Strymon: online platform for modelling data centres

Desislava Dimitrova (Desi)

dimitrova@inf.ethz.ch

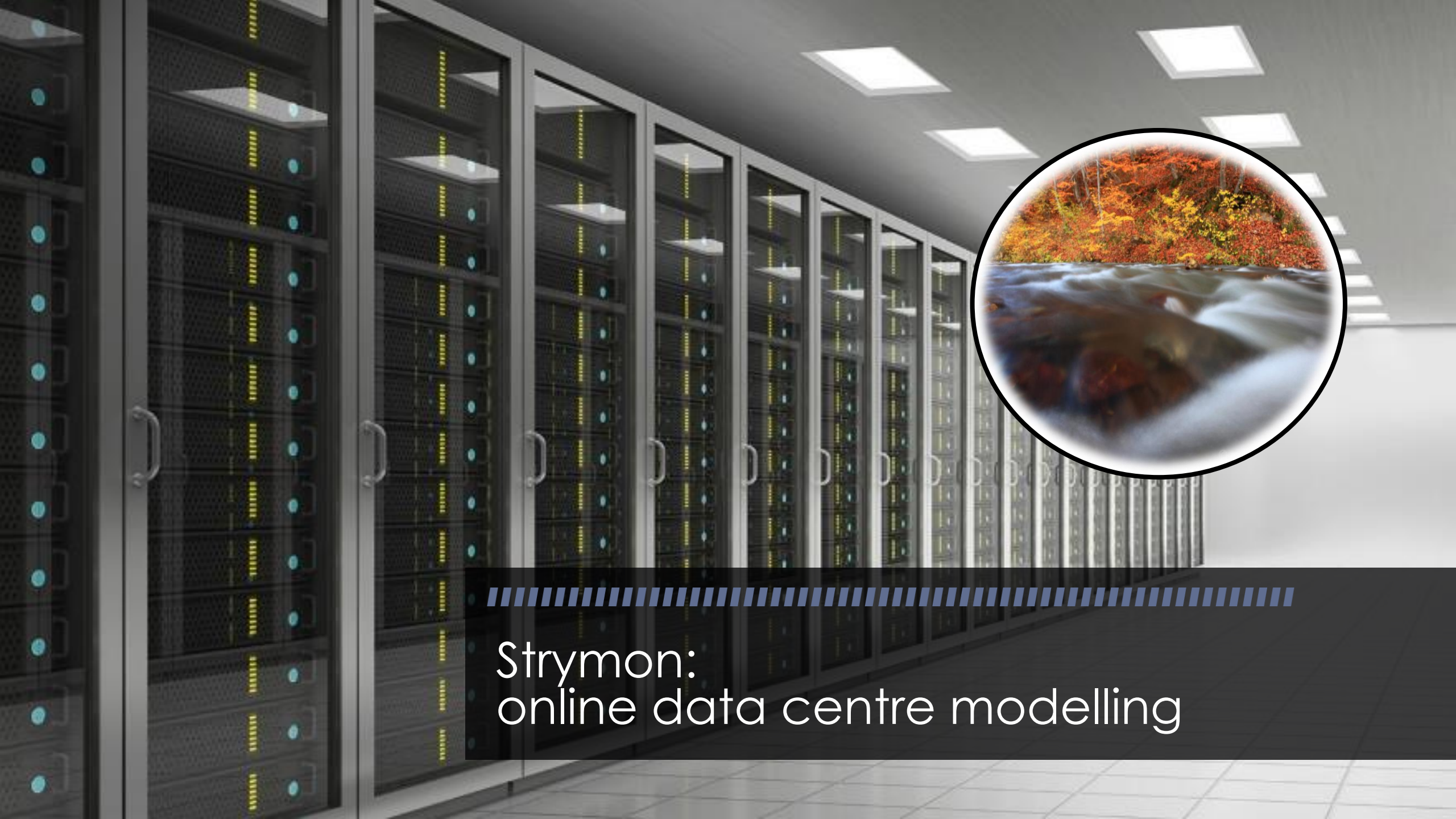
Systems Group, ETH Zurich

Today's agenda

Strymon:
online data centre
modelling

DeltaPath:
fast, scalable routing

A model for networking



Strymon:
online data centre modelling

Strymon supports data centre management

Operator queries

Data streams

Strymon

Online feedback

Online simulation

- Current system view
- Troubleshooting
- What-if analysis

Example use case

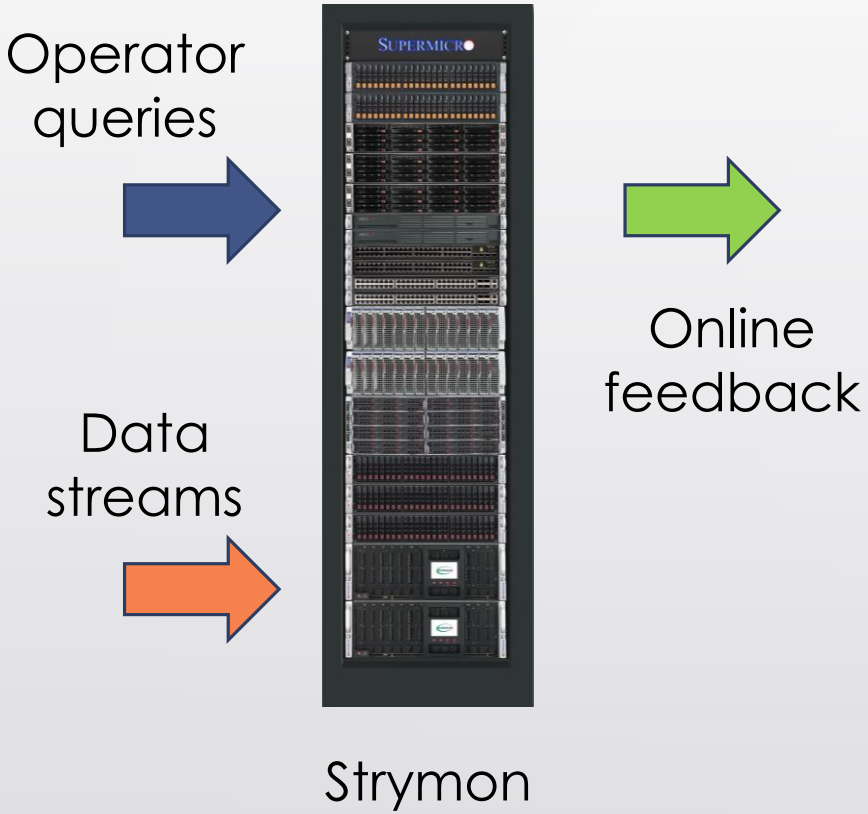
Alternative VLAN configuration

Application logs
Network topology

Strymon

Traffic matrices at application and network level

Requirements towards system design



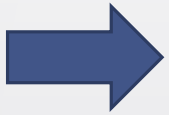
- Operate on live **streams**
- Low-latency, high-throughput computation
- Cross-layer data integration
- Modest operational cost

Strymon success stories

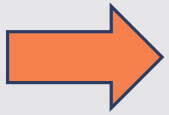
Analytics	Reconstructing transaction sessions Constructing traffic matrices
Profiling	Critical path of execution
Troubleshooting	Explaining outputs (provenance)

Strymon supports data centre management

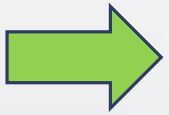
Policies



Data streams



Strymon



Configu rations

Actuation

- Configurations
- Control commands

DeltaPath: fast, scalable routing

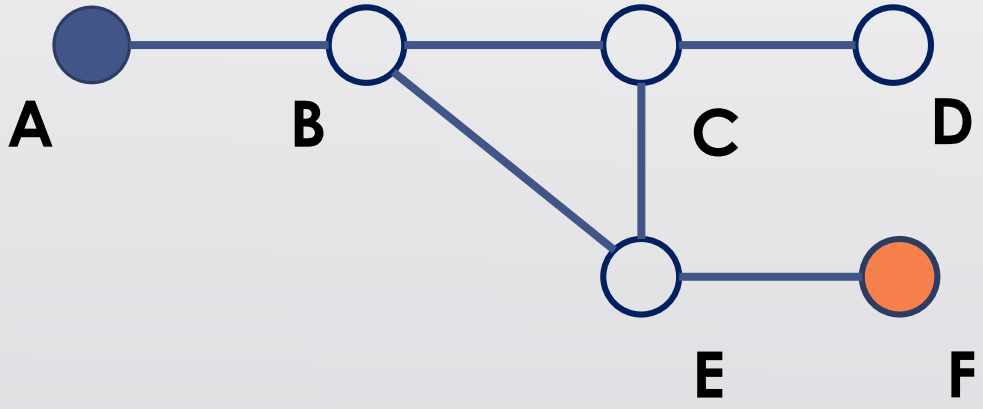
DeltaPath focuses on routing

Efficient execution of routing
in programmable networks is a challenge.

Routing in programmable networks

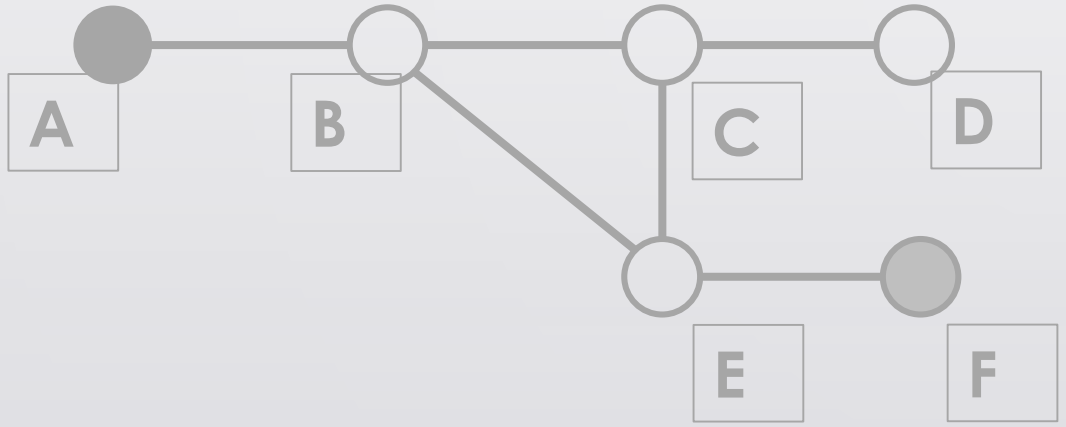
Routing in programmable networks

Routing logic

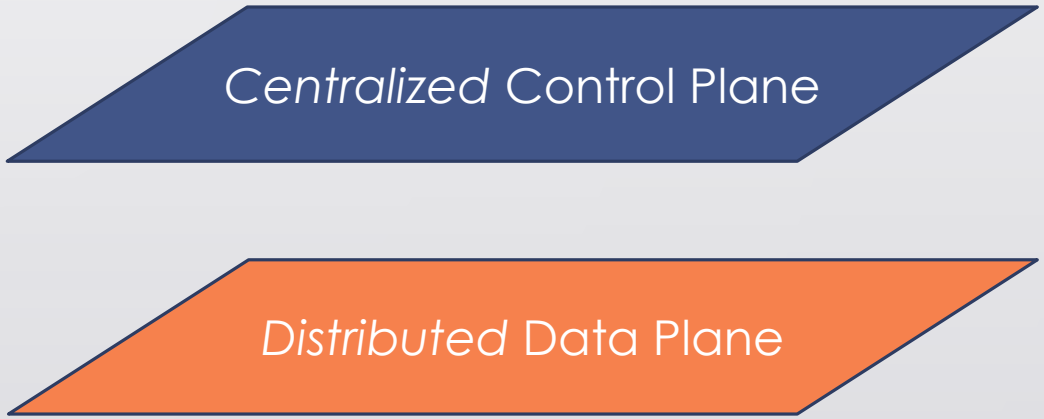


Routing in programmable networks

Routing logic



SDN's centralized control



Where challenges lay

Routing logic

- Topology changes
- Traffic changes
- Control policies change

SDN's centralized control

Where challenges lay

Routing logic

- Topology changes
- Traffic changes
- Control policies change

SDN's centralized control

Should deliver:

- Low-latency of operation
- High-throughput of handled events

A real-world example: ONOS

ONOS re-computes a single route in 36ms in a 32-port Fat Tree.

ONOS hangs with topologies bigger than 700 switches.

A real-world example: ONOS

ONOS re-computes a single route in 36ms in a 32-port Fat Tree.

DeltaPath updates all affected routes in 2.6ms.

ONOS hangs with topologies bigger than 700 switches.

DeltaPath handles a topology with at least 3k switches.

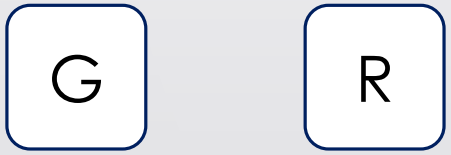
DeltaPath fusses programmable control and streaming systems

Proactive computation
of all-pairs shortest path

Incremental computation
on stream of graph changes

DeltaPath's algorithmic design

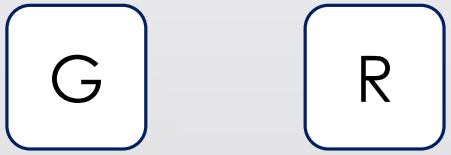
Proactive computation
of all-pairs shortest path



DeltaPath

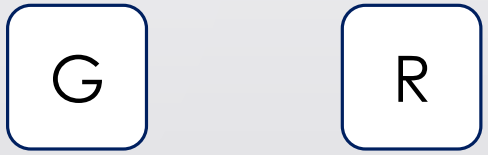
DeltaPath's algorithmic advantage to ONOS

Proactive computation of all-pairs shortest path



DeltaPath

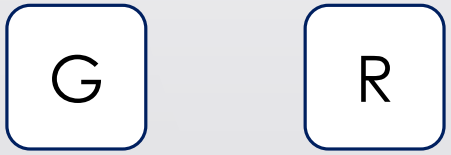
Reactive computation of single-source shortest path



ONOS

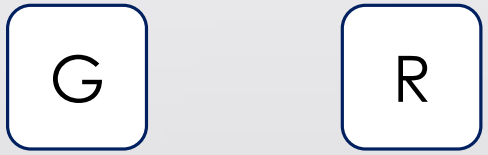
DeltaPath's algorithmic advantage to ONOS

Proactive computation of all-pairs shortest path



Single route look up in 0.1ms.

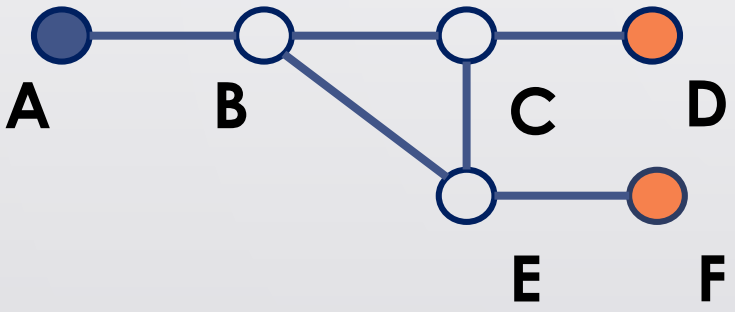
Reactive computation of single-source shortest path



Single route look up in 36ms.

DeltaPath' computational innovation

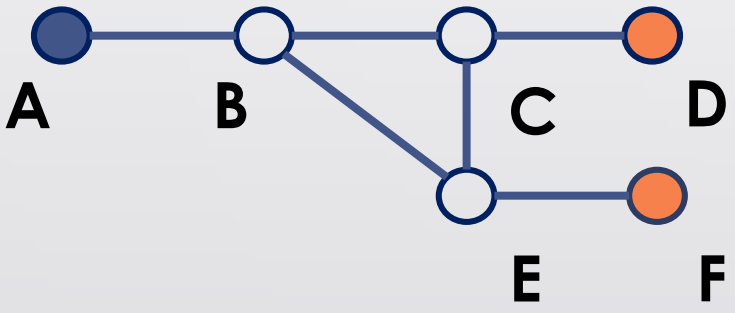
Incremental computation on stream of changes



DeltaPath

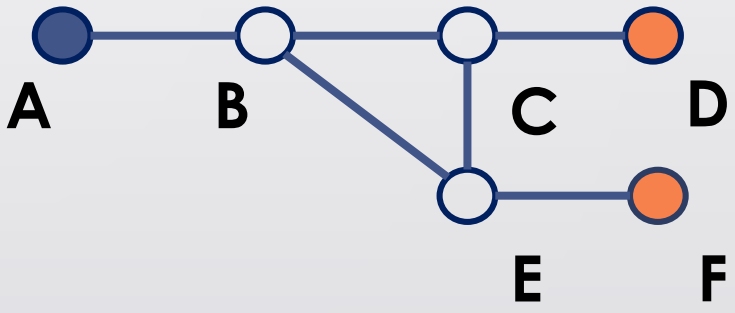
DeltaPath' computational innovation

Incremental computation on stream of changes



DeltaPath

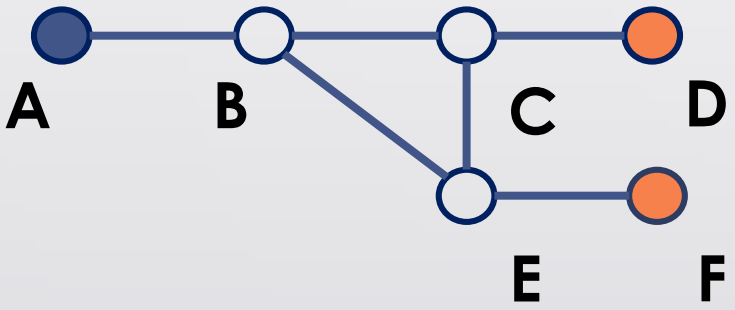
Full re-computation on graph changes



ONOS

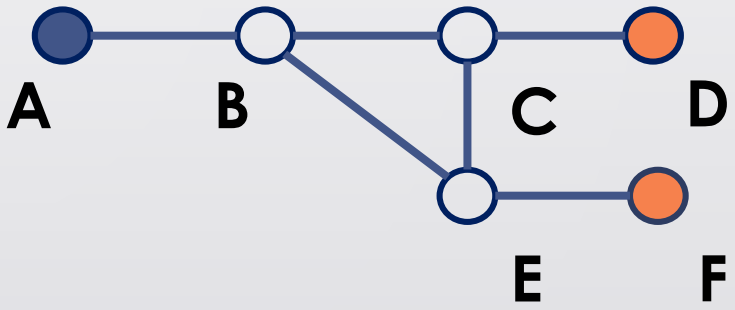
DeltaPath' computational innovation

Incremental computation on stream of changes



Re-routing is function of available paths

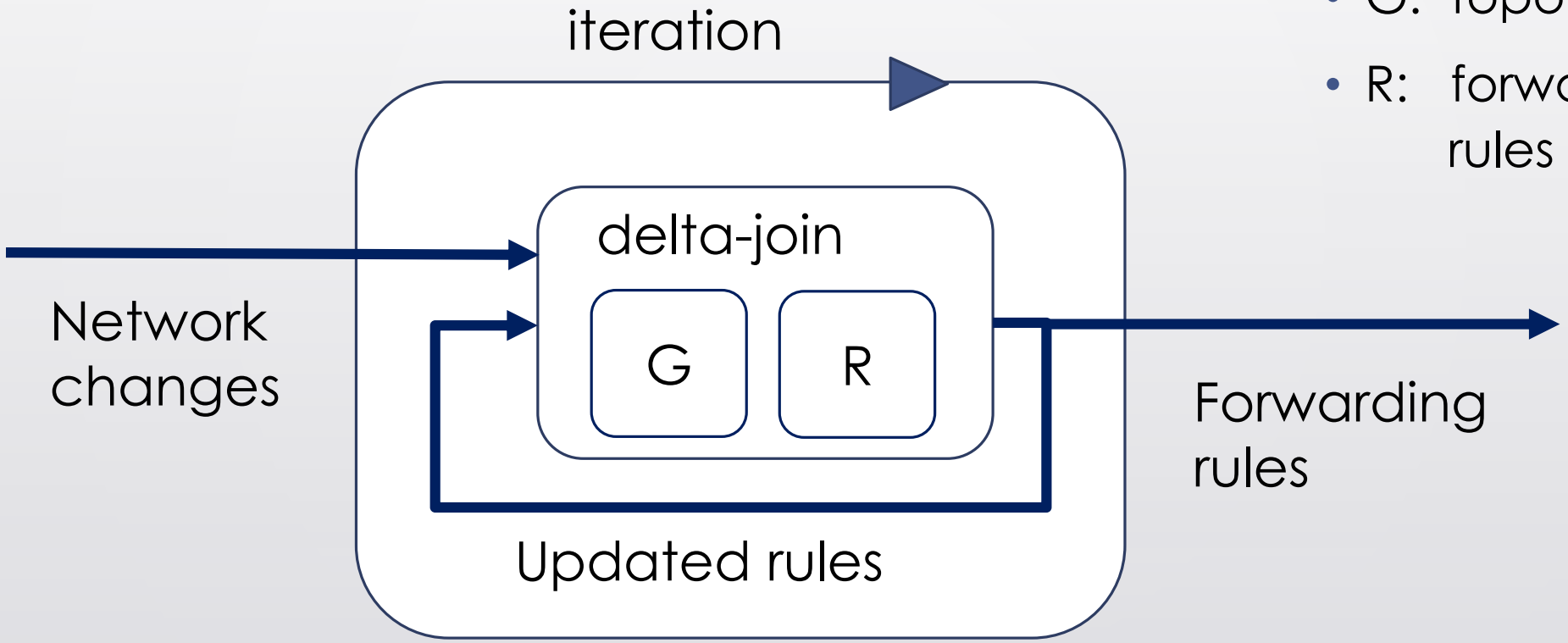
Full re-computation on graph changes



Re-routing is function of active flows

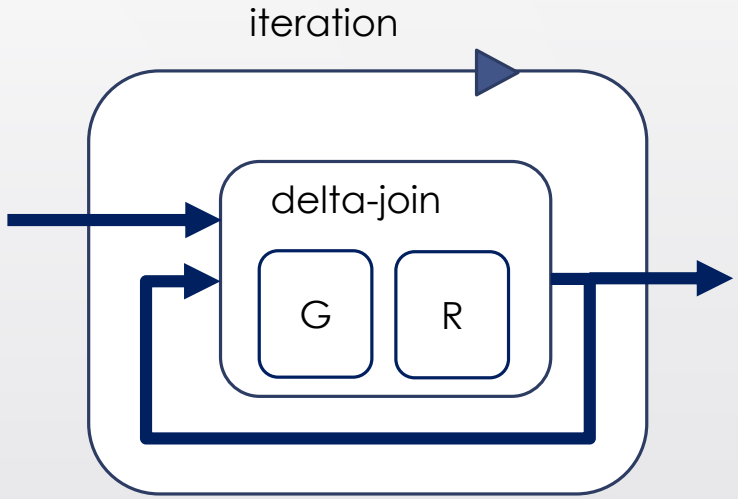
Incremental computation on stream of graph changes

- G: topology graph
- R: forwarding rules set



Custom operator in a streaming framework

- Timely Dataflow
  - Arbitrary cyclic dataflows
  - Logical timestamps (epochs)
  - Asynchronous execution
  - Low latency, modest resources



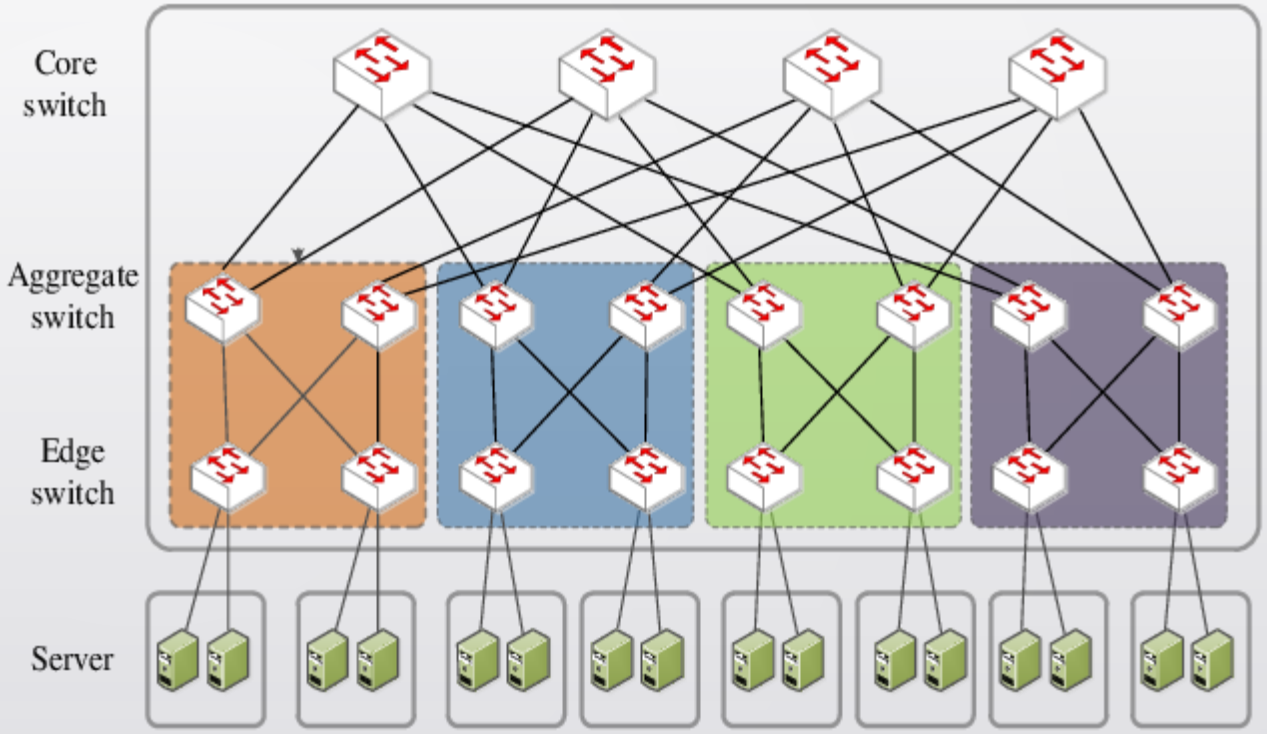
Rust implementation: **github.com/frankmcsherry/timely-dataflow**

D. Murray, F. McSherry, M. Isard, R. Isaacs, P. Barham, M. Abadi. ***Naiad: A Timely Dataflow System.*** *In SOSP, 2013.*

Did we just write the fastest SDN routing logic?

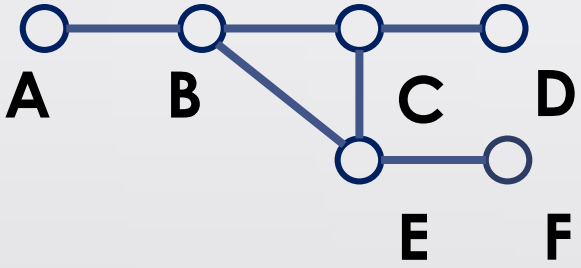
32 port Fat Tree

- 1280 switches
- 8k hosts
- random link weights

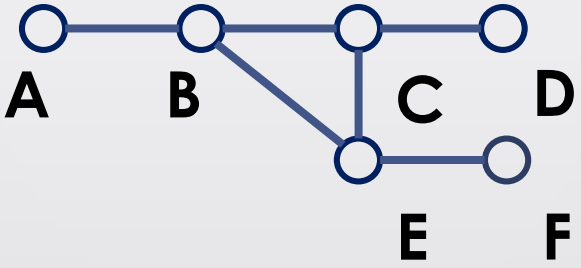


Quad-socket Intel Xeon E5-4640 with 512 GiB RAM and 8 2.40 GHz cores per socket

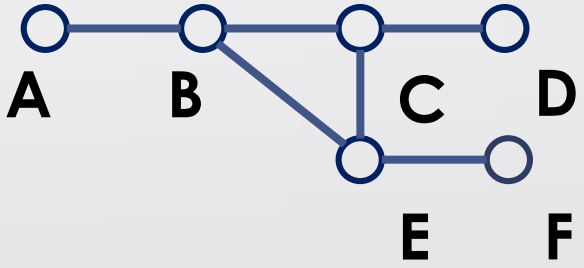
Evaluated events



Look-ups



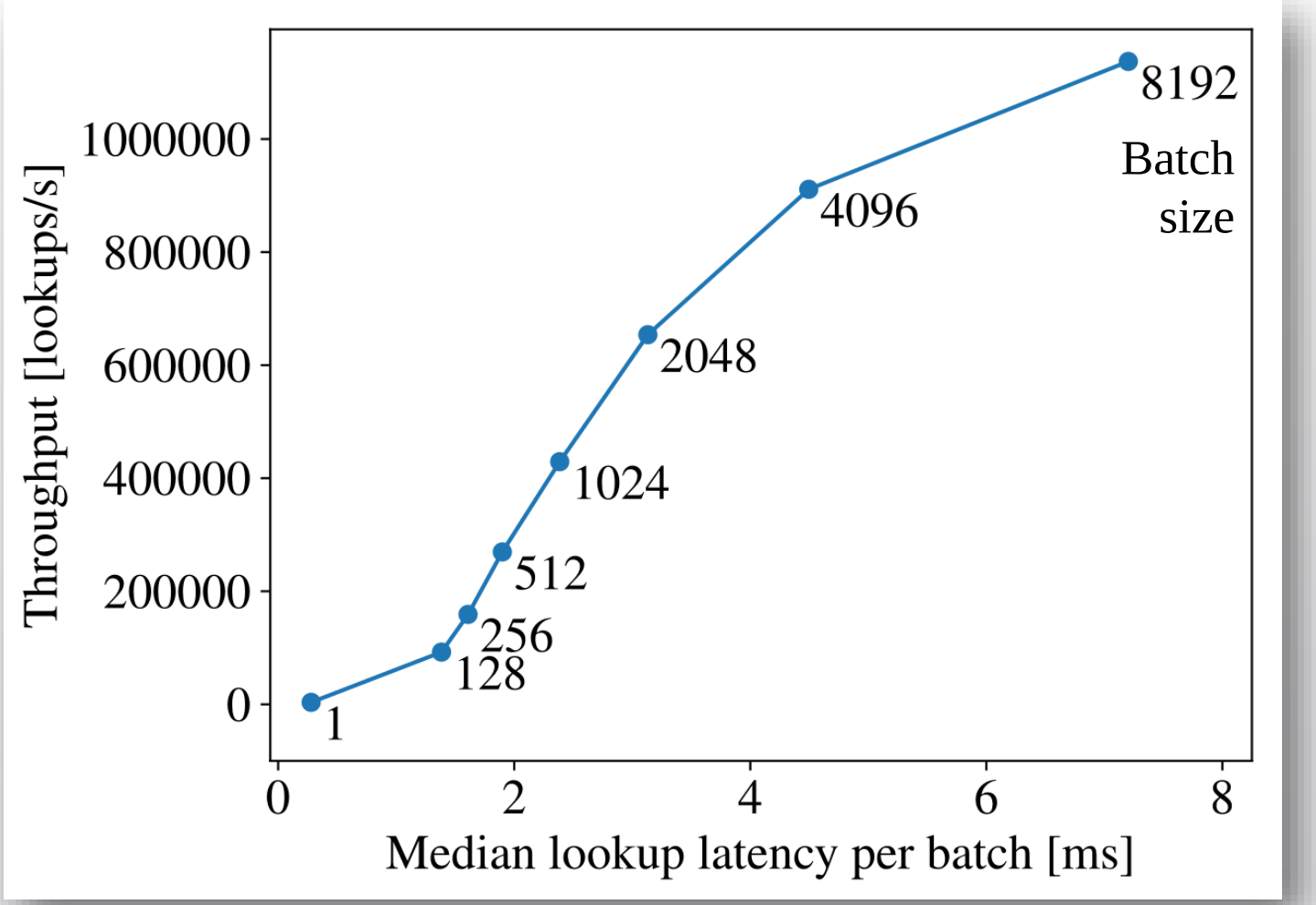
Failures



Updates

>1M path look-ups under 10ms

- Random source-destination pairs of access switches
- Increasing batch size: groups of look-up requests
- 8 threads



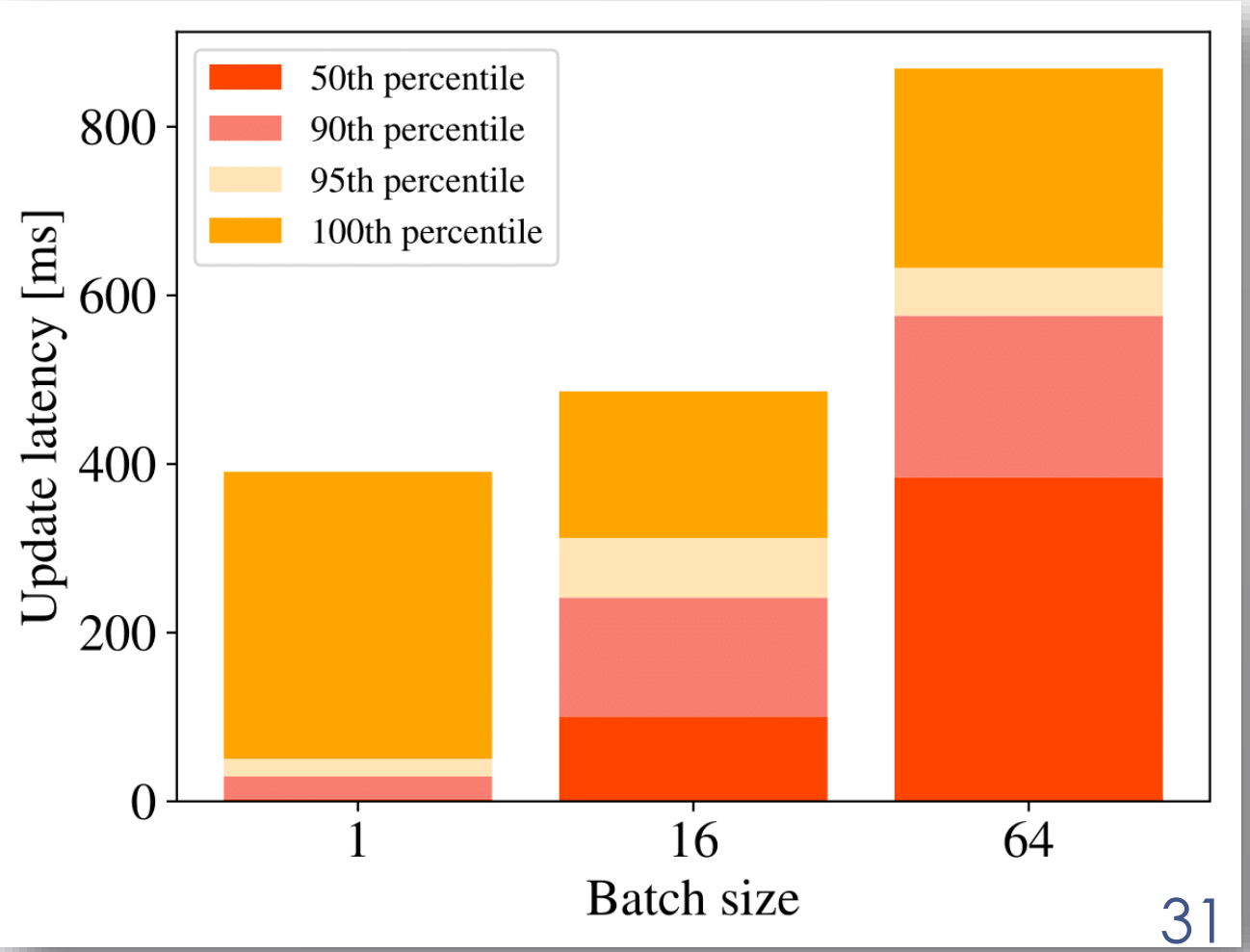
Median failure recovery in 2.6ms!

- Remove random link
- 1000 individual runs
- 8 threads

%	Latency (ms)
50	2.60
90	50.47
95	211.06
100	390.74

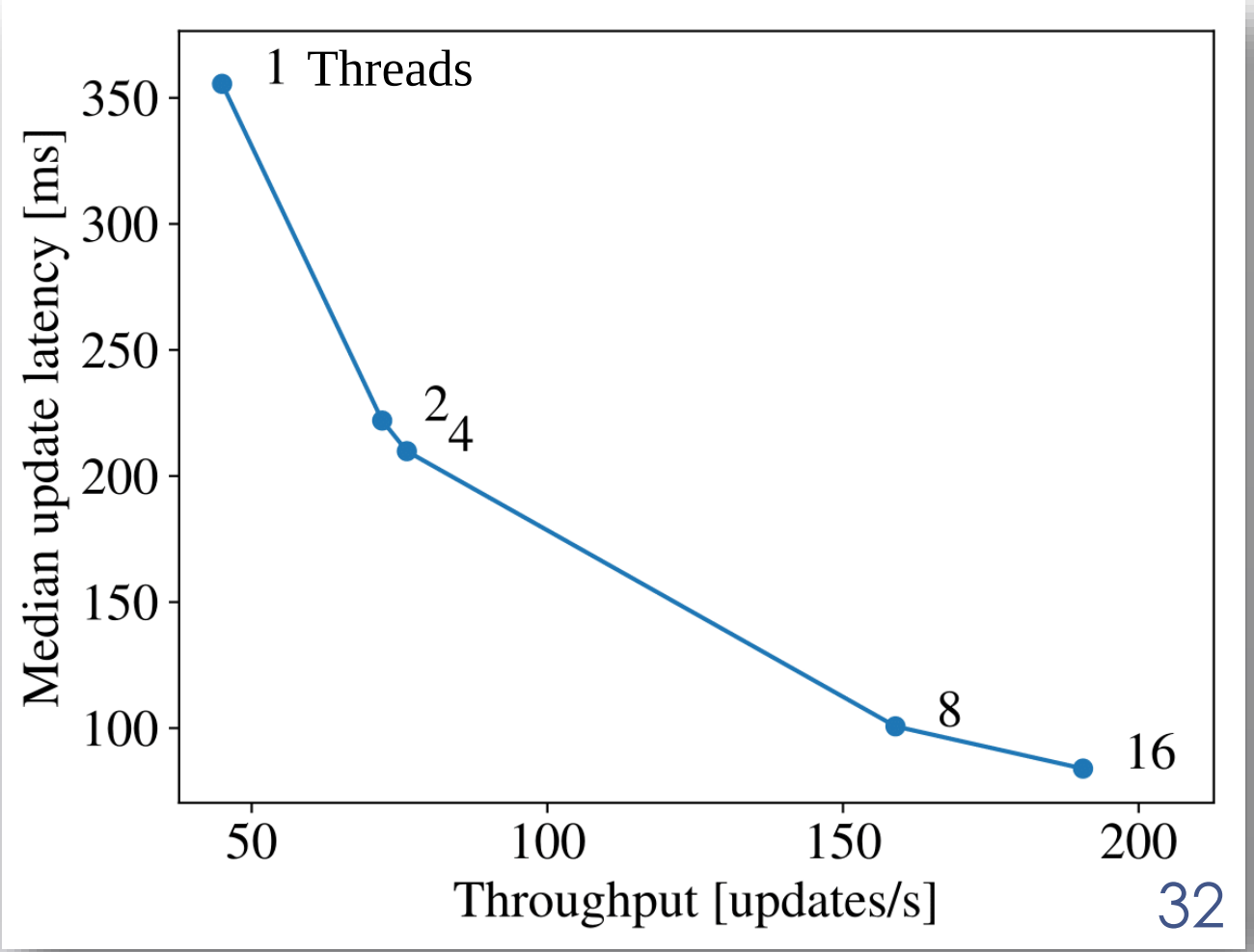
Adapts to fine-grained link changes

- Change weight of random links grouped in batches
- 1000 individual runs
- Vary batch size
- 8 threads



DeltaPath scales well

- Change weight of random links grouped in batches
- 1000 individual runs
- Batch size = 16
- Vary threads



DeltaPath outperforms open-source controllers

DeltaPath reacts fast.

DeltaPath scales.

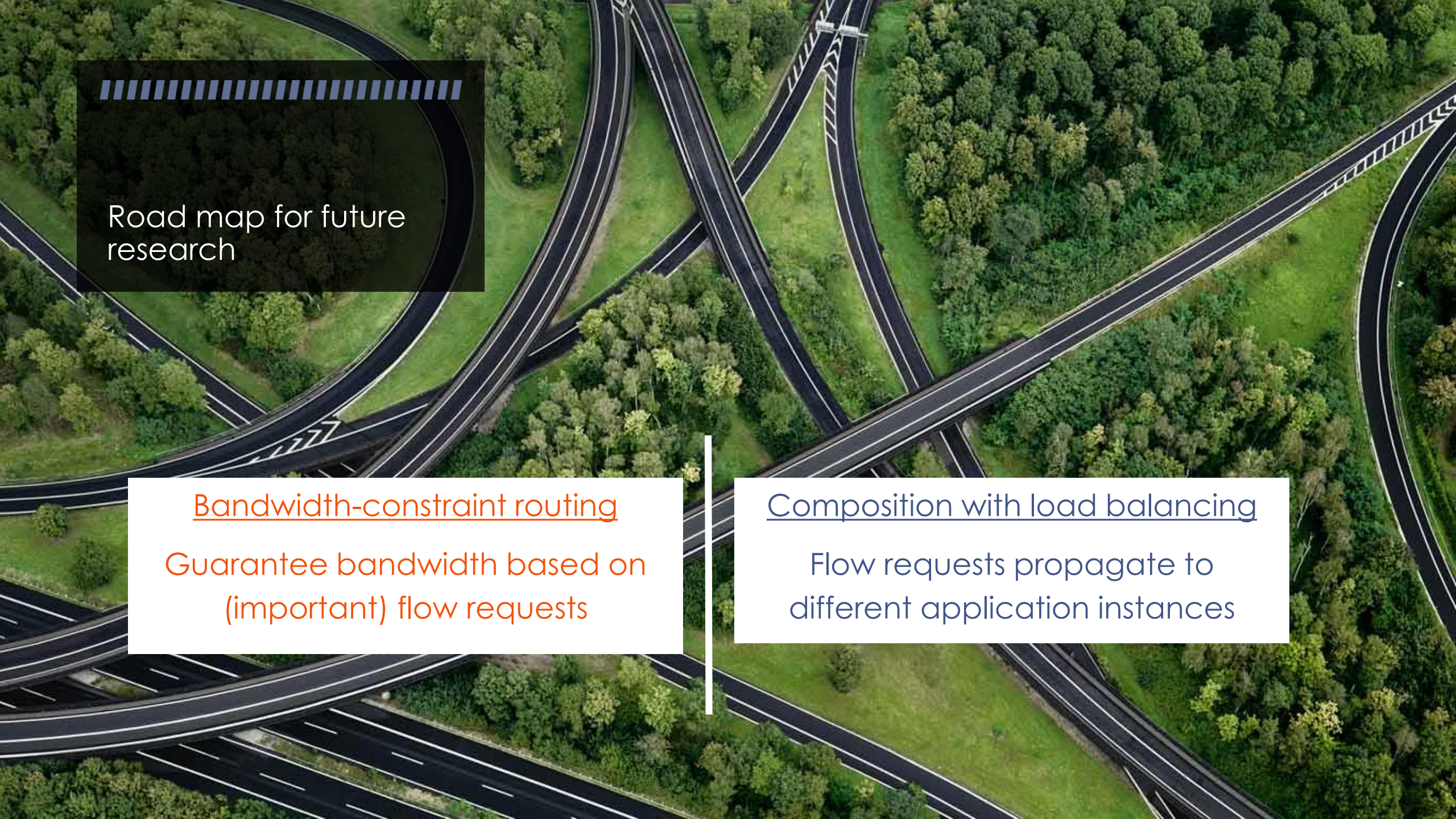
What else can we do?

Road map for future research

Road map for future research

Bandwidth-constraint routing

Guarantee bandwidth based on (important) flow requests



Road map for future research

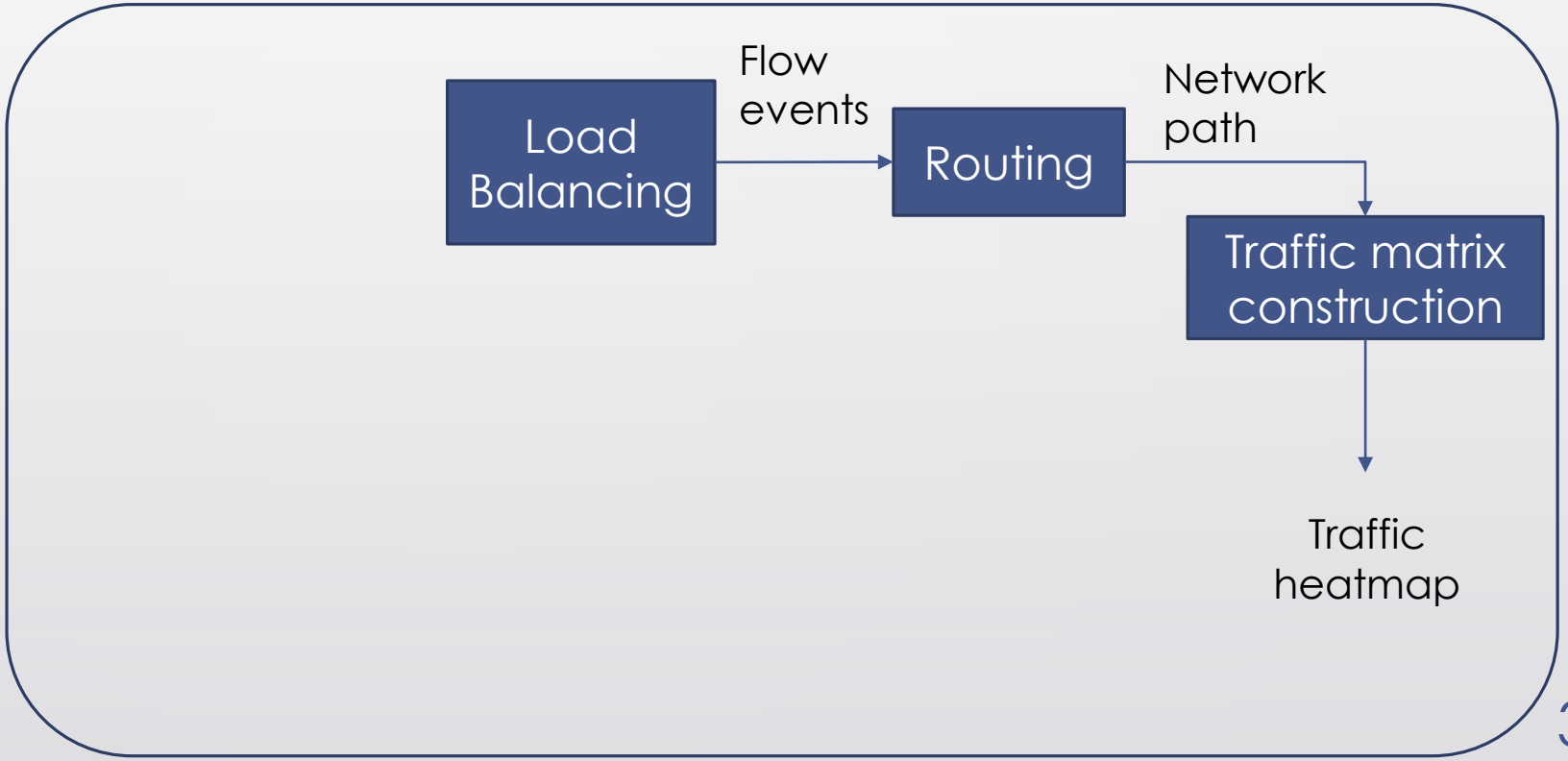
Bandwidth-constraint routing

Guarantee bandwidth based on (important) flow requests

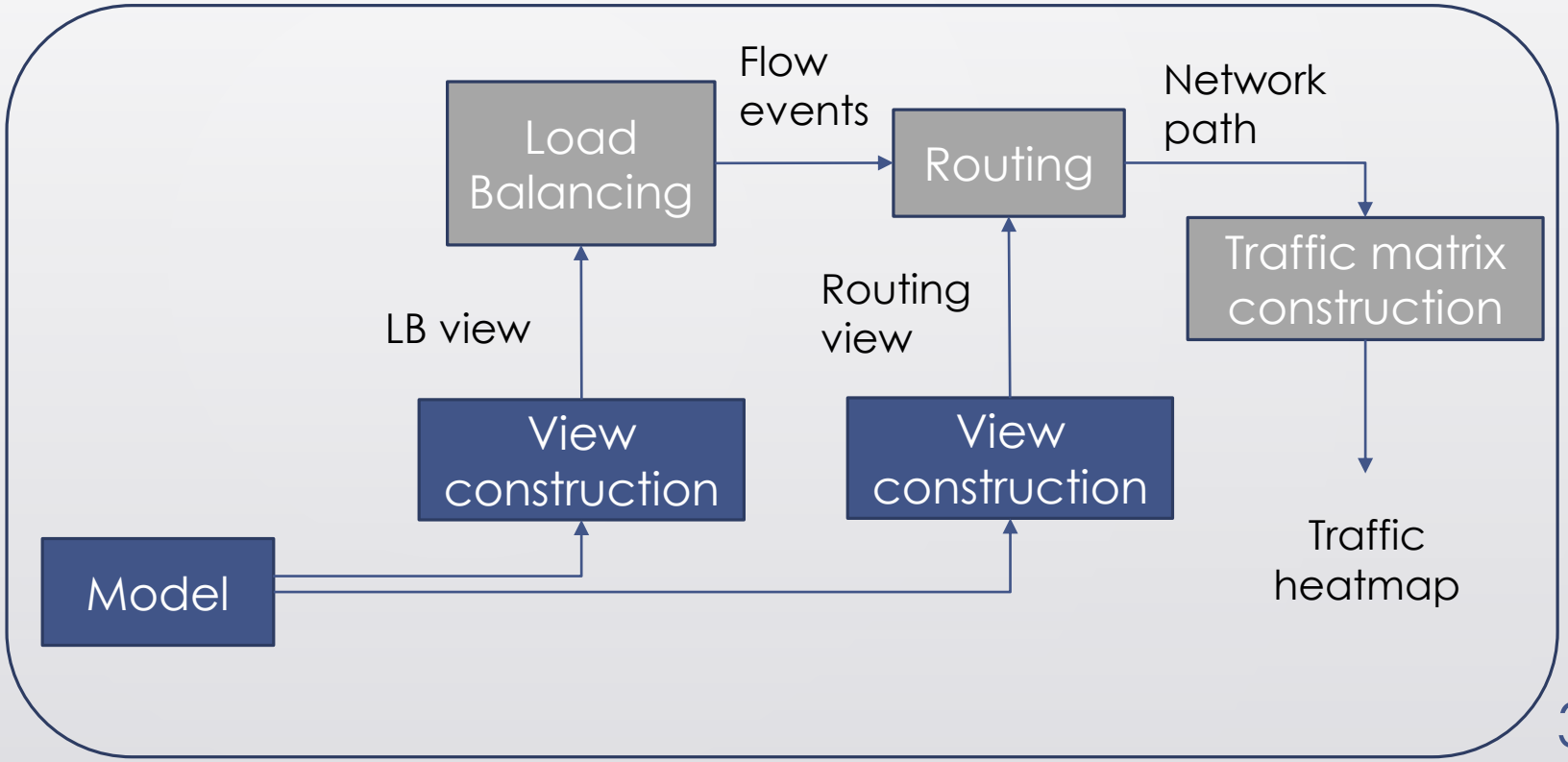
Composition with load balancing

Flow requests propagate to different application instances

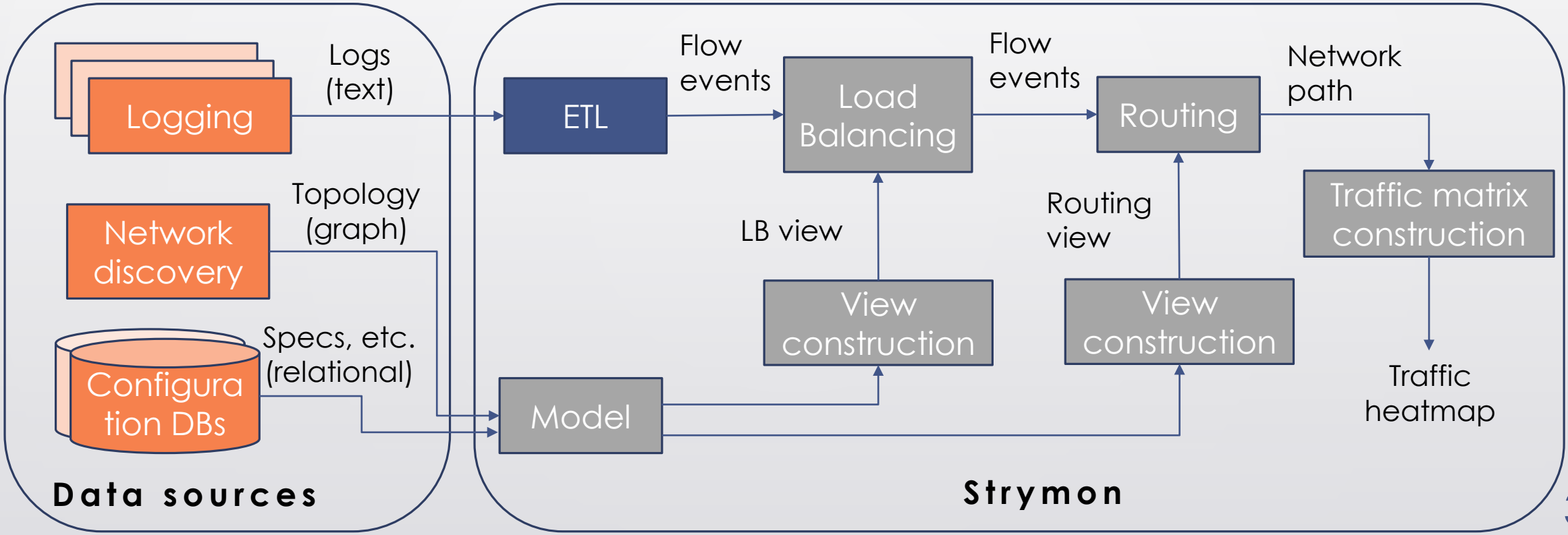
Complex networking scenarios



Complex networking scenarios require rich data model



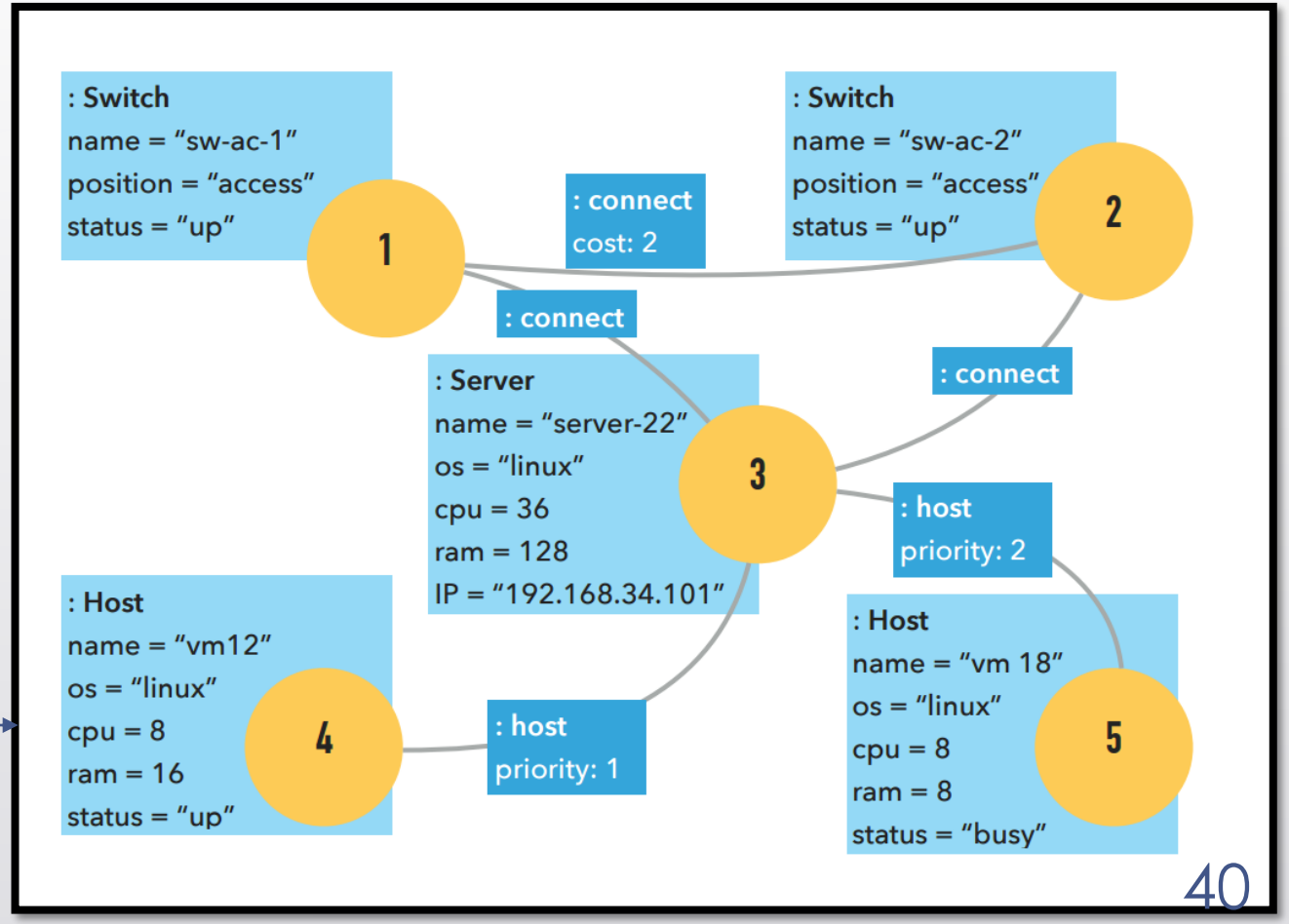
Rich data model requires data integration



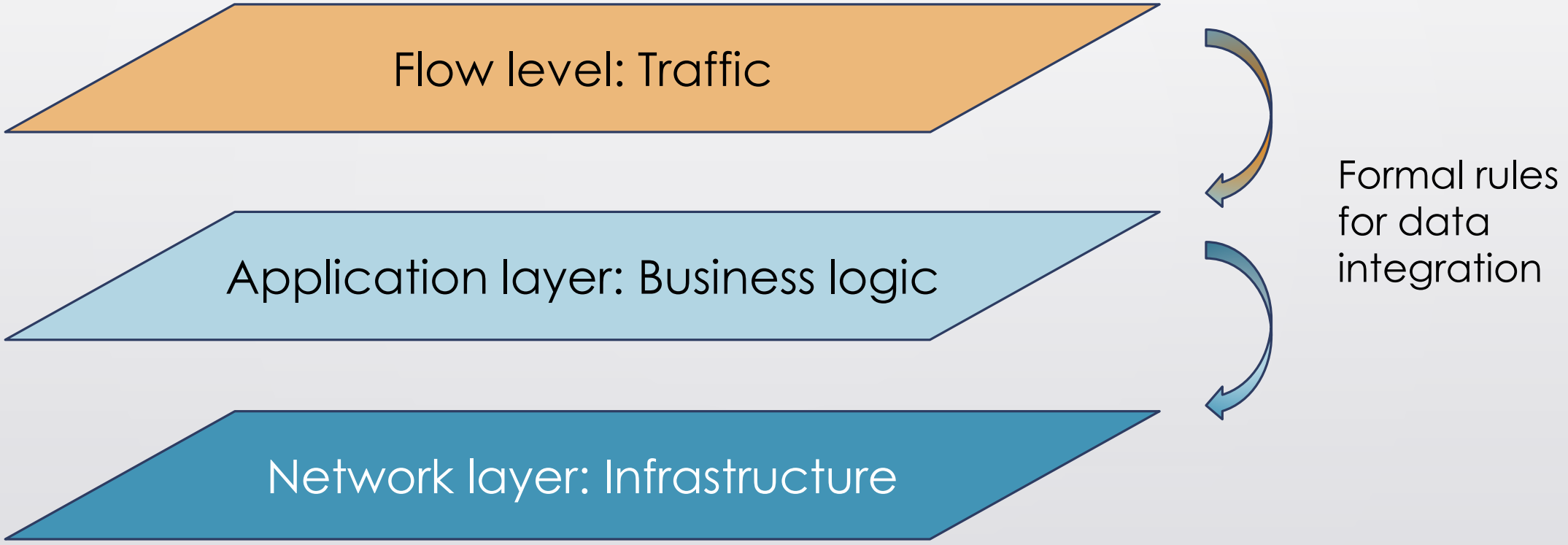
DeltaPath's data model: labelled property graph

Network discovery

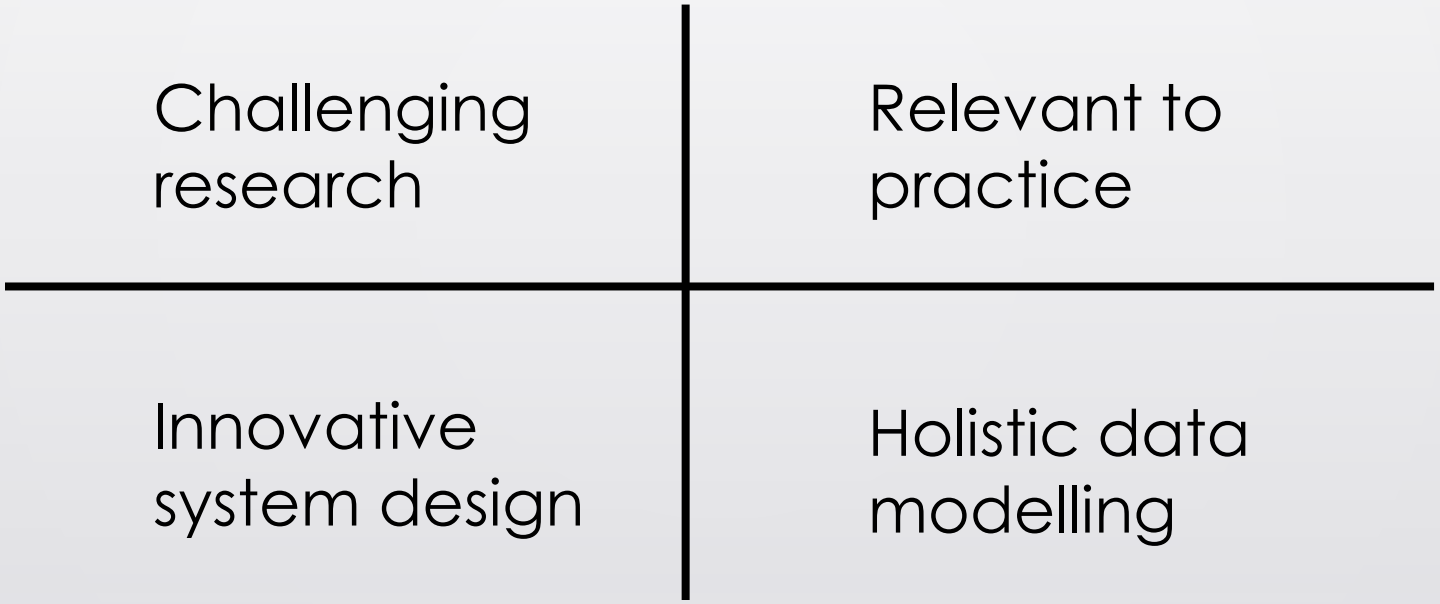
Topology (graph)



Multi-layered representation with formal data relations



Strymon: a future we believe in



We welcome exciting research challenges

Vasiliki Kalavri

http://strymon.systems.ethz.ch

sdn@inf.ethz.ch

John Liagouris

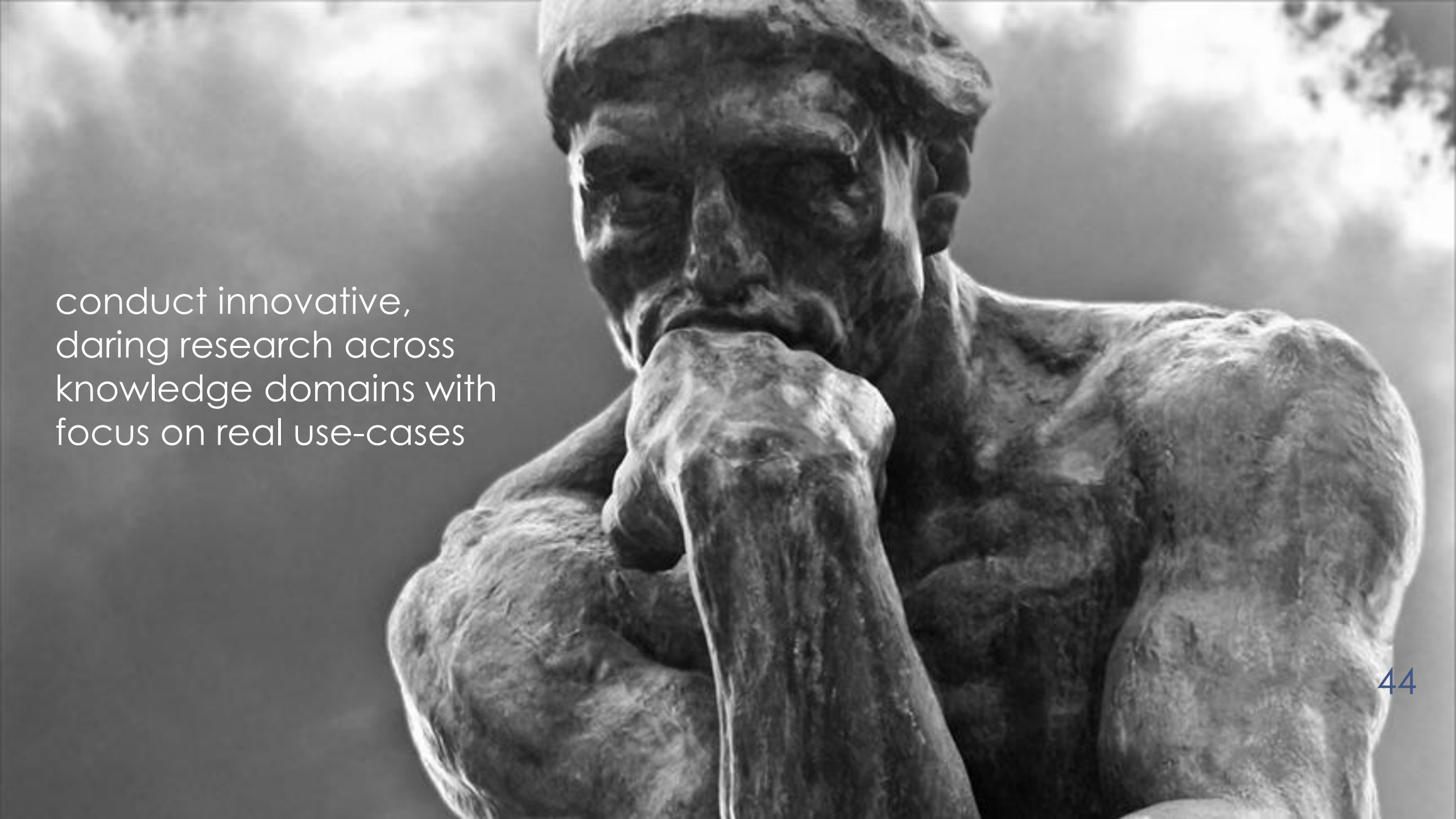
Andrea Lattuada

Zaheer Chothia

Moritz Hoffmann

Sebastian Wicki

Timothy Roscoe

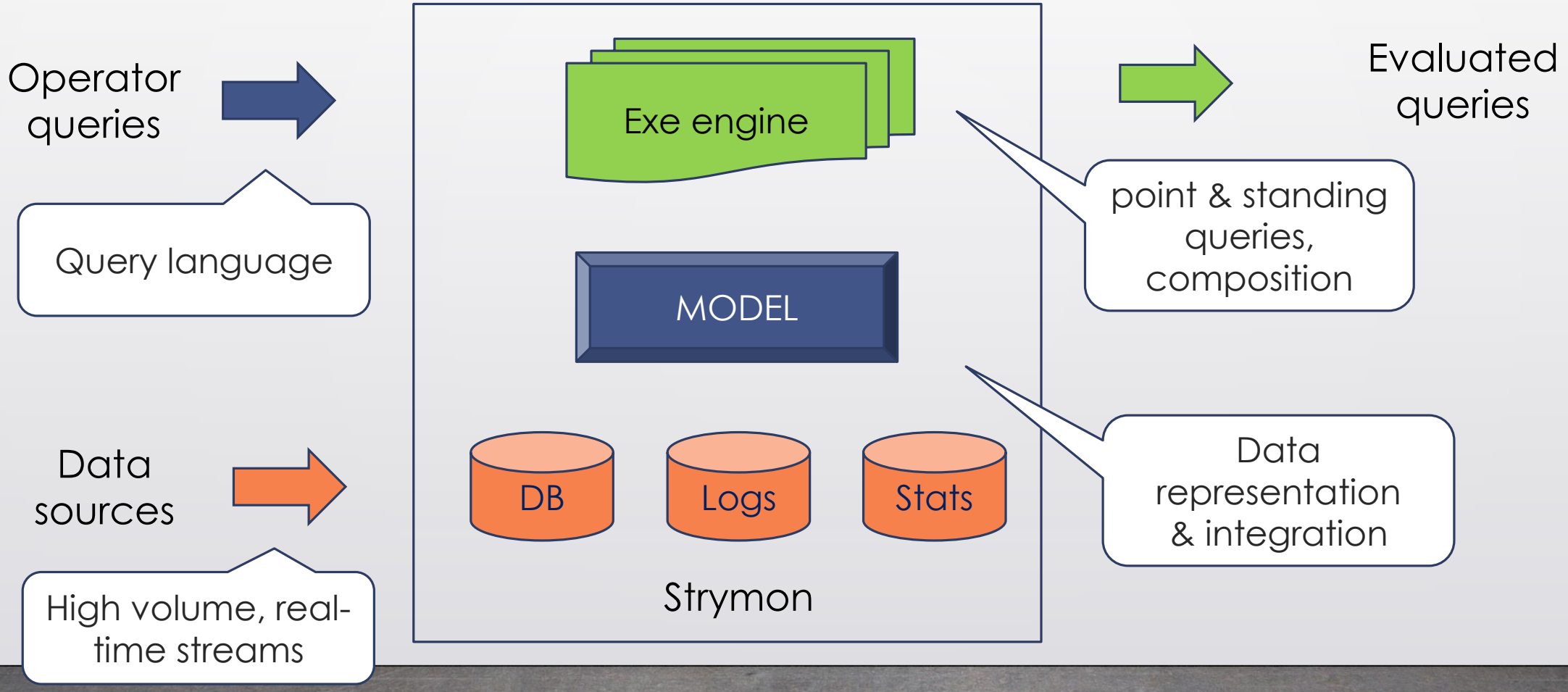

conduct innovative,
daring research across
knowledge domains with
focus on real use-cases

A unified approach towards ingesting data sources

- Data sources have different semantics and format
- Data sources may come and go to the system
- Reusability and automation are desirable

- Resource Description Framework as a common ground

Challenges in developing Strymon



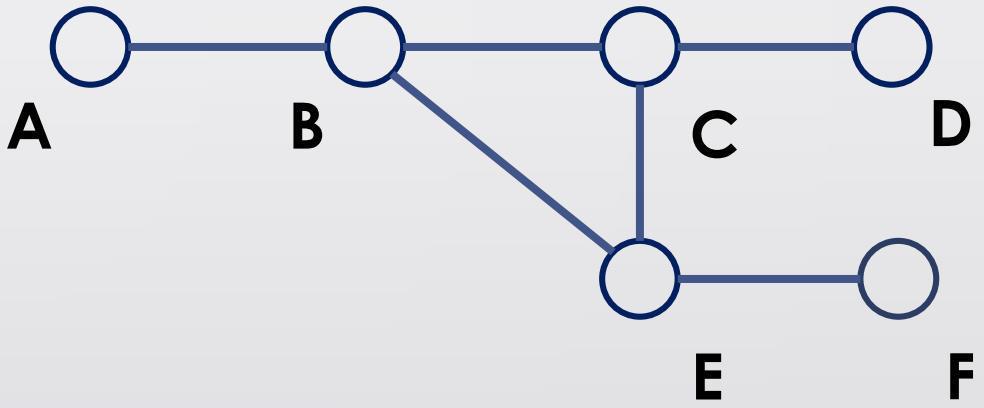
Key components to DeltaPath's performance

Routing as incremental, cyclic streaming computation on graphs

Proactive computation of all-pairs shortest path

Key components to DeltaPath's performance

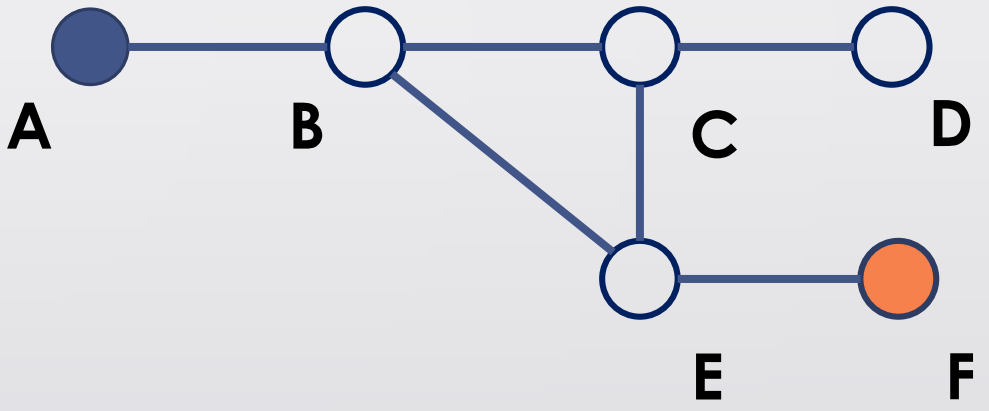
Routing as incremental, cyclic streaming computation on **graphs**



Key components to DeltaPath's performance

Routing as incremental, **cyclic** streaming computation on graphs

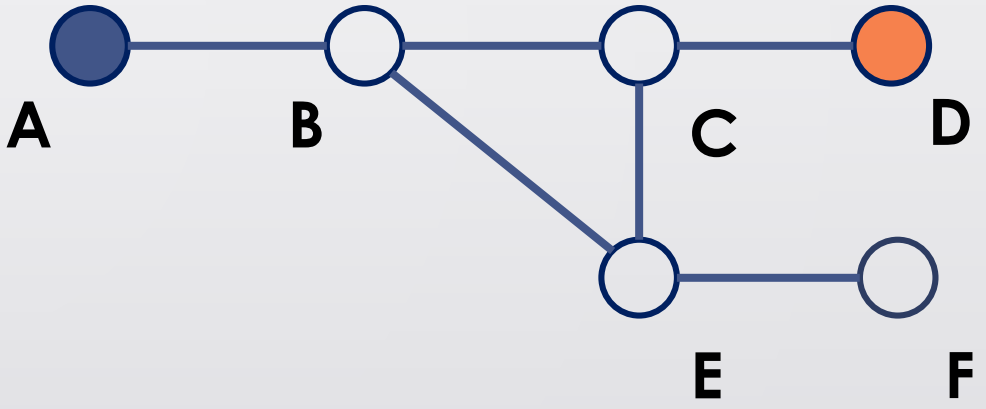
Dijkstra's algorithm for (A,F)



Key components to DeltaPath's performance

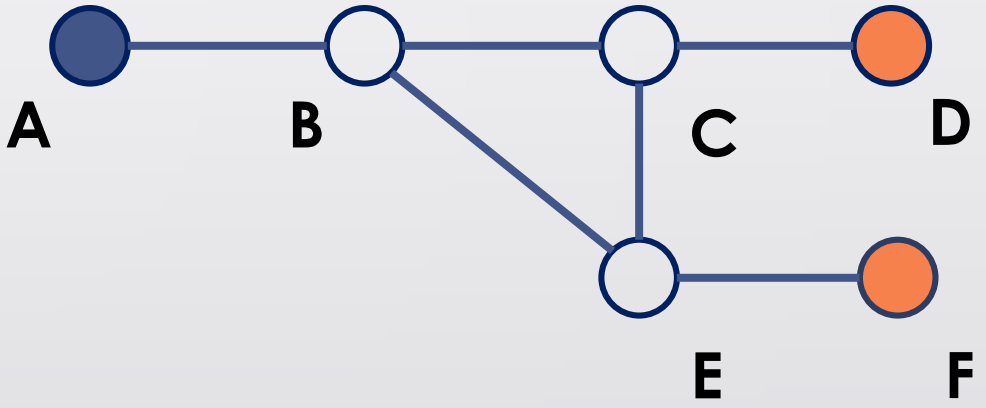
Routing as incremental, cyclic **streaming** computation on graphs

Stream: (A,F)(**A**,**D**)(D,F)…



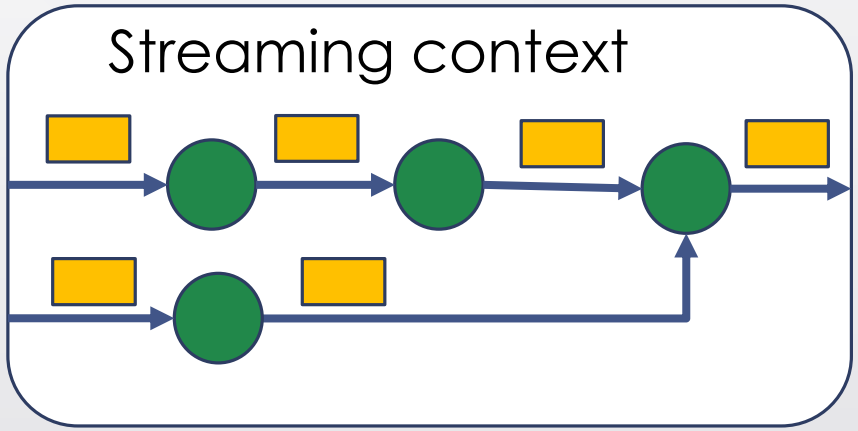
Key components to DeltaPath's performance

Routing as incremental, cyclic streaming computation on graphs



DeltaPath is natural fit to dataflow programming

- Computation is a graph of **operators**
- **Data** flows on graph edges



- DeltaPath's execution engine is is an operator in Timely

Where challenges lay

Routing logic

Everything changes and nothing remains still

SDN's centralized control

Where challenges lay

Routing logic

Everything changes and nothing remains still

SDN's centralized control

Scale is an issue