Data-centric network management, or network-centric data management?

Stratis D. Viglas
University of Edinburgh
sviglas@inf.ed.ac.uk

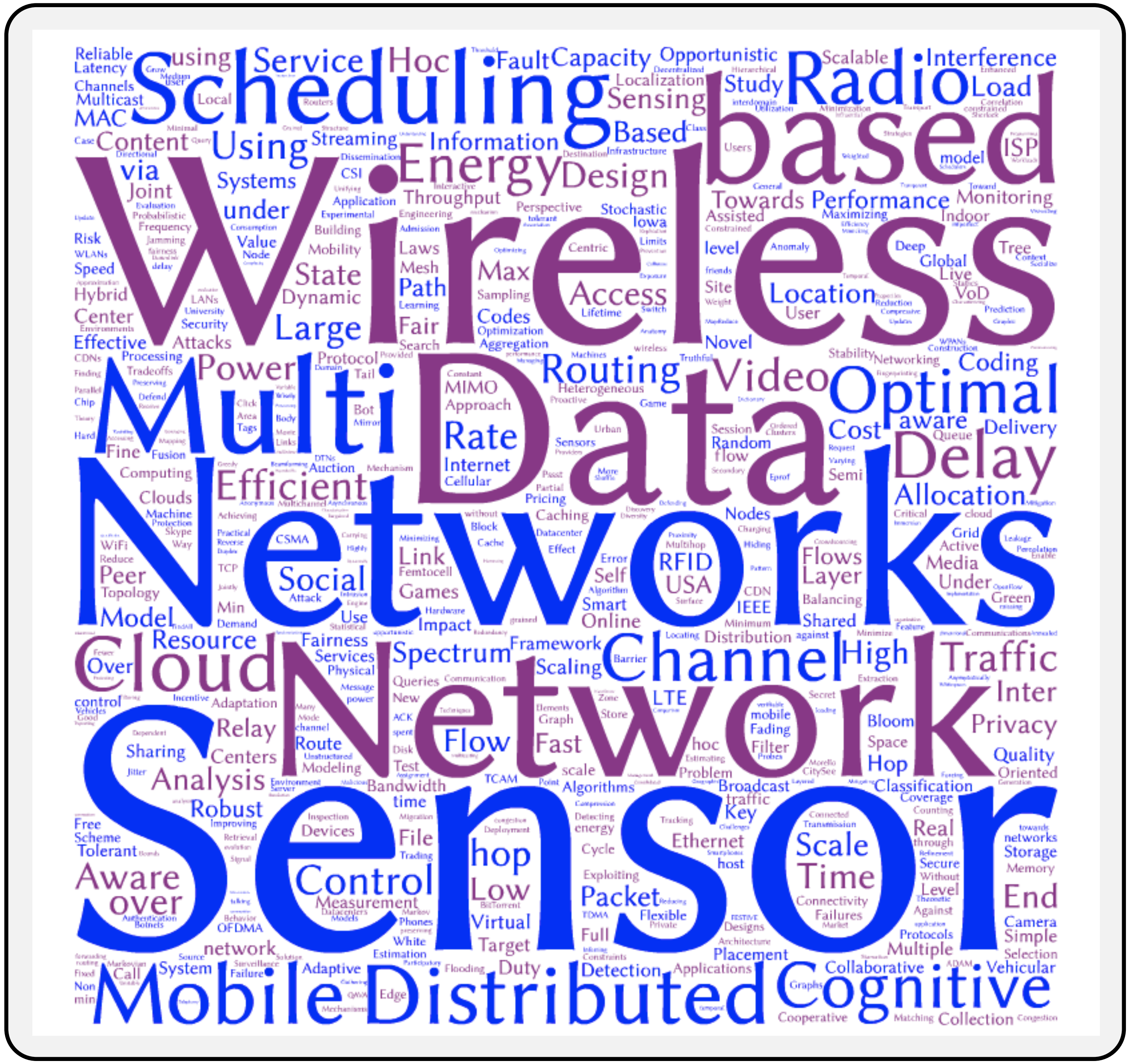
Setting (or rather, lowering) expectations

- What I will not do
    - Announce a project
    - Talk about a new grand research idea
    - Present fantastic recent results
- What I will attempt to do
    - Gauge the overlap between the two areas
    - Try to understand why we don’t talk to one another
- What I will most likely do
    - Showcase my ignorance on all things networking

But why?

- A working hypothesis
  - Networking people can help solve data management problems, and data management people can help solve networking problems
  - Solutions we can find in isolation may not be as optimal
  - In fact, solutions found in isolation may be in competition
- Some anecdotal evidence
  - This one time I tried to apply networking solutions to database problems, and I got a PhD in the end
  - I can understand your papers (some more than I can understand database ones!)

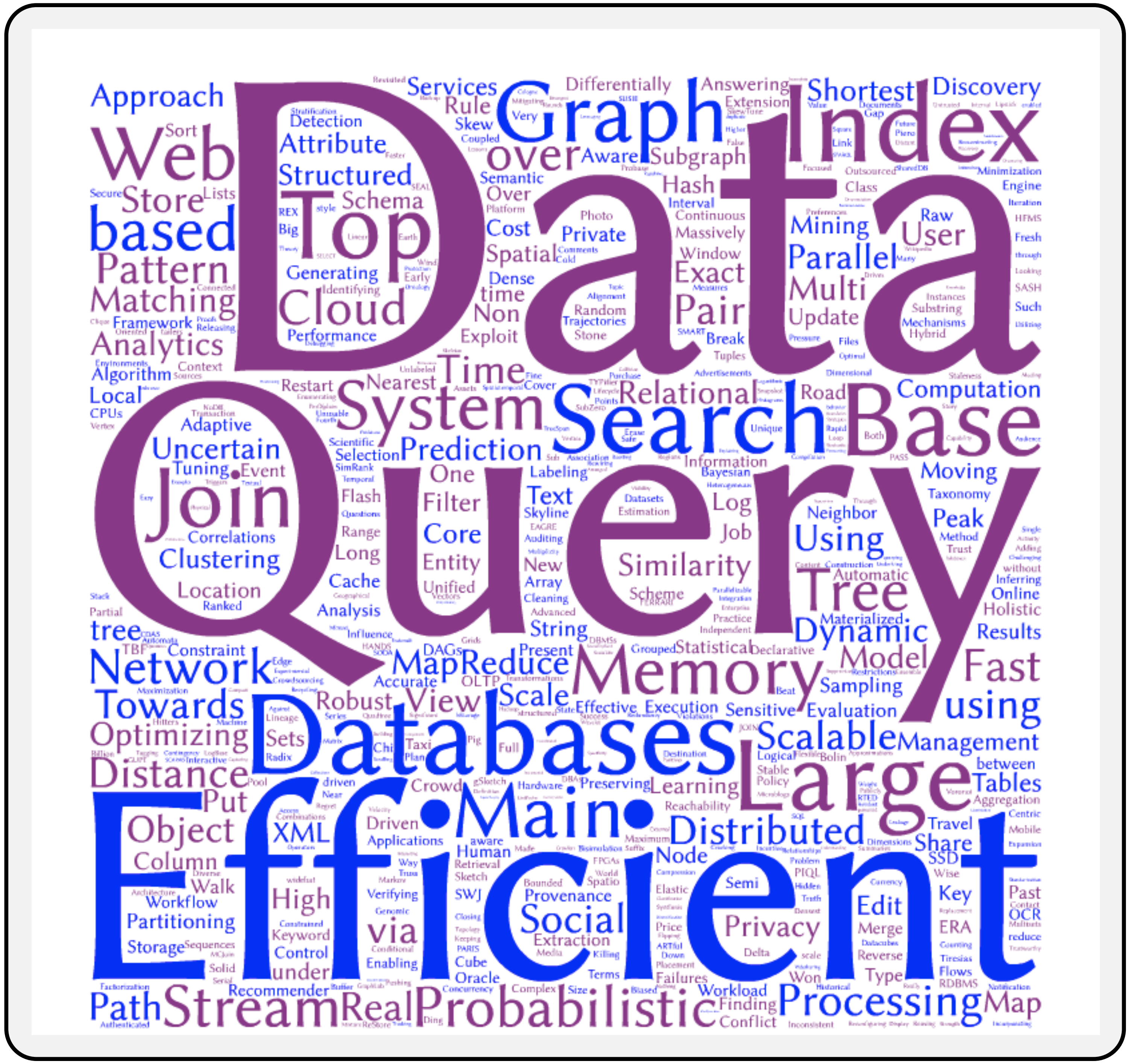
Networking tag cloud: SIGCOMM, Infocom, NSDI, CoNEXT



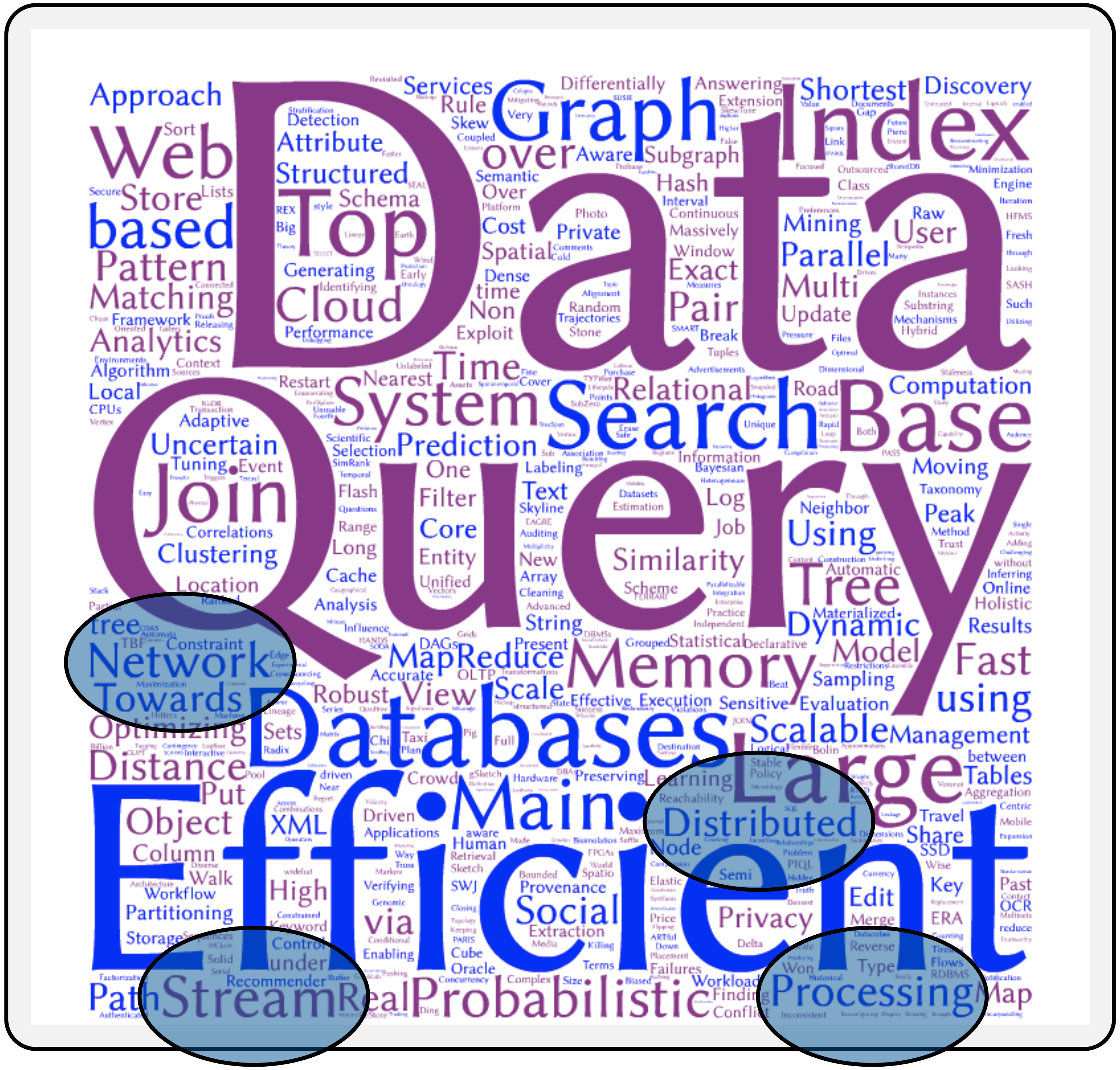
Networking tag cloud: SIGCOMM, Infocom, NSDI, CoNEXT

Wireless
Networks
Sensor
Data
Multi
Network
Scheduling
based
Radio
Channel
Distributed
Mobile
Cognitive
Optimal
Cloud
Energy
Efficient
Delay
Video
Routing
Access
Traffic
Control
Location
Allocation
Social
Spectrum
Power
Rate
Design
Analysis
Scale
Time
Flow
Service
Hoc
Capacity
Interference
Opportunistic
Load
Sensing
Streaming
Information
Performance
Monitoring
Large
State
Max
Path
Codes
Flows
Layer
RFID
USA
Link
Resource
Relay
Robust
Aware
over
Packet
hop
Low
Fast
Privacy
High
End
Storage
Real
Multiple
Collaborative
Vehicular
Duty
Detection
Applications
Content
Reliable
Latency
Channels
Multicast
MAC
Using
via
Joint
under
Systems
Peer
Model
Topology
Sharing
Centers
Free
Scheme
Tolerant
Ethernet
Cost
aware
Delivery
Coding
Random
Online
Smart
Self
Games
Impact
Fairness
Framework
Scaling
Quality
Classification
Broadcast
Key
Oriented
Inter
Bloom
Hop
Route
Modeling
Bandwidth
Measurement
Virtual
Target
Placement
Cooperative
Collection

Database tag cloud: SIGMOD, VLDB, ICDE



Database tag cloud: SIGMOD, VLDB, ICDE



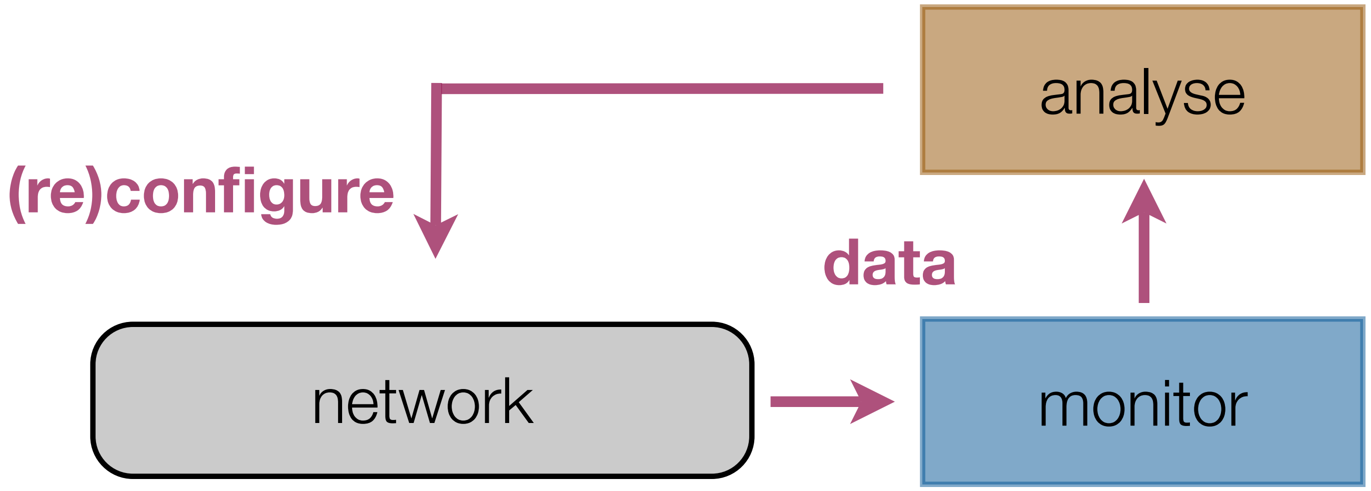
Data-centric network management

- Treated as a configuration problem
- “Data” in this case is well-defined
  - Measurements, performance indications, overall configuration of the network
- The key idea is to use the values in the data to configure and manage the network
  - For instance, identify bursts in traffic and allocate more resources
  - Identify collisions in packets and reallocate channels
- Reactive and proactive approaches
  - Either detect and fix, or estimate and evade

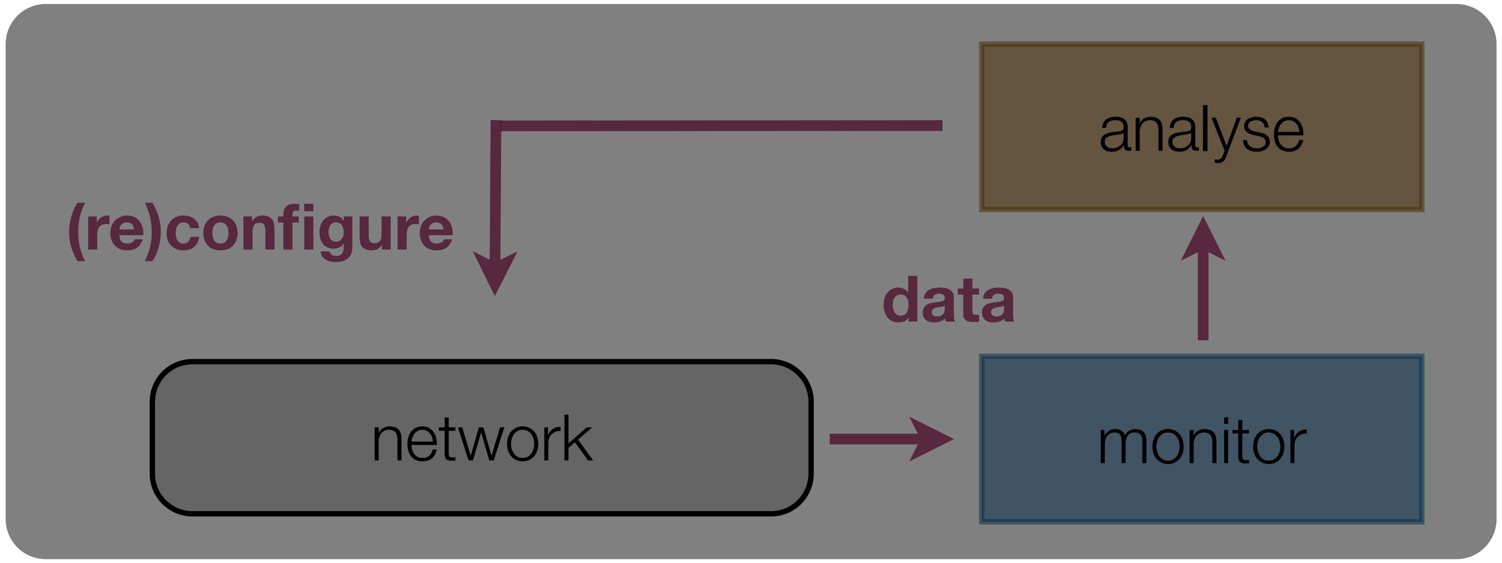
Network-centric data management

- A different configuration problem
- “Network” is now well-defined; but data can be reconfigured
  - Not data per se, but their management
  - Monitor the network and reconfigure data processing
- The key idea is to use not network-level measurements, but application-level measurement to detect problems
  - Once a problem is detected, change the data-flow
  - In some cases, depending on the semantics of the processing
- Again, reactive and proactive approaches

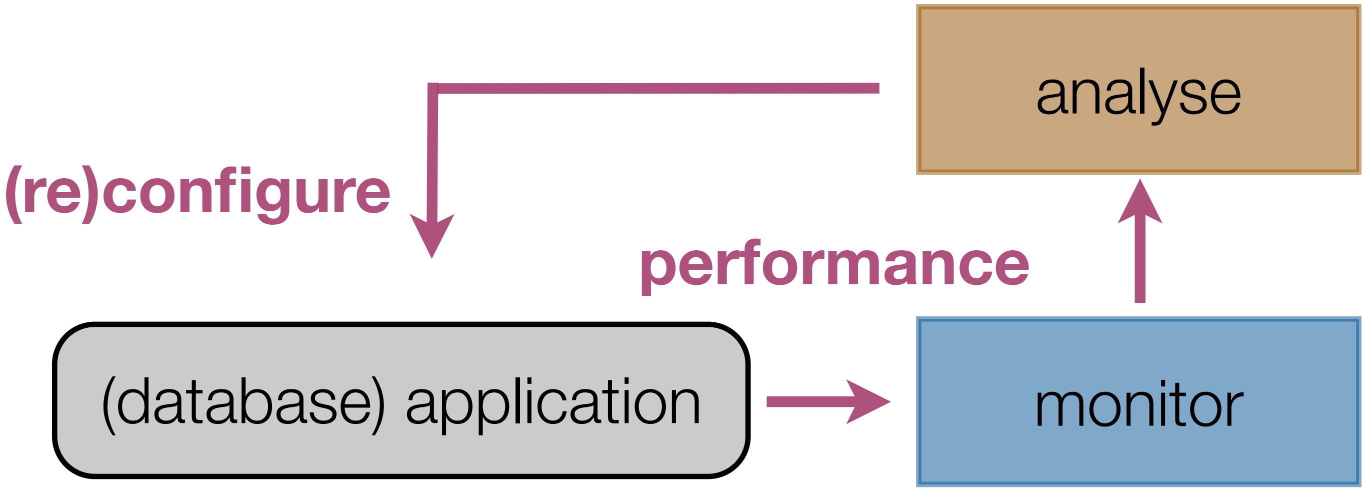
One black box too many?

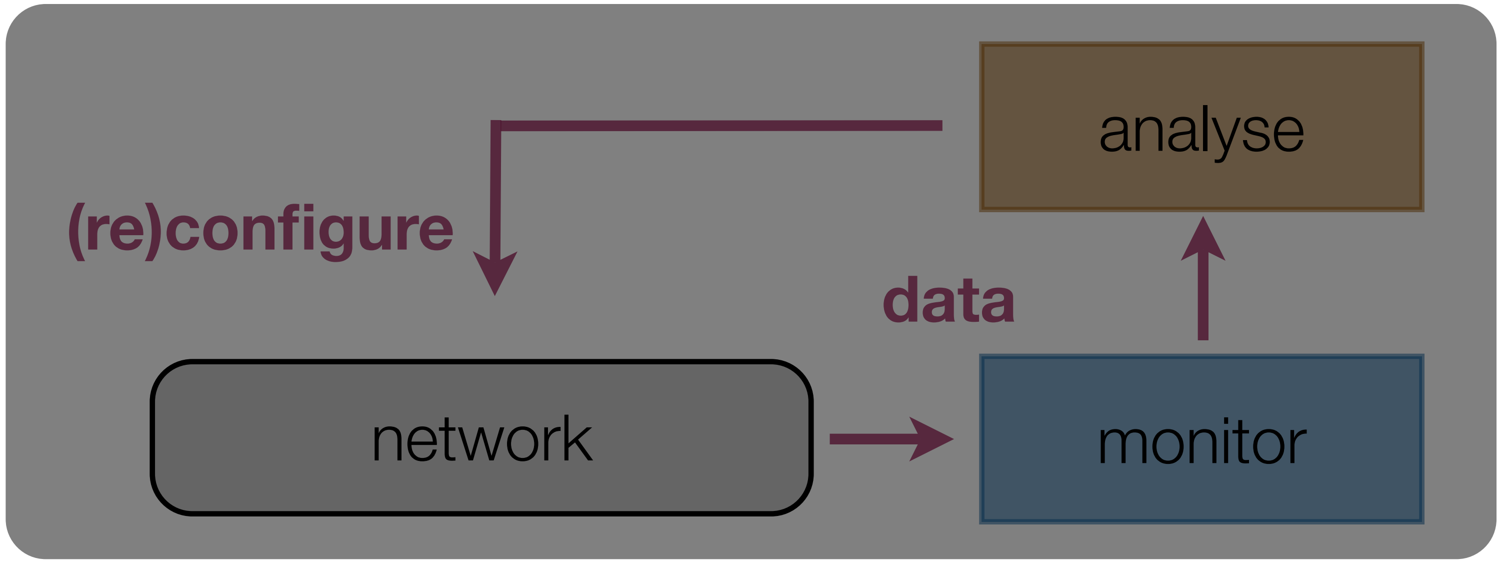


One black box too many?

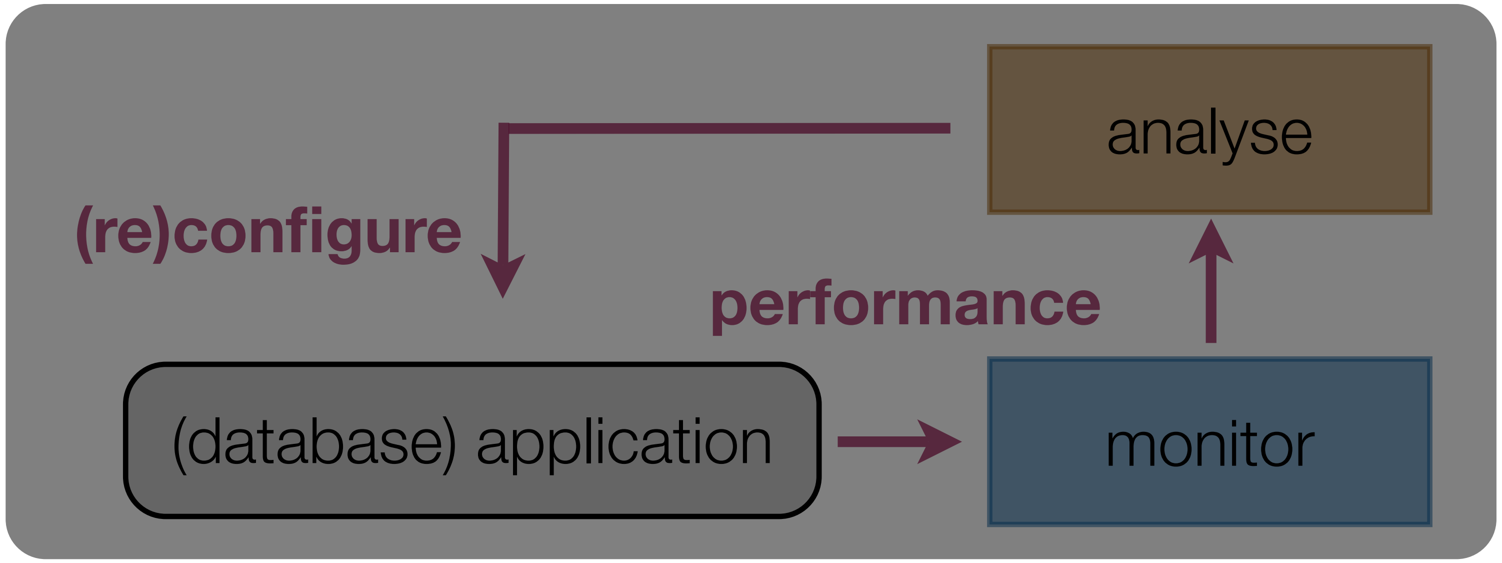


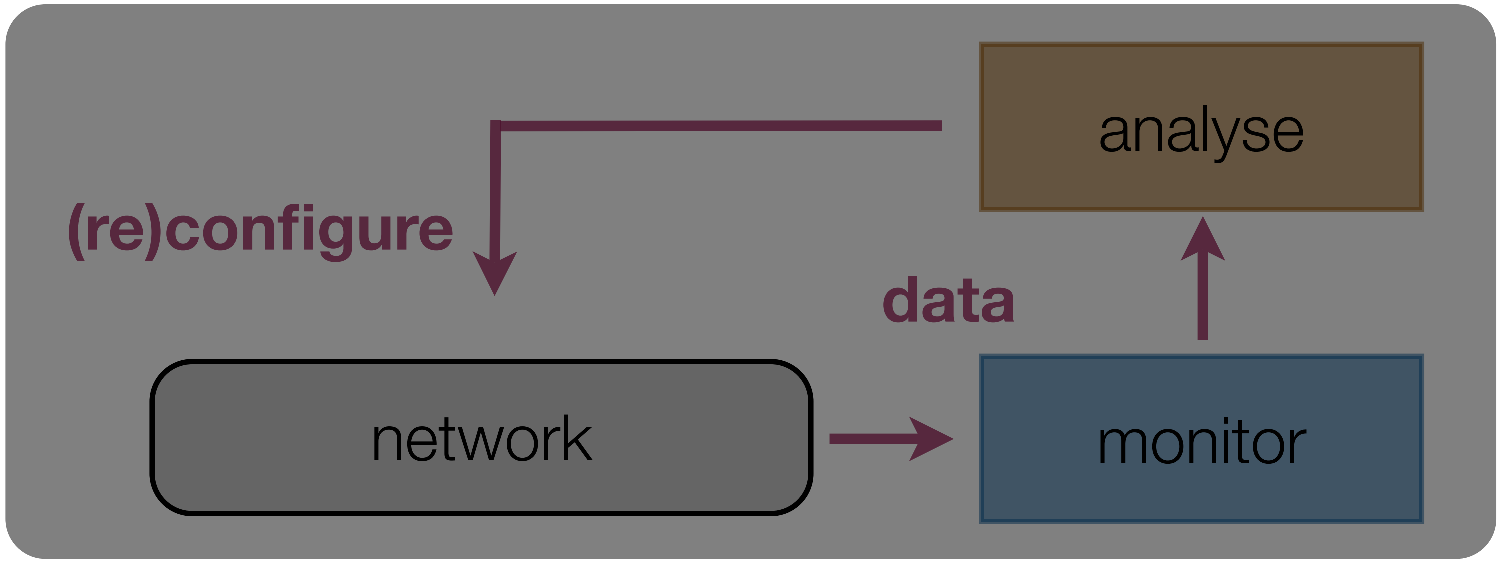
One black box too many?



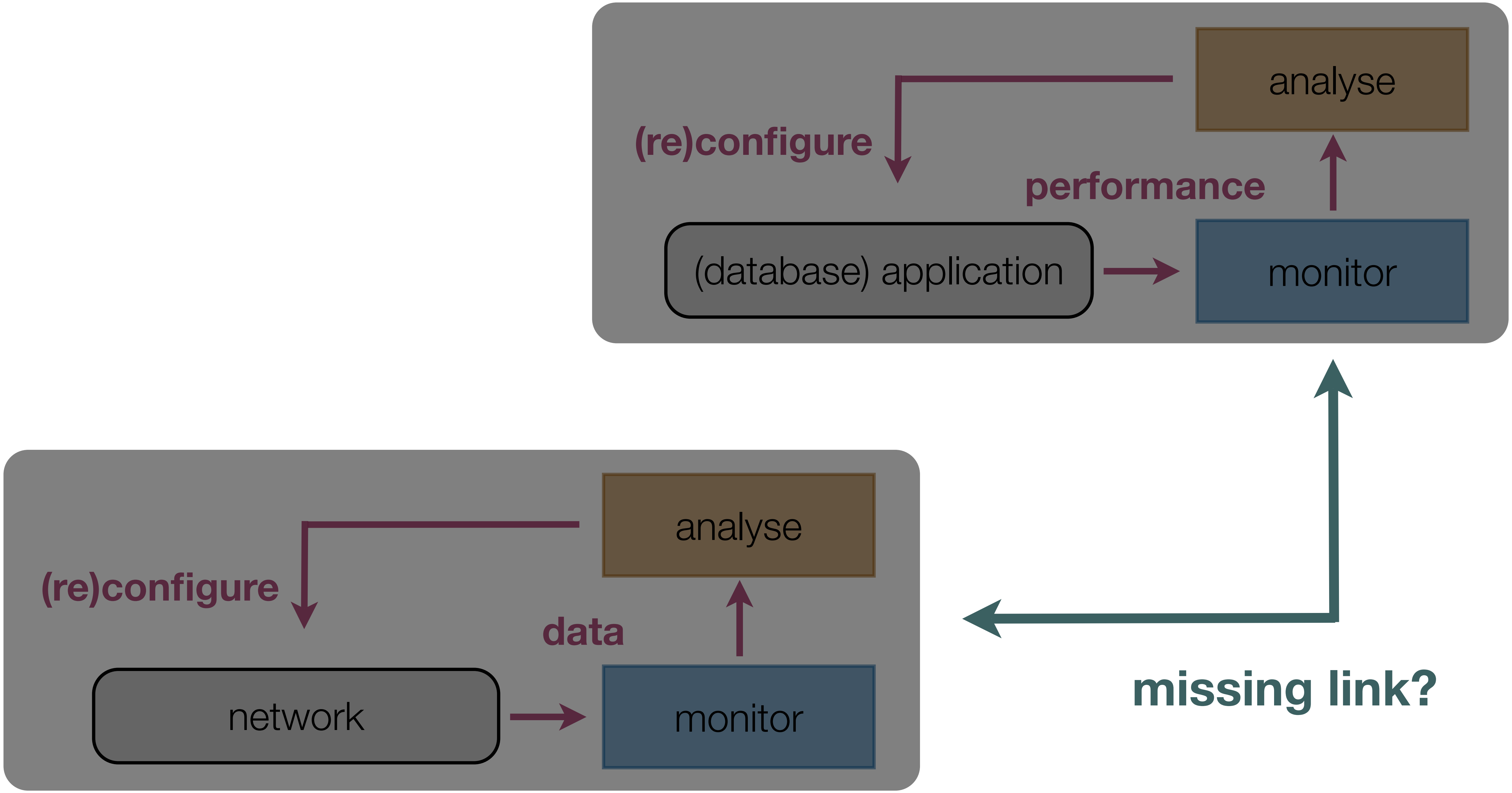


One black box too many?



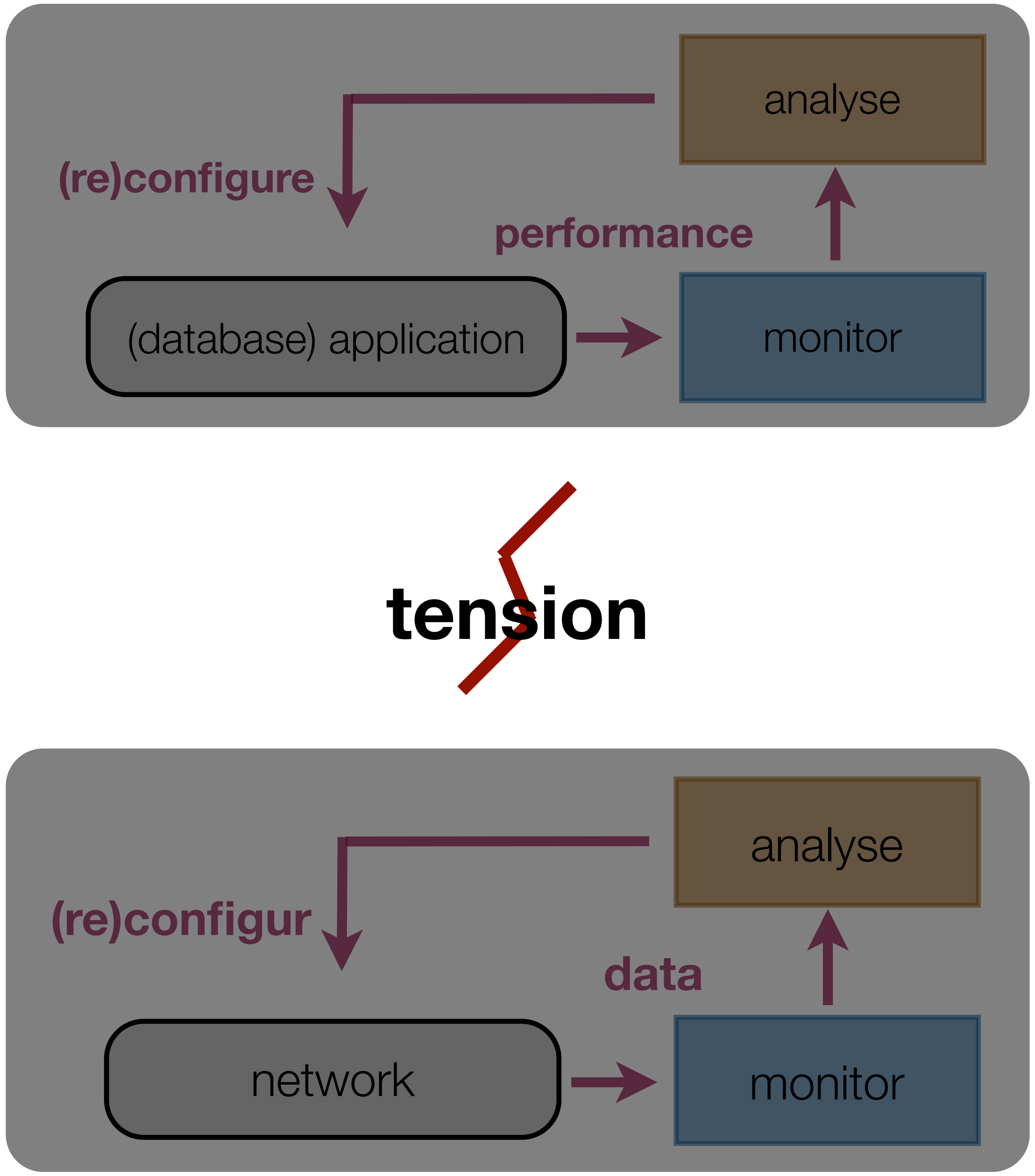


One black box too many?



The potential for tension

- Network configuration is application-agnostic
- Data configuration is network-agnostic
- By being agnostic, one caters for generality
    - But true optimisation is possible only if one customises
    - Data-level optimisation decisions may hinder network-level optimisation potential (and vice-versa)



Example of conflicting goals: data streams

- Networking perspective
  - Maximise throughput of network elements and thus delivery rate
- Data stream management perspective
  - Organise workflow to maximise throughput of output
- Caveat: data stream systems assume they can keep up with the input rate
  - But if the network is too optimised this may well not be true
- What happens then?
  - Load shedding
- If you are too good, you make me look bad!

What if the abstractions are collapsed?

- Using data semantics for network optimisation
  - Network configuration becomes data and operation-aware
- Expose network connectivity to data applications
  - Optimise the dataflow for the current state of the network
- No guess-work
- No stepping on each other’s toes
- Release the tension between the two substrates

Why is this a bad idea?

- Well-defined abstractions and isolation have been working just fine for a few decades
- Abstractions help solve problems
- Black boxes allow us to focus on doing what we know how to do best
  - Why try to fit a square peg in a round hole?
- Incremental improvement
- Independence assumptions are always correct because they are enforced by the abstraction

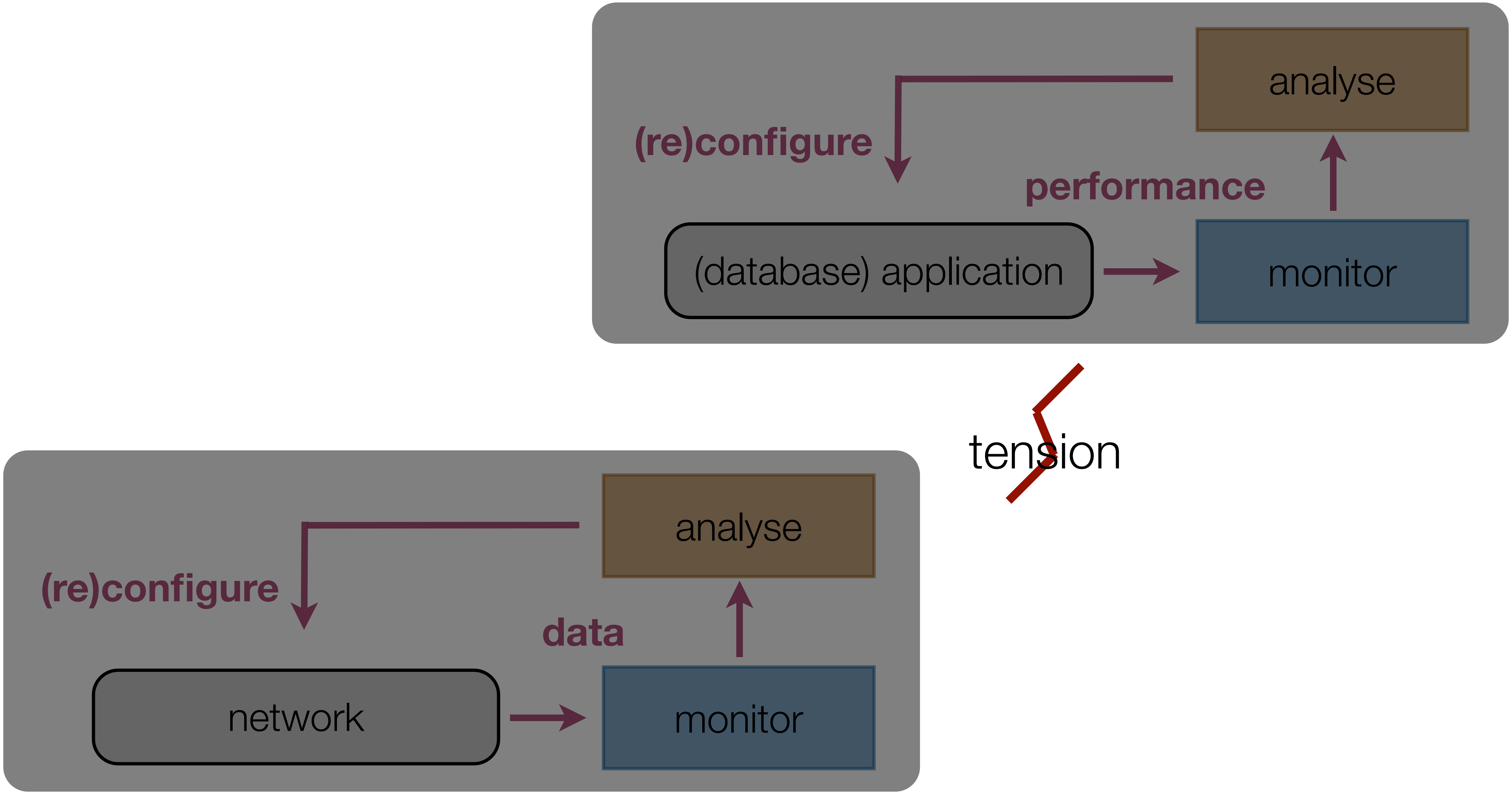
Why is this a good idea?

- New processing paradigms call for radically different workflows
  - Case in point: Google
    - Reinvented pretty much everything and optimised for the core business
  - They could afford to do that
    - Not everyone can, unless the system-level substrates provide them with ways of doing so
- What if the result can be greater than the sum of its parts?

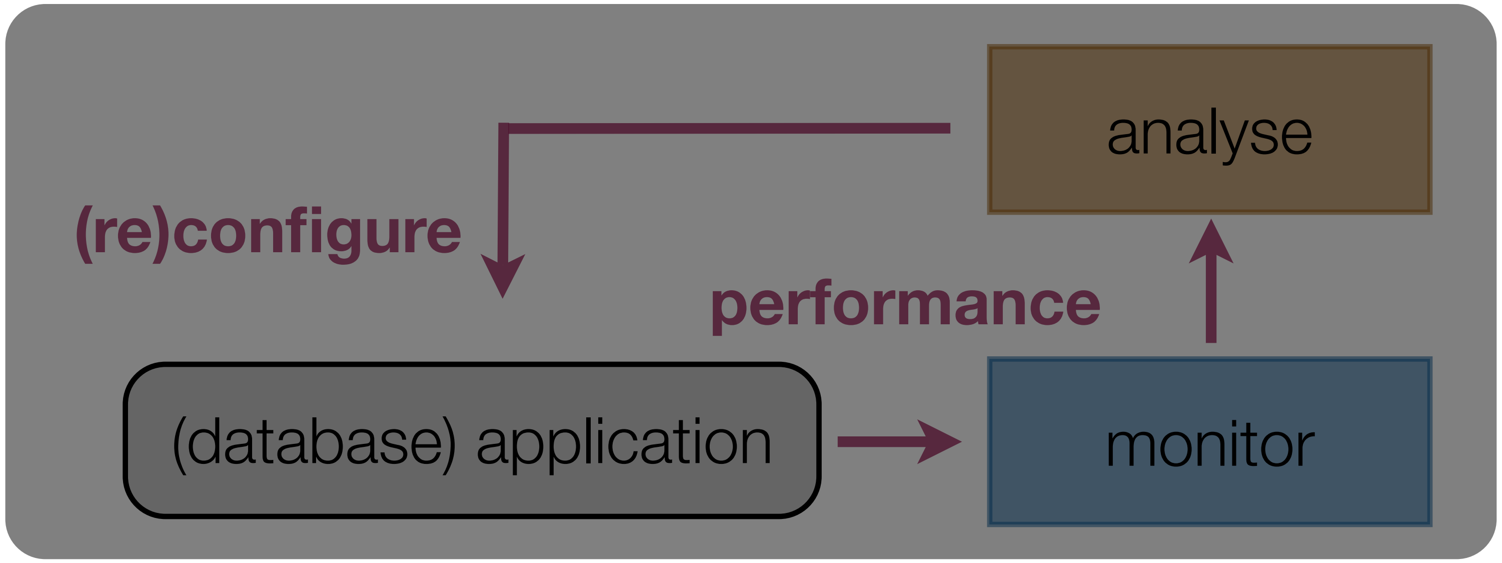
A modest proposal

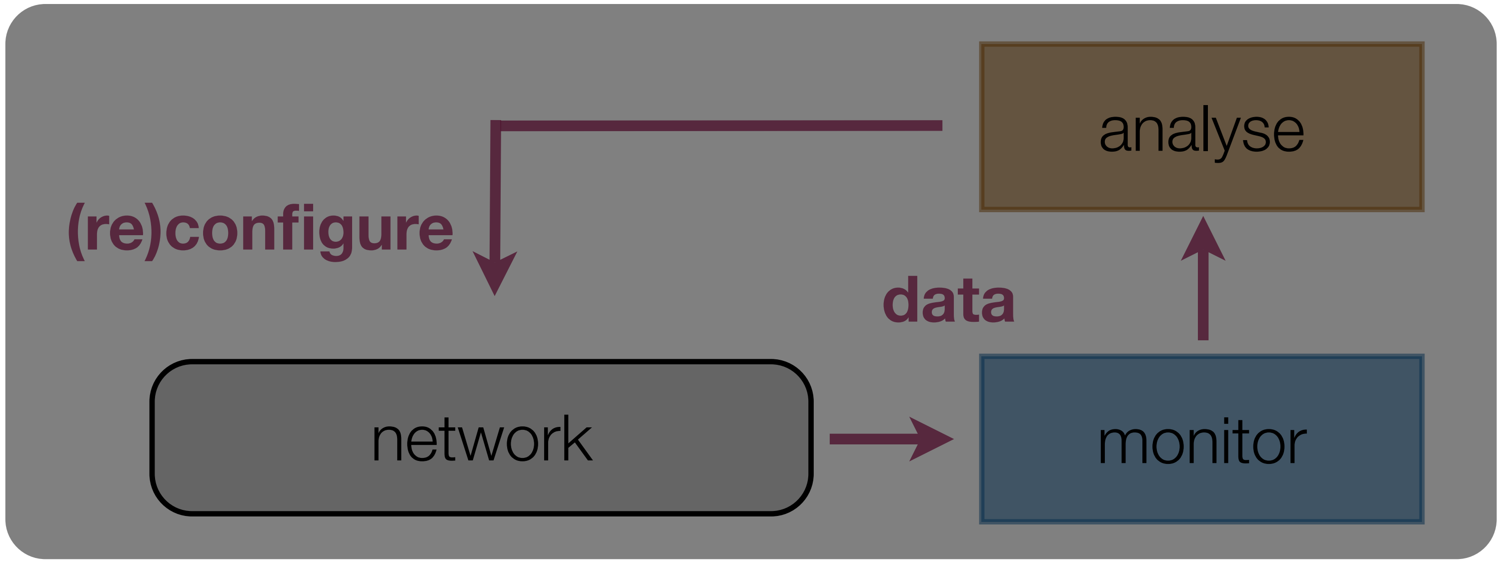
- Application-driven network configuration
- The networking service may need to be a bit more flexible
  - Much like the file system service is flexible
  - Need a richer interface, exposed to the application
- The application should also not take the networking service for granted
  - It can help customise and improve performance
- Application-driven networking implementations?
  - Databases seem a great starting point
  - Well-defined semantics, predictable performance

An all-around solution

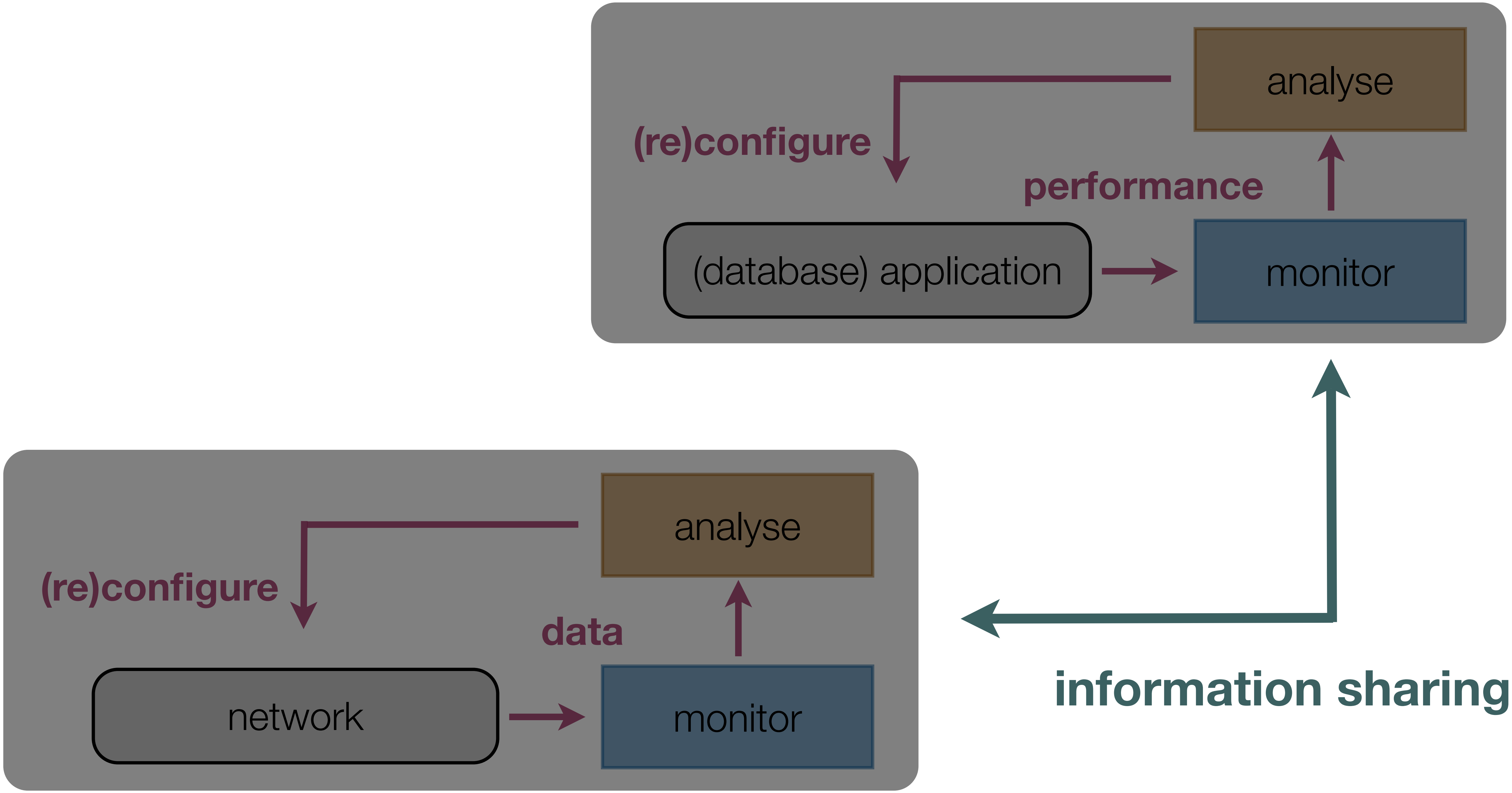


An all-around solution

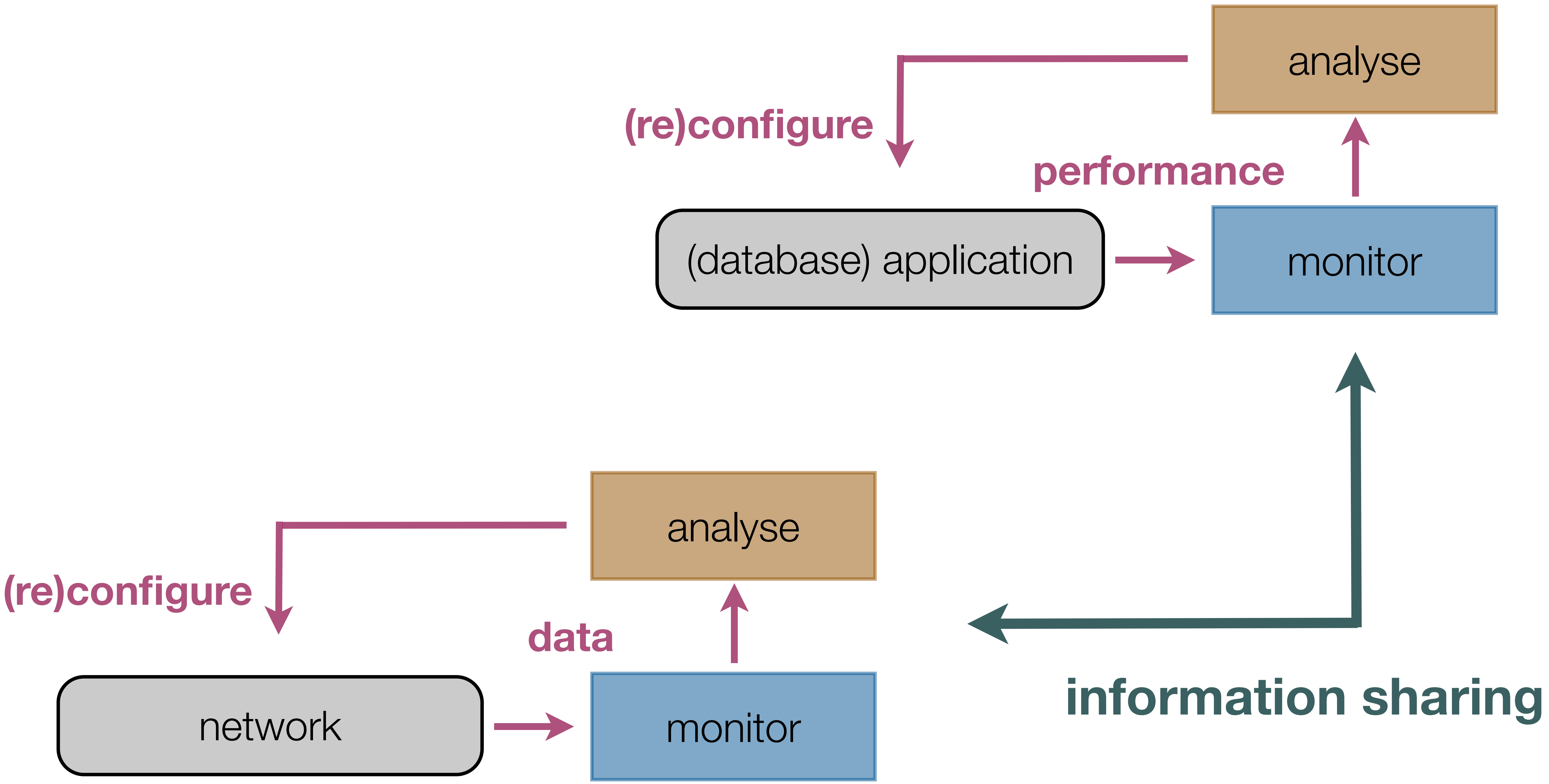




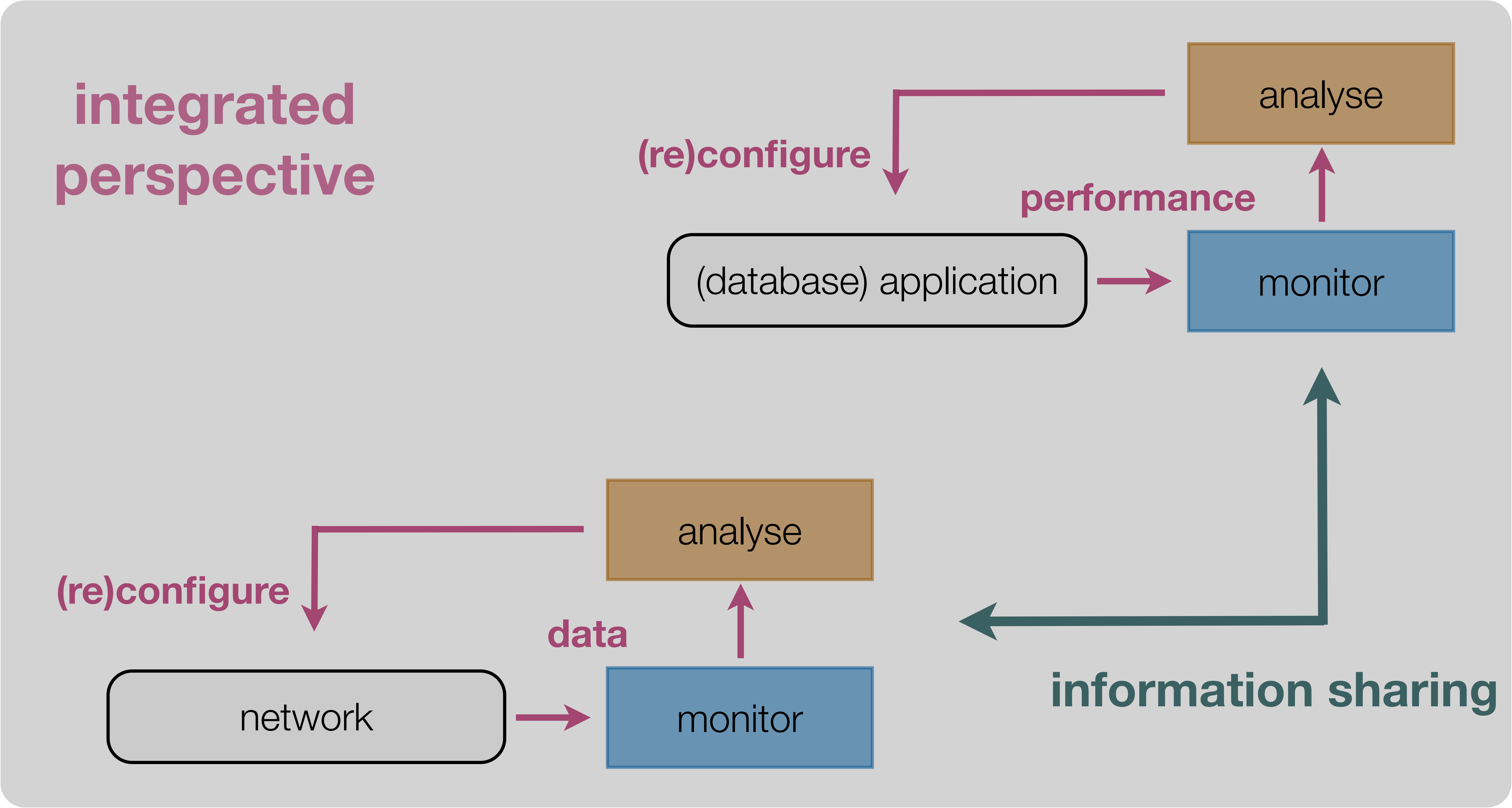
An all-around solution



An all-around solution



An all-around solution



integrated perspective

(re)configure

analyse

performance

(database) application

monitor

(re)configure

analyse

data

network

monitor

information sharing

Good composers borrow, great composers steal[1]

- All problems in computer science can be solved by another level of indirection
  - ...except for the problem of too many levels of indirection

David Wheeler

[1] Igor Stravinsky