Scaling data access in Visual Weather

Ján Valky
Innovation Department

Jozef Matula
CTO

16th Workshop on Meteorological Operational Systems
ECMWF Reading, United Kingdom, 1st–3rd March 2017
Introduction - Traditional data handling

Processing is close to data source
Why people love the Cloud?

Scalable processing power (hardware)

Better connectivity to Internet
Introduction - “Cloud” data handling

Supercomputer

NWP Model

How to transfer data in reasonable time?

Network

Datacenter

End user

WCS/WPS

How to transfer something what can be reused?
Real world physics

Typically constant

\[ t = \frac{d}{v} \]

Costs money
Digital data transfer
“physics”

\[ t \sim \frac{\text{size}}{\text{bandwidth}} \]

Costs money
Limited resource
We have more grid points than pixels on screen (NWP, satellite imagery)
Access degraded resolution
Access degraded resolution
Access degraded resolution

Network

Using smaller size resolution dataset
Next step - Tiling
Next step - Tiling
Tiling - Small area, middle resolution

Data tile can be cached by the client
Tiling - Bigger area, full resolution

Network

4 cacheable tiles
Different Data Access Intents
Single point query (server side interpolation)
Physical distance becomes important variable in the equation.

\[ t \sim \frac{\text{size}}{\text{bandwidth}} \]
Single point query (client side interpolation)
Time and vertical profiles

Would require transferring 100s-1000s of tiles
Point vertical query (client side interpolation)
Point vertical query (client side interpolation)

All vertical levels for a parameter
Point vertical query (client side interpolation)
Time series query (client side interpolation)

All forecast steps for a parameter
Time series query (client side interpolation)
• Understand what client does and what he may do next.

• Make data subsets cache-able.

• Expressed more technically:
  – Split source dataset into “tiles” on server side (considering resolution required by the user, make it transferrable, minimise request-response latencies)
  – Ensure data “tiles” or “tile sets” are cacheable and reusable for the client.
  – Minimise number of requests by building specialised “tile sets” for different access intents (geospatial maps vs. vertical profiles vs. time series etc.)
Summary - Some implications

• Built-in into Visual Weather’s WCS 2.x implementation, very “raw” extension.

• Correlation with OGC WC-Tile-S activities.

• Tiling has 2 important technical implications:
  – Data being processed is collocated in memory.
  – Allows better paralellisation in modern CPUs.
Questions?

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