Ubitus Cloud Computing with NVIDIA CUDA

ubitus

January 22, 2010

Agenda

Company overview

Challenges to FMC media sharing/ distribution services

How Ubitus technology can improve with CUDA-enabled platform

Case Study

Ubitus Inc.

Company Fact

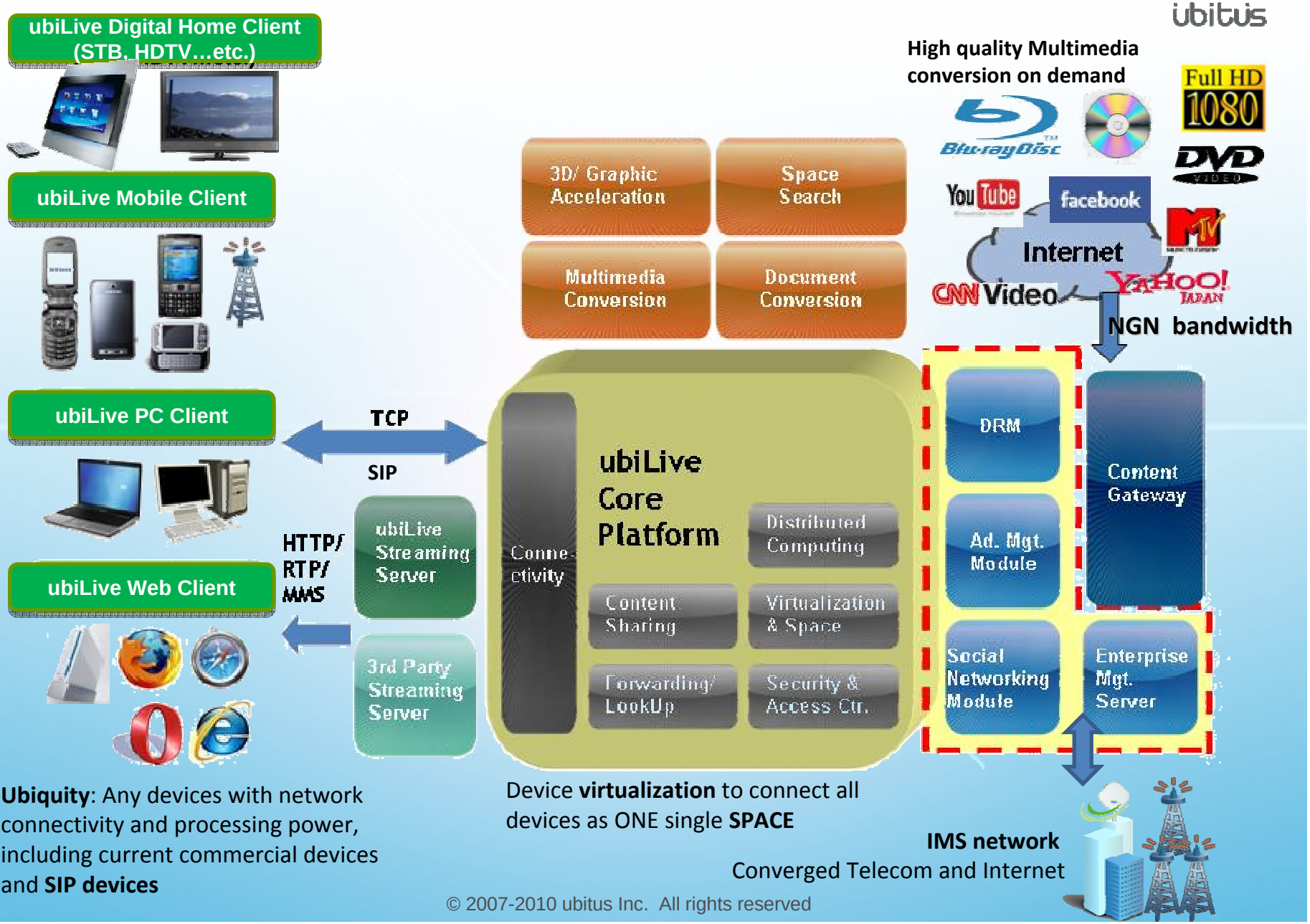
- Ubitus Inc. , a leading software platform provider for fixed-mobile convergence applications, develops a advanced cloud computing solution for telecom operators and service providers to drive consumer adoption of premium services relevant to sharing live video among friends as well as accessing digital media anytime, anywhere, without being restricted by the devices in use.

- Established in May 2007, the founding & leadership team is made up of members from
  - successful entrepreneurs who previously founded one embedded software company and merged with Japan listed company
  - professionals with deep industry and market knowledge and track records of delivery from international software company, management consulting, to investment banking
  - innovative engineering team with proven technology visions from several awards

- More than 65 employees (42 engineers with over 70% are PhD and master degrees focusing on Distributed Computing, Computer Networking, Multimedia and Embedded Systems)

- Office Locations: Taiwan, China, Korea, and Japan (in preparation)

Ubitus Product Overview



Ubiquity: Any devices with network connectivity and processing power, including current commercial devices and **SIP devices**

Device **virtualization** to connect all devices as ONE single **SPACE**

Multiple Mobile Platforms Support

Today, ubiLive is already available for major handset platforms in the market

- JavaME JTWI v1.0
- Windows Mobile v6.1/6.5
- Android v1.0/1.5
- iPhone v2.0/v3.0
- Symbian S60 v3.0/4.0

And will support other platforms soon

- Limo Linux v2.6/GTK+
- BREW v4.0

Agenda

Company overview

Challenges to FMC media sharing/ distribution services

How Ubitus technology can improve with CUDA-enabled platform

Case Study

Technology is democratizing content creation & experience sharing

Impact of new technology

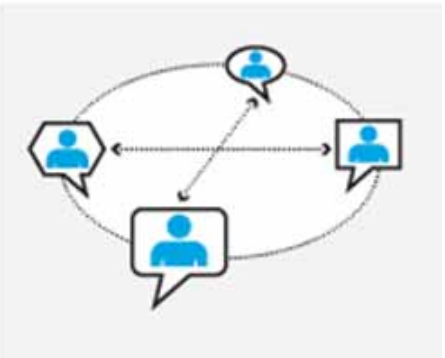
Content creation
(Cost of digital cameras / phones and editing SW Declining)

Content storage
(Not limited by physical shelf space, and storage costs are declining.)

Content distribution
(Growing no. of distribution channels (thru Internet). Bandwidth costs declining.)

Proliferation of User-Generated Content

Yesterday	Today	Tomorrow
Device-centric: Synchronized (Content is localized)	**Individual-centric:** connected (Content follows me)	**Interest-centric:** Ubiquitous community (My content determines my community)



Difficult to share content/resources among communities from various combination of heterogeneous networks & devices

Device	OS platform	Codec	Network	Media storage
Feature phone	Windows/WMP	3GP	GSM/EDGE/WCDMA	Hard disc
Smartphone	Symbian	FLV	HSDPA/ HSUPA	Flash/ SSD
UMPC	BREW/REX	H.263/H.264	CDMA 1x/ EV-DO	VCD/ DVD ROM
PC/ laptop	Linux	DivX	WiFi/ Mesh WiFi	HD/ Blu-ray Disk
Game console	Mac OS	AVI	WiMAX/ WiBRO	Internet space
Set-top-box	Nucleus	MPEG 2	DLNA/ UPnP	⋮
PND/GPS	⋮	Quicktime	⋮	
⋮		⋮		

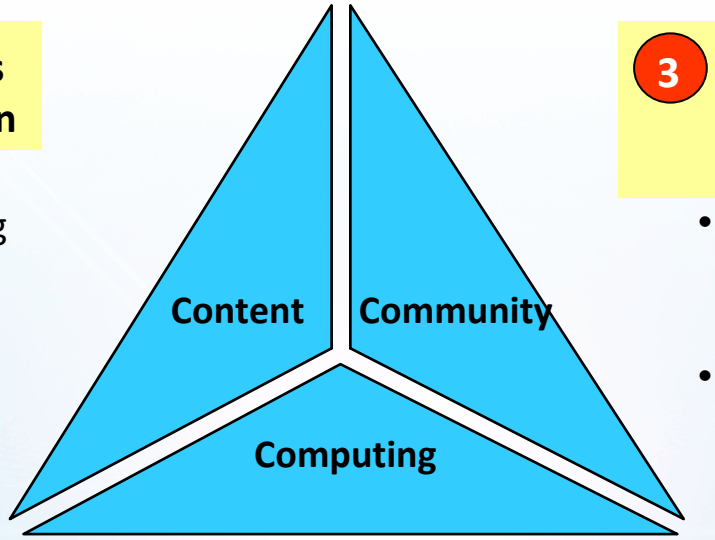
With the combination of various device, OS platform, codec, networking, media storage, and other factors, it makes user very difficult to just simply share/play/create internet content or their own data to different groups or communities

Challenges to FMC service over heterogeneous networks/devices

ubitus

1 Diversified content formats and storage source/location

- Various content formats and codecs makes content sharing a difficulty
- Device compatibility & hardware limitation prohibits the capability to play/ create contents for others
- Storage capacity of portable devices or mobile phones limits the easy access of rich media contents from open internet or online user generated content (UGC)



2 In-balanced computing power of mobile, PC and CE devices

- Limited CPU speed due to the compromise of battery life and computing power of mobile phone
- Diversified hardware difference among ARM, x86, MIPS, and GPU makes the performance of devices variant more than 100 times!
- Cannot leverage existing powerful “Cloud Computing” for scalability

3 Difficult to share content or resources among various communities

- Separated storages of content from HDD, Flash, DVD ROM, to the Internet
- Isolated user communities due to the disconnect among mobile phone, PC and CE devices without proper standard
- Fundamental differences of GSM/CDMA technology compared with IP-based network like WiFi/ WiMAX creates the barrier of communication for user community

Agenda

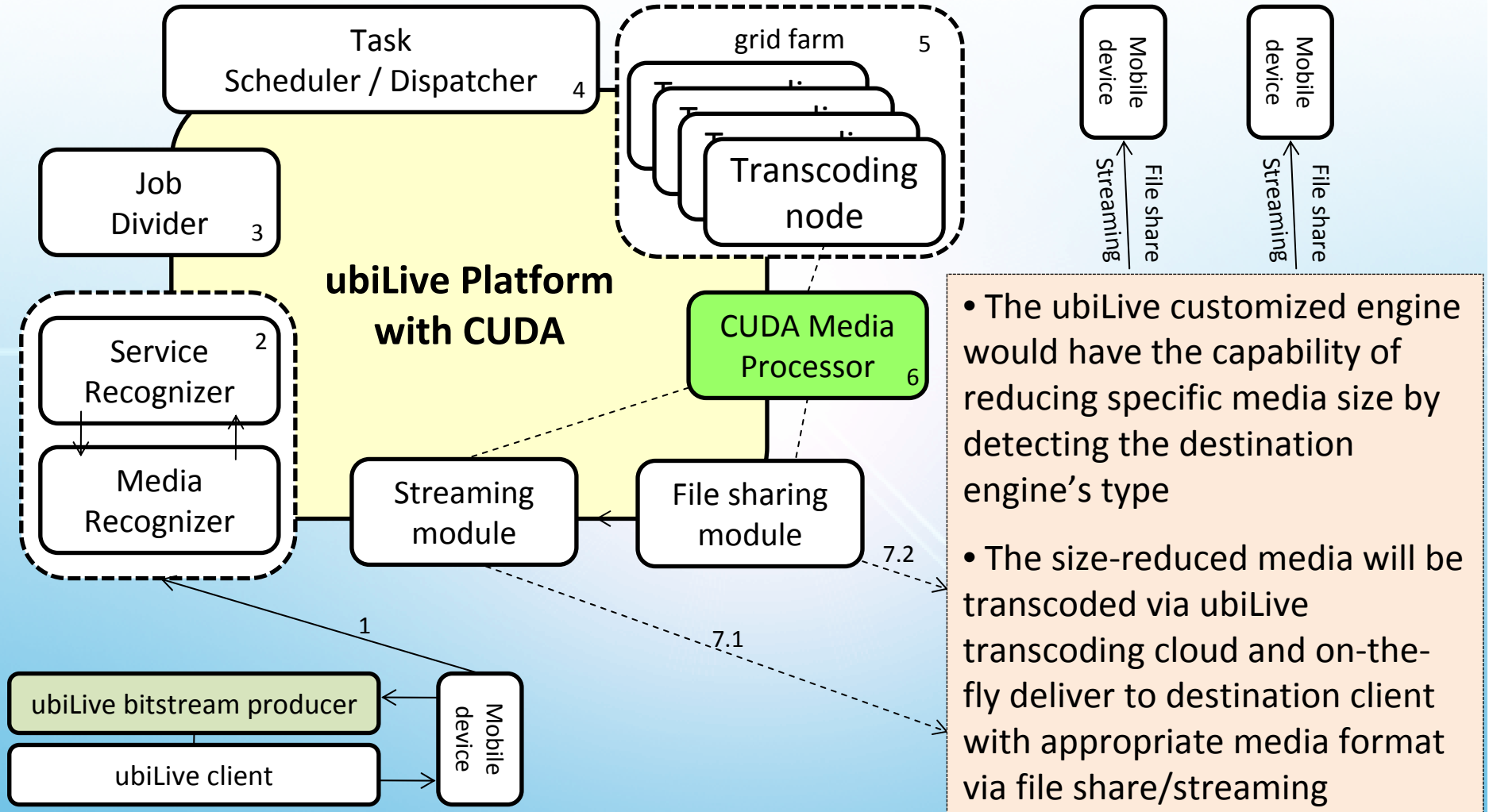
Company overview

Challenges to FMC media sharing/ distribution services

How Ubitus technology can improve with CUDA-enabled platform

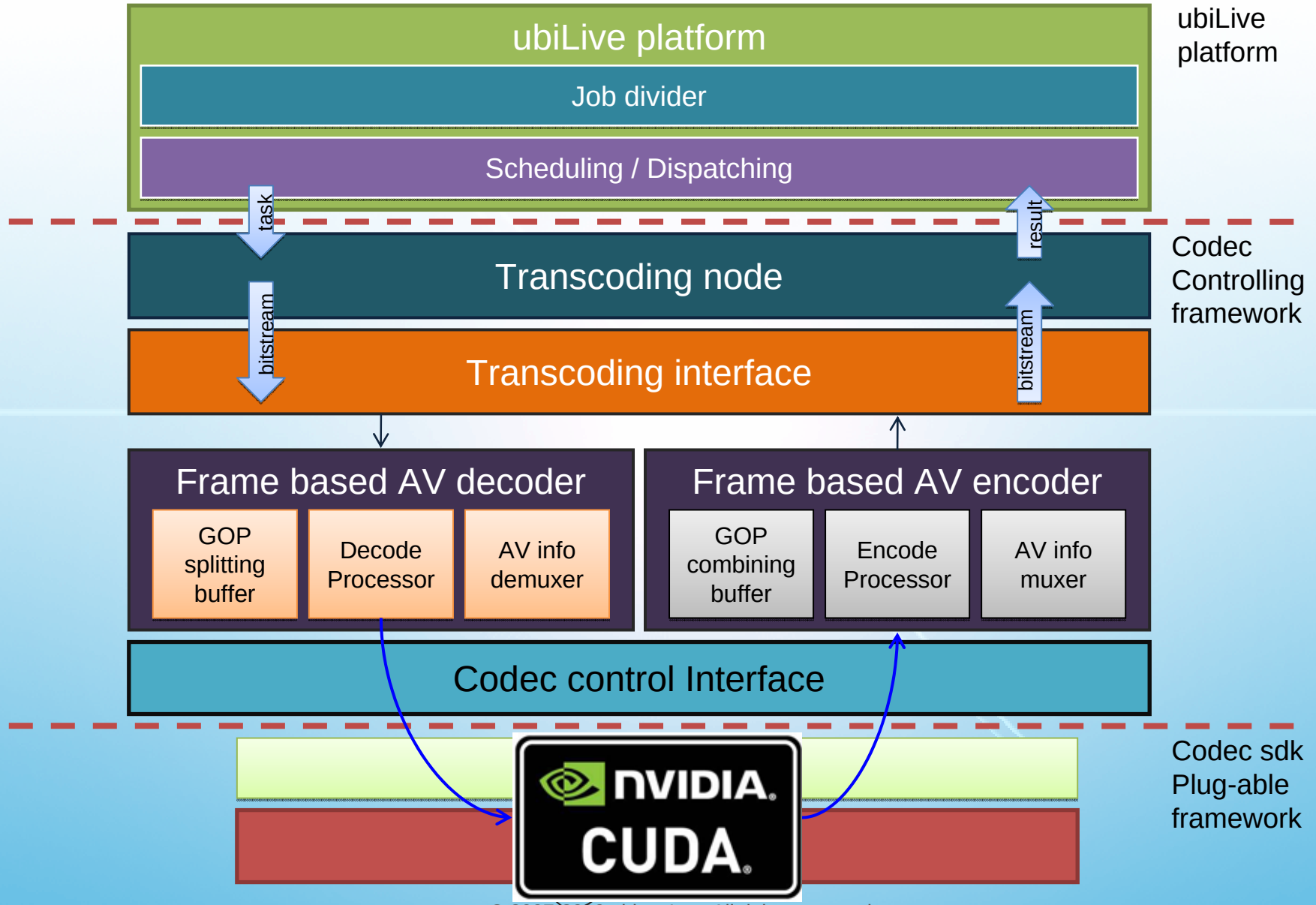
Case Study

ubiLive Cloud Platform with CUDA-enabled

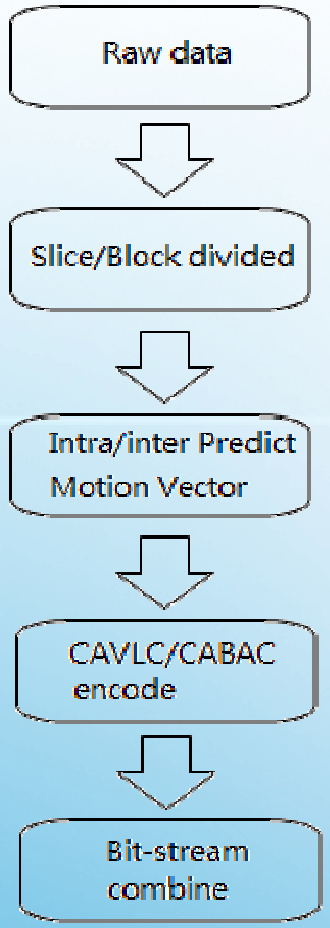


- The ubiLive customized engine would have the capability of reducing specific media size by detecting the destination engine's type
- The size-reduced media will be transcoded via ubiLive transcoding cloud and on-the-fly deliver to destination client with appropriate media format via file share/streaming

Faster Transcoding with nVIDIA CUDA Acceleration

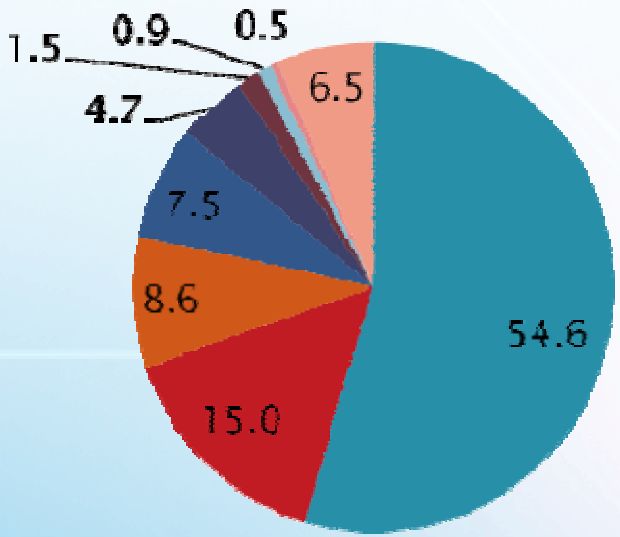


Typical Data Flow in H.264 Encoding



- Raw data
  - An image in M x N pixels
- Divided
  - Dividing image into slices and blocks
- Predict
  - Find the similar blocks/sub-blocks
- Entropy
  - Compute the residuals with transform
  - Encode the residuals (VLC)
- Bit-stream
  - Combine the headers and bit-stream into file

Computing Overhead in H.264 Encoding

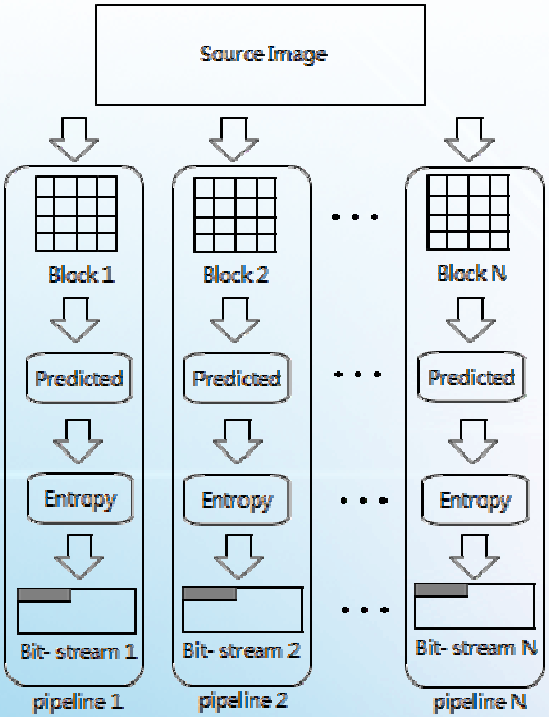


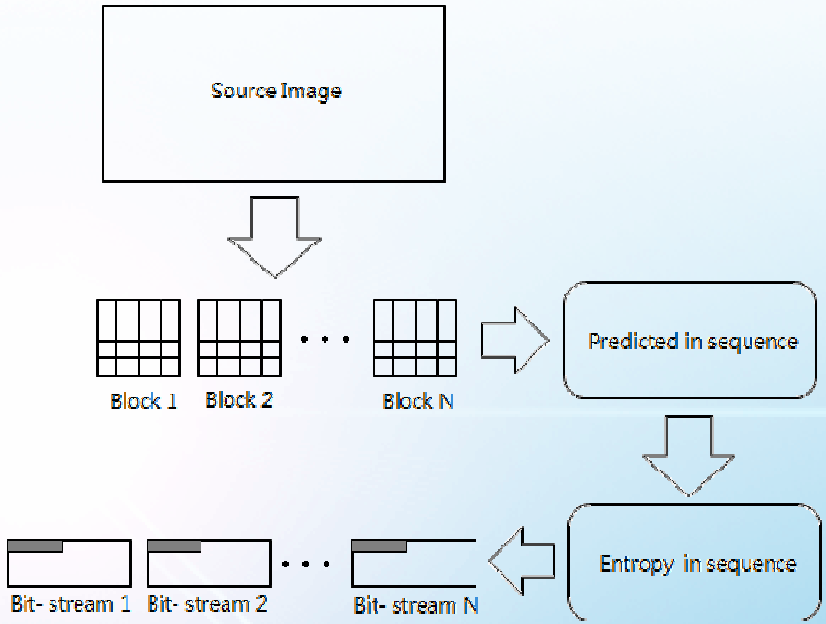
- Inter-predict
  - SAD (54.6%)
  - Sub-pixel interpolation (15%)
  - Sub-pixel MV prediction (8.6%)
  - SAD reduction (7.5%)
  - Intra-prediction (0.9%)
- Entropy
  - Residuals (4.7%)
  - CAVLC (1.5%)
- Others

•Predict(86.6%) + Entropy(6.2%) + Others(7.2%) = all operations (100 %)

[1] Lawrence Chan, Jae W. Lee, and Alex Rothberg. **Parallelizing H.264 Motion Estimation Algorithm using CUDA, *MIT*** *in IAP 2009*
[2] Karsten Shring. H.264/AVC Software Coordination. *http://iphome.hhi.de/suehring/tml/.*

CUDA v.s. General Purpose CPU





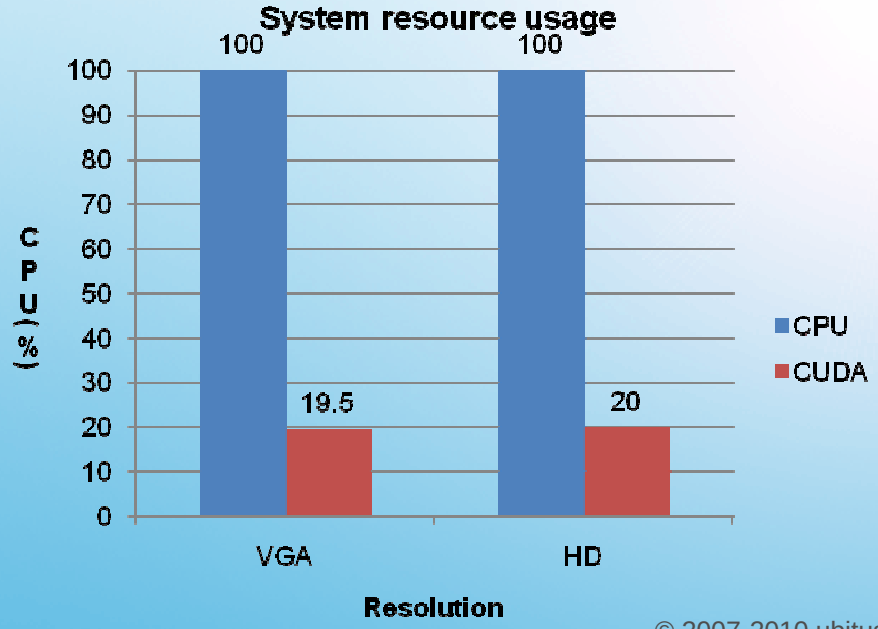
CUDA	General purpose CPU
•Multiple pipelined arch •Parallel computing •Full optimization (floating arithmetic) •Much accuracy, quality promise	•Single pipelined arch •Optimized only with special machine instruction set (MMX/SSE2,3,4) •Less accuracy (integer arithmetic)

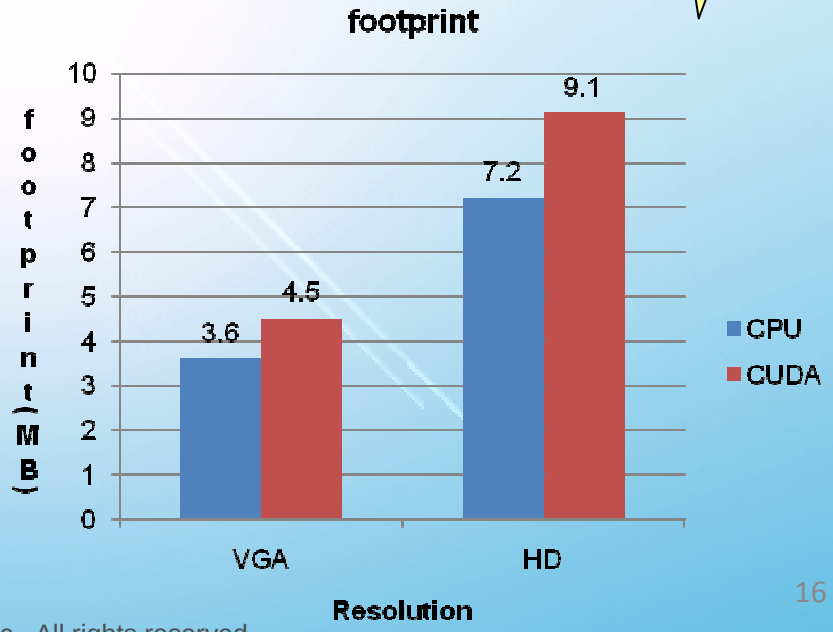
Overall Throughput Comparison

based on current experiment result, still has room for more optimization

Resolution		CUDA	General purpose CPU
VGA	predict	5.1ms/frame	64.2ms/frame
	entropy	0.8ms/frame	7.0ms/frame
HD	predict	23.5ms/frame	174.6ms/frame
	entropy	2.1ms/frame	15.3ms/frame

6~12 times faster





10 technologies Ubitus can enhance FMC richmedia services

1. **Better transcoding performance** with faster response time
2. **Dynamic bitrate** and resolution transcoding
3. **Micro transcoding** on mobile for better QoS and 3G network transmission
4. **Upscaling** of video quality (distributed super resolution)
5. **Flexible multimedia framework** to add/upgrade or swtich different codecs no matter software, hardware (DSP or GPU), or hybrid modes.
6. **Phone-to-phone sharing** without cable, easily switch between 3G & WiFi environment
7. Content gateway with **universal video search** and on-the-fly transcoding
8. **Browser widget** (i.e. AJAX support), it will allow consumers to access the service not only just from PC browser, but also mobile browser or TV browser.
9. Inner **devices virtualization** and content search
10. **Comminuty function** compatible with major IM softwares, like Yahoo messanger, Google Talk etc.

Agenda

Company overview

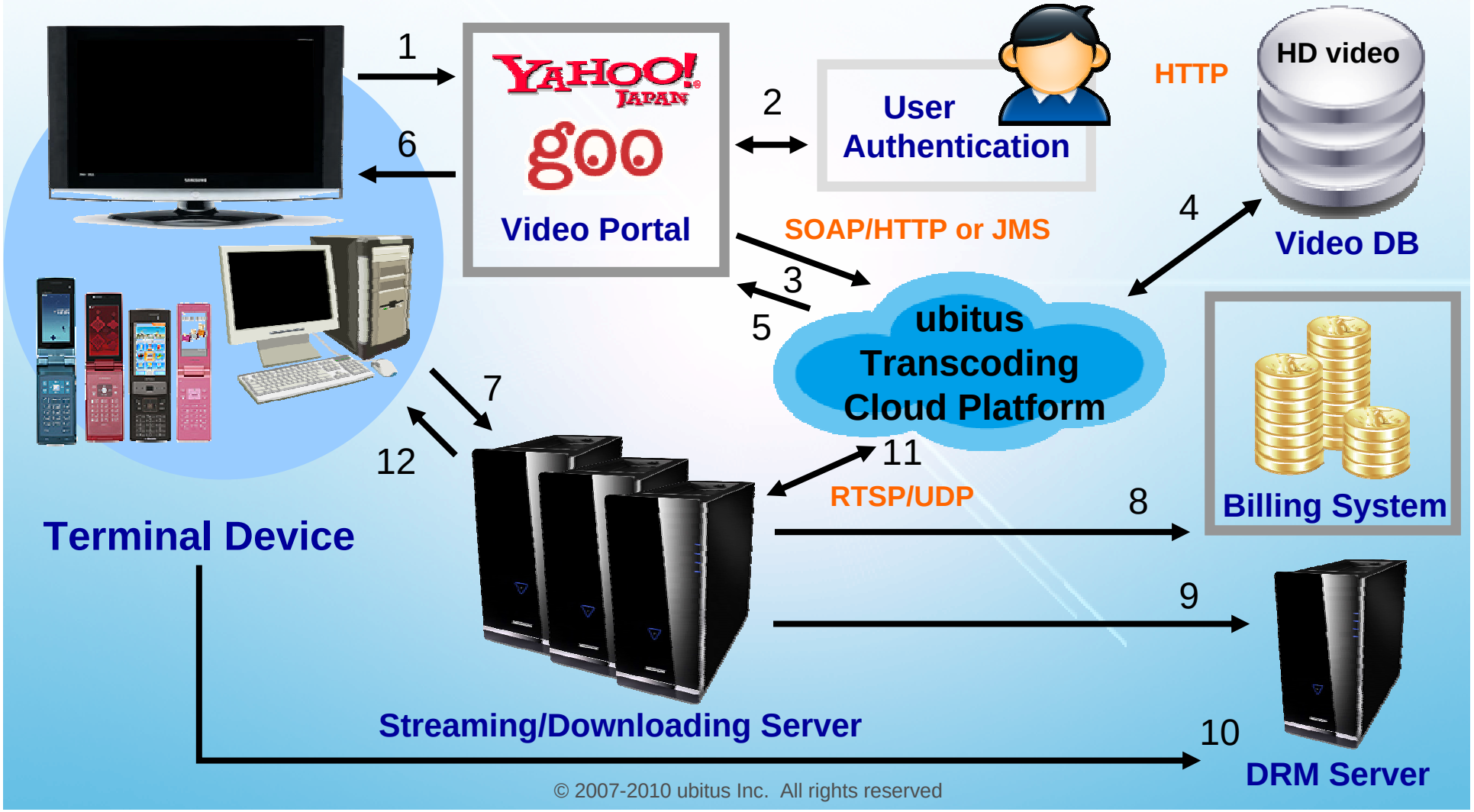
Challenges to FMC media sharing/ distribution services

How Ubitus technology can improve with CUDA-enabled platform

Case Study

Case study – Video distribution cloud over NGN/LTE

Ubitus builds a service platform using cloud technology to deliver Full HD rich contents from single source DB to multiple terminals with different resolution, codec, and bit-rate.



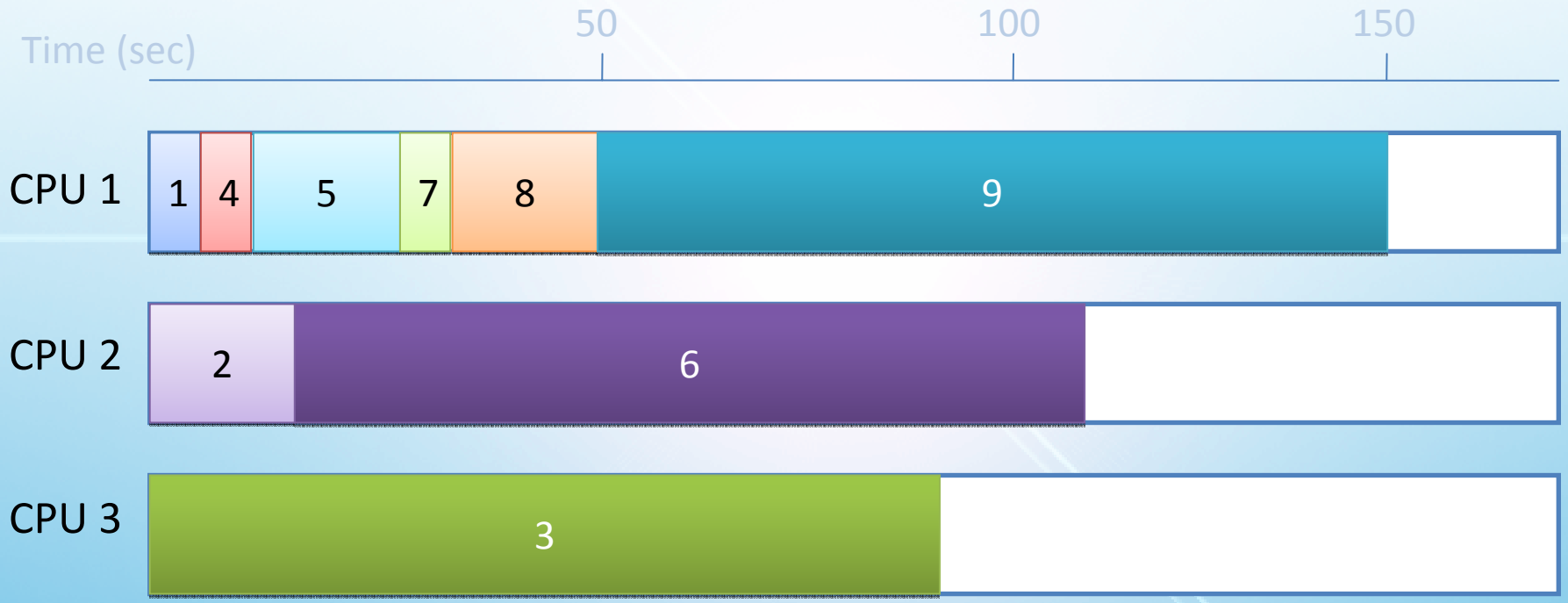

Thank you!

Faster Performance with Cloud Architecture (1/2)

Traditional transcoding –

Concurrent 9 connections invoked

- CPU: 3x transcoding Speed, 3 Videos:
  1, 4, 7: 10sec 2, 5, 8: 60sec 3, 6, 9: 5min



- Average Response Time:
  (3.33+6.66+26.66+30+50+150+20+120+100)/9 = **56.29 sec**
  Longest Response Time: **150 sec**

Faster Performance with Cloud Architecture (2/2)

Cloud computing transcoding –

Concurrent 9 connections invoked

- CPU: 3x transcoding Speed, 3 Videos:
  1, 4, 7: 10sec 2, 5, 8: 60sec 3, 6, 9: 5min

- Average Response Time:
  (1.67*3+3.34*3+5*3)/9 = **3.33 sec**
  Longest Response Time: **5 sec**