

# Cloud-Scale Application Performance Monitoring with SDN and NFV

Guyue Liu Timothy Wood

George Washington University

An aerial photograph of a large, modern data center building. The building is a long, rectangular structure with a flat roof. On the roof, there are numerous rows of white server racks. The building is surrounded by a parking lot with several cars and a few construction vehicles. In the background, there is a large, open area of land, possibly a construction site or a field.

**. Data Centers are BIG!**  
**- 100,000+ servers**

A photograph of a server room with rows of server racks. The racks are filled with server units, and the room is illuminated with blue light. The perspective is from a low angle, looking down a central aisle between the racks.

**. Applications are DIVERSE!**  
**- More than just LAMP**



**. Networks are COMPLEX**

**- More than just cables  
and switches**

# Common Questions

- . Why is my multi-tier web application running slowly?
- . Which types of requests are fast or slow?
- . What is my most popular content?
- . Which tier is my bottleneck?
- . Where is network congestion affecting performance?

*very efficiently*

How to understand the **performance** and **behavior** of these large scale systems?

# Emerging Trends

## . Big Data

- Hadoop/Map Reduce: scalable batch processing
- Real-time, streaming analytics

## . Software-based Networks

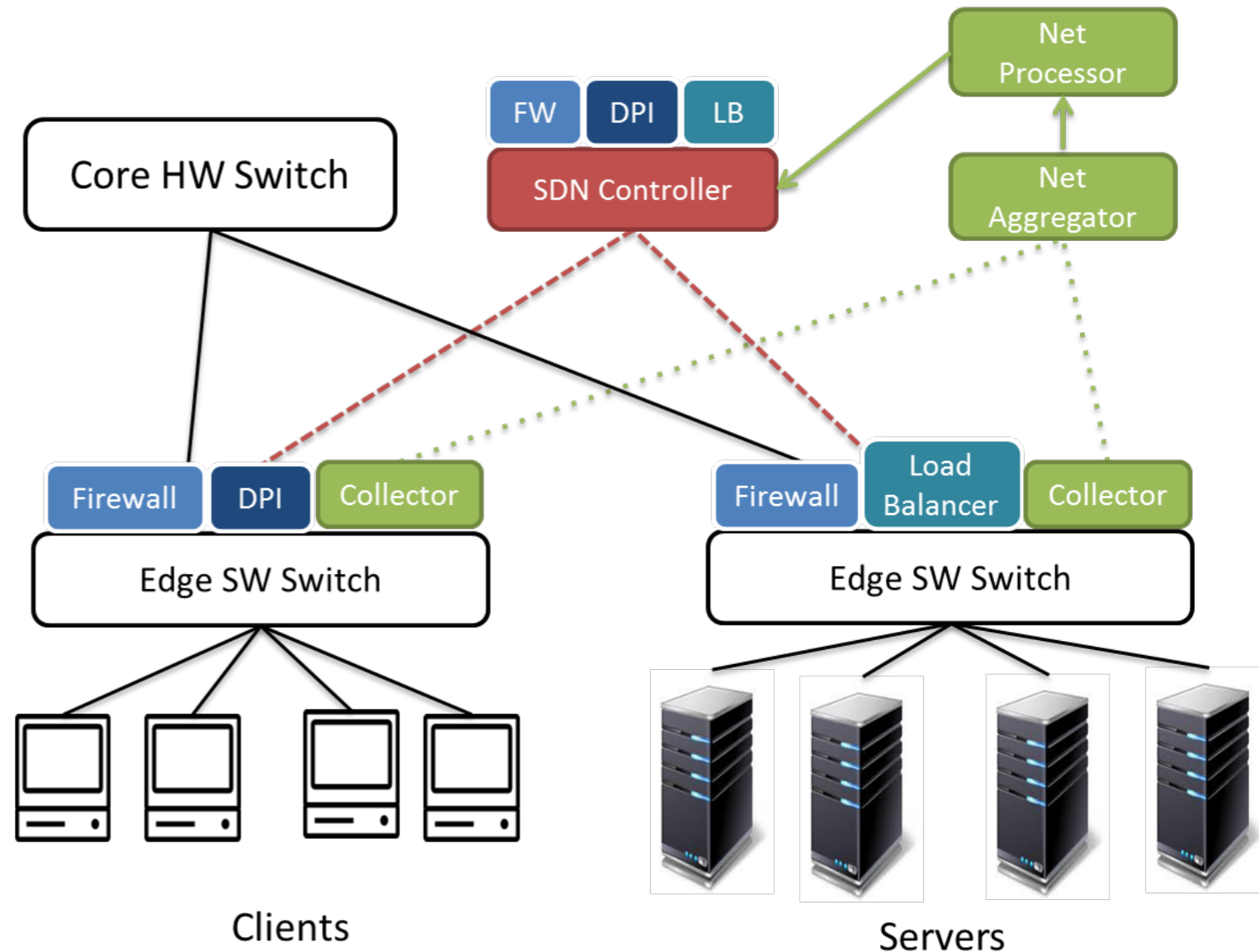
- Software Defined Networks: flexible control of packet flows
- Network Function Virtualization: efficient packet processing software

#1: Tools to analyze large amounts of data with low latency

#2: Flexible network architectures that can gather network state

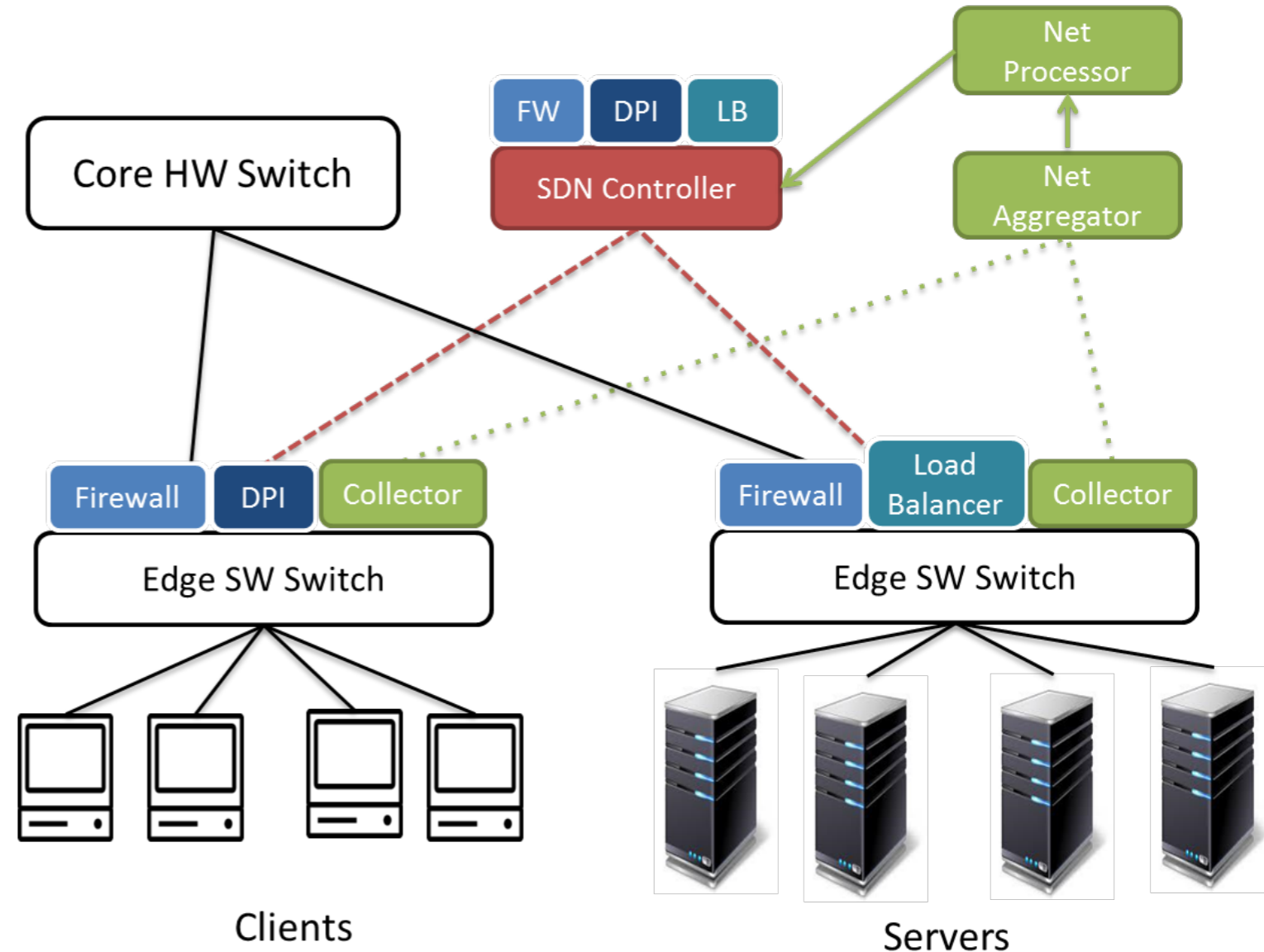
# NetAlytics Overview

- . **Monitors:**
  - . Virtual machines running on top of software switches.
  - . Efficiently gather data by observing traffic
- . **Collectors:** Collect data from the local monitors and may perform minimal processing



# NetAlytics Overview

- . **Aggregators:**
- . Queuing services that store measurements from one or more collectors
- . **Processors:** Scalable analytics engine to provide real time insights into network and application behavior





# Why focus on the network?

- . NetAlytics gathers **network state** for analysis
- . No application modifications
  - Deployable as a “cloud service”
- . Can infer a lot
  - Measure web response times by observing TCP SYN/FIN
  - Differentiate between network and end-host bottlenecks
  - Detect anomalous behavior, suspicious packets
- . Can easily be combined with existing VM-monitoring solutions to provide network + end-host data

# NetAlytics: Control

- . Analytic results should drive the cloud manager
- . Existing cloud management systems use simple data
  - resource consumption on each host
  - available network bandwidth
- . NetAlytics can expose much more:
  - anomalous traffic behavior
  - clustered response times
  - packet-level security analysis

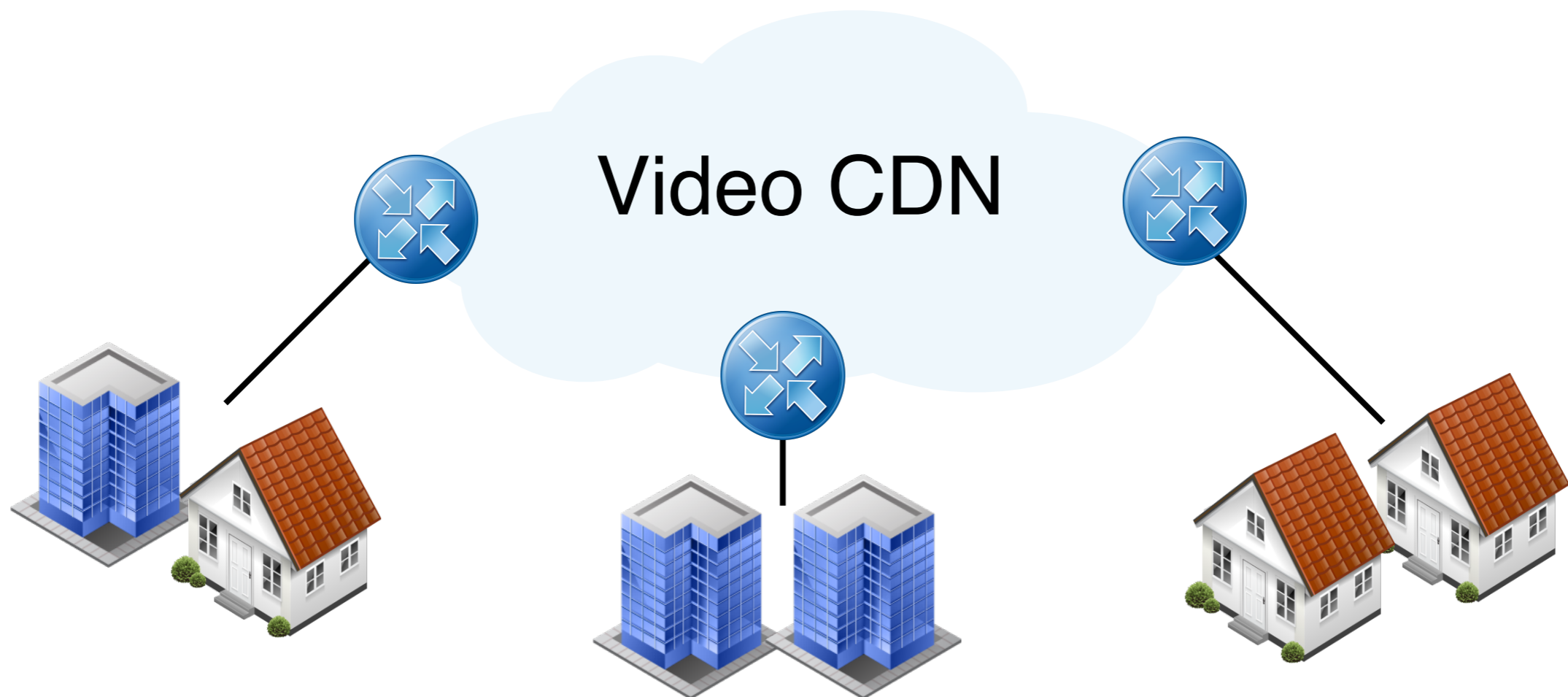
# Implementation

- . **NetVM monitors** [NSDI 14]
  - High performance NFV platform (50M+ pps)
  - Zero-copy data transfer to and between VMs
- . **Apache Kafka aggregators**
  - High throughput, distributed messaging system
- . **Apache Storm processors**
  - Distributed real time computation system
- . **Managed by POX SDN controller**
  - Configures packet routing and port mirroring



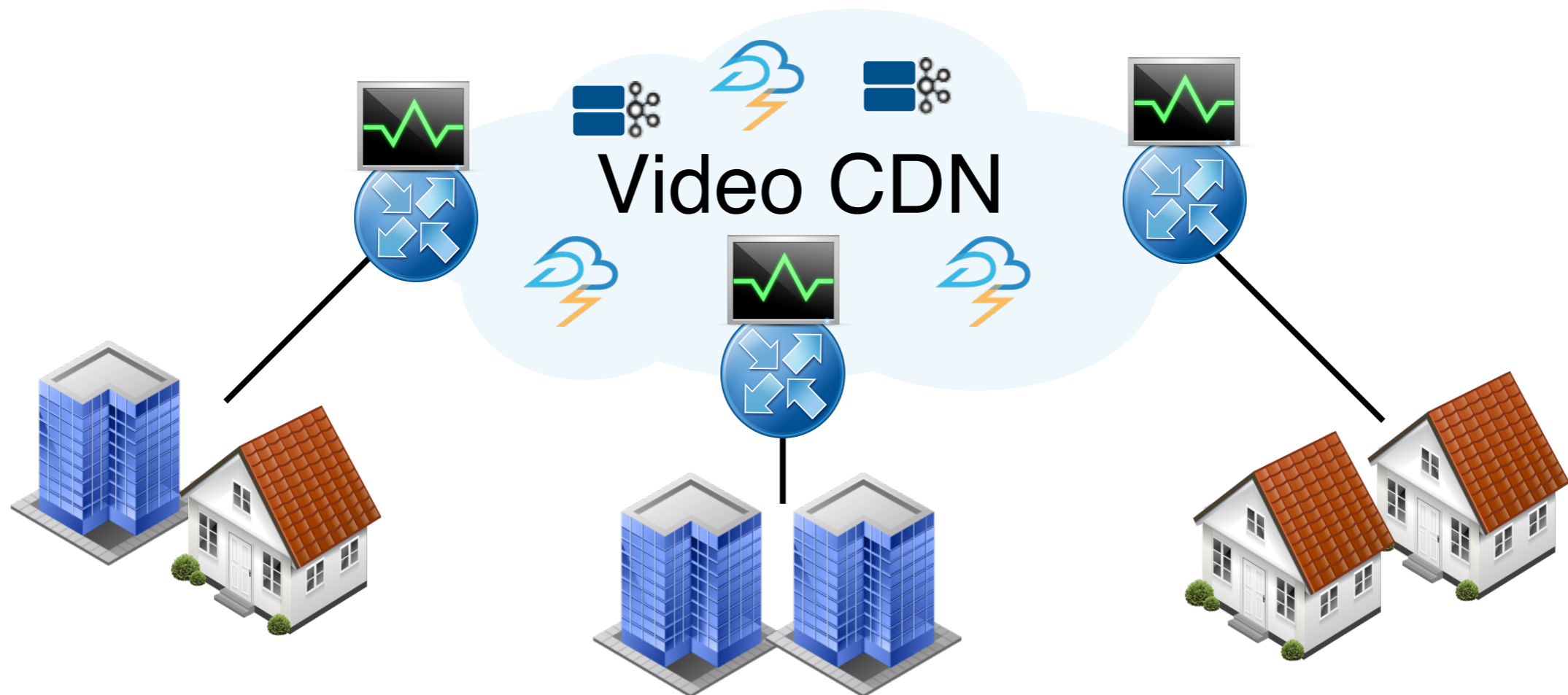
# Live Popularity Detection and Caching

- . What is the most popular content *right now*?
- . How to allocate cache resources for this content?



# Live Popularity Detection and Caching

- . Monitors parse HTTP requests for requested URL
- . Processors use **top-k** streaming analytics algorithm
- . Feedback video popularity to replication system



# Conclusion

- . NetAlytics is a platform for real time analysis of data center networks
  - SDNs redirect certain traffic flows so they can be monitored
  - Monitors use NFV to provide efficient, software measurements
  - Aggregators and Processors provide scalable message queuing and real time stream processing
- . On-going work
  - Developing a query language that automatically deploys the monitors and processing elements
  - Combining the real-time analytics with a cloud manager