



AWS Innovation at Scale

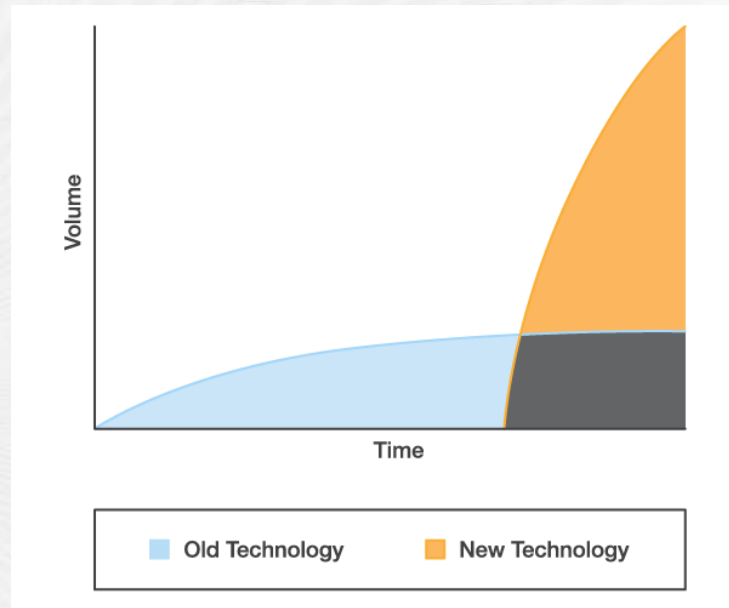
James Hamilton, AWS VP & Distinguished Engineer

SPOT301: November 12, 2014 | Las Vegas, NV



The Pace Quickens

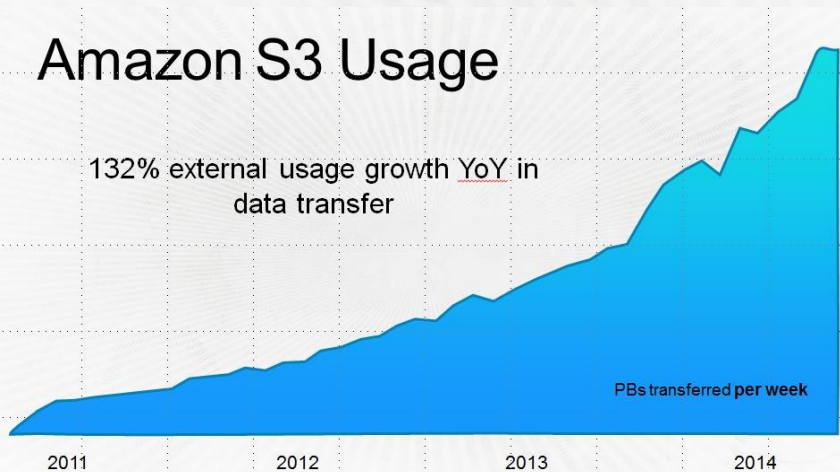
- Industry generational changes rare
 - Only when economics far superior
 - Mainframes to UNIX Super Servers
 - UNIX Super Servers to x86 Servers
- It's happening again
 - x86 on premise servers to the cloud
 - Past transitions have taken a decade+
 - What's different this time is speed of change
- Bigger customer gains drive faster industry transitions



AWS Growth Accelerates

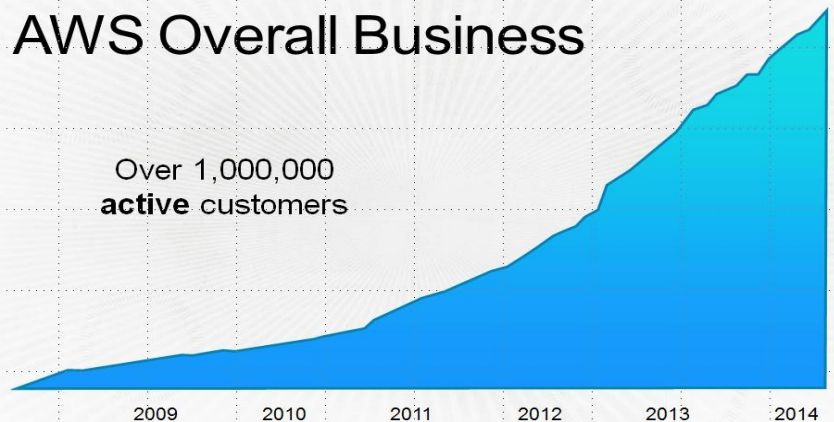
Amazon S3 Usage

132% external usage growth YoY in data transfer



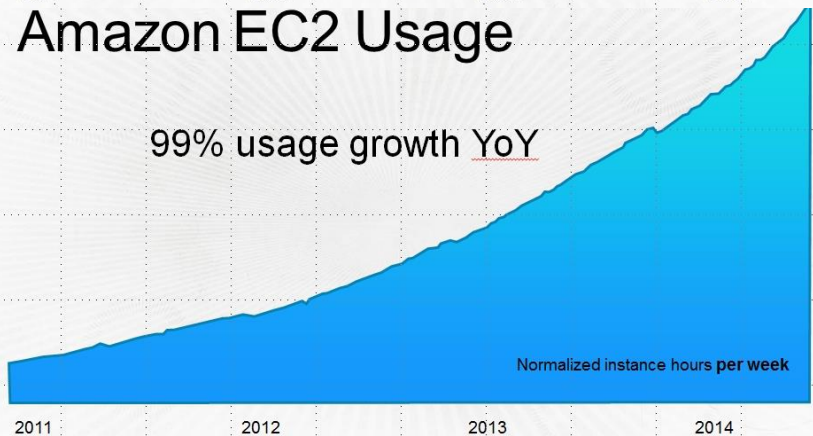
AWS Overall Business

Over 1,000,000 **active** customers



Amazon EC2 Usage

99% usage growth YoY



“5X the cloud capacity in use than the aggregate total of the other 14 providers”

Gartner®

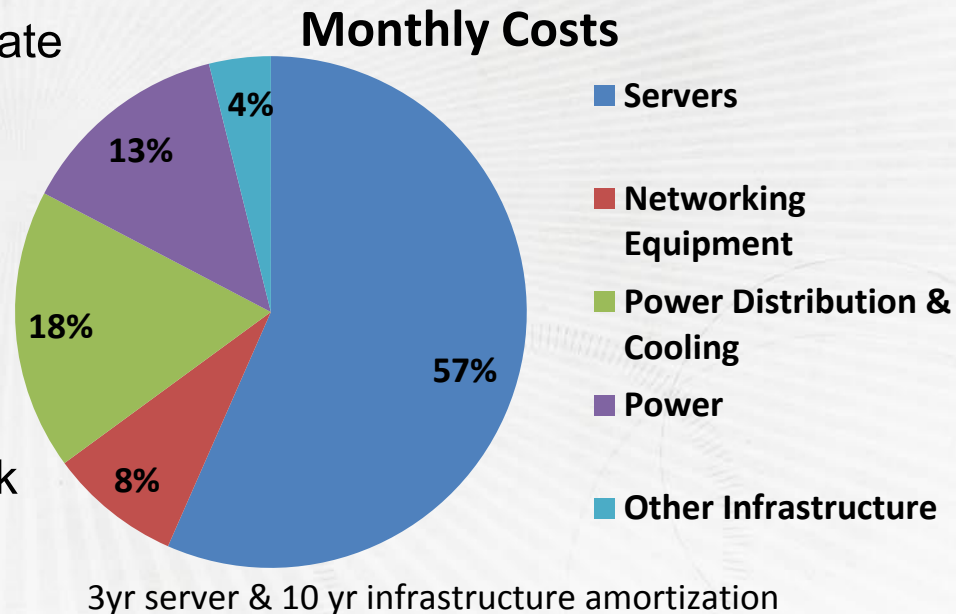
Perspective on Scaling



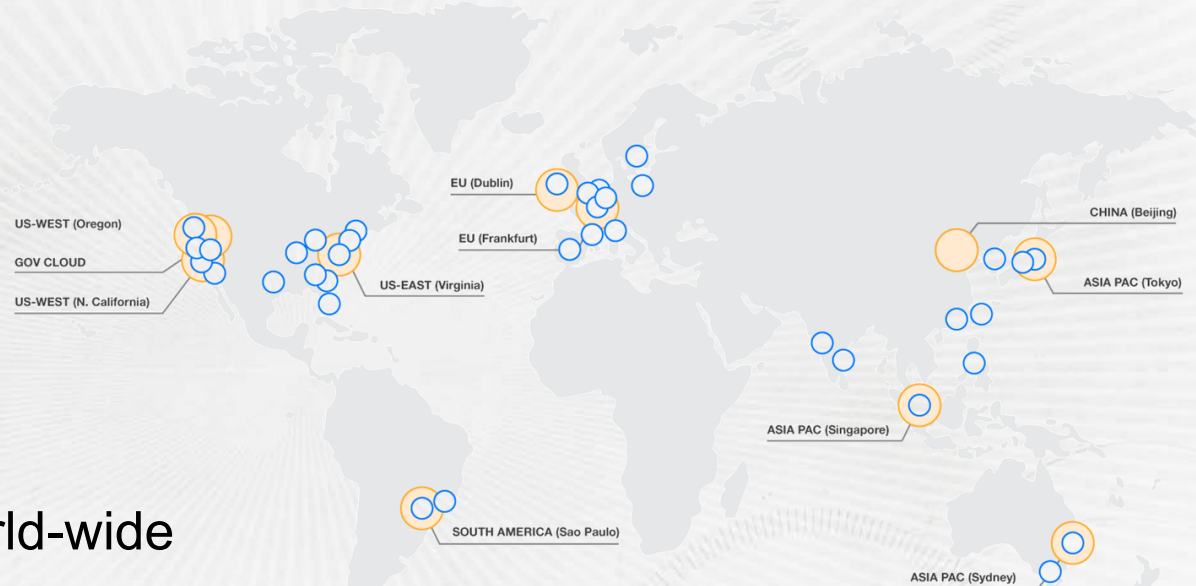
Every day, AWS adds enough new server capacity to support all of Amazon's global infrastructure when it was a \$7B annual revenue enterprise

Get Networks Out of the Way

- Relative cost of networking increasing quickly
 - Server & storage prices falling fast
 - Network costs trending to dominate
- Networking frozen in time
 - Vertically integrated ecosystem
 - Indefensible profit margins
- AWS solution:
 - Custom net H/W & protocol stack
 - Private long haul links

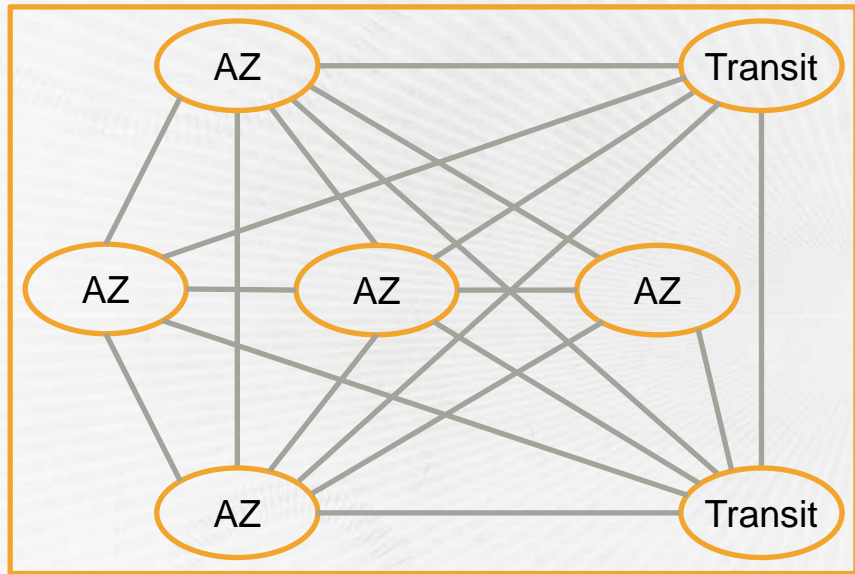


AWS Worldwide Network Backbone

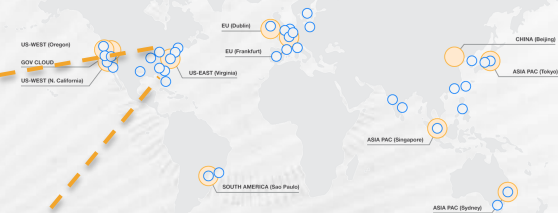


- 11 AWS Regions world-wide
- Compute & storage to customers & Users or required jurisdictional boundaries
- Private AWS fiber links interconnect all major regions
 - Increased availability, higher performance, lower jitter, & reduced costs

Example AWS Region

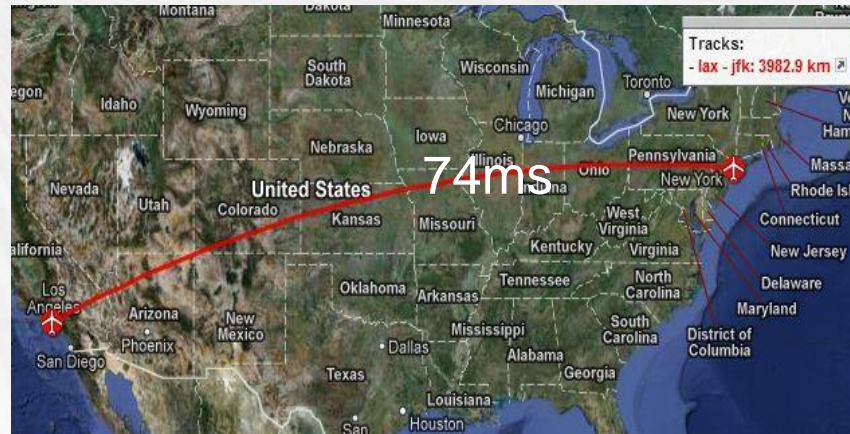
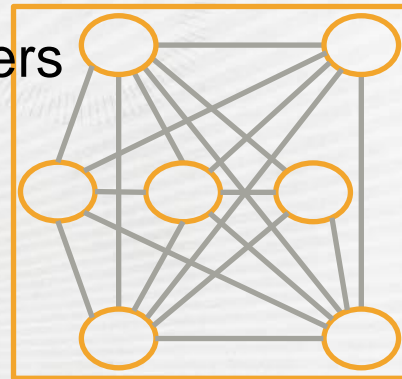


- 1 of 11 AWS world-wide AWS regions
- Redundant paths to transit centers
- Transit centers connect to:
 - Private links to other AWS regions
 - Private links to Direct Connect customers
 - Internet through peering & paid transit
- Metro-area DWDM links between AZs
- 82,864 fiber strands in region
- AZs <2ms apart & usually <1ms
- 25Tbps peak inter-AZs traffic

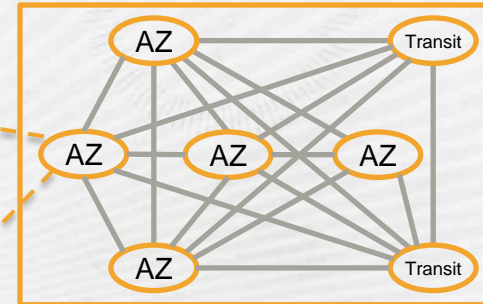
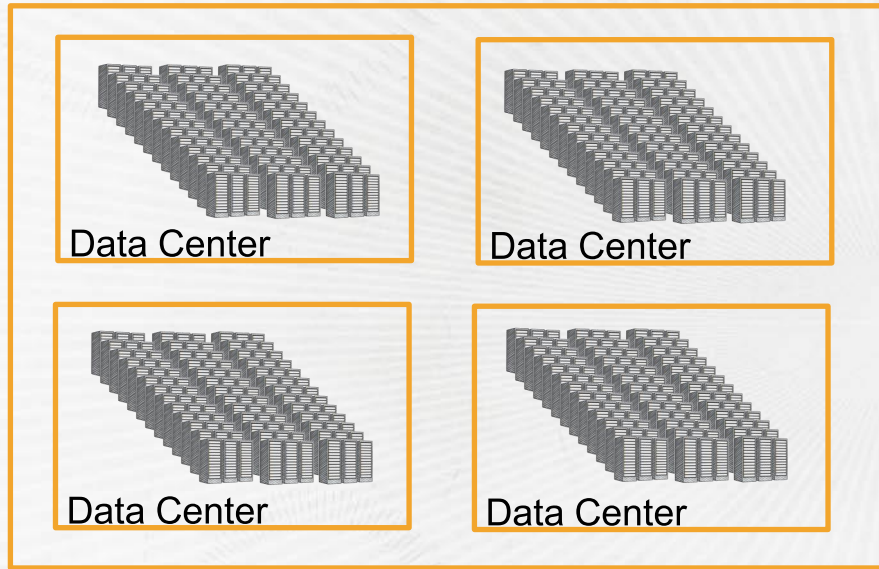


Why Does AWS Offer AZs?

- Asynchronous replication between distant data centers
 - Committing to an SSD order 1 to 2ms
 - But, LA to New York is 74ms round trip
 - You can't wait 74ms to commit a transaction
- On failure, difficult & high skill decision:
 - Fail-over & lose transactions, or
 - Or don't fail-over & lose availability
 - Difficult choice
- Use AZs for no-admin failover
 - Sync works when $< 2\text{ms}$
 - Can be combined with regional replication for very high availability

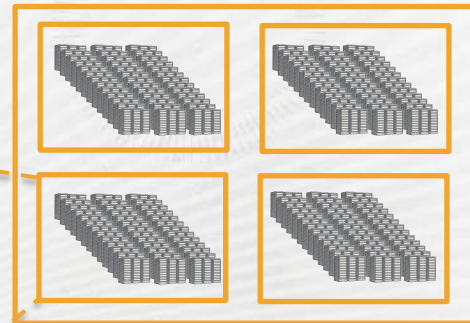


Example AWS Availability Zone



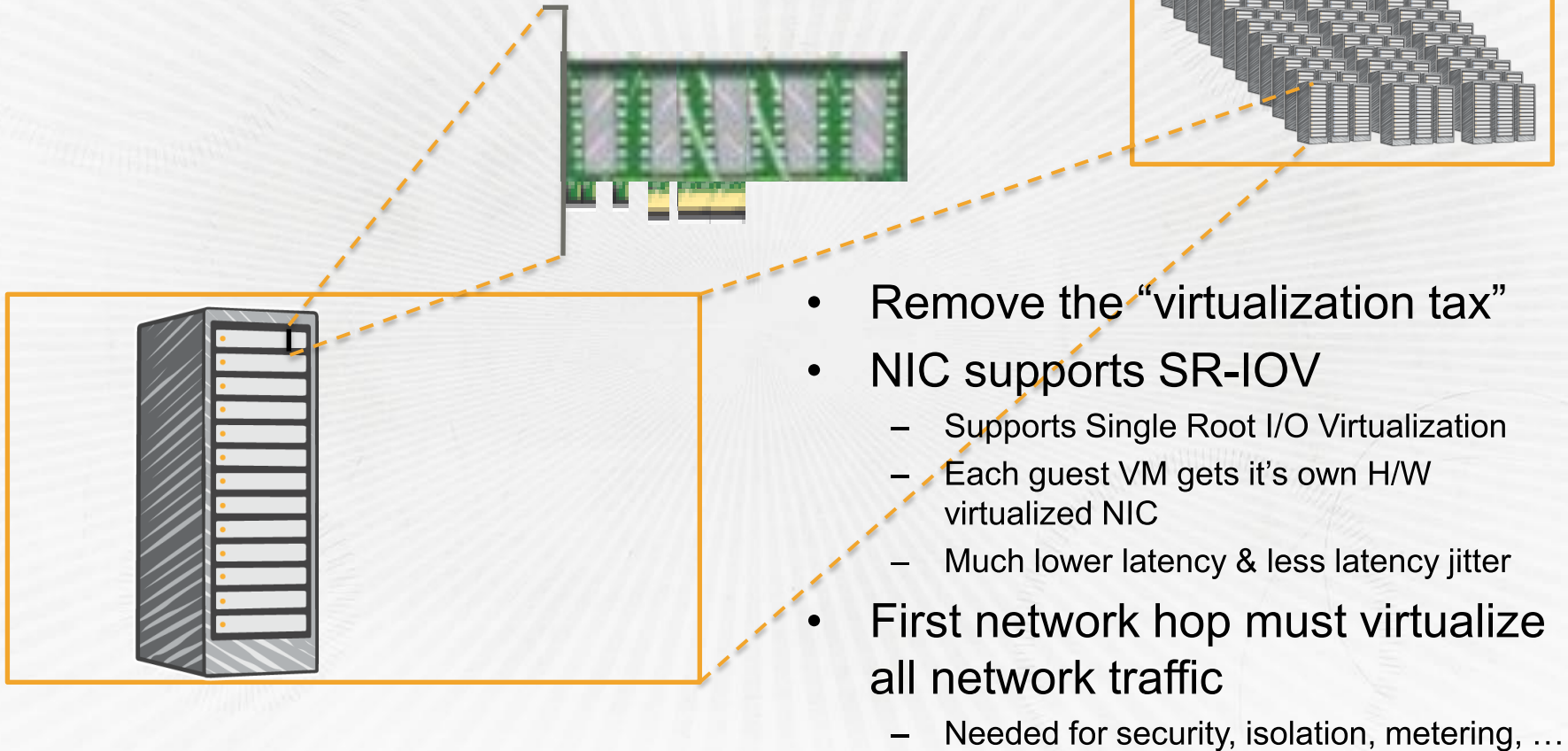
- 1 of 28 AZs world-wide
- All regions have 2 or more AZs
- Each AZ is 1 or more DC
 - No data center is in two AZs
 - Some AZs have as many as 6 DCs
- DCs in AZ less than $\frac{1}{4}$ ms apart
 - Don't need inter-AZ independence
 - Do require low latency & full B/W

Example AWS Data Center



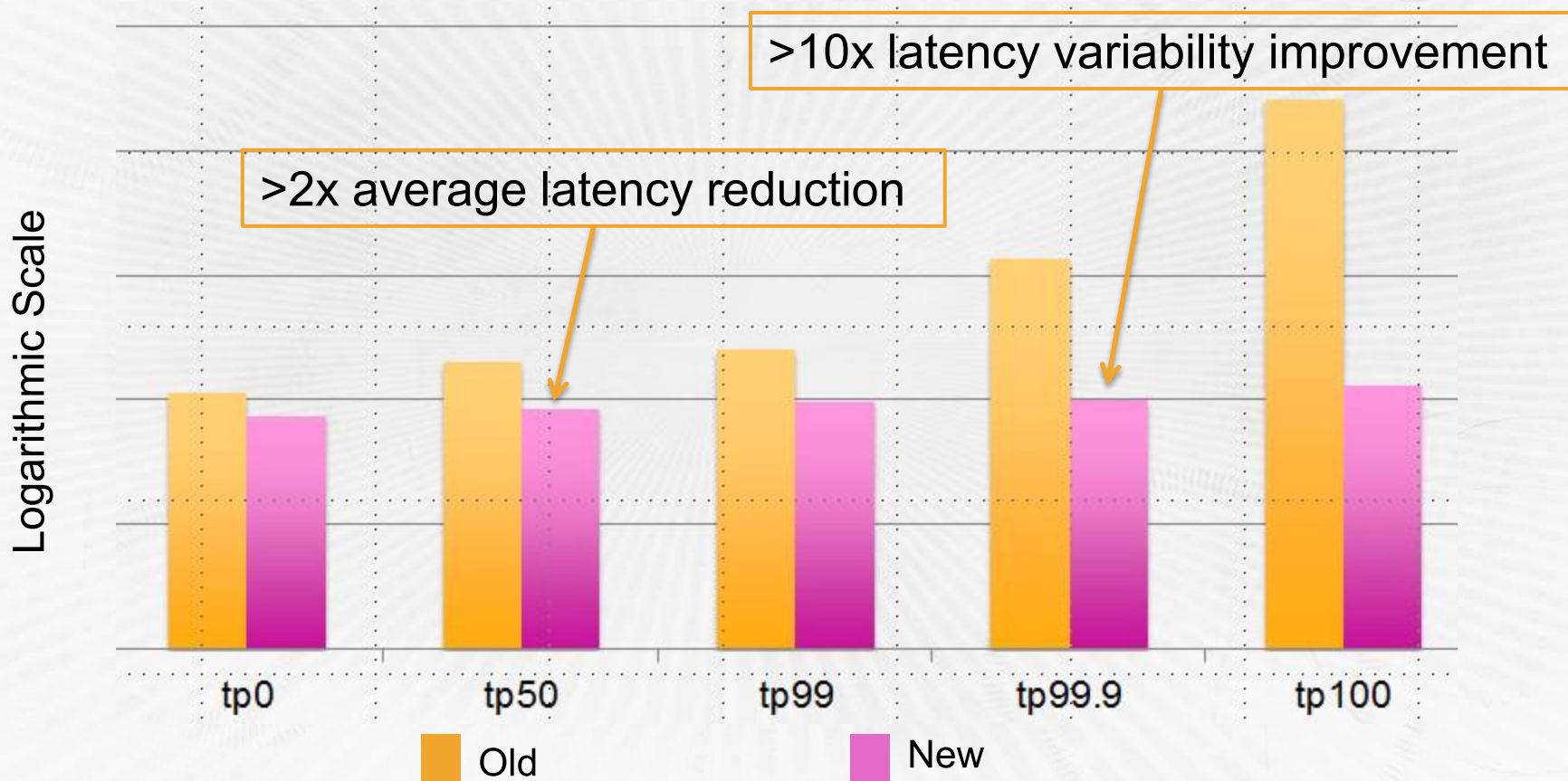
- Single DC typically over 50,000 servers & often over 80,000
 - Larger DCs undesirable (blast radius)
- Up to 102Tbps provisioned to a single DC
- AWS custom network equipment:
 - Multi-ODM sourced
 - Amazon custom network protocol stack

Example Rack, Server & NIC



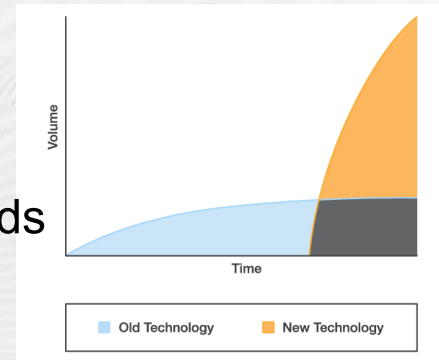
- Remove the “virtualization tax”
- NIC supports SR-IOV
 - Supports Single Root I/O Virtualization
 - Each guest VM gets it's own H/W virtualized NIC
 - Much lower latency & less latency jitter
- First network hop must virtualize all network traffic
 - Needed for security, isolation, metering, ...

Network Latency & Variability



AWS Custom Server & Storage Designs

- OEM server ecosystem:
 - Very general designs able to run wide variety of workloads
 - Vast, expensive, world-wide distribution network
- AWS custom servers & storage:
 - Specialized servers optimized for a specific workload
 - Move hot s/w kernels to hardware implementations
 - Custom Intel procs beyond commercially available clock rates
 - DCs, servers, networking, storage designed to integrated specs



Example Storage Rack

- 864 disks, 2,350lb

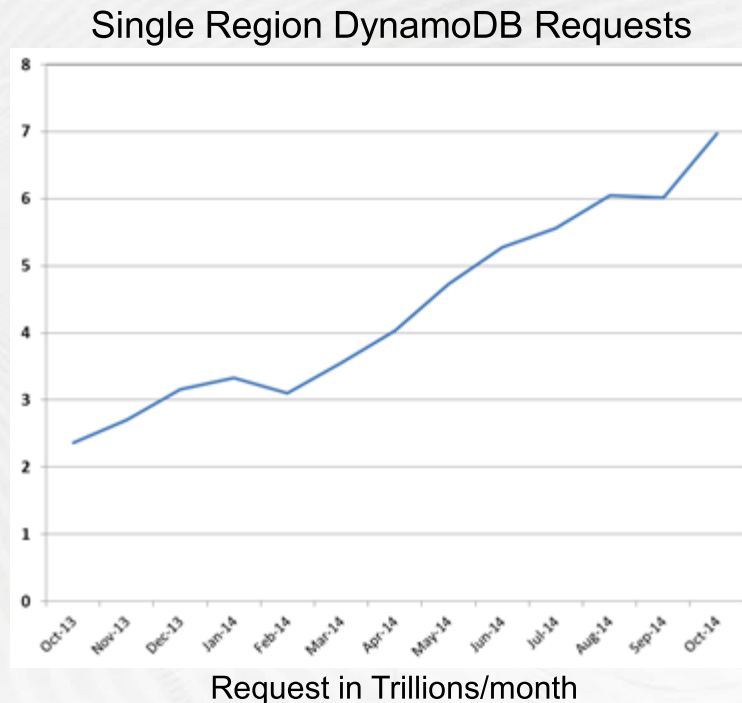
Relational Database Expensive & Hard

- Relational Database dominated by “big 3”
 - Oracle, SQL Server, & DB2
- Expensive, hard to administer, don't scale, & just about impossible to switch
- “No SQL” scales & relieves some administrative burden
 - e.g. MongoDB
- Cloud NoSQL both scales & virtually eliminates classic DB admin Issue
 - e.g. Amazon DynamoDB

The Oracle logo, featuring the word "ORACLE" in a bold, red, sans-serif font.

Amazon DynamoDB

- Cloud NoSQL database optimized for latency & scale
- 3x request growth last year
 - Single digit ms response times
 - Still same low & predictable jitter
 - 4x storage growth over same period
- Key new features
 - JSON Support
 - Up to 400KB items
 - Global Secondary Indexes
 - DynamoDB Streams
 - Cross Region replication



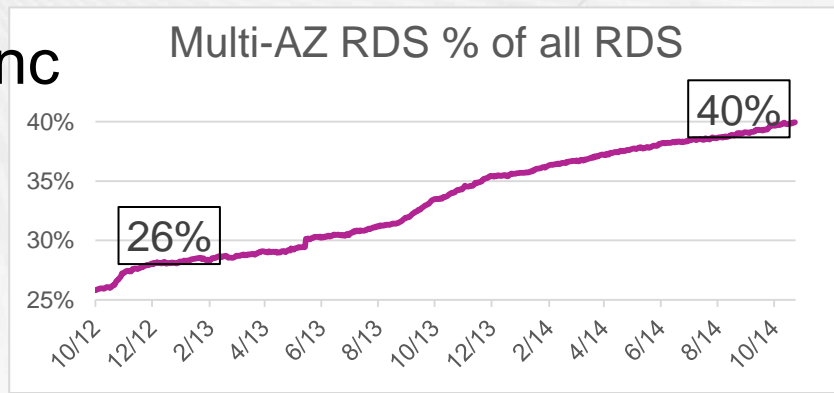
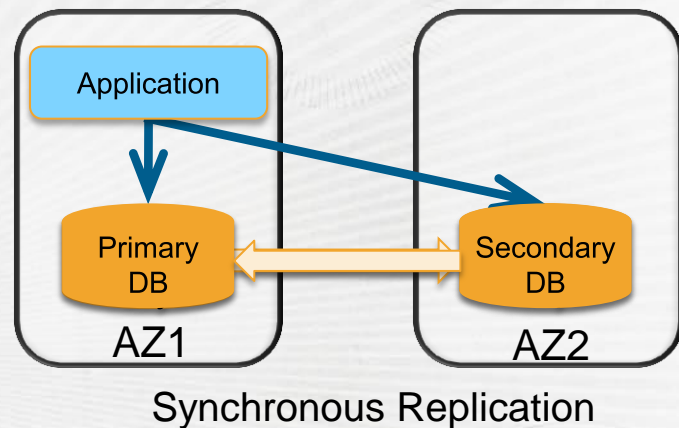
Addressing RDB Administrative Challenge

- Relational easy to use, feature rich, but admin intensive
 - RDBs still the core of many applications
 - Also largest single largest driver of downtime & lost sleep
- RDS MySQL admin complexity & cost breakthrough
 - Addresses the administrative complexity issue
 - Amazon RDS MySQL, Oracle, SQL Server, & PostgreSQL
- Cloud managed alone doesn't address RDBMS cost, availability, performance or scaling limitations



RDS Multi-AZ Relational Availability

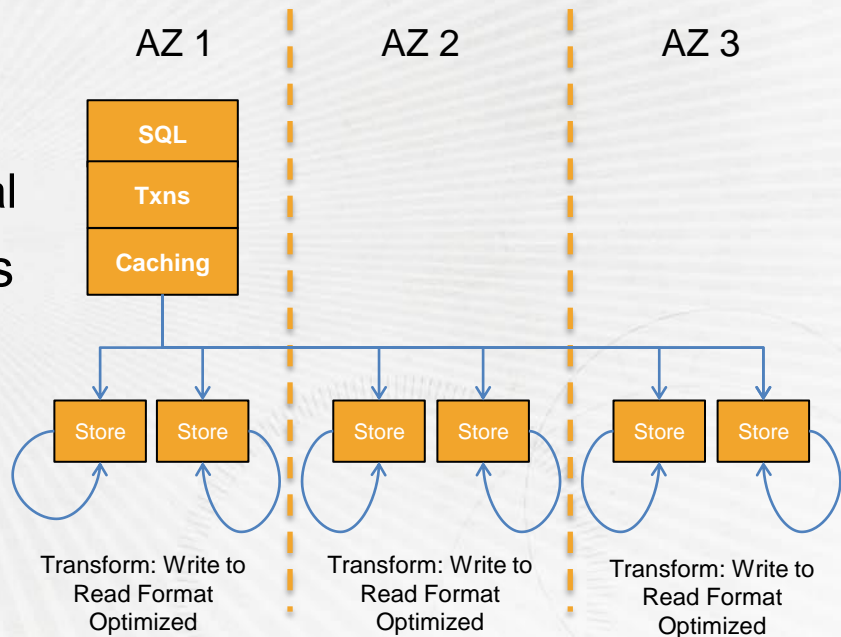
- Hard to reliably get beyond 3 9s in single building deployments:
 - RDS MySQL Multi-AZ Synchronous replication
- Not new technology
 - EMC SRDF/S & Oracle Fast Start Failover
 - But all come with “enterprise” pricing
- RDS MySQL Multi-AZ makes sync replication inexpensive & easy
 - More application 9s & way more sleep



Amazon Aurora

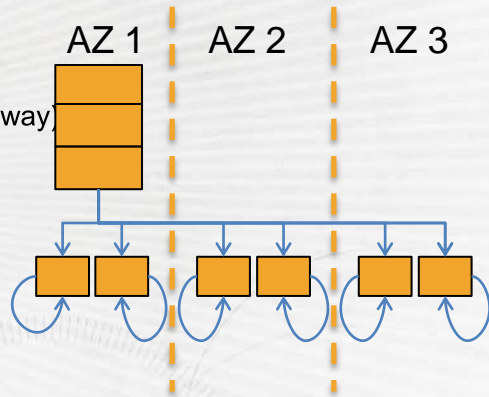


- Custom AWS MySQL Storage Engine
 - Enterprise DB features at cloud pricing
 - Drop in compatible with MySQL apps
 - Storage engine separate from relational
- Triple AZ storage engine handles faults W/O read or write pause
 - e.g. Entire DC can go down at same time as a disk or server failure
- “Impossible” faults such as loss of 2 DCs still don’t lose data
 - Synchronous multi-DC replication



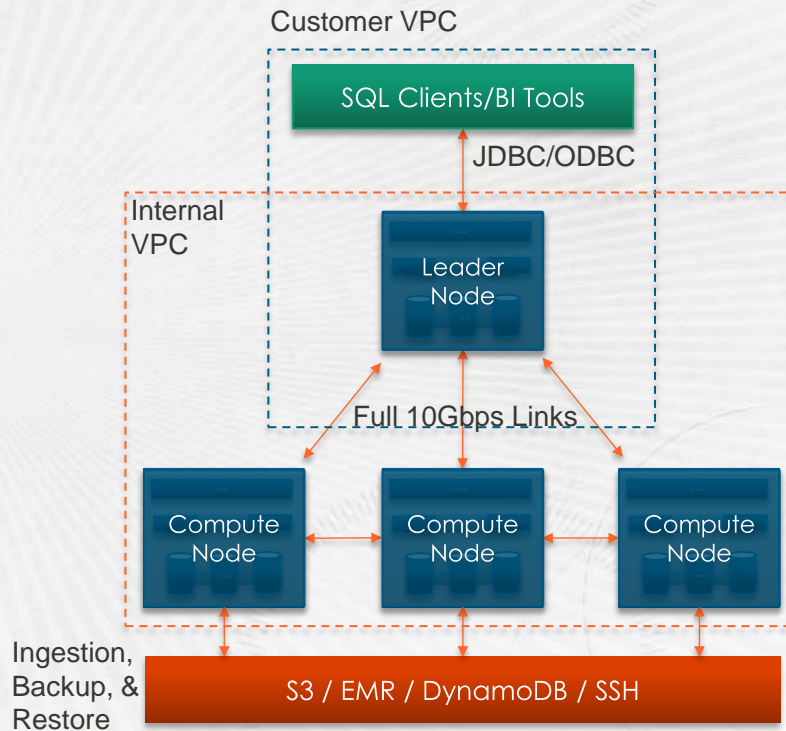
Amazon Aurora Performance

- MySQL updates combined with new storage manager
 - 3x write performance_(sysbench)
 - 5x read performance_(sysbench)
- Supports up to 16-way read replicas (RDS MySQL: 5-way)
 - 400x less lag (2,000ms vs 5ms)
- Supports up to 64TB tables (RDS/MySQL: 3TB)
- Near instant fail-over (no database crash recovery time)
- Auto-recovery from storage faults
 - Auto-data page patch or full disk loss recovery without operational impact



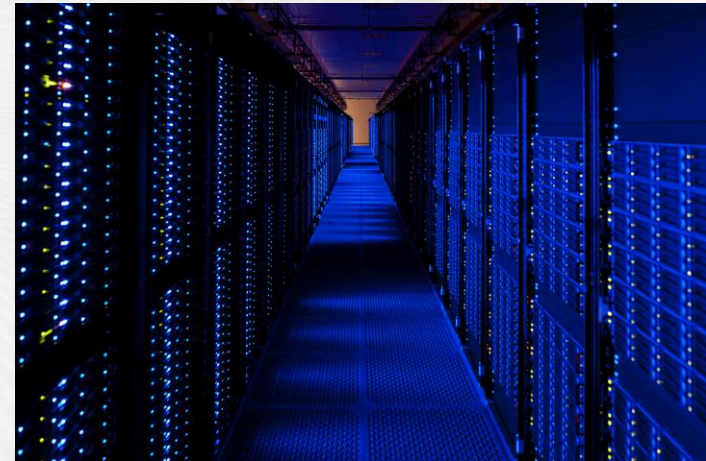
Redshift Parallel SQL Data Warehouse

- Up to 128 server parallel SQL DB
 - Columnar data warehouse
- Disruptive cost \$1,000/TB/Year
 - Fastest growing AWS service
 - Already 1000s of customers
 - Multiple PB+ clusters in production
- Automated provisioning, patching, security, resize, backup/restore
- Massive data scaling
 - DW1: HDD; scale from 2TB to 2PB
 - DW2: SSD; scale from 160GB to 326TB

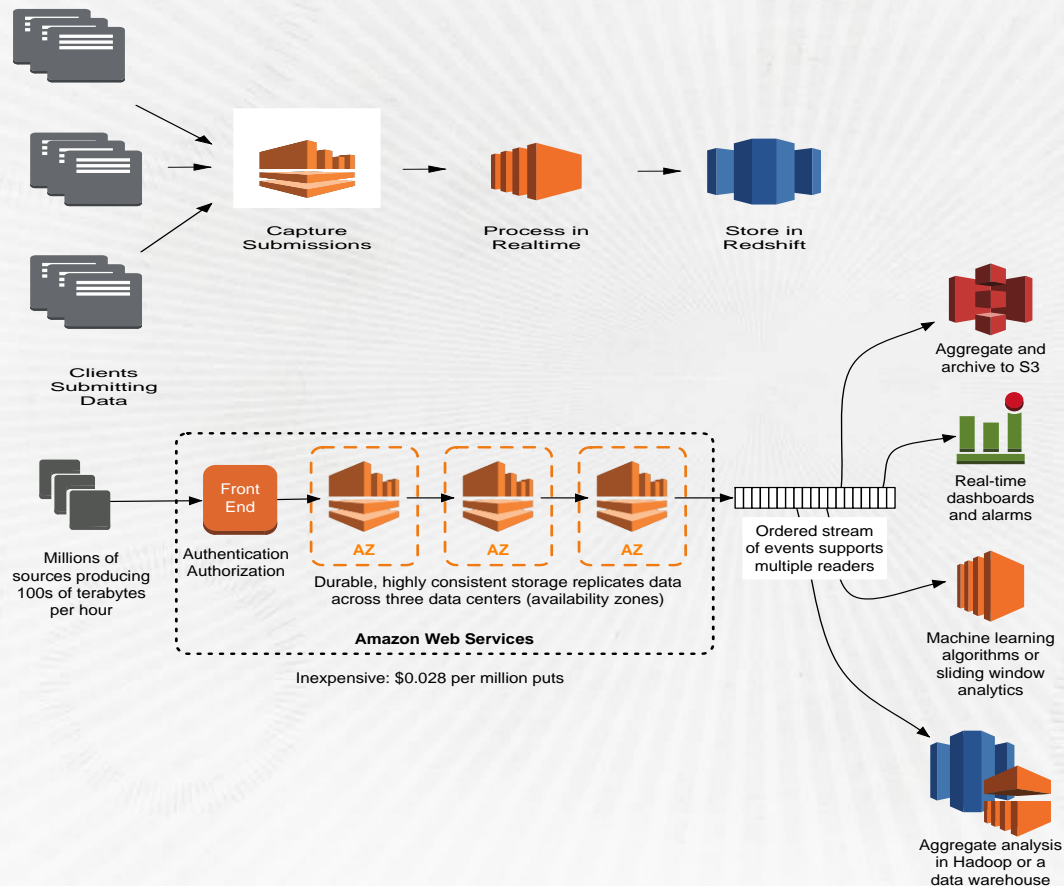


EBS at 20,000 IOPS

- Provisioned IOPS (SSD)
 - Max volume to 16TB (From:1TB)
 - Max I/O rate to 20,000 IOPS (From:4k IOPS)
 - Max throughput to 320MB/s (From:180MB/s)
- General Purpose (SSD)
 - Max volume size to 16TB (From:1TB)
 - Max I/O rate to 10,000 IOPS (From:3k IOPS)
 - Max throughput to 160MB/s (From:128 MB/s)



Internal Challenge to External Service



AWS Metering:

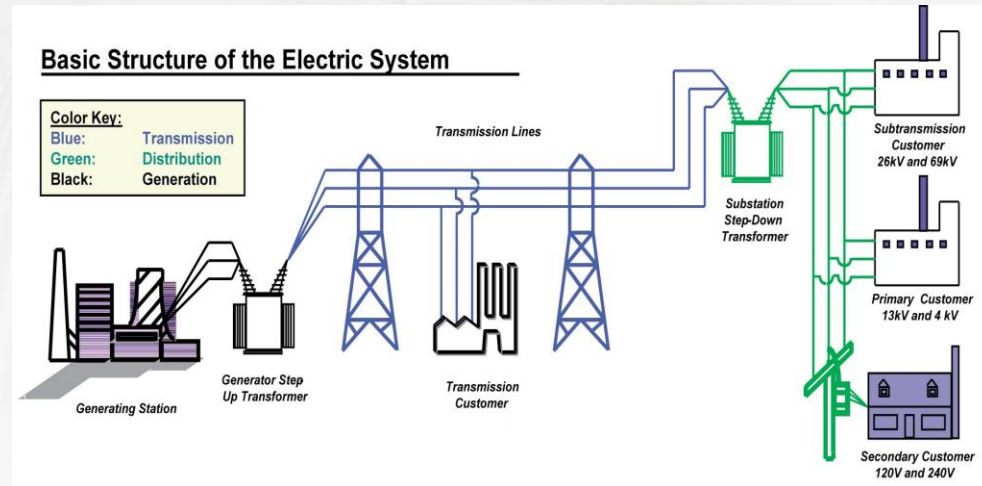
- Tens of millions records/sec
- Multiple TB per hour
- 100,000s of internal sources
- Scales, low-cost, auditable, with real time alerting

AWS Kinesis:

- Producers call put
- Sequence # returned
- Distributed over shards
- Scales per shard at 1 MB/s & 1000 TPS

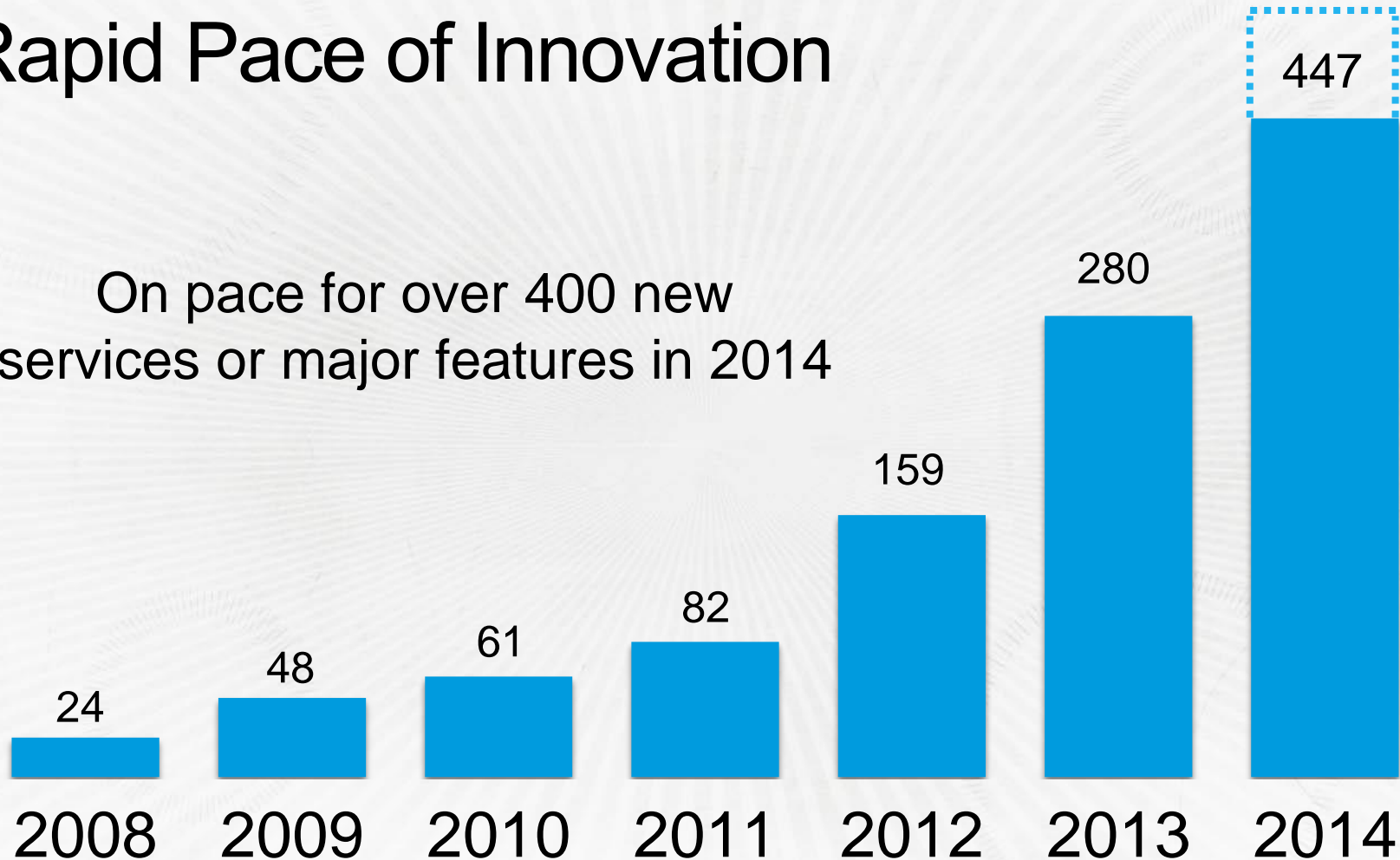
Power Infrastructure

- Some DCs with custom power sub-stations
 - Lower power cost & build more quickly
- Negotiated power purchasing agreements
- Custom switchgear firmware
- 3 100% carbon neutral regions:
 - US-West (Oregon)
 - US-Gov-west (US)
 - EU-Central-1 (Frankfurt)



Rapid Pace of Innovation

On pace for over 400 new services or major features in 2014



AWS re:Invent

Please give us your feedback on this
presentation

SPOT301

Thank You!



Join the conversation on Twitter with **#reinvent**