Big Data: Not Just Big, But Different - Part 2

Cisco IT Insights Series - Network Architecture for Internet of Everything

This five-part series describes a new network architecture for collecting and analyzing big data in the Internet of Everything era. You can read the parts in any order. Part 1 explains the role of the network in the Internet of Everything era. Part 2 summarizes how big data is different from other data (volume, variety, and velocity). Part 3 describes the seven types of network intelligence to act on big data. Part 4 presents use cases for analyzing network data with big-data techniques. Part 5 presents a four-layer architecture: resource, distributed repository, processing, and application.

What You Will Learn

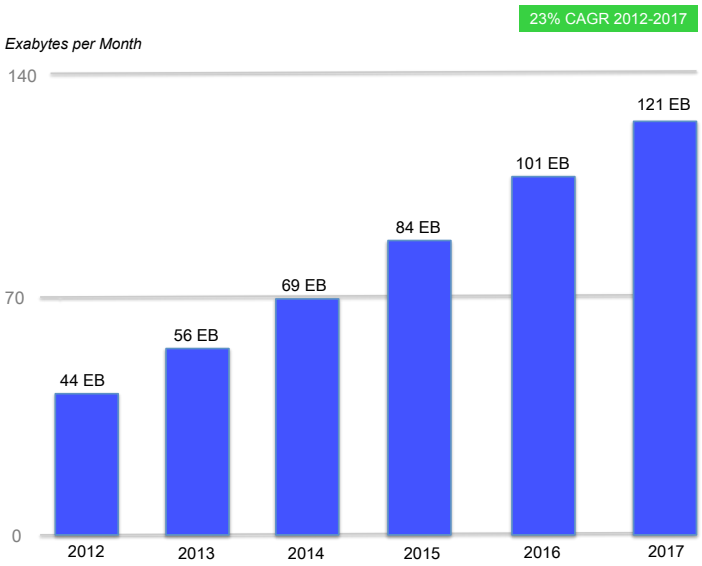
The first step in designing a network for the Internet of Everything era is to understand the traffic it carries. Big data is fundamentally different from traditional data in three ways: volume, variety, or velocity.

Volume

Volume is the "big" aspect of big data. The Cisco enterprise network generates about 1TB of logs and 4TB of security data every day. And that's just us. According to research by Cisco and industry analysts:

- Total IP traffic will grow to 120.6 exabytes monthly by 2017 (Figure 1).
- Global mobile data traffic will reach 190 exabytes a year in 2018. The estimate includes 42 trillion images and 4 trillion video clips.
- Data centers will host 7.7 zettabytes in 2017. That's the equivalent of one three-hour high-definition video every day for everyone on the planet.
- Sixty-four percent of data will be in the cloud in 2017.

Figure 1. Projected IP Traffic Growth Through 2017 (source: Cisco VNI)



Variety

An estimated 90 percent of all data is semi-structured or unstructured. Examples include social media, mobile communications, customer service records, warranties, census reports, sensors, phone records, email, and web logs.

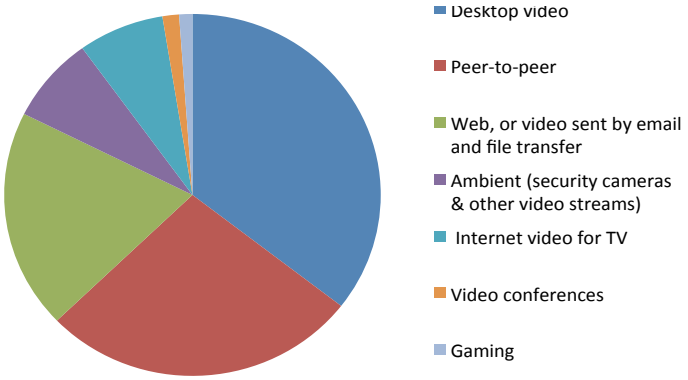
Table 1. Intelligent Switches Detect and Respond to Events in Real-Time

Structured Data (10% of All Data)	Semi-Structured or Unstructured Data (90% of All Data)
Databases	Microsoft Office documents
XML data	Email messages
Data warehouses	RSS feeds
Customer Relationship Management systems	Audio files
Enterprise Resource Planning systems	Video
	Social media data
	Sensor output

Unstructured documents typically contain a lot of text, and they sometimes contain dates, numbers, and facts. This data is harder for computer systems to understand than data that is stored in database fields or semantically tagged in documents.

A growing portion of unstructured data is video. Video represented about 70 percent of all Internet traffic by the end of 2013. Figure 2 shows the types of video as a portion of the total. We project that mobile video will soon account for 70 percent of all mobile traffic.

Figure 2. Types of Video Traffic as a Portion of the Total



Velocity

Velocity refers partly to how fast the Internet of Everything produces big data. Twitter users generate nearly 100,000 tweets every 60 seconds. During that same minute, Facebook users post 510 comments, update 293,000 statuses, and upload 136,000 photos (source: The Social Skinny). Network infrastructure also generates torrents of data, and sensors connected to the Internet of Everything can produce multiple terabytes a day.

Velocity also refers to the speed at which the network needs to analyze big data to make a decision. For instance, an intelligent switch needs to respond to unusual network activity in real-time in order to prevent or minimize security breaches. Similarly, online advertisers need to very quickly determine which ad is most likely to appeal to you. An eye blink's delay might mean you've left the page, and the advertiser has lost an opportunity.

There is not enough time to store, clean, and integrate big data before using it in analyses. Therefore, systems need the intelligence to quickly figure out what's important and what's not. Extrapolating from historical data is not good enough because the information may be out of date before the calculation is complete. It's more effective to embed predictive analytics in systems that can initiate real-time action.

For More Information

To read other papers in the "Network Architecture for Internet of Everything" series, as well as IT case studies, visit www.cisco.com/go/ciscoit.