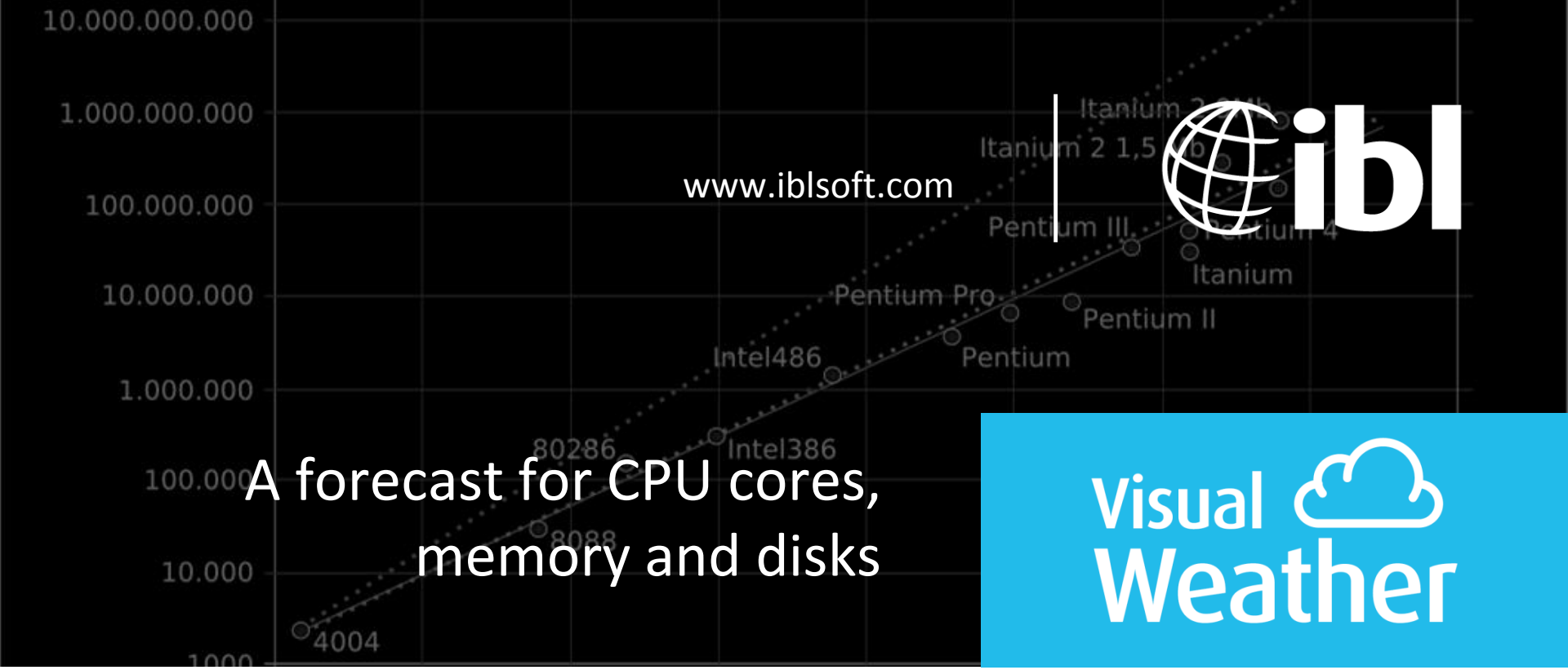


www.iblsoft.com

ibl

A forecast for CPU cores, memory and disks

Visual Weather

Jozef Matula
Visual Weather Team Lead

Michal Weis
Managing Director

ECMWF Visualisation in Meteorology week 2015
26th European Working Group on Operational Meteorological Workstations, 30st October 2015

Paradigms to be questioned

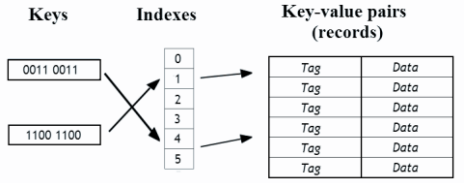
Visual Weather

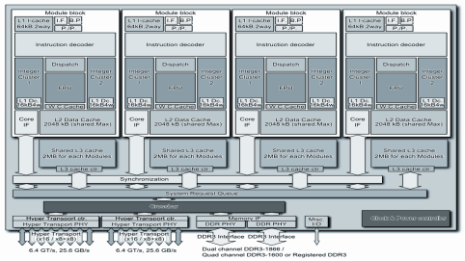
In-memory cache is easy and good thing to accelerate your code.

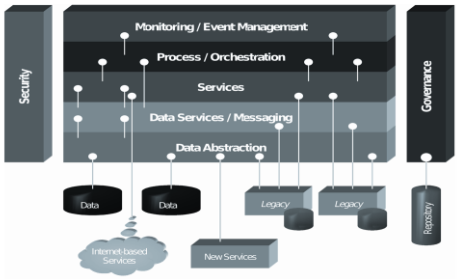
Buying more CPU cores helps to improve performance.

Memory (RAM) is fast.

Web Services and/or SOA help to solve performance by increasing scalability.







ibl

Background

Visual Weather

- Visual Weather has a growing number of Web Service users, in particular demanding:
  - OGC Web Map Service (weather maps) and Web Coverage Service (weather data)
  - Custom data extractions and processing in Python
- Operating on wide variety of hardware:
  - slowest known 2 cores, 4GB RAM
  - fastest standalone 120 cores, 512GB RAM
  - load balancing clusters
- Over the past year or two we experienced several “scalability” issues which challenged our IT knowledge.

ibl

Visual Weather Web Service Architecture

Visual Weather

- Native in C++, closely coupled to maximise performance:
  - HTTP(S) server itself
  - HTTP(S) request handling code
- Multi threading model
  - No subprocess invocation cost
  - Minimum interprocess communication cost
  - Ability to share memory caches between threads - no network cost
- Typical response time
  - Δ ~100ms including encoding and transfer (WMS, WCS)
  - ≦ ~10ms for cached products (e.g. map tiles)
- Various deployment patterns (mostly Linux):
  - Standalone server systems
  - Clusters

ibl

Open Your Minds :-)

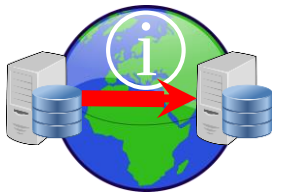
Visual Weather

Real world exercise:

How to transfer high resolution model, ~2TB model run, from UK to IBL office as quickly as possible?

1st option - Internet

- IBL downlink - 50MBit/s = 5.6MB/s (theoretical)
- Expected ideal transfer time ~4.3 days

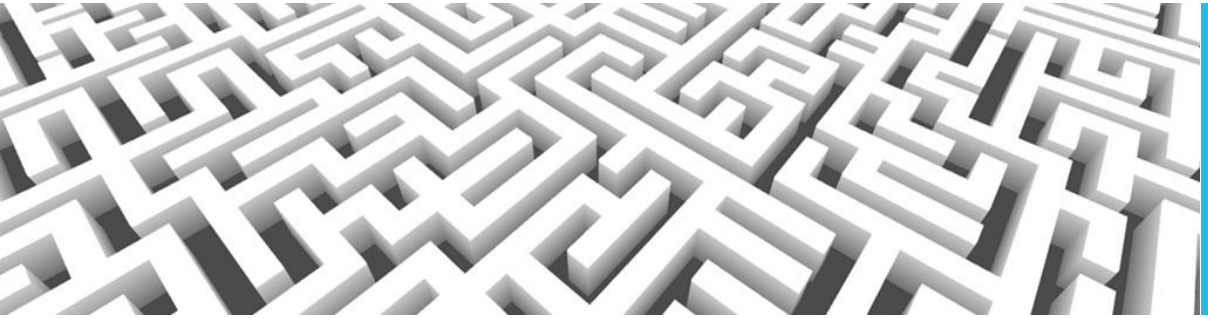


2nd option - UK Royal Mail

- Ideal delivery time ~2 days = ~48 hours
- Transfer speed ~12.1 MB/s

10 hours to copy data to portable disk over 1GBit LAN!

ibl



Visual Weather

Horizontal Scaling Paradox

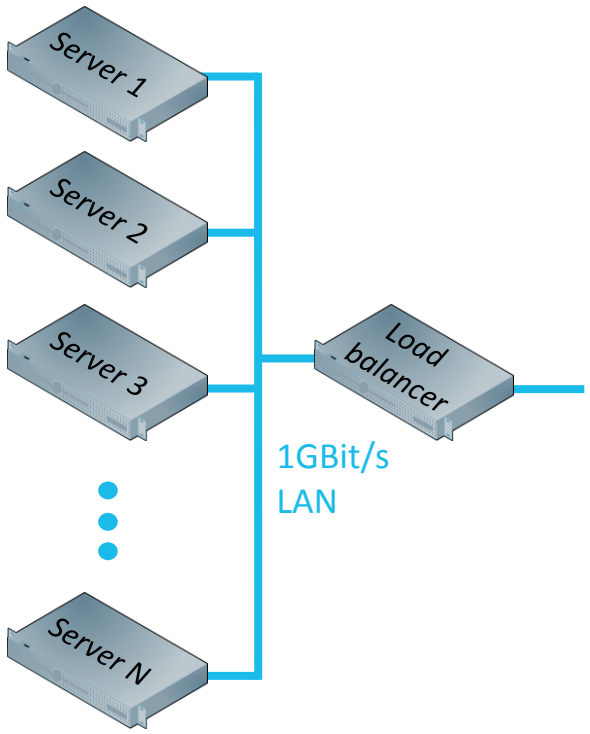
ibl

Horizontal Scaling Paradox

Visual Weather

If one computer is slow, can you just add more? Once upon a time we benchmark and our target was:

- Handle 90000 WCS requests in less than 30 minutes.
- ~50 requests per second... sounds easy. But we only approached this limit!



Forgot to mention: output data volume 250GB

That means 142 MB/s ⇒ 1GBit/s network is not fast enough.

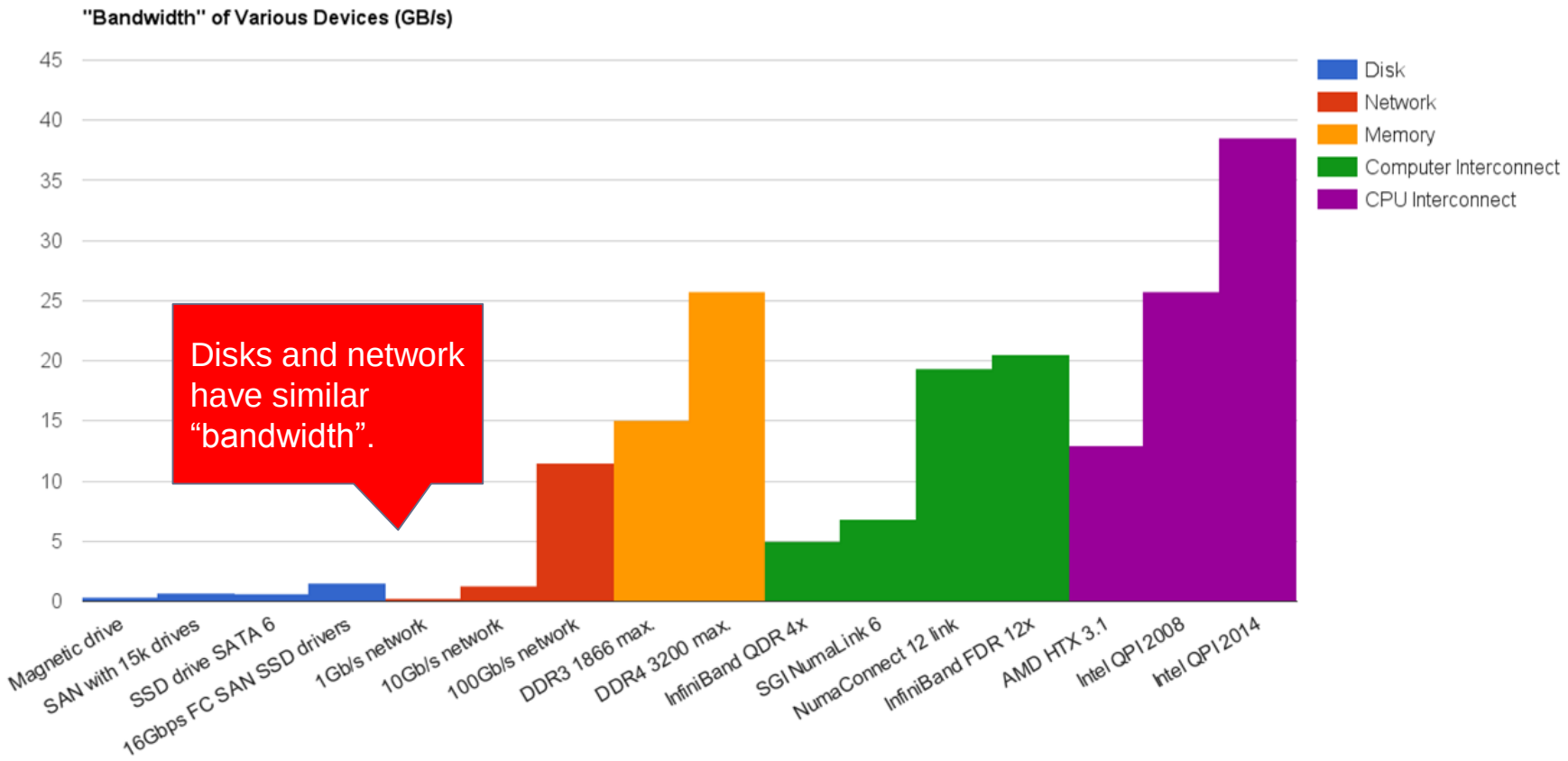
ibl

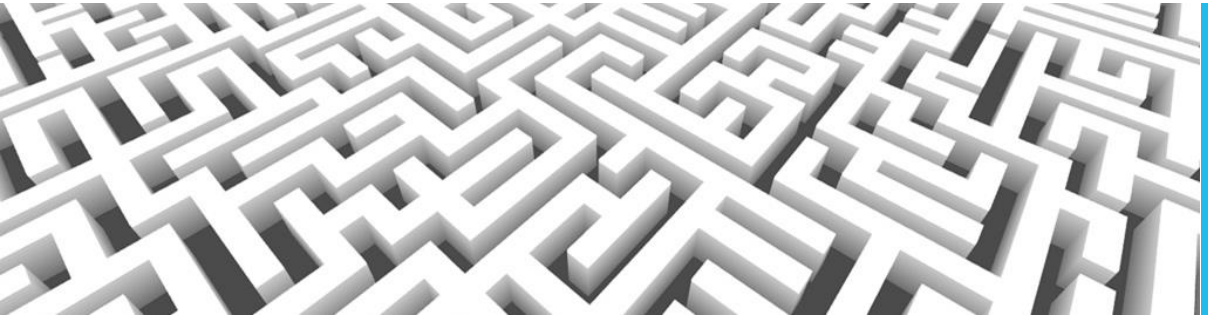
Horizontal Scaling Paradox

Visual Weather

Building SOA (or basically any web service based architecture) is an easy way to run out of your network bandwidth.

If you can keep your computing problem in memory, do so.





Visual Weather

Massive Memory Cache Paradox

ibl

Massive Memory Cache Paradox

Visual Weather

Once upon a time we benchmarked on a server with “only”:

240 CPU cores (20 physical processors)

4TB of RAM

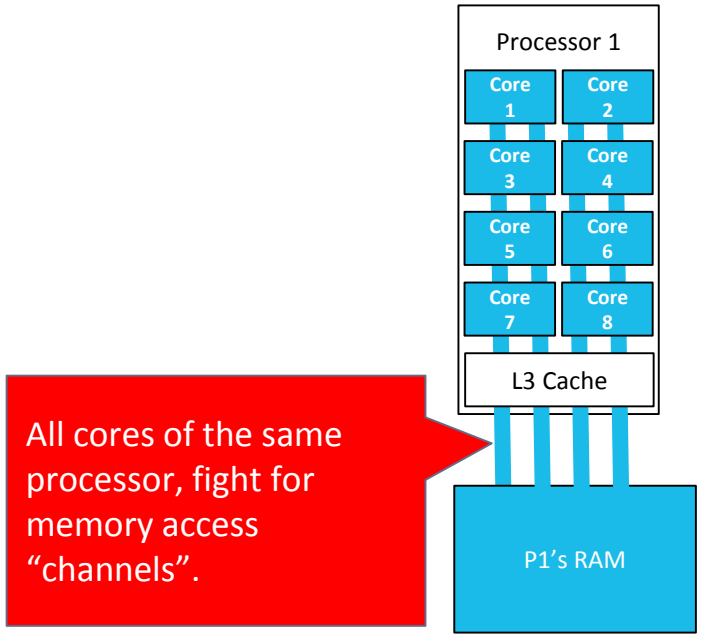
*Almost ideal scaling up to 10 cores, much worse up to 20 and actual **performance degradation** up to 200 cores!*

There were several reasons for this:

- Bad luck
- Massive in memory cache for NWP grids
- Large NUMA (Non-Uniform Memory Access) system composed of multiple blade computers targeted for HPC

Massive Memory Cache Paradox

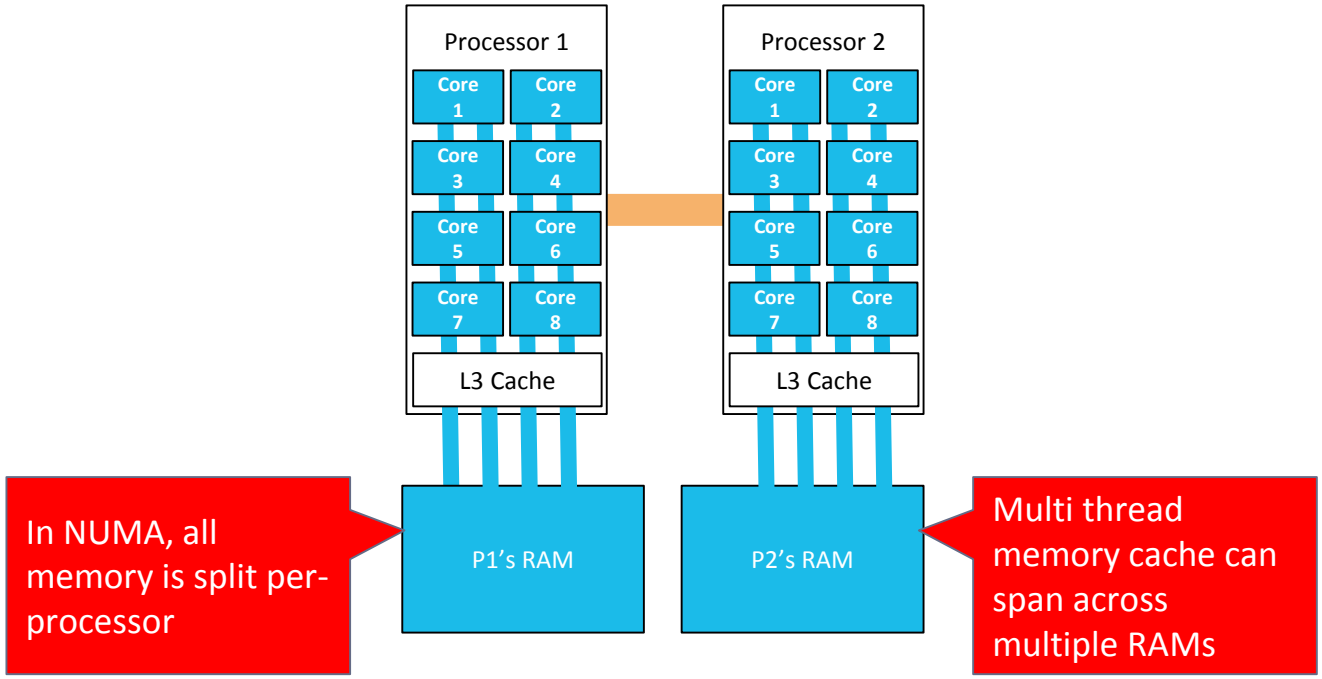
Visual Weather



All cores of the same processor, fight for memory access "channels".

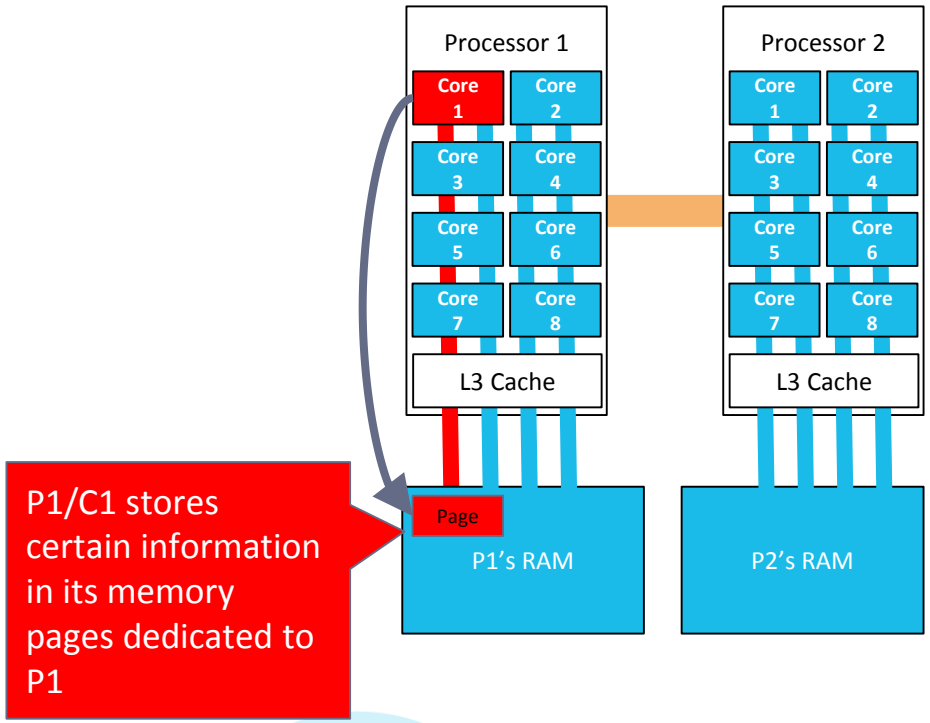
Massive Memory Cache Paradox

Visual Weather



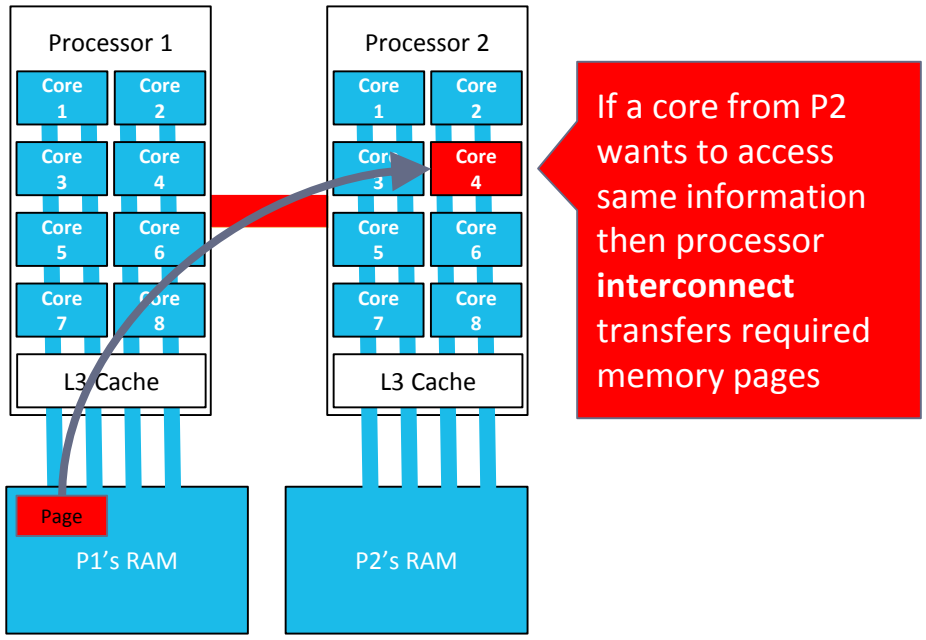
Massive Memory Cache Paradox

Visual Weather



Massive Memory Cache Paradox

Visual Weather



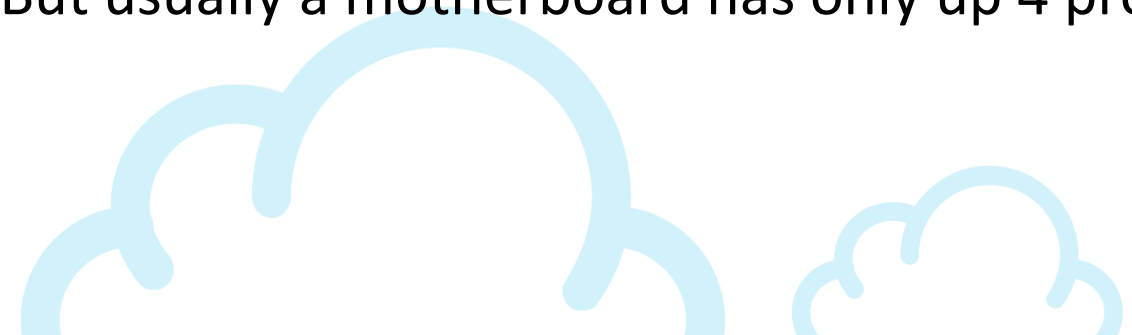
NUMA is only as fast as the interconnect.

Massive Memory Cache Paradox

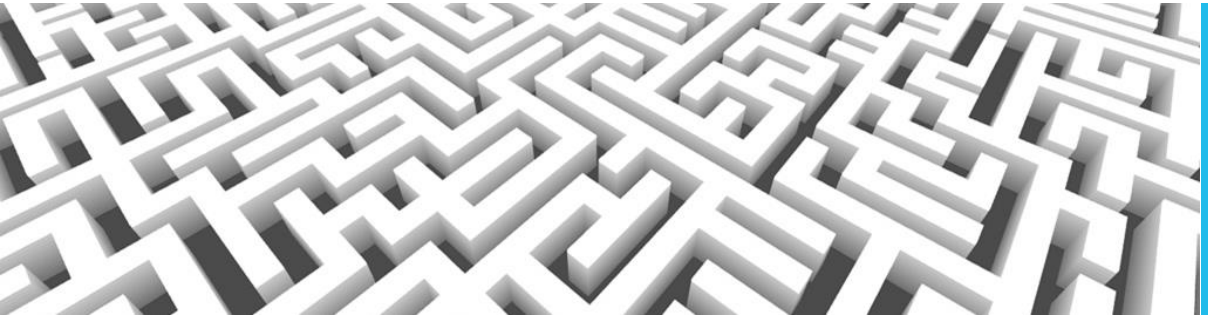
Visual Weather

NUMA systems are optimised for multiprocessing but not for multithreading where memory can be unexpectedly shared between processors.

- Typical processor has only up to 4 memory channels shared between all its cores. Having lot of core does not necessarily help with big data set.
- Processor NUMA interconnects on motherboard is much faster than interconnect between individual computers which form NUMA system.
  But usually a motherboard has only up 4 processor sockets.



ibl



Visual Weather

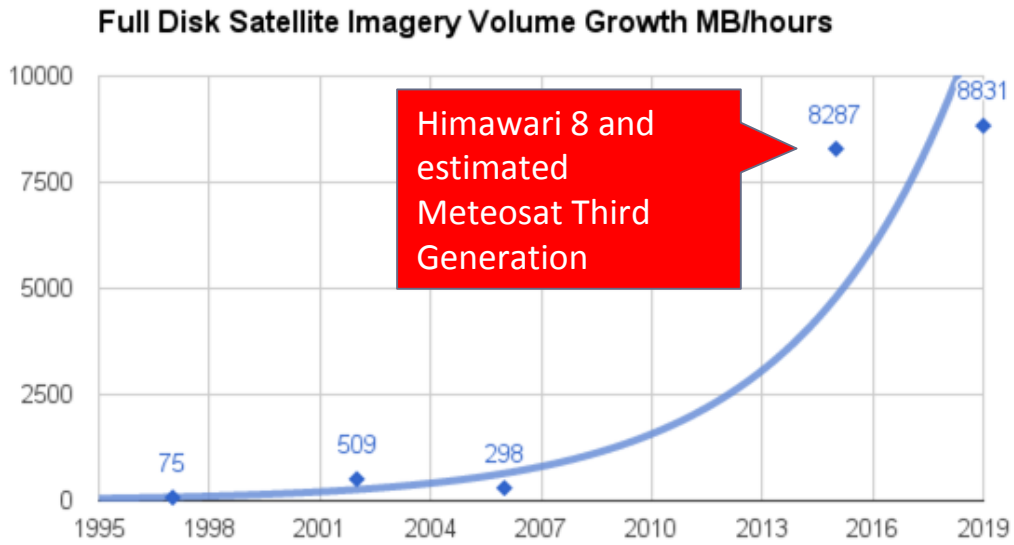
Future and Forecast

ibl

Meteorological Data Volume Growth

Visual Weather

- 2 biggest contributors chasing each other:
  - NWP volume ~30 times bigger in last 10 years
  - Satellite Imagery ~30 times bigger in last 10 years
- Interesting is the growth of a single product size especially with satellite imagery exceeding 500MB per channel every 10 minutes.



Data transfer time becomes the limiting factor.

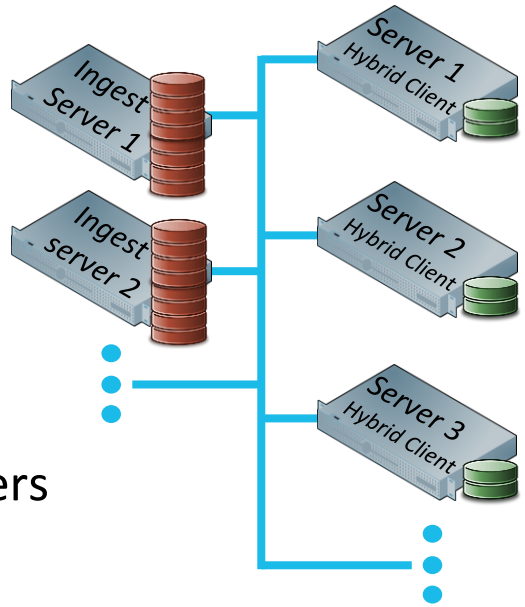
ibl

Impact on Visual Weather Architecture

Visual Weather

Horizontal scalability bottleneck is the database

- Single server will not be able to ingest all data anymore because speed of disk does not increase over time that much.
- Only a percentage of all ingested data is actually processed.
- *Hybrid Client* - a client system which:
  - Unifies data from multiple ingest servers
  - Provides all database services as ingest servers
  - Caches data which have been accessed
  - Preloads frequently used data
  - Does not suffer on low latency networks (WAN)



ibl

Forecast?

Visual Weather

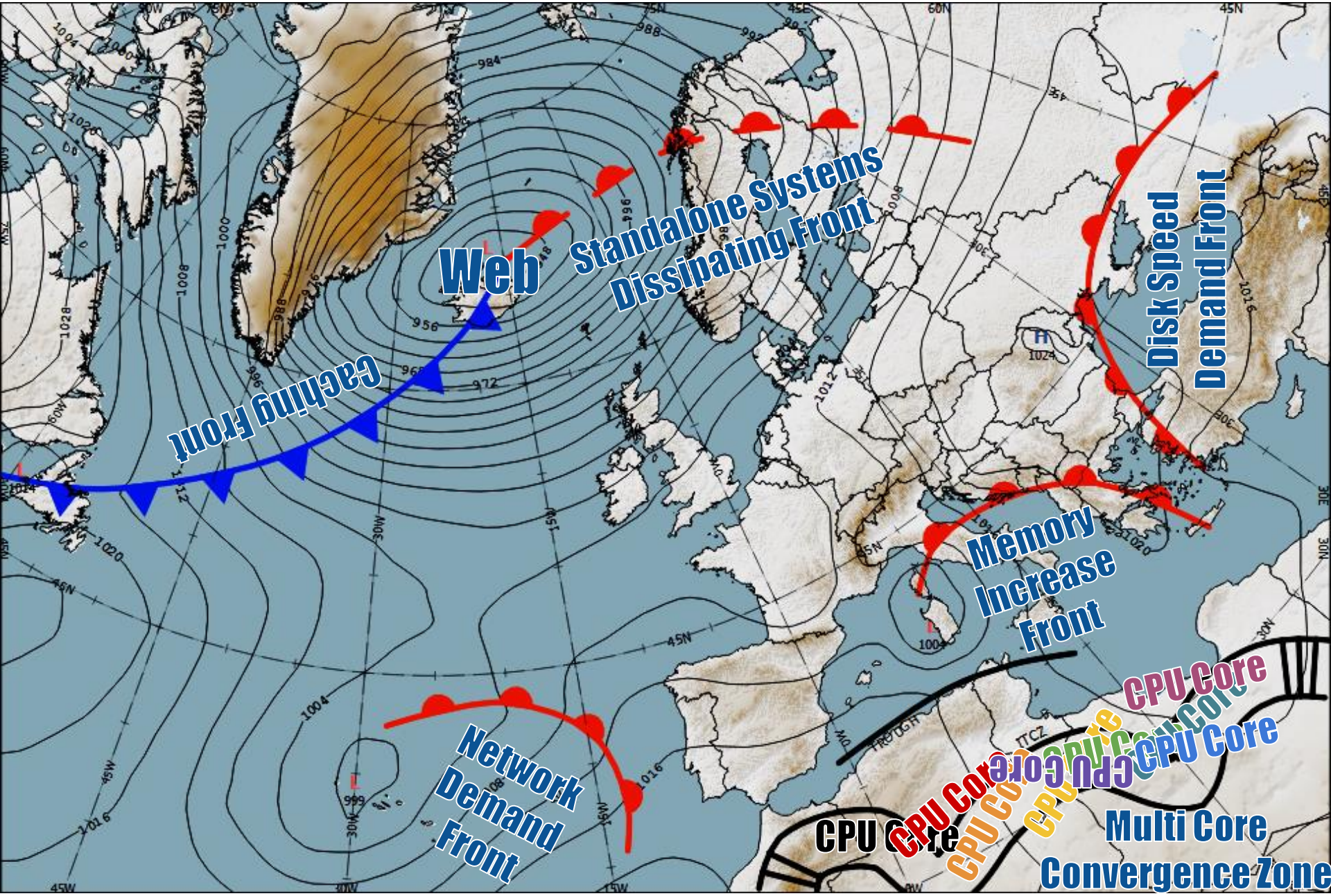
Forecasting visualisation will slowly transition to web

CPU cores number will grow, but only redesign of of applications to properly support this paradigm will allow their full utilisation.

Memory amount will become important, but more important will be network bandwidth needed to populate the memory with actual data.

Disks speed will increase latency of uncached data access because of growth in data feed volumes. Also the operational database will be connected to several servers.

ibl


Web
Standalone Systems
Dissipating Front
Caching Front
Disk Speed
Demand Front
Memory
Increase
Front
Network
Demand
Front
CPU Core
CPU Core
CPU Core
CPU Core
CPU Core
CPU Core
CPU Core
Multi Core
Convergence Zone

Visual Weather

Questions?

Jozef.Matula@iblsoft.com • www.iblsoft.com

ibl