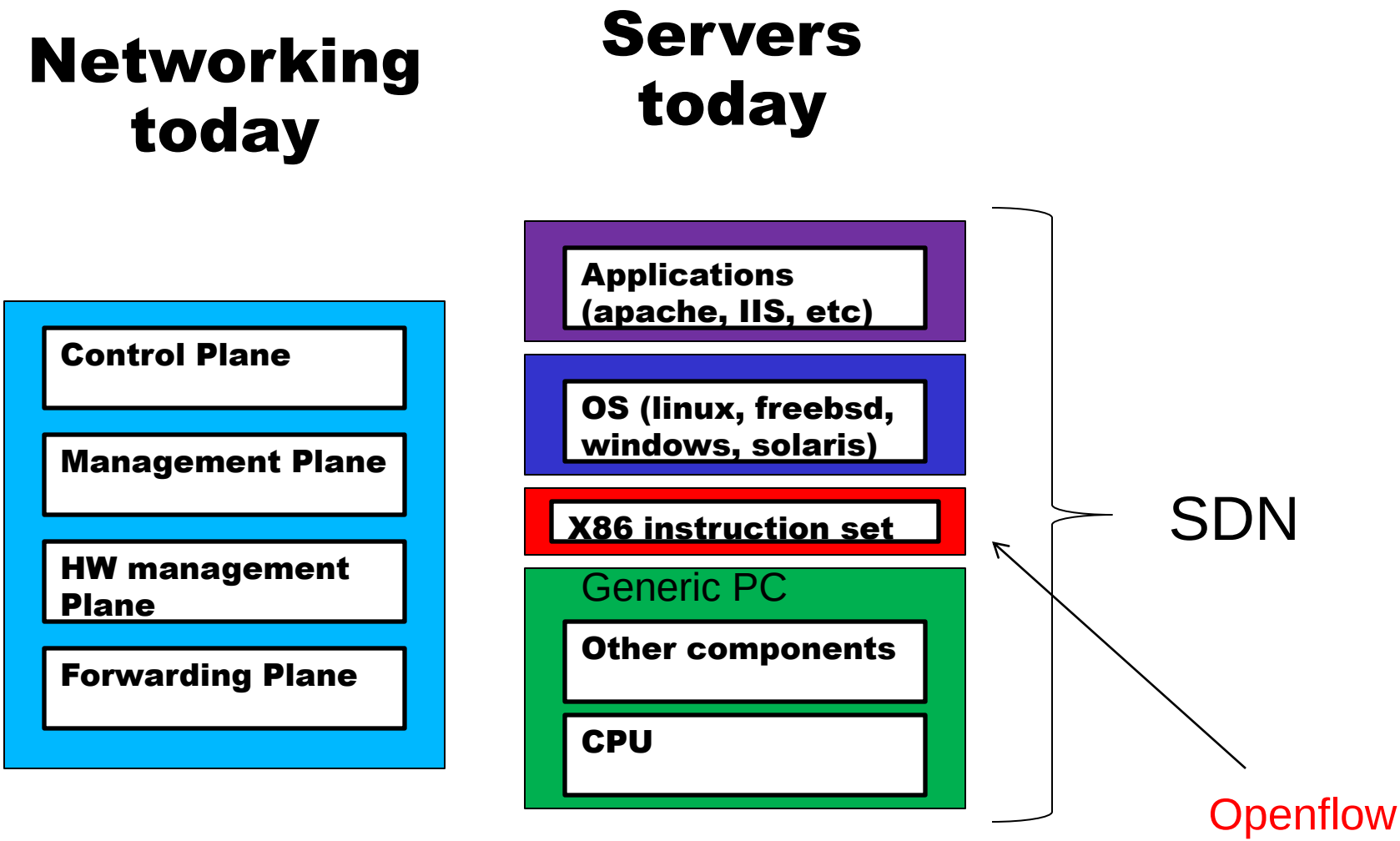
SDN in Warehouse Scale Datacenters v2.0

Igor Gashinsky
igor@yahoo-inc.com
Principal Architect
Yahoo!
April 17th, 2012

Some Terminology

What is SDN and OpenFlow



What stayed the same?

Datacenter Virtualization

Why SDN for Virtualization?

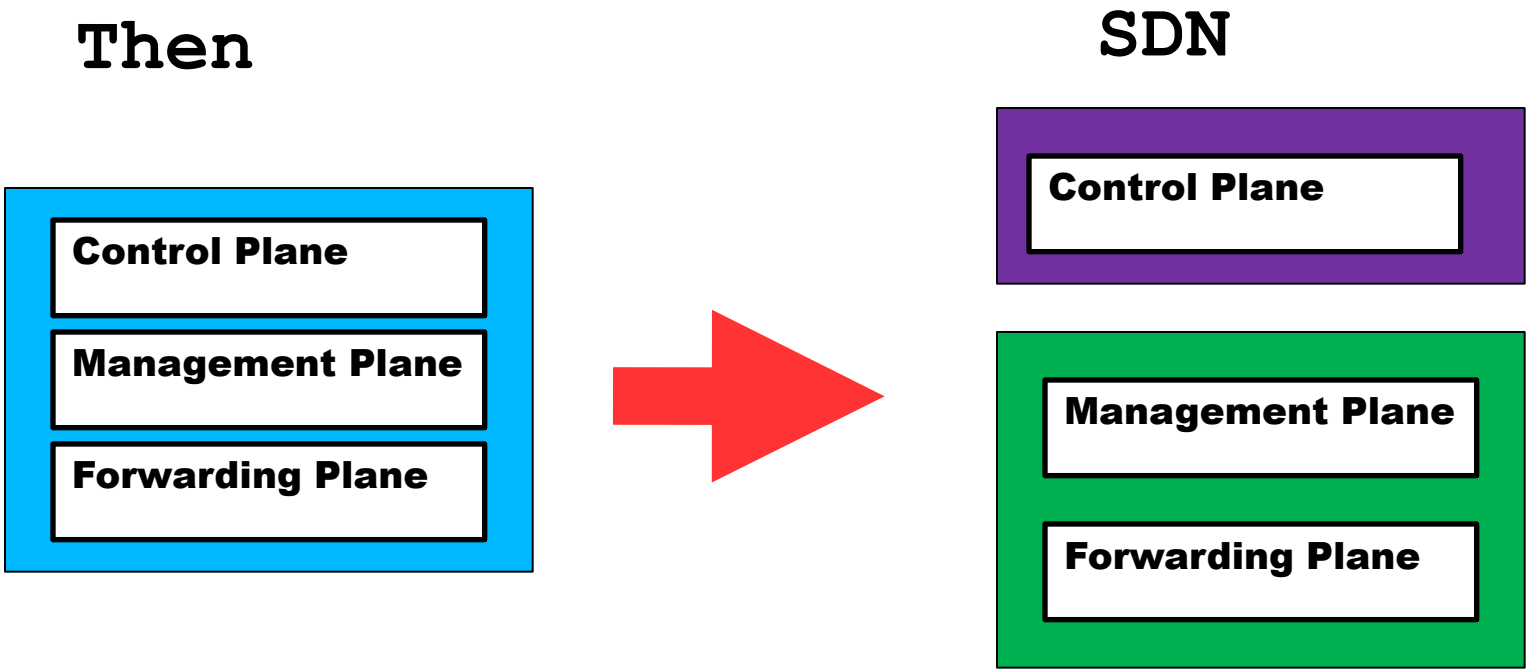
- **Requirements:**
  - 20k servers per cluster = 400k VM's
  - Full any to any communication
  - Place any VM anywhere, anytime
  - VM migration (sub-second)
  - Guaranteed Consistency Model

- **Problem:**
  - How do you keep 20k devices in sync w/ 400k+ entities each?

- **Solutions:**
  - Current hardware can't keep up with FIB requirements
  - Current routing protocols don't do this well
    - Lack of a consistency model
  - Flood and (s)pray doesn't work very well!
  - **Program the vSwitch from a central, distributed database!**

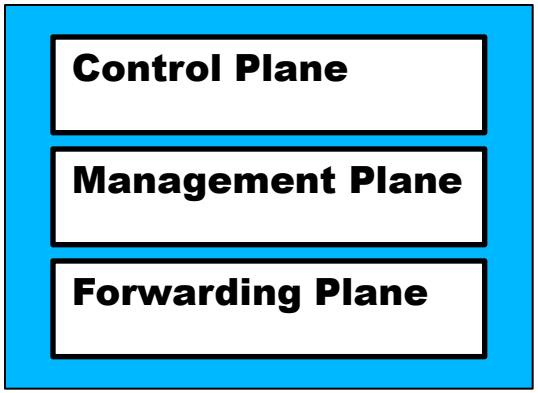
How has the market & vision evolved?

Predictions from 6 months ago:

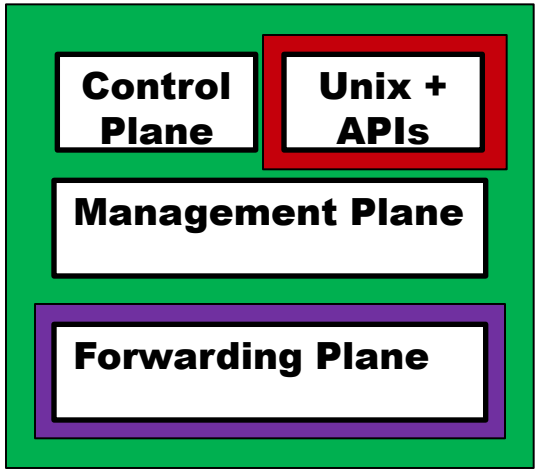


What we see now

Then

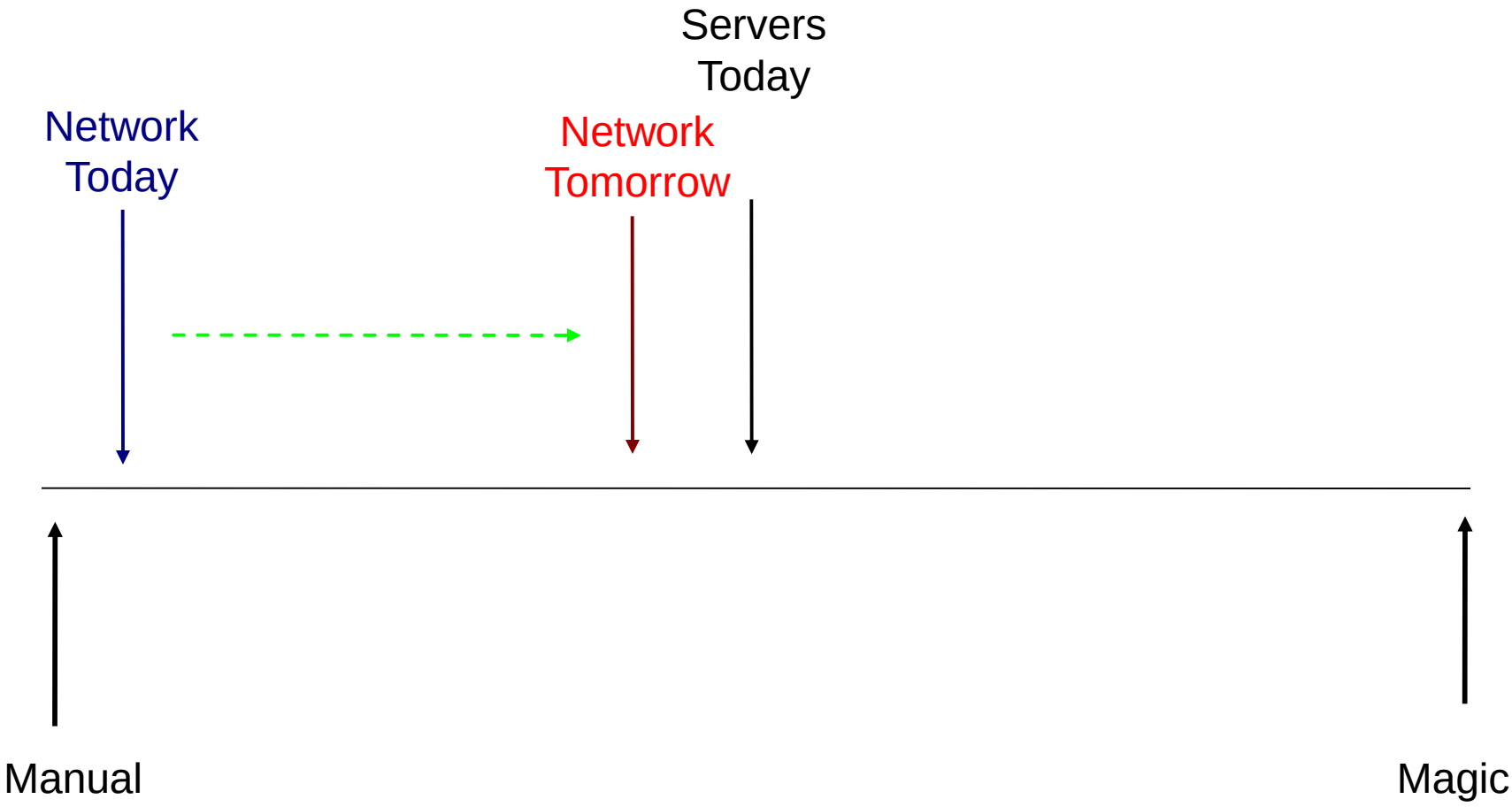


SDN



Why is that important?

Configuration & Deployment Automation



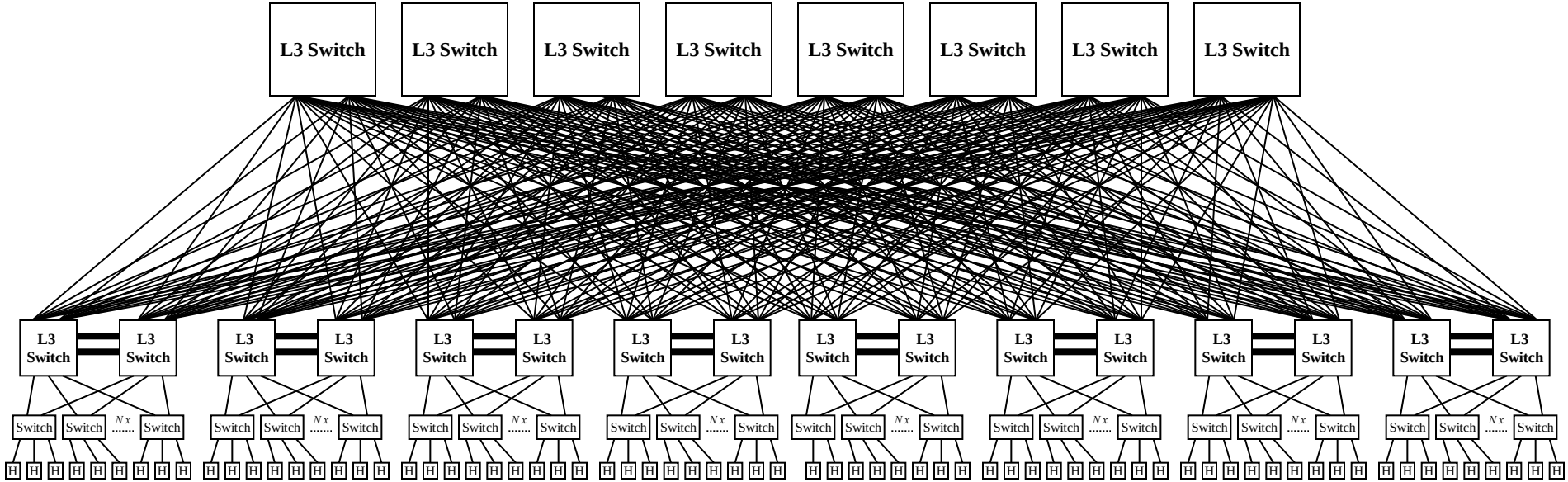
Self-Healing Fabrics (and a pony!)

Blast from the past (Y! Presentation to HSSG in 2007)

** 8 way ECMP w/ 2x10GE LAGs**

** Way too many paths **

** Way too many cables **



L3 Switch <10GE> L3 Switch
L3 Switch <GE> Switch
Host <GE> Switch

Today's Topologies

- **CLOS-like**
  - **20k server cluster ~= 16k internal links**
    - This means upto **1024** distinct links between a pair of hosts
    - How do you troubleshoot this (for packetloss, etc)?
      - # of links to test = 1024/2 = 512
      - 30 seconds/test
      - **256 man-minutes for most-basic troubleshooting!**
  - **Is that acceptable?**
  - **Really? :)**

Enter SDN

- **Local Testing Agent**
  - Looks at interface counters (errors, etc)
  - Performs interface/RIB/FIB health checking
- **Local Repair Agent**
  - Perform local repairs (ie FIB consistency check)
  - If certain conditions are met, automatically remove failed link(s)
  - If unsure of a safe action, ask the controller
- **Global Controller**
  - Has full visibility of the entire network
  - Can initiate repair/fixup actions of it's own

Where's my pony?

Thank you!

Questions?

igor@yahoo-inc.com