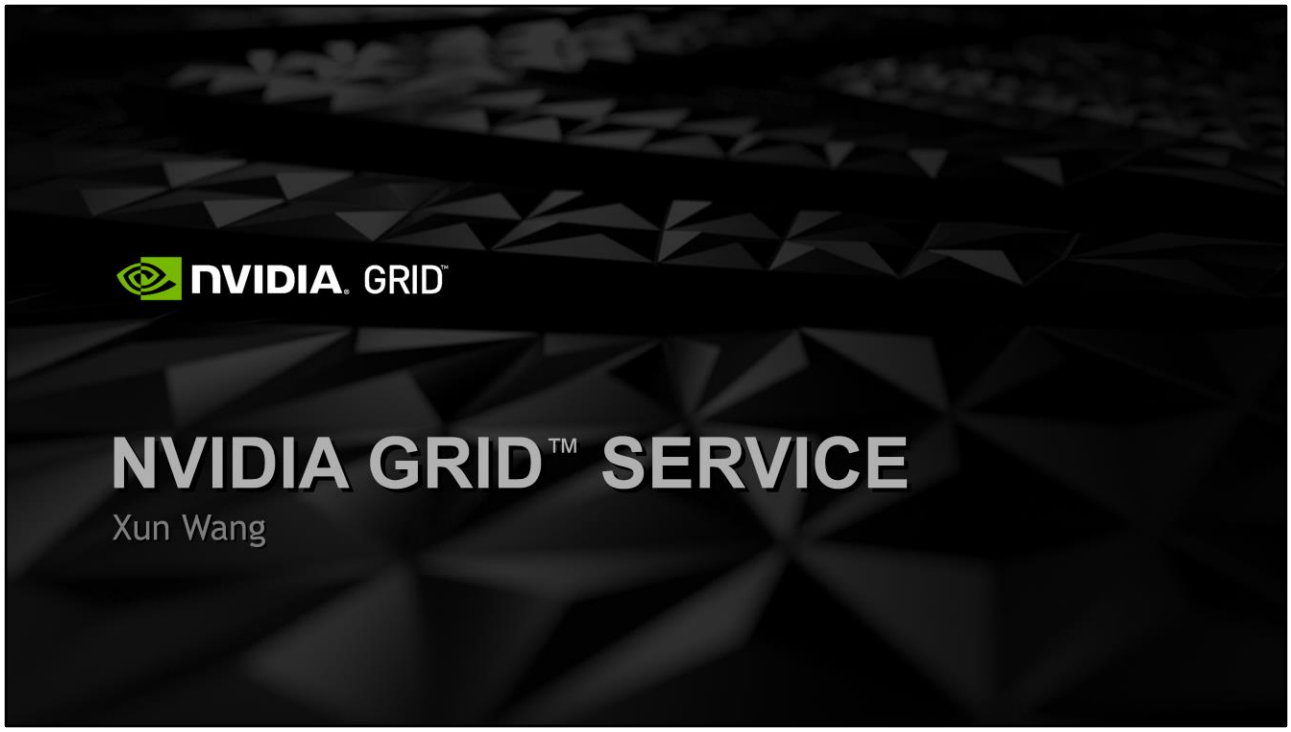
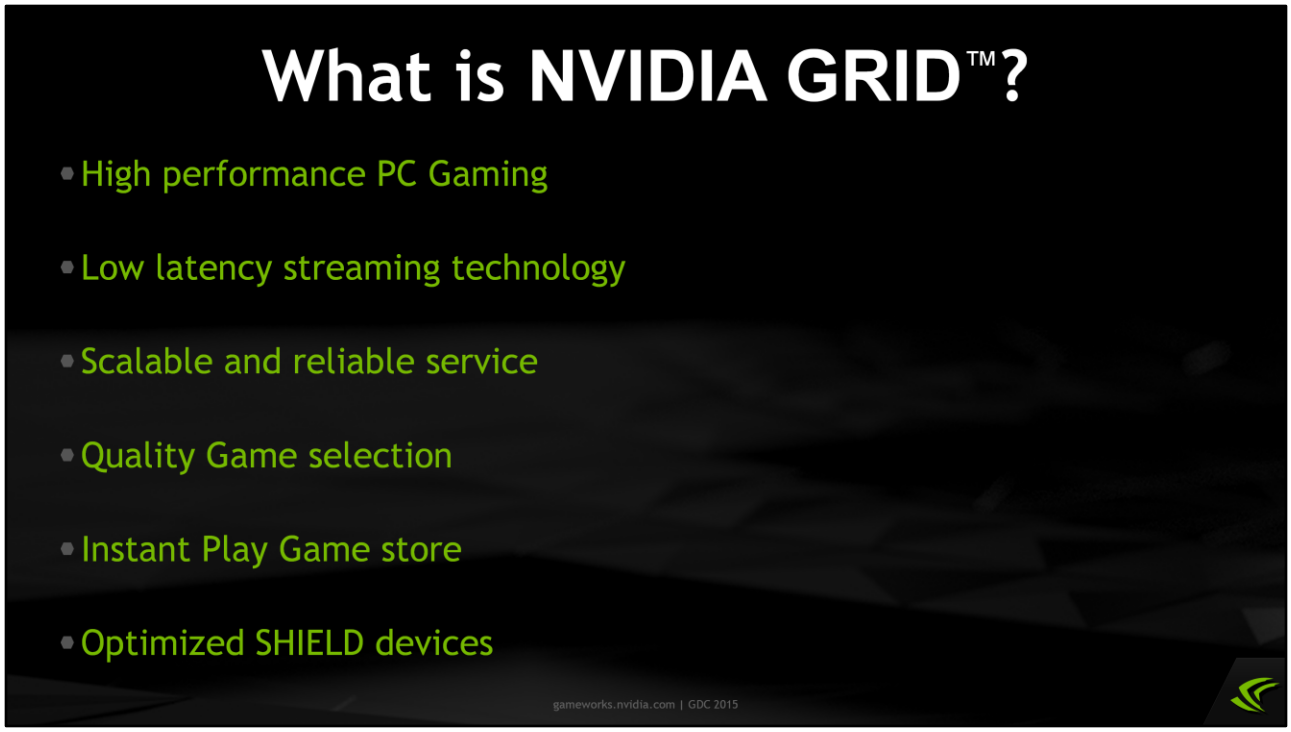



Welcome everyone, my name is Xun and today I'm very proud and excited to be giving you guys a glimpse of the technology behind the Nvidia GRID Service.

First a bit about myself. I'm from Canada originally (anyone here from the cold north?) class of 2001 from Waterloo.. I head up the engineering team that built the GRID Service. I've actually been working at Nvidia for about 15 years now and I consider GRID the most exciting project that I've been apart of.




At it's core, GRID is about high performance gaming; delivering stunning 3d rendered visual quality through the cloud. A system that aligns with the core values of the Nvidia gaming brand. Leveraging all of the greatness and technology that Nvidia has to offer.

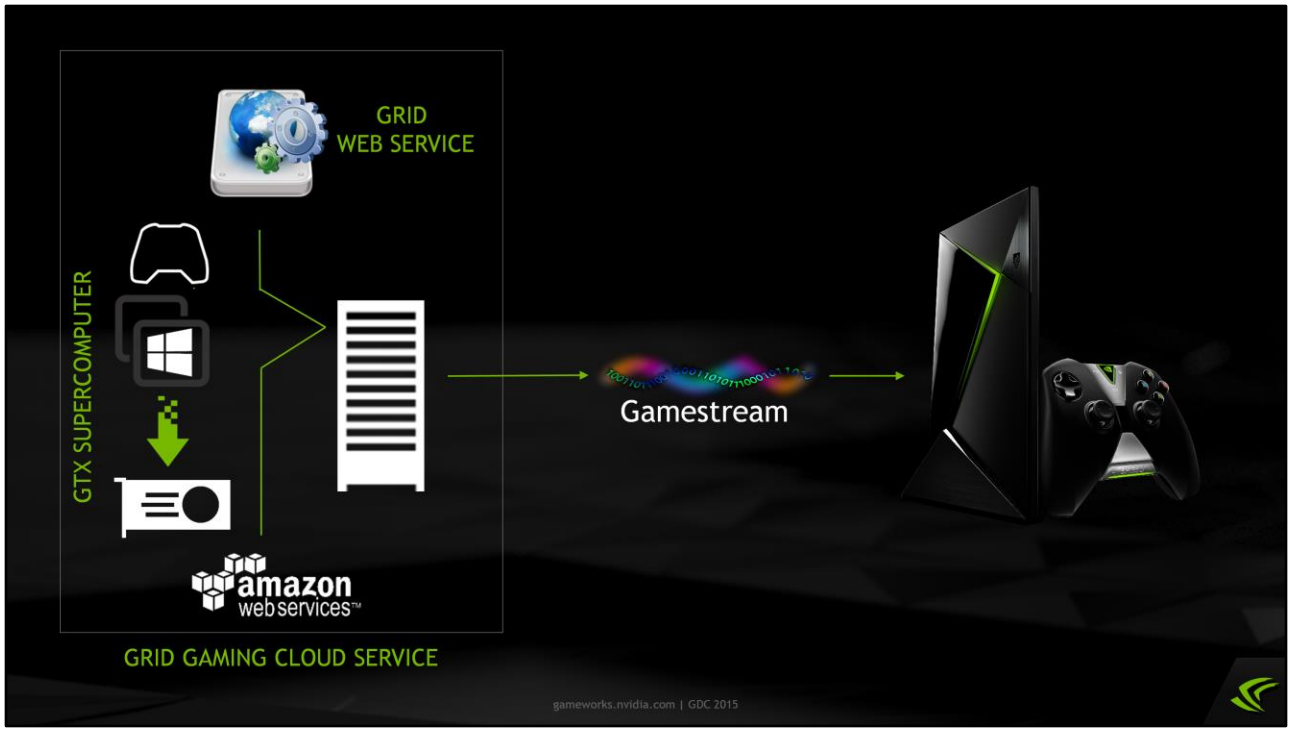
Now bringing that gaming excellence to the cloud required us to develop a highly optimized low latency streaming technology that we call Gamestream

And then merging that technology that together with a scalable and reliable web service, accessible from anywhere on the internet.

When these two components combine; this platform gives us an opportunity to deliver **your** best content in a convenient and beautiful way to the gamers…

For the first run premium content, we provide a way to help publishers monetize by providing an integrated Game Store: buy individual games and play instantly.

And the final piece is the SHIELD devices. Optimized to be the receiver of the streaming content. We've tuned the entire pipeline, from rendering to encoding to decoding and display. And by controlling that, is how we can deliver the highest quality streaming experience.

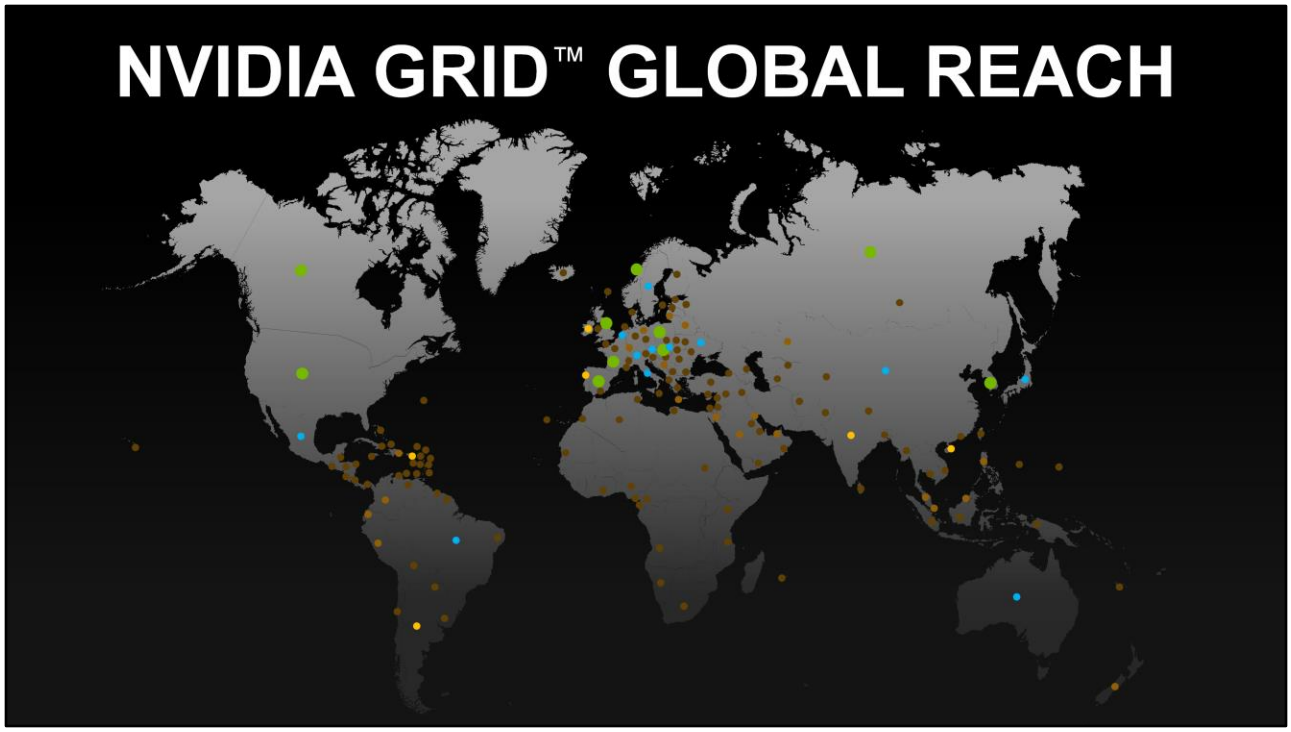



At a high level, this is what our architecture looks like:

At it's core.. our platform consists of specially designed highly performant and highly reliable GRID GPUs, hosted in server class systems. This runs on a custom virtualization SW stack, that enables us to utilize our high quality geforce drivers, and geforce experience

Each game runs in it's own virtualization environment, safe from other tenants on the cloud. A web service ties together a cluster of this rendering power. And our low latency streaming technology that we call Gamestream delivers that to the SHIELD devices.

All of this scaled out, made reliable, monitored and secure on the Amazon web services platform.




Over the past year GRID has reached many players all over the world. We have been learning through data from the system and feedback from users and game development partners to continually improve our service.

Today, we're going to open up the hood a little of our platform.

Explain the technology behind streaming , how we scaled out our service in the cloud. Show some learnings we gathered from the data.

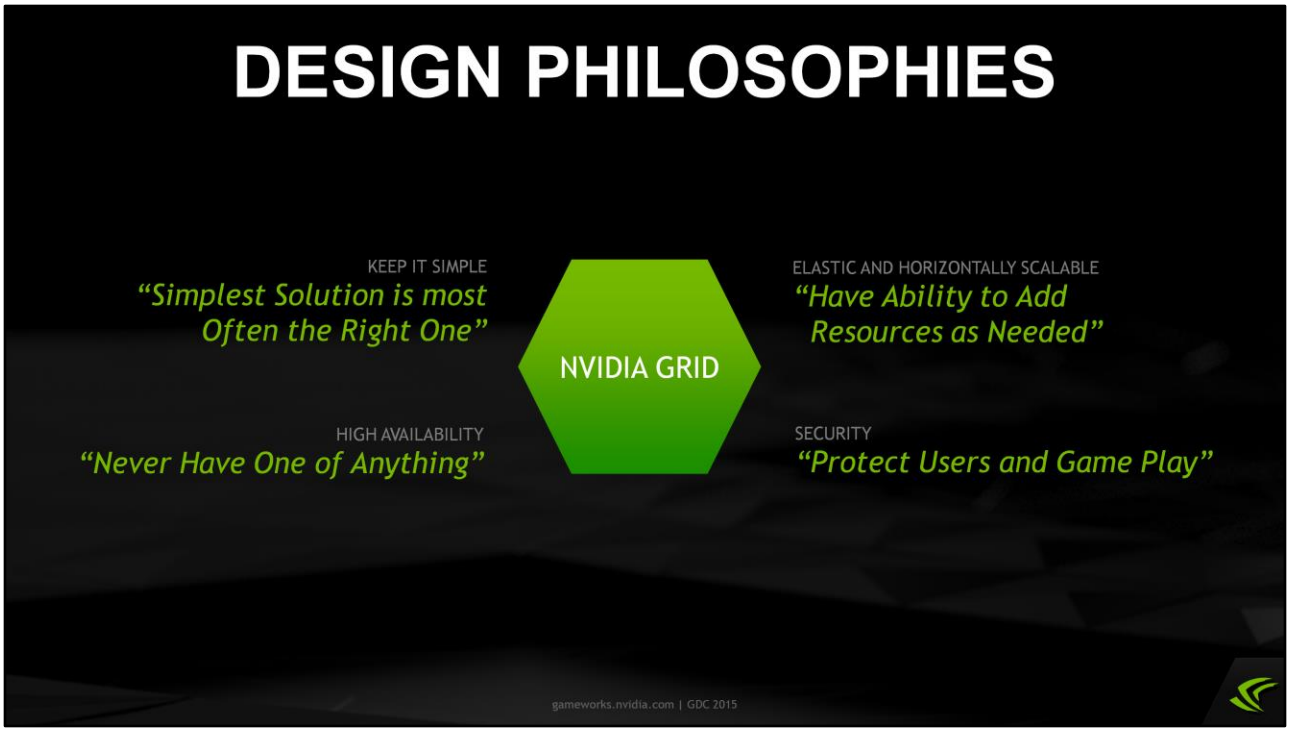
And then go into details of how you guys, the game developers can optimize your games for GRID.

So thank you for attending we are all very excited to be here today.. we will be reserving the questions for the end of this session.

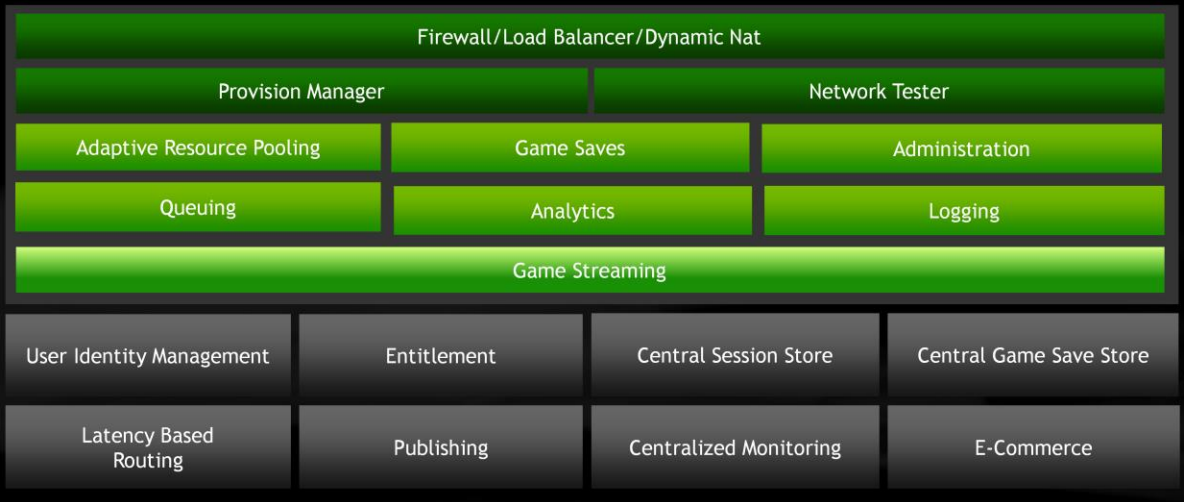
And I'd like to introduce Bojan, our chief architect for the service. He'll be diving deeper into the architecture.


NVIDIA

# DESIGN PHILOSOPHIES



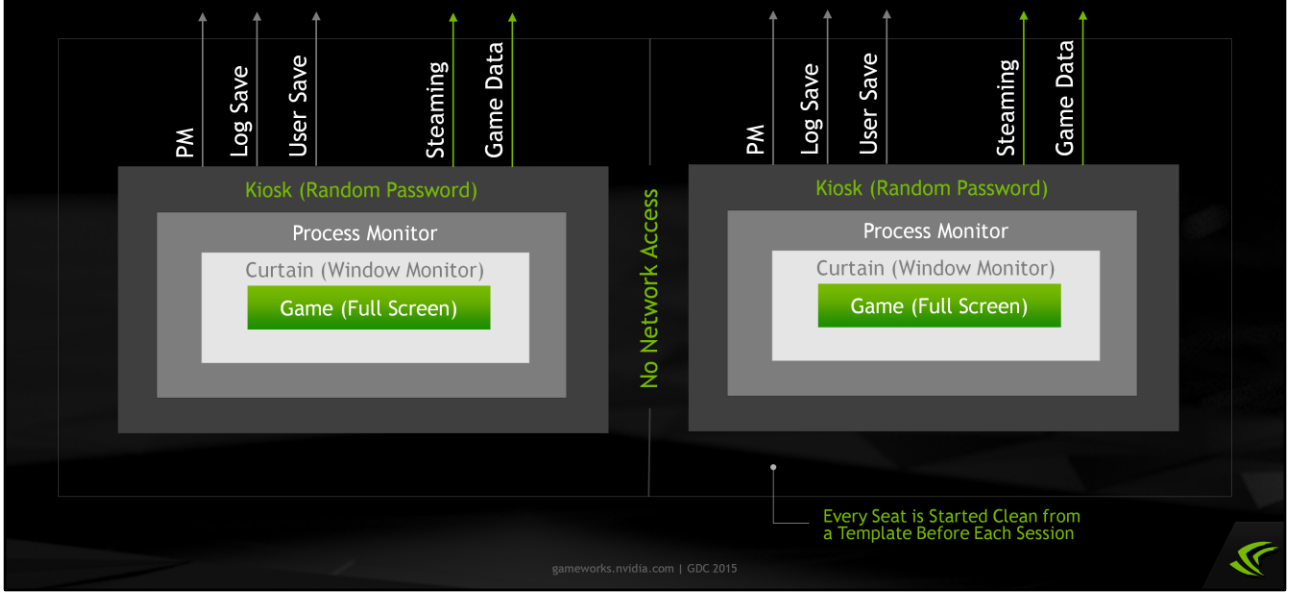
# SERVICES



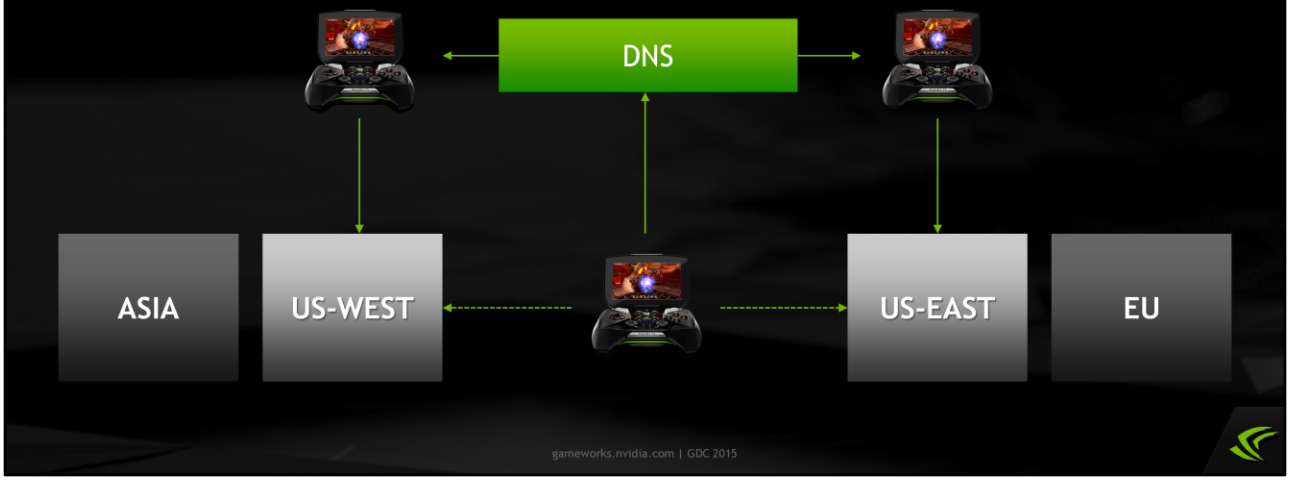

User Identity Management | Entitlement | Central Session Store | Central Game Save Store

Latency Based Routing | Publishing | Centralized Monitoring | E-Commerce

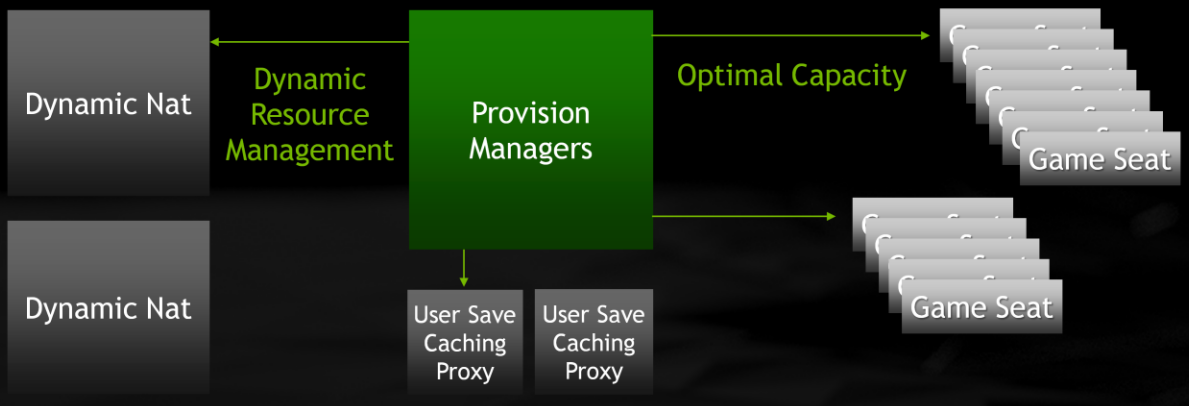
# GAME SESSION SERVER SECURITY



# GLOBAL REACH AND LATENCY BASED ROUTING



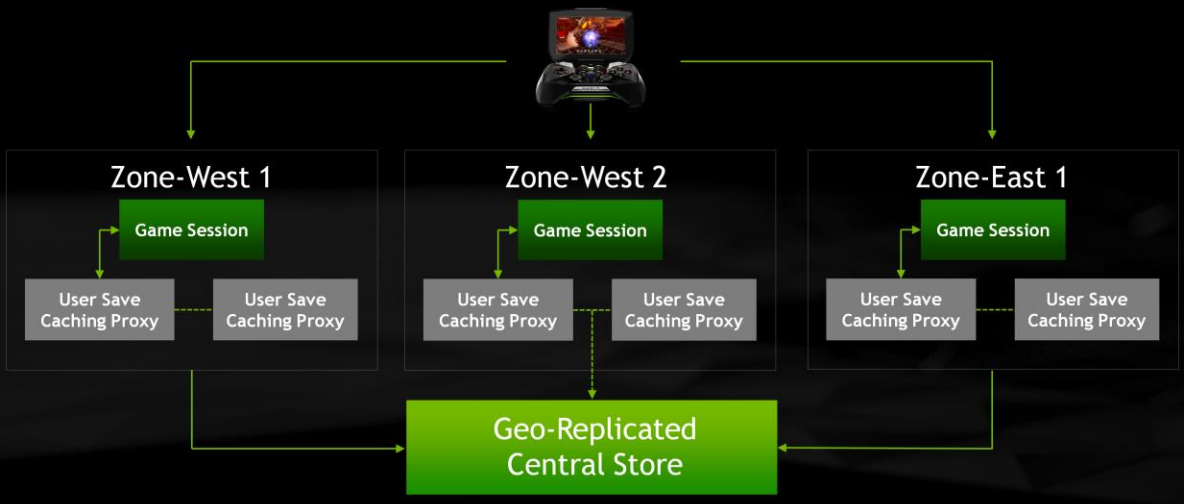
# ELASTICITY




Stateless Provision Managers are Responsible for Resource Management in One Zone

Add or Remove Resources Based on the Session Influx and Resource Availability

# GLOBAL GAME SAVES




NVIDIA GRID
GAMESTREAM
Ultra Low Latency Streaming Technology Overview

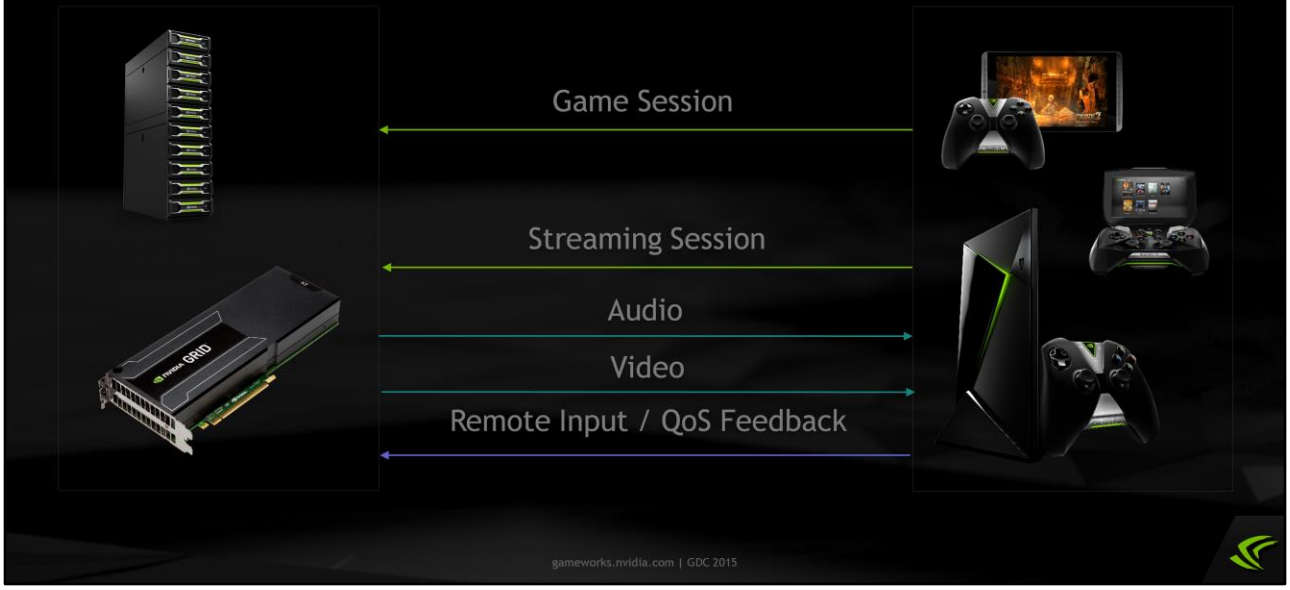
NVIDIA

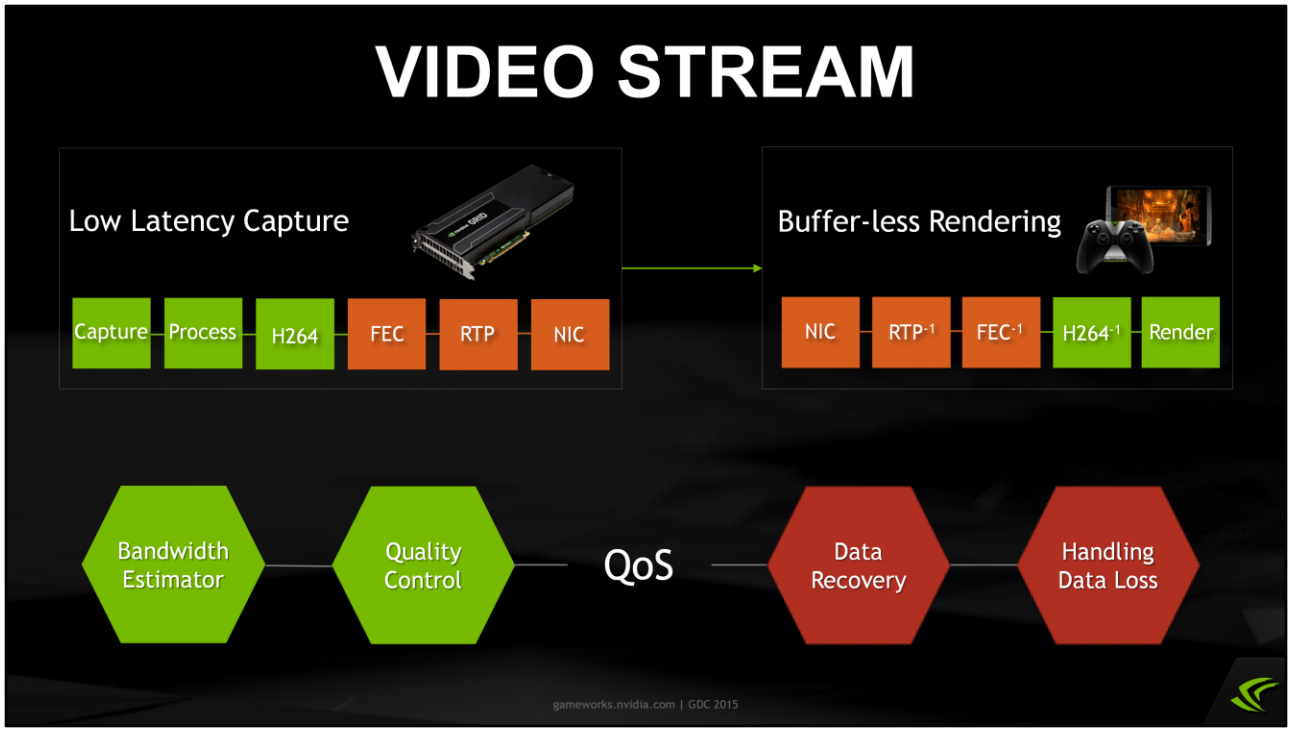
# INTERACTIVE STREAMING TECH

| | Netflix | GRID |
|---|---|---|
| Stream Type | Static / pre-recorded | Interactive / live |
| Buffering | Yes | No |
| Reliable Network Protocols | Yes | **No** |
| Encoding | Multi-pass, pre-encoded | **Real time** |
| Latency | High | **Low** |



# OVERVIEW






The absolute lowest latency possible server pipeline: full screen capture, post processing, color space conversion, scaling, encoding – all in hardware bypassing system memory

No buffering on the rendering side. Decoding and rendering in hardware

RTP stream on the network

To prevent data loss

Bandwidth Estimator to deal with limited (and constantly changing) network bandwidth

Quality Control to maintain good quality given the estimated channel capacity

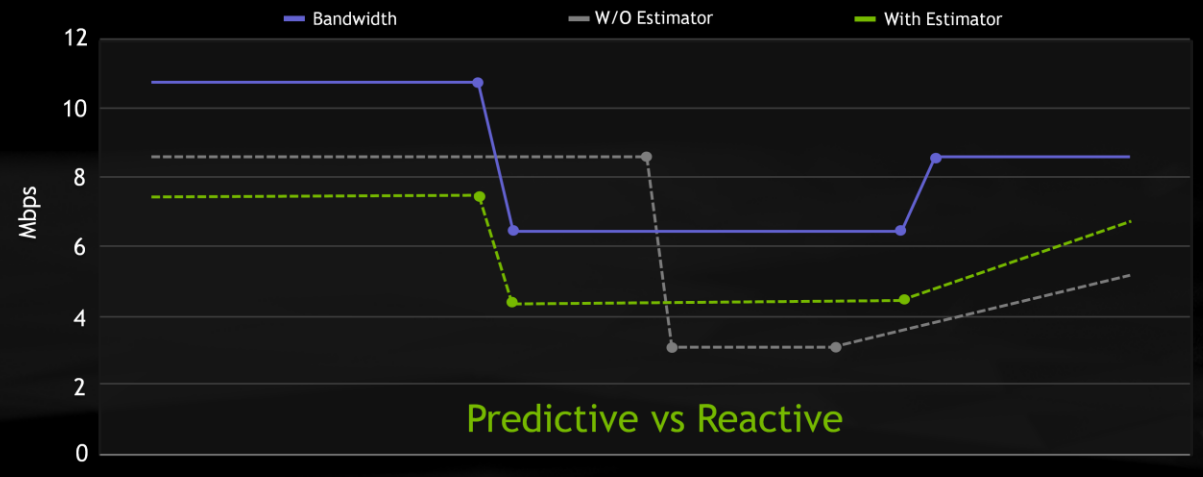
To deal with data loss

Data recovery using redundant data (FEC)

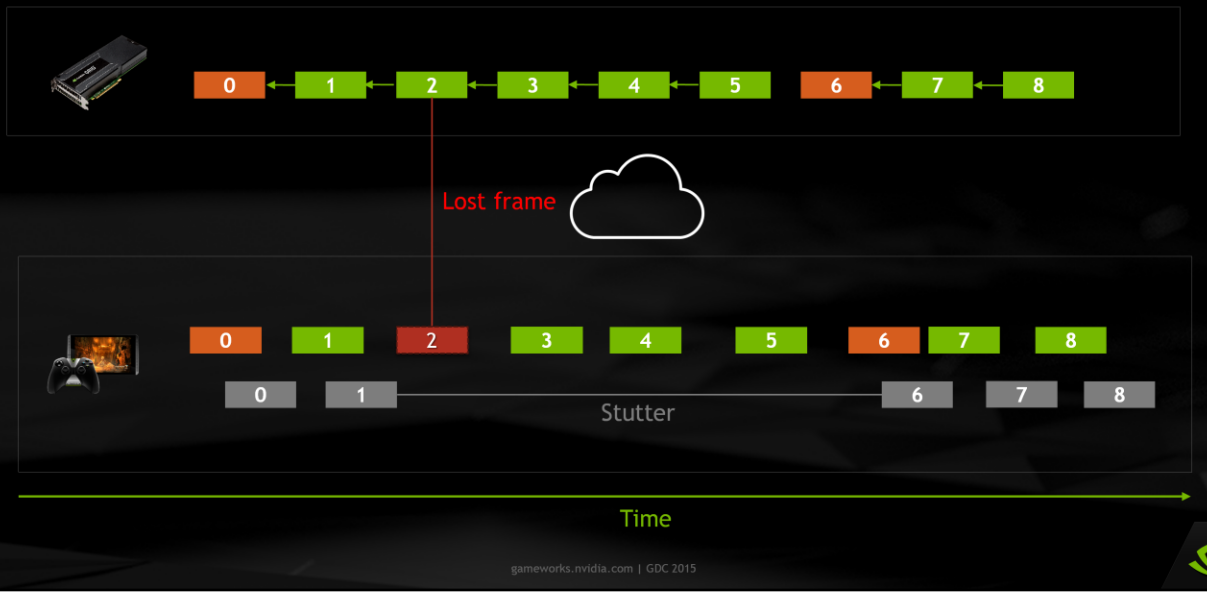
Handling data loss by requesting new reference frame

# QOS: BANDWIDTH AND QUALITY

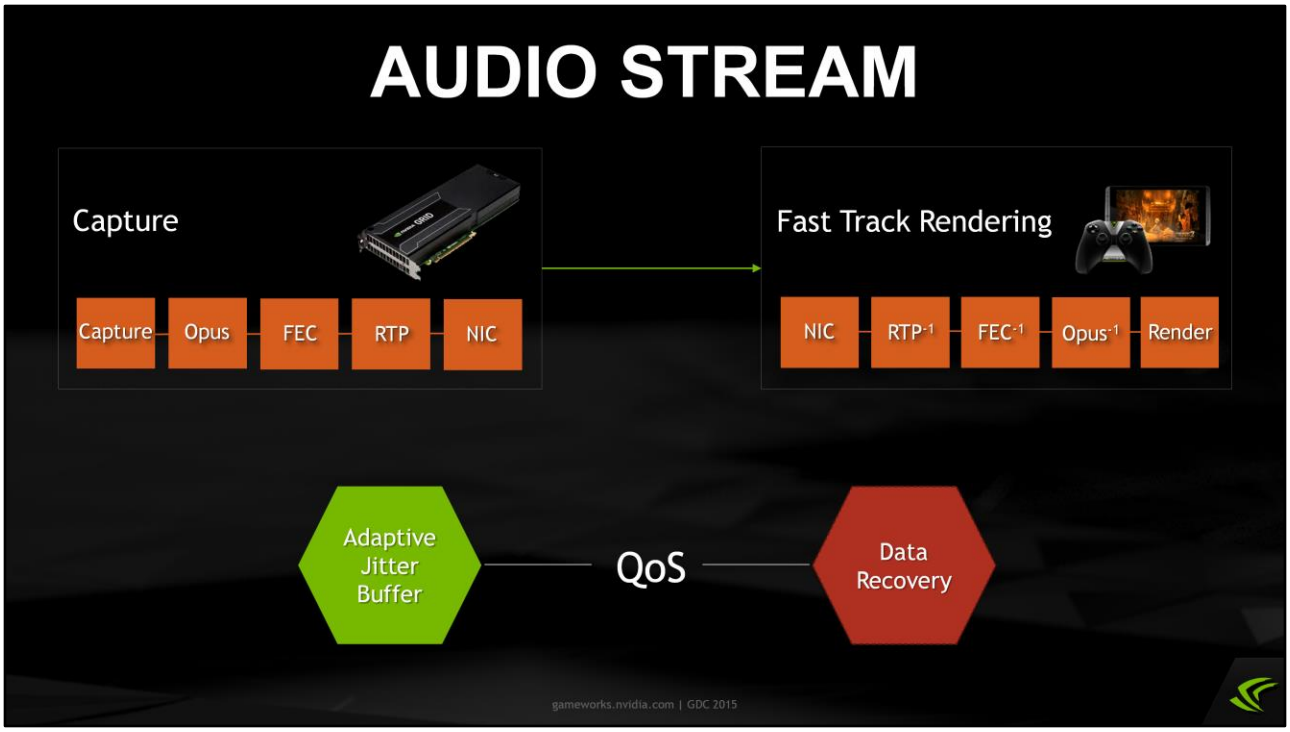
## Estimate bandwidth and react to changes immediately



# QOS: DATA LOSS



# AUDIO STREAM

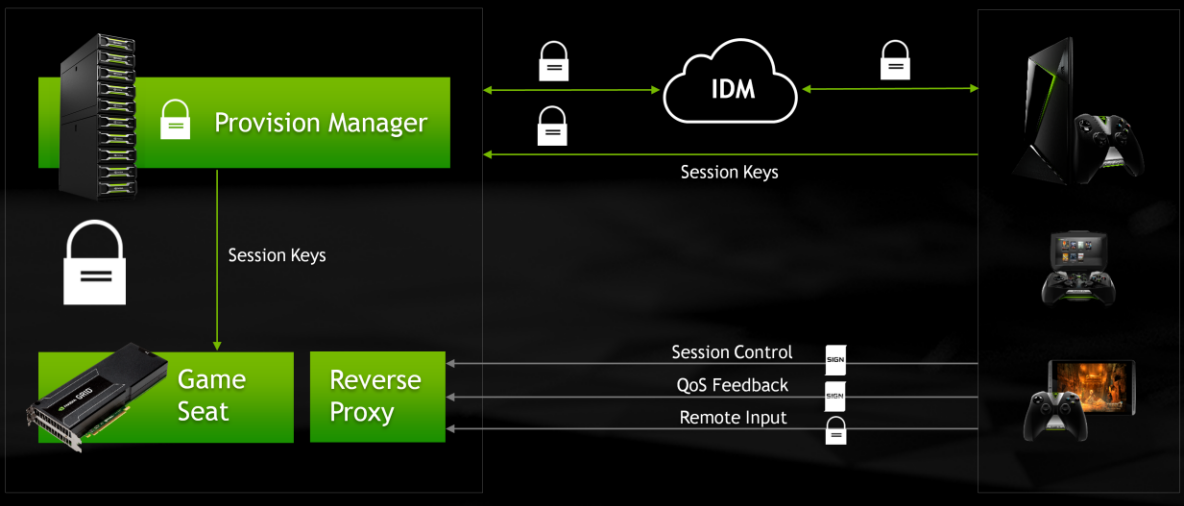


# SECURITY

- Encrypt for Remote Input Packets
- Sign Session Requests
- Sign QoS Feedback Messages
- Audio and Video are in the Clear



# SECURITY


Provision Manager
IDM
Session Keys
Session Keys
Game Seat
Reverse Proxy
Session Control
QoS Feedback
Remote Input