Characterizing network paths in and out of the Clouds

by Igor Sfiligoi, UCSD/SDSC (isfiligoi@sdsc.edu)
co-authors John Graham and Frank Wuerthwein, both UCSD
Who cares about clouds?

• Like it or not, Clouds are gaining popularity
  • Especially great for prototyping
  • But great for urgent compute activities, too

• Funding agencies are taking notice
  • See NSF E-CAS award(s) as an example
  • SDSC Expanse has Cloud explicitly in the proposal text
  • NSF-funded CloudBank
  • NSF also recently awarded an EAGER award for exploratory work
Characterizing Cloud Resources

• When choosing which resources to use (and pay for) knowing what to expect out of them is important

• The compute part of Cloud resources is very well documented
  • They all use pretty standard CPUs and GPUs
  • And the models used are relatively well documented:
    https://aws.amazon.com/ec2/instance-types/
    https://docs.microsoft.com/en-us/azure/virtual-machines/linux/sizes-general
    https://cloud.google.com/compute/docs/machine-types

• The networking is instead hardly documented at all
  • Yet networking can make a big difference when doing distributed computing
Structure of this talk

• What Cloud networking is capable of
  • Using Storage as a proxy
• Getting data in and out of the cloud
• The cost of Cloud networking
In-Cloud Networking to their Storage
Why storage? How does it relate to networks?

• All Cloud providers operate large pools of distributed storage
• Fetching data from remote storage moves data over the network
• Did not really hit an upper bound in testing locally (in single-region)
  • Network thus the bottleneck for remote transfers
• Characterizing Cloud Storage is an interesting topic in itself
In-Cloud Storage Summary

• All cloud providers are capable of delivering **well over 1Tbps** in bandwidth to their storage
  • Did not try to push much over

• All cloud providers have **multi-100Gbps** networking between major regions
  • And around 100Gbps for lesser region connections
  • Note: Google **tops 1Tbps** pretty much across all tested connections

• Details on following slide, if you are interested
  • Note that for all connections over 150Gpbs I probably could get more if I tried harder
# Single Region Object Store Performance

Measured values, may not be max achievable

<table>
<thead>
<tr>
<th>Service</th>
<th>Gbps</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS S3</td>
<td>1650 Gbps</td>
</tr>
<tr>
<td>MS Azure BLOB</td>
<td>1300 Gbps</td>
</tr>
<tr>
<td>GCP</td>
<td>1800 Gbps</td>
</tr>
</tbody>
</table>

Note: All Cloud storage uses HTTP[S] as the underlaying protocol (HTTP can be used for anonymous access, HTTPS for either anonymous or authenticated)
AWS S3 bandwidth

Marginally lower bandwidth to US West from EU, compared to US East

Comparable bandwidth to Ireland and Germany (from US)

Measured values, may not be max achievable

100 Gbps

440 Gbps

460 Gbps

65 Gbps

80 Gbps

90 Gbps

Nov 4th, 2019

CHEP2019 - Adelaide, Australia
MS Azure BLOB bandwidth

Measured values, may not be max achievable

110 Gbps
190 Gbps
185 Gbps
100 Gbps
88 Gbps
Google Cloud Storage bandwidth

Measured values, may not be max achievable

940 Gbps

1060 Gbps

980 Gbps
AWS ping times between regions
MS Azure ping times between regions

Nov 4th, 2019
CHEP2019 - Adelaide, Australia
Networking Between On-Prem equipment and Cloud Storage

On-prem equipment all part of the PRP/TNRK Kubernetes cluster
Getting data out of the Clouds

• Few people these days live 100% in the Clouds
  • Often the Cloud compute is used to generate data, but final analysis happens on-prem
  • Another very important Cloud use case seems to be to park data in Cloud storage and access it on as-needed basis

• Understanding how long will it take to extract the data is thus essential

• In-cloud Storage bandwidth not a worry
  • As shown in previous slides
  • Networking between science network backbones and the Cloud the limit
Data Export Summary

• In the US, we can easily exceed
  • 30Gbps in the same region
  • 20Gbps coast-to-coast

• Assuming that on-prem node and campus have 100Gbps networking

• But finding mis-configured nodes is not unusual
  • As an example, a routing problem in Internet2 East Coast DTN resulted in less than 2Gbps bandwidth
  • After notifying the appropriate network teams, the transfer times dropped down dramatically
Reading from AWS S3 (HTTPS)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Orange edges Cloud
Reading from AWS S3 (HTTPS)

Early October test results

US West 2

US East

35 Gbps

1.3 Gbps

1.6 Gbps

23 Gbps

1.5 Gbps

1.3 Gbps

Legend:
Blue dots are on-prem
Orange edges Cloud

Measured values, may not be max achievable

Nov 4th, 2019
CHEP2019 - Adelaide, Australia
Reading from Azure BLOB (HTTPS)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Navy edges Cloud
Reading from GCP Storage (HTTPS)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Green edges Cloud

Nov 4th, 2019
CHEP2019 - Adelaide, Australia
Networking Between Cloud and On-Prem equipment

On-prem equipment all part of the PRP/TNRP Kubernetes cluster
Getting data into the Clouds

• The other side of the equation is importing data into the Cloud in support of the compute jobs
  • Either asynchronously by uploading into Cloud Storage
  • Or by just-in-time streaming
• We focused on using OSG-supported StashCache
  • Basically xrootd with HTTP as the transfer protocol (which happens to be the same protocol used by Cloud storage)
Data Import Summary

• Unsurprisingly, in the US we observe similar network bandwidth results to what we saw for fetching data from the Cloud
  • I.e. most pairs fall in the 20GBps to 40 GBps range
  • Although I was pleasantly surprised to see over 60Gbps from UCSD

• Also assessed network performance using both ”local” and ”remote” caches
  • The situation is generally good, with most Cloud regions having at least a 10Gbps to an existing StashCache node, with the worst being 6.3Gbps
  • But it also showed that UCSD is currently almost always faster than “local” caches!
AWS reading from StashCache using HTTP

(From US Regions)

Legend:
Blue dots are on-prem
Orange edges Cloud

Measured values, may not be max achievable

Nov 4th, 2019
CHEP2019 - Adelaide, Australia
Azure reading from StashCache using HTTP
(From US Regions)

Legend:
Blue dots are on-prem
Navy edges Cloud

Measured values, may not be max achievable
AWS reading from StashCache using HTTP

(From non-US Regions)

The Illinois StashCache node performs very similarly to the Virginia node.

Measured values, may not be max achievable.

Legend:
Blue dots are on-prem
Orange edges Cloud
Measured values, may not be max achievable.

The Illinois StashCache node performs very similarly to the Virginia node.

Legend:
Blue dots are on-prem
Navy edges Cloud

Azure reading from StashCache using HTTP
(From non-US Regions)

East Austr. 14 Gbps
Korea 14 Gbps
EU West 13 Gbps

Legend:
Blue dots are on-prem
Navy edges Cloud
AWS ping times to On-Prem nodes (From US Regions)

Legend:
Blue dots are on-prem
Orange edges Cloud
MS Azure ping times to On-Prem nodes
(From US Regions)

Legend:
Blue dots are on-prem
Navy edges Cloud
AWS ping times to On-Prem nodes

(From non-US Regions)

Legend:
Blue dots are on-prem
Orange edges Cloud

159ms
135ms
201ms
152ms
91ms
7ms
5ms

AP Korea
AP Austr.
EU Frankfurt
MS Azure ping times to On-Prem nodes
(From non-US Regions)

Legend:
Blue dots are on-prem
Navy edges Cloud
Single-node GridFTP Bandwidth Between Cloud and On-Prem equipment

On-prem equipment all part of the PRP/TNRP Kubernetes cluster
The importance of GridFTP and Summary

• Many communities still rely on GridFTP
  • Both for bulk file transfers between DTNs
  • And for moving data to and from compute jobs

• Mostly focused on the DTN-to-DTN use case
  • i.e. when only a single node is involved at each end

• In the single-node setup
  • In-Cloud-Region performance comparable to on-prem results (30Gbps)
  • Close-by on-prem nodes can deliver 10Gbps into the Cloud
  • Over WAN hard to exceed 8Gbps, with several regions in the 5Gbps range
AWS Single-Node to GridFTP Server
(From US Regions)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Orange edges Cloud

US West 2
8.4 Gbps
9.3 Gbps
8.4 Gbps
11 Gbps
US East
7.8 Gbps

30 Gbps – inside AWS cloud Region

43 Gbps – SDSU to SDSC

Nov 4th, 2019
CHEP 2019 - Adelaide, Australia
MS Azure Single-Node to GridFTP Server
(From US Regions)

Legend:
Blue dots are on-prem
Navy edges Cloud

Measured values, may not be max achievable

Nov 4th, 2019
CHEP2019 - Adelaide, Australia
Google Cloud Single-Node to GridFTP Server
(From US Regions)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Green edges Cloud

US West 1 → US East
- 7.8 Gbps
- 7.7 Gbps
- 1.2 Gbps
- 6.9 Gbps
- 1.1 Gbps

Chep2019 - Adelaide, Australia
Nov 4th, 2019
AWS Single-Node to GridFTP Server
(From non-US Regions)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Orange edges Cloud
MS Azure Single-Node to GridFTP Server

(From non-US Regions)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Navy edges Cloud
Google Cloud Single-Node to GridFTP Server
(From non-US Regions)

Measured values, may not be max achievable

Legend:
Blue dots are on-prem
Green edges Cloud

AP Japan
AP Austr.
EU NL
11 Gbps
4.0 Gbps
1.1 Gbps
6.9 Gbps
7.5 Gbps

Nov 4th, 2019
CHEP2019 - Adelaide, Australia
And now let’s talk money

• Many people get surprised that they have to pay for networking in the Clouds
  • Science networks do not charge final users a dime!

• All cloud providers have a similar model
  • Incoming network traffic is free
  • Most in-Cloud-zone network traffic is free
  • Data owners must pay for any data leaving the Cloud
  • Moving data between different Cloud regions also requires payment, although generally much less expensive
How expensive?

• Ballpark numbers
  • $70/TB egress from Cloud
  • $20/TB data movement between Cloud regions
• Note that most Cloud providers have agreements with most research institutions to waive network costs **up to 15% of the total bill**
  • So, if you compute a lot in the Cloud, network may be actually free
  • E.g. using AWS spot pricing, you get 1GB of out-networking free:
    • every 12h of a single CPU core
    • every 1h of a 18-core/36-thread CPU instance
    • every 30 mins of a V100 GPU
Incoming Network Traffic is Free

AWS example – MS Azure and Google have same policy

All in-region networking free, too

Bandwidth

| $0.000 per GB - data transfer in per month | 15,048.134 GB | $0.00 |

AWS Data Transfer USW2-USE1-AWS-In-Bytes

| $0.00 per GB - US West (Oregon) data transfer from US East (Northern Virginia) | 4,425.836 GB | $0.00 |

Nov 4th, 2019

CHEP2019 - Adelaide, Australia
## Outbound Network Traffic is Charged

AWS example – MS Azure and Google have same policy

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>Amount</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>$0.090 per GB - first 10 TB / month data transfer out beyond the global free tier</td>
<td>281.436 GB</td>
<td>$25.33</td>
</tr>
<tr>
<td>AWS Data Transfer USE1-USW2-AWS-Out-Bytes</td>
<td>$91.31</td>
<td></td>
</tr>
<tr>
<td>$0.02 per GB - US East (Northern Virginia) data transfer to US West (Oregon)</td>
<td>4,565,513 GB</td>
<td>$91.31</td>
</tr>
</tbody>
</table>

General Internet is more expensive than traffic between Cloud Regions.
Outbound Network Traffic is Charged

AWS example – MS Azure and Google have same policy

For research institutions, Network fees waived if under 15% of total monthly bill

General internet more expensive than traffic between Cloud Regions

Bandwidth

$0.090 per GB - first 10 TB / month
$0.02 per GB - US East (Northern Virginia) data transfer to US West (Oregon)

281,436 GB $25.33

4,565,513 GB $91.31

Great for compute heavy tasks, can be a problem for pure network testing.
Outbound Network Traffic is Charged

AWS example – MS Azure and Google have same policy
- General internet more expensive than traffic between Cloud Regions
- For research institutions, Network fees waived if under 15% of total monthly bill

15% of total bill means, 1GB of out-networking:
- every 12h of a single CPU core
- every 1h of a 18-core/36-thread CPU instance
- every 30 mins of a V100 GPU (using Oct’19 AWS spot pricing)

Great for compute heavy tasks, can be a problem for pure network testing.
Network prices clearly advertised

### Data Transfer OUT From Amazon EC2 To Internet

<table>
<thead>
<tr>
<th>Bandwidth</th>
<th>Price</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up to 1 GB / Month</td>
<td>$0.00 per GB</td>
</tr>
<tr>
<td>Next 9.999 TB / Month</td>
<td>$0.09 per GB</td>
</tr>
<tr>
<td>Next 40 TB / Month</td>
<td>$0.085 per GB</td>
</tr>
<tr>
<td>Next 100 TB / Month</td>
<td>$0.07 per GB</td>
</tr>
<tr>
<td>Greater than 150 TB / Month</td>
<td>$0.05 per GB</td>
</tr>
</tbody>
</table>

If the Data Transfer per month is greater than 500 TB / month, please [contact us](https://aws.amazon.com/ec2/pricing/on-demand/).

### Data Transfer OUT From Amazon EC2 To

#### MS Azure

<table>
<thead>
<tr>
<th>Outbound Data Transfers</th>
<th>Zone 1*</th>
</tr>
</thead>
<tbody>
<tr>
<td>First 5 GB /Month ¹</td>
<td>Free</td>
</tr>
<tr>
<td>5 GB - 10 TB ² /Month</td>
<td>$0.087 per GB</td>
</tr>
<tr>
<td>Next 40 TB (10 - 50 TB) /Month</td>
<td>$0.083 per GB</td>
</tr>
<tr>
<td>Next 100 TB (50 - 150 TB) /Month</td>
<td>$0.07 per GB</td>
</tr>
<tr>
<td>Next 350 TB (150 - 500 TB) /Month</td>
<td>$0.05 per GB</td>
</tr>
<tr>
<td>Over 500 TB /Month</td>
<td>Contact us</td>
</tr>
</tbody>
</table>


[https://cloud.google.com/storage/pricing#network-egress](https://cloud.google.com/storage/pricing#network-egress)
# AWS test example costs – US East to West

<table>
<thead>
<tr>
<th>Amazon Elastic Compute Cloud running Linux/UNIX Spot Instances</th>
<th>8.647 Hours</th>
<th>$4.91</th>
<th>AWS Data Transfer USE1-USW2-AWS-Out-Bytes</th>
<th>$91.31</th>
</tr>
</thead>
<tbody>
<tr>
<td>c5.4xlarge Linux/UNIX Spot Instance-hour in US West (Oregon) in VPC Zone #4 (blended rate: 0.264)</td>
<td>0.586 Hrs</td>
<td>$2.28</td>
<td>$0.02 per GB - US East (Northern Virginia) data transfer to US West (Oregon)</td>
<td></td>
</tr>
<tr>
<td>c5n.2xlarge Linux/UNIX Spot Instance-hour in US West (Oregon) in VPC Zone #1 (blended rate: 0.129)</td>
<td>4.817 Hrs</td>
<td>$0.70</td>
<td>$0.00 per GB - US West (Oregon) data transfer from US East (Northern Virginia)</td>
<td>$0.00</td>
</tr>
<tr>
<td>c5n.2xlarge Linux/UNIX Spot Instance-hour in US West (Oregon) in VPC Zone #2 (blended rate: 0.144)</td>
<td>10.945 Hrs</td>
<td>$1.86</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### AWS test example costs – US to Brazil

<table>
<thead>
<tr>
<th>Service</th>
<th>Cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>AWS Data Transfer USE1-USE1-AWS-Out-Bytes</td>
<td>$37.03</td>
</tr>
<tr>
<td>$0.02 per GB - US East (Northern Virginia) data transfer to South America (Sao Paulo)</td>
<td></td>
</tr>
<tr>
<td>AWS Data Transfer SAE1-USE1-AWS-In-Bytes</td>
<td>$0.00</td>
</tr>
<tr>
<td>$0.00 per GB - South America (Sao Paulo) data transfer from US East (Northern Virginia)</td>
<td></td>
</tr>
<tr>
<td>AWS Data Transfer SAE1-USE1-AWS-Out-Bytes</td>
<td>$2.19</td>
</tr>
<tr>
<td>$0.16 per GB - South America (Sao Paulo) data transfer to US East (Northern Virginia) (blended rate: 0.160)</td>
<td></td>
</tr>
<tr>
<td>AWS Data Transfer USW2-USE1-AWS-Out-Bytes</td>
<td>$46.67</td>
</tr>
<tr>
<td>$0.02 per GB - US West (Oregon) data transfer to South America (Sao Paulo)</td>
<td></td>
</tr>
<tr>
<td>AWS Data Transfer SAE1-USW2-AWS-In-Bytes</td>
<td>$0.00</td>
</tr>
<tr>
<td>$0.00 per GB - South America (Sao Paulo) data transfer from US West (Oregon)</td>
<td></td>
</tr>
<tr>
<td>AWS Data Transfer SAE1-USW2-AWS-Out-Bytes</td>
<td>$3.11</td>
</tr>
<tr>
<td>$0.16 per GB - South America (Sao Paulo) data transfer to US West (Oregon) (blended rate: 0.160)</td>
<td></td>
</tr>
</tbody>
</table>
Conclusion

• Cloud computing is starting to be an important part of many science domains
  • And networking has traditionally been the least well understood part of the Cloud model

• We show that Cloud providers invested heavily in networking, and in-Cloud networking is a non-issue
  • The largest science network providers also have mostly good connectivity to the Clouds

• Data egress can be expensive
  • But not a big issue if paired with Cloud computing, too
Acknowledgements

- This work has been partially sponsored by NSF grants OAC-1826967, OAC-1541349, OAC-1841530, OAC-1836650, MPS-1148698 and OAC-1941481.

- We kindly thank both Amazon and Microsoft for providing Cloud credits that covered a portion of the incurred Cloud costs.
External material used

- Background images downloaded from Wikimedia Commons
  - Lokal_Profil [Public domain]
  - Theshibboleth [CC BY-SA 3.0 (http://creativecommons.org/licenses/by-sa/3.0/)]
  - maix [CC BY-SA 2.5 (https://creativecommons.org/licenses/by-sa/2.5)]
- Cloud background downloaded from Pixabay