Disruptive Architecture – 2016 Edition

OOP Munich

3 February, 2016

Stefan Tilkov, @stilkov

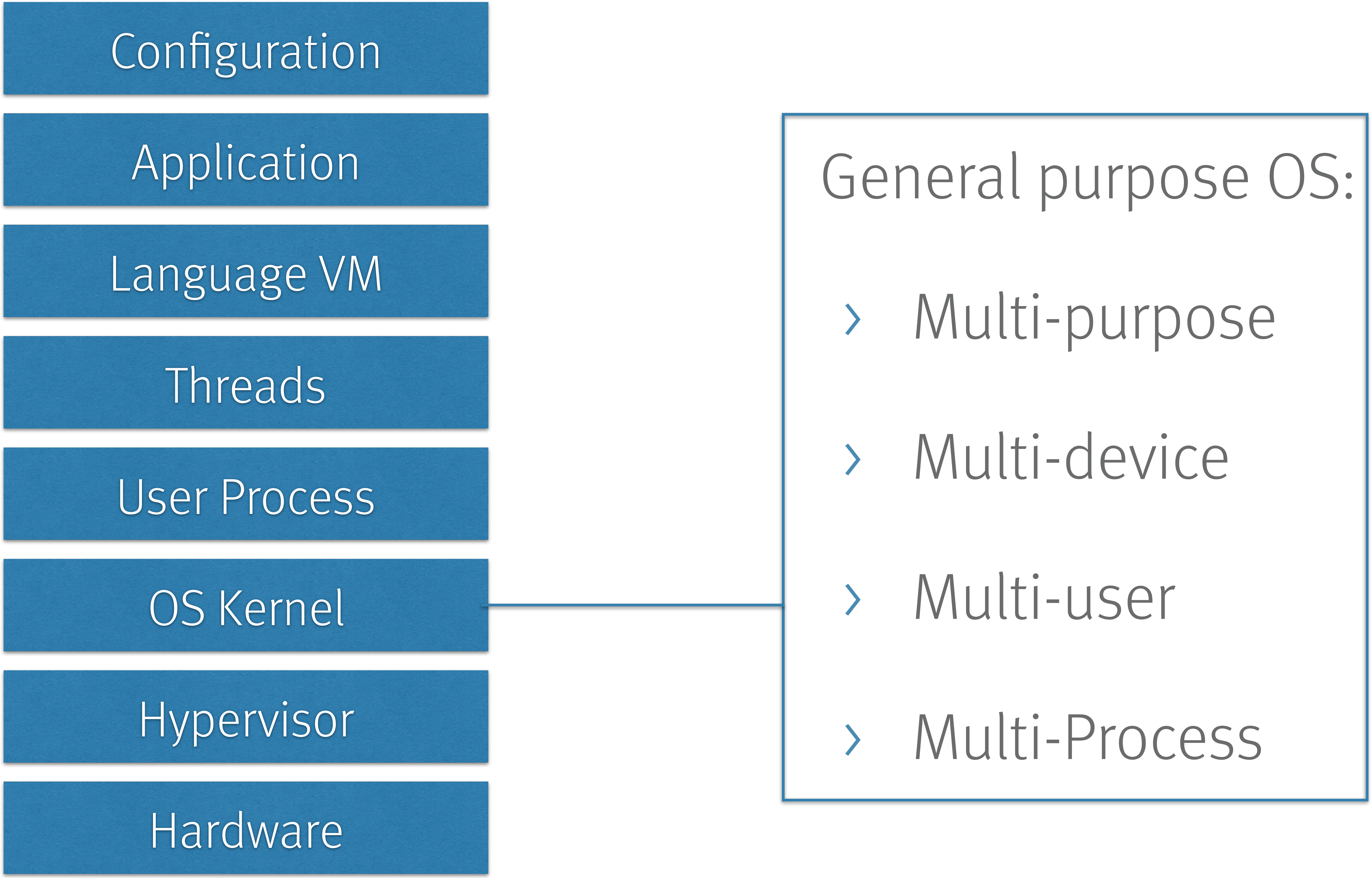
innoQ Deutschland GmbH

innoQ

Things we’ll look at

- Unikernels
- SCM
- In-memory computing
- SDx
- Serverless architecture

Unikernels



Configuration

Application

Language VM

Threads

User Process

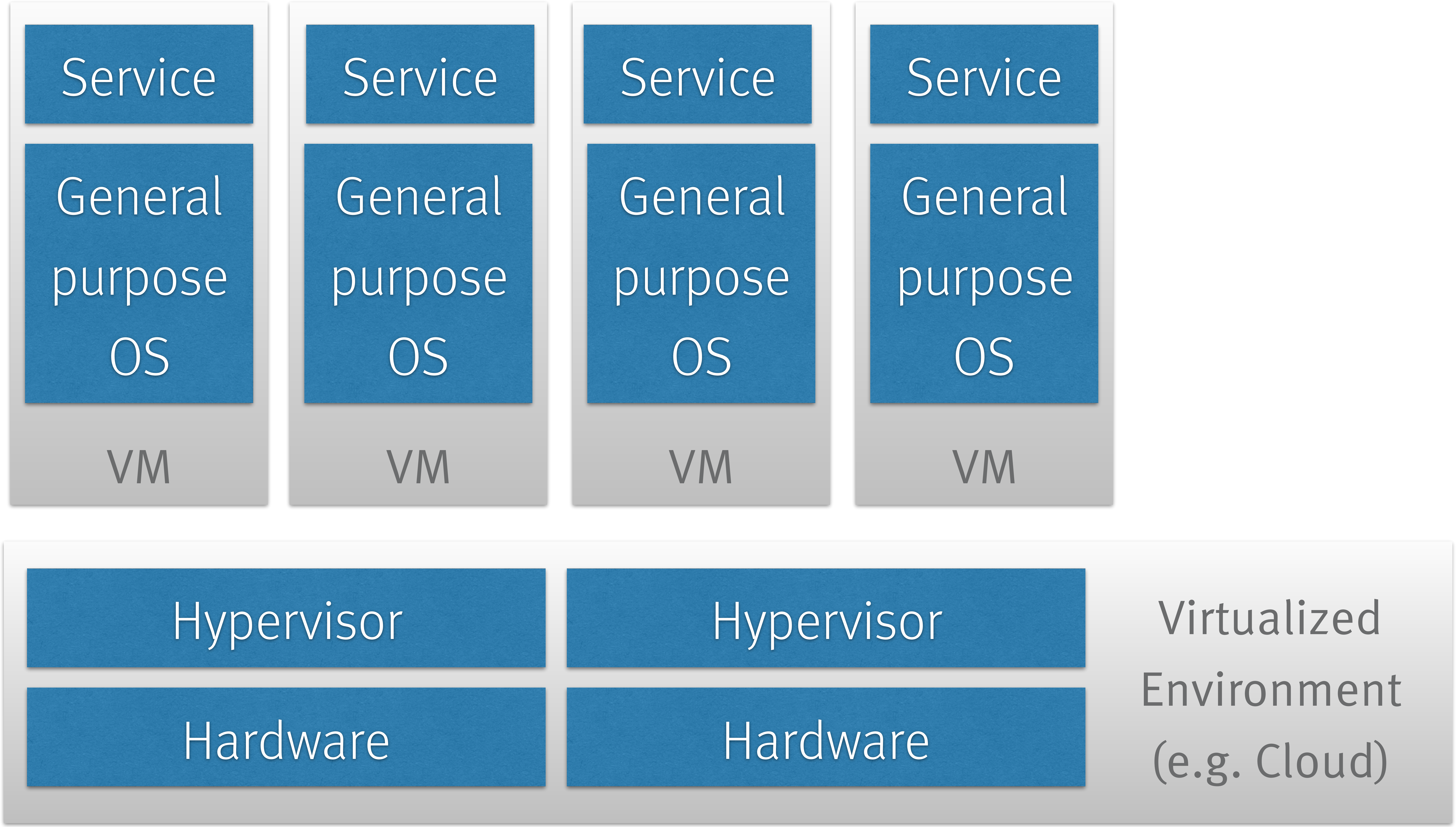
OS Kernel

Hypervisor

Hardware

General purpose OS:

- Multi-purpose
- Multi-device
- Multi-user
- Multi-Process



Service

General purpose OS

VM

Service

General purpose OS

VM

Service

General purpose OS

VM

Service

General purpose OS

VM

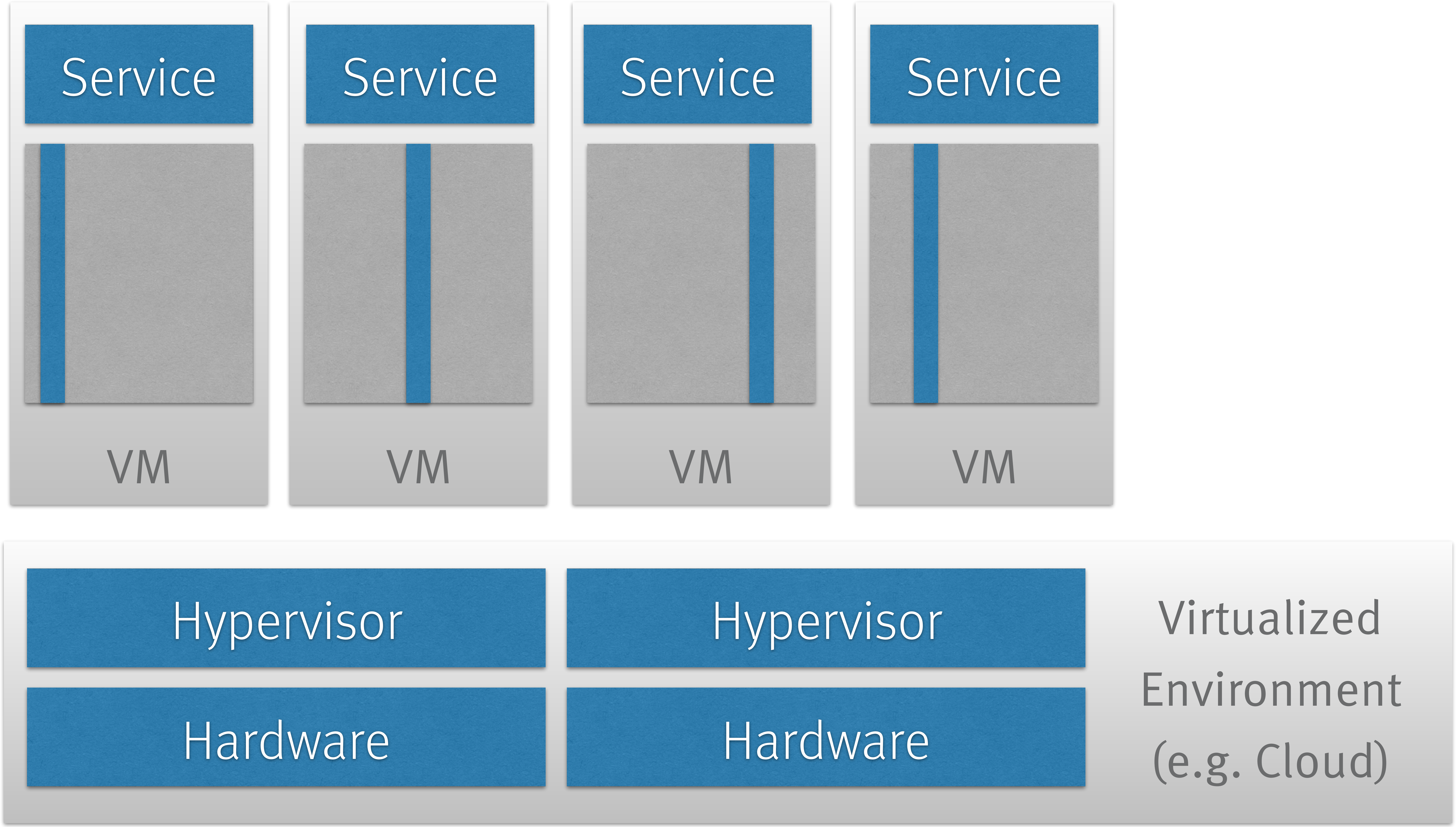
Hypervisor

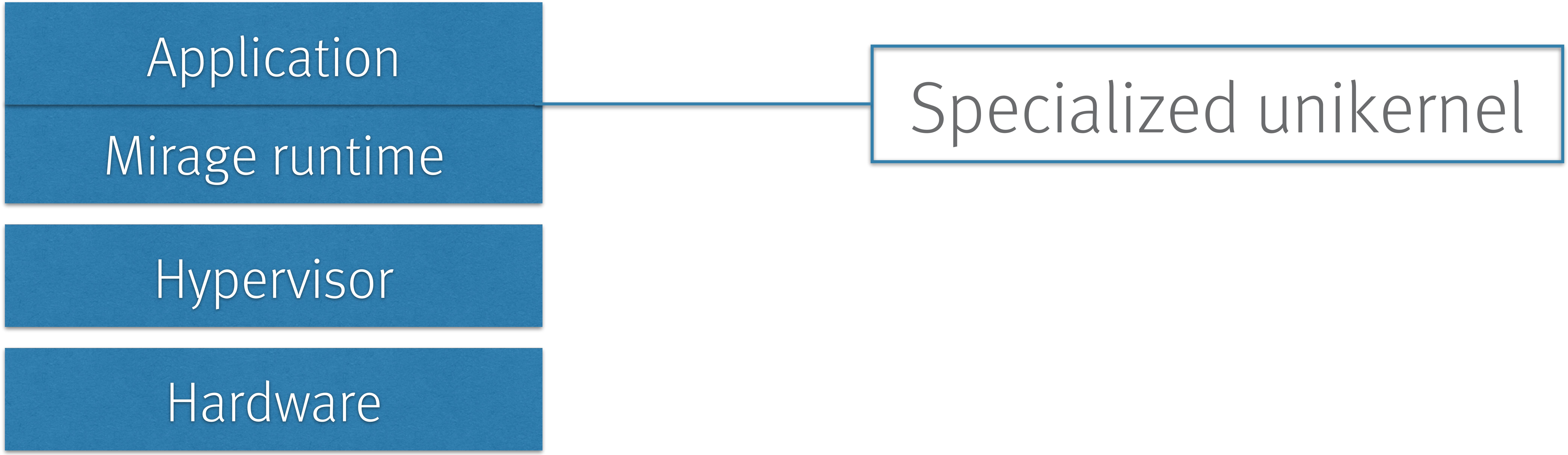
Hardware

Hypervisor

Hardware

Virtualized Environment (e.g. Cloud)


Service
Service
Service
Service
VM
VM
VM
VM
Hypervisor
Hypervisor
Hardware
Hardware
Virtualized
Environment
(e.g. Cloud)



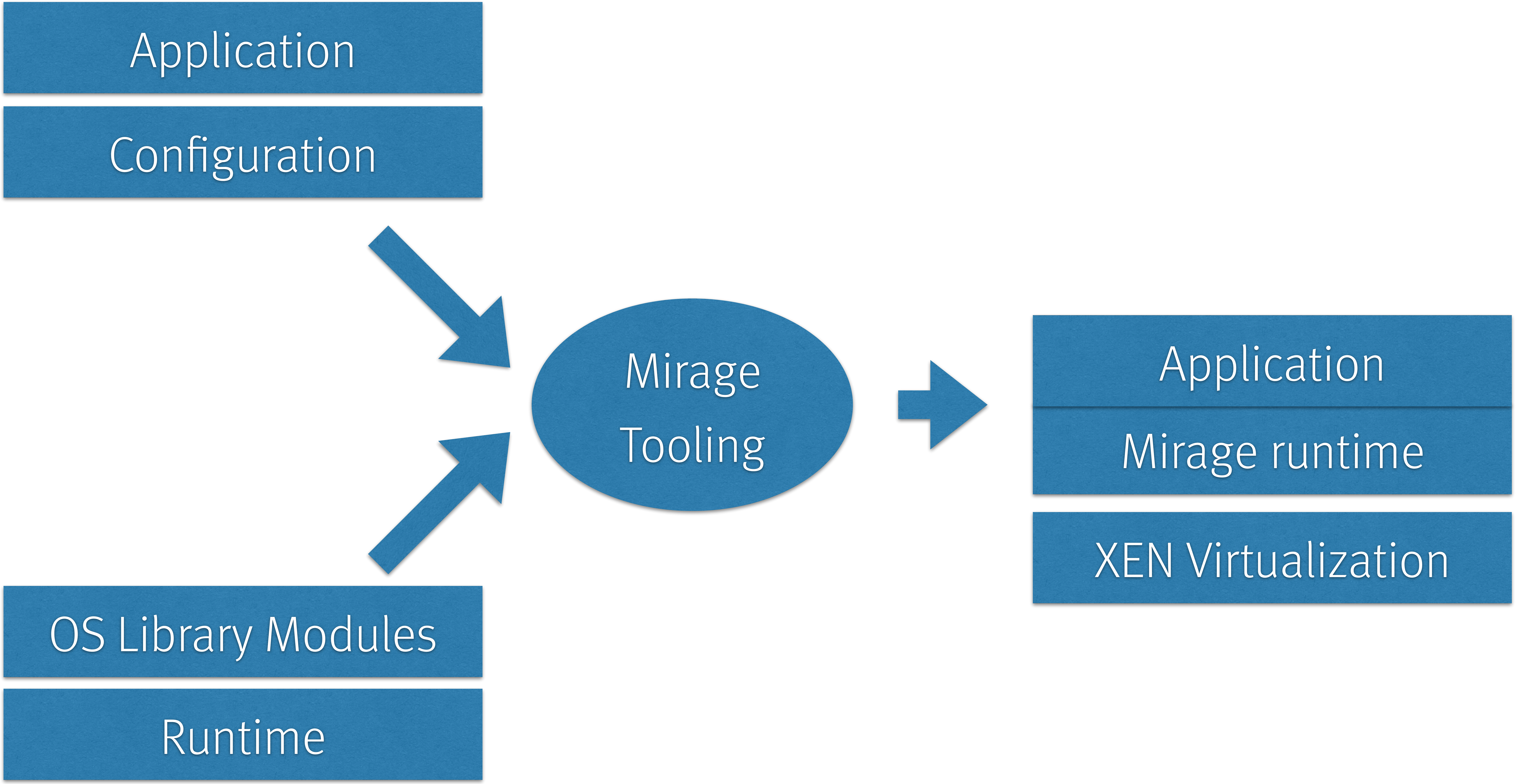
Application

Mirage runtime

Hypervisor

Hardware

Specialized unikernel



Application

Configuration

OS Library Modules

Runtime

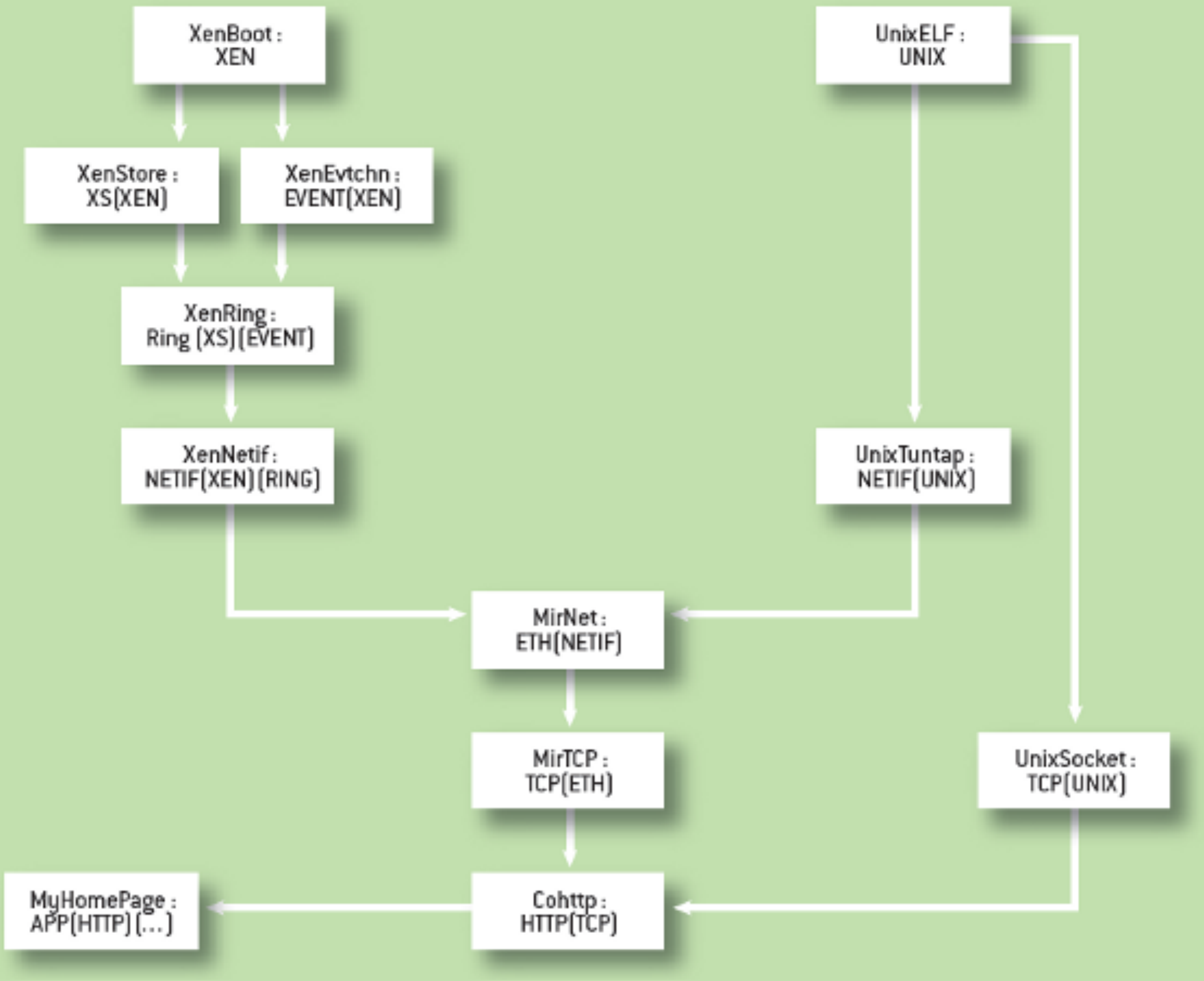
Mirage Tooling

Application

Mirage runtime

XEN Virtualization

Module Graph



OCaml

- Very fast
- Statically typed
- OO & functional
- Single-threaded
- Minimalistic

Source: Unikernels: Rise of the Virtual Library Operating System, Madhavapeddy & Scott, http://queue.acm.org/detail.cfm?id=2566628

Alternatives

HALVM (Haskell)

Erlang on Xen (LING)

Rumprun (general purpose)

Update 21 Jan 2016:

Unikernel Systems (makers of Mirage and Rumprun) aquired by Docker

Storage Class Memory

See: https://queue.acm.org/detail.cfm?id=2874238

CPU + RAM = Fast

Disk = Slow

- Keep CPUs busy while waiting for I/O
- Async/Evented I/O
- Keep working set in RAM (e.g. caching)
- Ensure access to RAM is fast
- Minimize disk access (e.g. dedup, compression)

Enter: SSDs

- Compatibility with existing infrastructure
- SAS or SATA connection, spinning HD form factor
- Significant speed-up
- Some architectural change
  (e.g. network/SSD faster than local HD)

Enter SCM

Storage Class Memory (SCM)

- Flash memory
- PCIe instead of SAS/SATA
- 25x price increase over HDs
- 1000 times faster than spinning disks
- 100000 IO operations/second
- Storage 1 million times faster, network 1000 times

CPU + RAM = Fast

Barely fast enough

Disk = Slow

Fast

- 10 microseconds to process one I/O request
- Less if network involved
- Entirely new problem: Saturating “disk”
- RAM needed for buffering might result in swapping!

Strategies

- Balanced Systems
- Contention-Free I/O-centric Scheduling
- Horizontal Scaling and Placement Awareness
- Workload-aware Storage Tiering

In-Memory Computing

- Efficient, horizontally scaling caches (Varnish, Memcached, …)
- In-memory databases (Hana, Redis, Hazelcast, Coherence, …)
- … all done using existing DC infrastructure

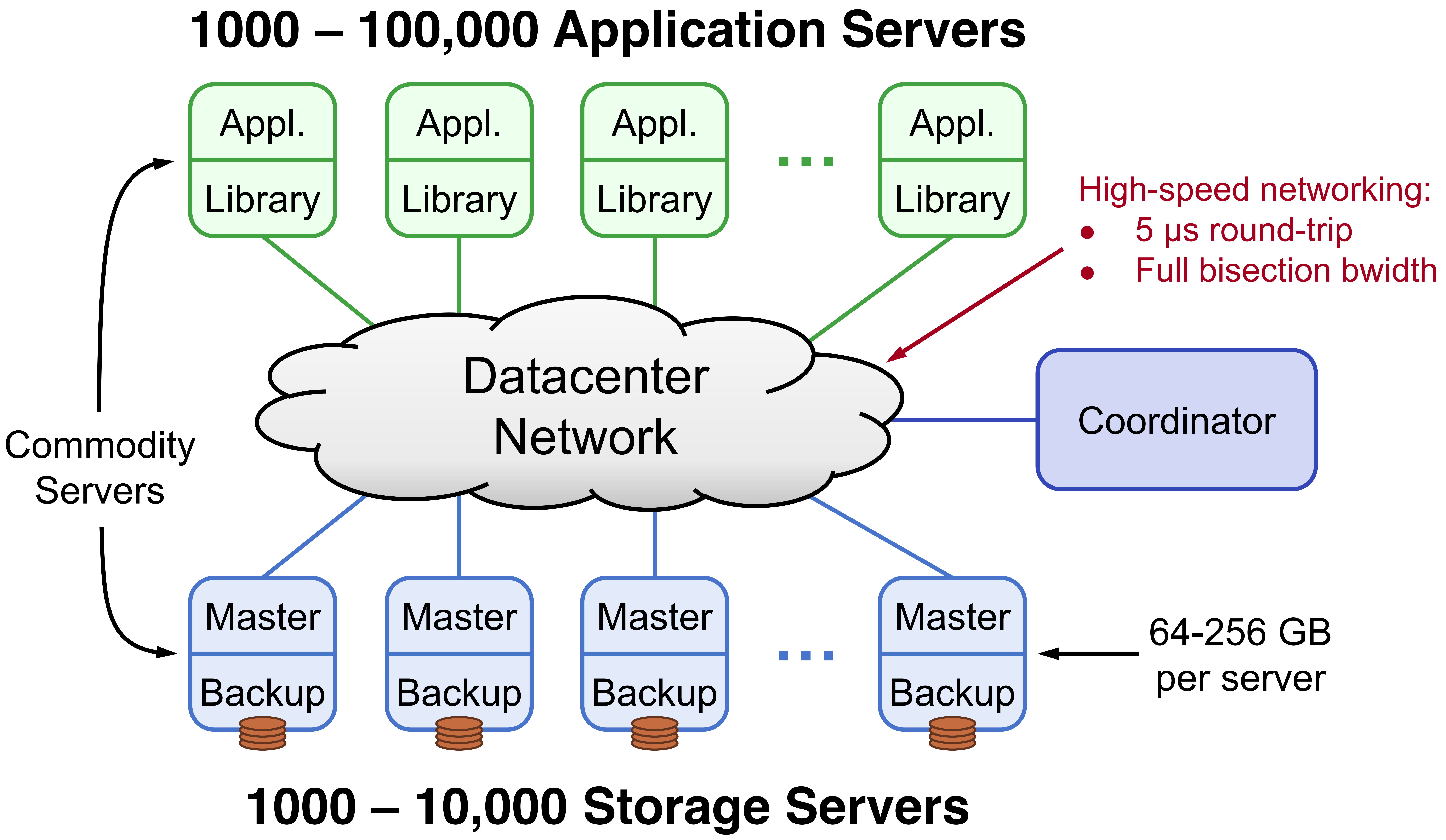
RAMCloud

https://ramcloud.atlassian.net/wiki/display/RAM/RAMCloud

http://storageconference.us/2014/Presentations/Ousterhout.pdf

- Keep all data in DRAM
- Persisted to disk/flash
- Read + write, no caching
- 5-10µs for remote RAM access
- 15µs for writes

RAMCloud Architecture



Source: http://storageconference.us/2014/Presentations/Ousterhout.pdf

RAMCloud Availability

- Open Source, “production ready”
  see https://github.com/PlatformLab/RAMCloud
- Check first: https://ramcloud.atlassian.net/wiki/display/RAM/Deciding+Whether+to+Use+RAMCloud
- Related research: FaRM (Fast Remote Memory), Microsoft
  see http://blog.acolyer.org/2015/05/20/farm-fast-remote-memory/

SDx:
Software-defined Everything

- Datacenter tradition of boxes and wires – Hardware vendors, embedded, special purpose software
- Increase in commodity hardware
- Software + services
- API-driven infrastructure service providers
- The end of shipping boxes?

We expect APIs for …

- Configuring virtual machines
- Setting up networks between arbitrary machines
- Defining firewall rules
- Assigning storage and other resources
- Auditing and compliance

Google's Jupiter

"From relatively humble beginnings, and after a misstep or two, we've built and deployed five generations of datacenter network infrastructure. Our latest-generation Jupiter network has improved capacity by more than 100x relative to our first generation network, delivering more than 1 petabit/sec of total bisection bandwidth. ***This means that each of 100,000 servers can communicate with one another in an arbitrary pattern at 10Gb/s.****"*

http://googleresearch.blogspot.de/2015/08/pulling-back-curtain-on-googles-network.html

Serverless architecture

Serverless architecture

- Write focused, small functions
- Use services provided by the platform
- Deploy to hosted cloud infrastructure
- Automatically scale on demand
- “Microservice platform as a service”

Amazon AWS Lambda

API Gateway

IAM

Cognito

Lambda

SNS

SQS

…

DynamoDB

S3

AWS Lambda

- Functions written in NodeJS, Python, Java
- Invoked by outside requests
- Triggered by integration with AWS services
- Framework & tools emerging (e.g. JAWS/Serverless framework)
- Books in progress (e.g. Obie Fernandez' "Serverless")
  https://leanpub.com/serverless

Alternatives (sort of)

- Parse (aquired by Facebook in 2013, shut down Jan 2016)
- Firebase (aquired by Google in 2014)
- Microsoft Azure App/Service Fabric
- Google Cloud

Summary

Disclaimer first:

Maybe you need none of this

Processing Power

Amazon EMR vs. “|”

- 2 million chess games, 1.75GB data
- Hadoop using 7 c1.medium EC2 instances: 26 minutes (1,14MB/s)
- Laptop using find, xargs, (m)awk: 12 seconds (270MB/s)
- 235 times faster

http://aadrake.com/command-line-tools-can-be-235x-faster-than-your-hadoop-cluster.html

Simplicity

Independents’ Tool Stacks

- A few cheap servers
- Backend written in PHP
- *Instapaper* and *Overcast* (Marco Arment)
- *Pinboard* (Maciej Ceglowski)
- *stack overflow*, written in .NET, runs on ~30 servers

  http://meta.stackexchange.com/questions/10369/which-tools-and-technologies-are-used-to-build-the-stack-exchange-network

… but if you do:

Prepare to change your DC strategy …

- Move from hardware towards software
- Automation and self-service(*)
- New economics in networking, storage, memory, CPU
- Not for the faint of heart
- Essentially, become a Cloud provider

(*) see: https://scs-architecture.org

… or use services of someone who has done so.

Maybe That Cloud Thing is for You, After All

- Dramatic change in acceptance
- Regional offerings
  (e.g. Frankfurt a.M., Germany)
- Improvements in legal aspects
  (e.g. Microsoft/T-Systems trustee arrangement)
- No silver bullet

Thank you.

Questions?

Comments?

@stilkov

Stefan Tilkov

stefan.tilkov@innoq.com

Phone: +49 170 471 2625

innoQ

www.innoq.com

innoQ Deutschland GmbH

Krischerstr. 100
40789 Monheim am Rhein
Germany
Phone: +49 2173 3366-0

Ohlauer Straße 43
10999 Berlin
Germany
Phone: +49 2173 3366-0

Ludwigstr. 180E
63067 Offenbach
Germany
Phone: +49 2173 3366-0

Kreuzstraße 16
80331 München
Germany
Phone: +49 2173 3366-0

innoQ Schweiz GmbH

Gewerbestr. 11
CH-6330 Cham
Switzerland
Phone: +41 41 743 0116