Deploying Splunk Enterprise On Microsoft Azure

Pramit Gupta
Senior Software Engineer, Microsoft Corporation

Roy Arsan
Solutions Architect, Splunk
Disclaimer

During the course of this presentation, we may make forward looking statements regarding future events or the expected performance of the company. We caution you that such statements reflect our current expectations and estimates based on factors currently known to us and that actual events or results could differ materially. For important factors that may cause actual results to differ from those contained in our forward-looking statements, please review our filings with the SEC. The forward-looking statements made in the this presentation are being made as of the time and date of its live presentation. If reviewed after its live presentation, this presentation may not contain current or accurate information. We do not assume any obligation to update any forward looking statements we may make. In addition, any information about our roadmap outlines our general product direction and is subject to change at any time without notice. It is for informational purposes only and shall not, be incorporated into any contract or other commitment. Splunk undertakes no obligation either to develop the features or functionality described or to include any such feature or functionality in a future release.
Objective:

Deploy, manage & integrate your Splunk Enterprise deployment in Azure
Roy Arsan

- 4 Years @ Splunk
- Roles in:
  - Product Engineering
  - Cloud Architecture
  - Partner Integrations
- Currently focused on cloud partner ecosystem

Pramit Gupta

- 14 Years @ Microsoft
- Roles in:
  - SharePoint Enterprise Content Management
  - Office.com Portal
  - Office Client Monitoring and Telemetry
- Currently focused on cloud services and data analysis
Agenda

- Azure IaaS
- Splunk Azure Deployment @ Microsoft Office
- Provisioning & Automation
- Azure Best Practices
- Splunk & Azure Integrations
Agenda

- Azure IaaS
- Splunk Azure Deployment @ Microsoft Office
- Provisioning & Automation
- Azure Best Practices
- Splunk & Azure Integrations
Azure Infrastructure
Why Azure?

- Top IaaS market leaders
- 120k+ new Azure customer sub/month
- 5M organizations using Azure AD
- Commercial cloud exceeds $10B annual run rate, $20B+ by 2018

Gartner IaaS Magic Quadrant, 2016
Azure Virtual Machines (VM)

- Available in 24 Regions

- Billing:
  - Pay-As-You-Go or Prepaid (5% discount)
  - Per-minute basis
Azure VM Selection

- **VM Image:**
  - Linux & Windows
  - Extra $ for Windows
    (additional cost of 40% to 90%)

- **VM Disks:**
  - Local temporary disk
  - Network-attached persistent disks (VHD)
    ‣ 1 OS disk
    ‣ 1+ data disks (up to 64)

- **VM Size Recommendation:**
  - 8+ CPU cores
  - 14GB+ RAM
  - Multiple data disks (6+)
  - Compute-optimized:
    ‣ **Dv2-series** (& DSv2-series)
    ‣ **F-series** (& Fs-series) – Best !/$
# Azure VM Selection

## Indexers

<table>
<thead>
<tr>
<th>VM Size</th>
<th>Daily Indexing Volume (GB)</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard_F8(s)</td>
<td>Up to 100</td>
<td>Good</td>
</tr>
<tr>
<td>Standard_F16(s)</td>
<td>100-150</td>
<td>Better</td>
</tr>
<tr>
<td>Standard_D(S)15_v2</td>
<td>150-250</td>
<td>Best</td>
</tr>
</tbody>
</table>

## Search Heads

<table>
<thead>
<tr>
<th>VM Size</th>
<th>Concurrent Users</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard_F16(s)</td>
<td>Up to 8</td>
<td>Good</td>
</tr>
<tr>
<td>Standard_D(S)15_v2</td>
<td>Up to 16</td>
<td>Better</td>
</tr>
</tbody>
</table>

## Deployment Server, License or Cluster Master

<table>
<thead>
<tr>
<th>VM Size</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard_F4(s)</td>
<td>Good</td>
</tr>
<tr>
<td>Standard_F8(s)</td>
<td>Better</td>
</tr>
</tbody>
</table>
Disk Storage Selection

- VHD stored as Page Blob in Azure Storage
- High durability & availability
- Scalable
- Two types of VHDs:
  - Standard Storage (HDD)
    - Up to 1TB
  - Premium Storage (SSD)
    - 3 sizes: 128GB, 512GB, 1TB
Disk Storage Selection

- **Premium storage recommended**
  - Consistent high throughput and low latency

- **Standard storage also feasible....if configured correctly**
  - Minimum 6+ disks striped in RAID0
  - Enable “ReadOnly” host caching
  - More economical

<table>
<thead>
<tr>
<th>Disk Type</th>
<th>IOPS</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>500 (8KB)</td>
<td>Up to 60MB/s</td>
</tr>
<tr>
<td>Premium</td>
<td>5,000 (256KB)</td>
<td>Up to 200MB/s</td>
</tr>
</tbody>
</table>
Technical Brief - Splunk On Azure

Agenda

- Azure IaaS
- **Splunk Azure Deployment @ Microsoft Office**
- Provisioning & Automation
- Azure Best Practices
- Splunk & Azure Integrations
Splunk Azure Deployment @ Microsoft Office
Office Client Telemetry - Challenges

• Delight customers, improve satisfaction
  – Respond quickly to feedback and fix bugs effectively
  – Move from multi year cycle to rapid releases

• Office client applications collect 100s of TB diagnostics data per day
  – Enabling engineers to browse through this much data isn’t easy
  – Regression risk, huge legacy complex code base, shared components
Office Client Telemetry - Splunk

Splunk provides near real time search and diagnostics capability to Office engineers

Examples

- Observe crash rates of an application
- Adoption of a new feature, a new button in the Office ribbon
- Dashboard to quickly monitor and alert on key metrics
Office Client Telemetry - Goals

- Data access
  - 24hrs+ down to 30min

- MTTD
  - Several days down to 6hrs
Office Client Splunk - Architecture
Office Client Splunk - Architecture

Office Apps

Billion devices
Office Client Splunk - Architecture

Office Apps

Nexus

Nexus Western US

Payload Processor

Azure Blob Storage
Office Client Splunk - Architecture

Office Apps

Nexus Global Stamp

Nexus Western US
Payload Processor
Azure Blob Storage
Nexus South East Asia
Nexus Eastern Europe
......
......

10 Azure regions worldwide
Office Client Splunk - Architecture

Office Apps

Nexus Global Stamp

Nexus Western US

Payload Processor

Zipped Batch

Azure Blob Storage

Nexus South East Asia

Nexus Eastern Europe

......
Office Client Splunk - Architecture

Office Apps

Nexus Global Stamp

- Nexus Western US
  - Payload Processor
    - Token
    - Zipped Batch
- Azure Blob Storage
- Azure Event Hub
- Nexus South East Asia
- Nexus Eastern Europe
- .....
Office Client Splunk - Architecture

Office Apps

Nexus Global Stamp

Payload Processor

Token

Zipped Batch

Azure Blob Storage

Nexus Western US

Nexus South East Asia

Nexus Eastern Europe

......

......

Azure Event Hub

Data Retriever
Office Client Splunk - Architecture

Office Apps

Nexus Global Stamp

- Payload Processor
  - Token
  - Zipped Batch
- Azure Blob Storage
- Nexus Western US
- Nexus South East Asia
- Nexus Eastern Europe
- .......

Splunk

- Universal Forwarder
- Data Retriever
Office Client Splunk - Architecture

Office Apps

TRE

Office Apps

Nexus Global Stamp

Payload Processor

Token

Zipped Batch

Azure Blob Storage

Azure Event Hub

Universal Forwarder

Data Retriever

Splunk Scripted Input

Splunk

Nexus Western US

Nexus South East Asia

Nexus Eastern Europe

......

......
Office Client Splunk - Architecture

- **Nitix Western US**: Payload Processor
- **Nitix South East Asia**: Data Retriever
- **Nitix Eastern Europe**: Data Retriever
- **Nitix Eastern Asia**: Data Retriever
- **Nitix South East Asia**: Data Retriever

- **Azure Blob Storage**
- **Azure Event Hub**
- **Universal Forwarders**
  - 6.2.3 Universal Forwarders
  - 100 instances of D12’s
  - 4 cores, 28 GB RAM, local SSD

- **Office Apps**
- **Nexus Global Stamp**
- **Splunk**
Office Client Splunk - Architecture

Nexus Western US
- Payload Processor
- Azure Blob Storage
- Token
- Zipped Batch

Nexus South East Asia
- Nexus Eastern Europe
- ......
- ......

Azure Event Hub
- Universal Forwarder
- Data Retriever

Indexer Cluster
- Indexer

Office Apps
- TRE

Nexus Global Stamp

Splunk
Office Client Splunk - Architecture

- **Nexus Western US**
  - Payload Processor
  - Token
  - Zipped Batch
  - Azure Blob Storage

- **Nexus South East Asia**
- **Nexus Eastern Europe**
  - ......

- **Azure Event Hub**

- **Universal Forwarder**
  - Data Retriever

- **Indexer Cluster**
  - Indexer

- **Splunk**
  - 50 instances of G4’s
  - 16 cores, 224 GB RAM, 3TB SSD

Office Apps

Nexus Global Stamp

Azure Blob Storage
Office Client Splunk - Architecture

- Removed forwarder tier
- Ran connector exe directly on indexers
Office Client Splunk - Architecture

- Ensured uniform data distribution
- 600 partitions
Office Client Splunk - Architecture

- Nexus Western US
  - Payload Processor
  - Token
  - Zipped Batch
  - Azure Blob Storage

- Nexus South East Asia
  - Nexus Eastern Europe
  - Indexer Cluster
  - Indexer
  - Data Retriever

- Azure Event Hub

- Single cluster master with RF = 1, SF = 1
- 300 instances of DS14’s
- 16 cores, 112 GB RAM
- Cluster bundle push problem

Office Apps

Nexus Global Stamp

Splunk
Office Client Splunk - Architecture

Office Apps

Nexus Global Stamp

Payload Processor

Token

Zipped Batch

Azure Blob Storage

Nexus Western US

Nexus South East Asia

Nexus Eastern Europe

......

......

Azure Event Hub

Indexer Cluster

Indexer

Data Retriever

Splunk

• Sub-optimal search performance with Azure standard storage
Office Client Splunk - Architecture

- **Nexus Western US**
  - Payload Processor
  - Token
  - Zipped Batch
  - Azure Blob Storage

- **Nexus South East Asia**
- **Nexus Eastern Europe**
- **……**
- **……**

- **Azure Event Hub**

- **Indexer Cluster**
  - Indexer
  - Data Retriever

- **Splunk**
  - 4TB Premium Storage, (Raid0, 4 discs * 1TB each)
  - Indexing 2000+ MB/s
  - 25,000 IOPS
Office Client Splunk - Architecture

Nexus Western US
- Payload Processor
  - Token
  - Zipped Batch
- Azure Blob Storage

Nexus South East Asia

Nexus Eastern Europe

......

......

Azure Event Hub

Indexer Cluster
- Indexer
  - Data Retriever
- Search Head Cluster
  - Search Head

- 30 instances of D14v2’s
- 16 cores, 112 GB RAM

Office Apps

Nexus Global Stamp

Splunk

Payload Processor

Zipped Batch

Token

Azure Event Hub

Data Retriever

Indexer

Search Head

Splunk

Office Apps

Nexus Global Stamp

Splunk

Payload Processor

Zipped Batch

Token

Azure Event Hub

Data Retriever

Indexer

Search Head

Splunk
Office Client Splunk - Architecture

Office Apps

Nexus Global Stamp

- Payload Processor
  - Token
  - Zipped Batch
- Azure Blob Storage
- Nexus Western US
- Nexus South East Asia
- Nexus Eastern Europe
- ....
- ....

Splunk

- Indexer Cluster
  - Indexer
  - Data Retriever
- Search Head Cluster
  - Search Head
  - Search Deployer used for cluster configuration

Azure Event Hub
Office Client Splunk - Architecture

- Nexus Western US
  - Payload Processor
  - Token
  - Zipped Batch
  - Azure Blob Storage
- Nexus South East Asia
- Nexus Eastern Europe
- Azure Event Hub
- Indexer Cluster
  - Indexer
  - Data Retriever
- Search Head Cluster
  - Search Head

- Office Apps
- Nexus Global Stamp
- Splunk
  - Connected to corpnet for Office engineers in Seattle
  - Incorporating Azure Active Directory (AAD) in Splunk 6.4
Office Client Splunk - Metrics

Data access: 24hrs+ down to 30min, most of the time < 5min
MTTD: Several days down to 6hrs, most of the time < 30min
Agenda

- Azure IaaS
- Splunk Azure Deployment @ Microsoft Office
- Provisioning & Automation
- Azure Best Practices
- Splunk & Azure Integrations
Provisioning & Automation
Cloud Provisioning Tools

- Powershell, Chef, Puppet for machine provisioning

- ARM templates for deployment provisioning
  - Available via Portal, CLI or Powershell
Windows Azure Powershell

PS used extensively to manage Azure instances
Import “Azure PowerShell” module

- Deployment: Remotely provision various Splunk roles
- Configuration: Modify Splunk system local configuration files
- Manage: Install critical Windows Updates as well as planned maintenance
- Storage: Attach Premium Storage disks, format & create single volume
Windows Azure Powershell - Example

Storage
Attach premium storage disks, format & create single volume

- $PhysicalDisks = Get-StorageSubSystem -FriendlyName "Storage Spaces*" | Get-PhysicalDisk -CanPool $True | Where-Object {$_.FriendlyName -ne "PhysicalDisk0"}

- New-StoragePool -FriendlyName "SplunkPool" -StorageSubsystemFriendlyName "Storage Spaces*" -PhysicalDisks $PhysicalDisks | New-VirtualDisk -FriendlyName "SplunkDisk" -Interleave $StripeSize -NumberOfColumns $DiskCount -ResiliencySettingName simple -UseMaximumSize | Initialize-Disk -PartitionStyle GPT -PassThru | New-Partition -DriveLetter H -UseMaximumSize

Azure ARM Templates

- Automated deployment provisioning via ARM templates
- Deploy via Azure portal, Azure PowerShell or Azure CLI:

  $ azure group create -n SplunkRG -l "East US"

  $ azure group deployment create -n SplunkCluster -g SplunkRG \
  -f azuredeploy.json -e azuredeploy.parameters.json
Splunk Enterprise in Azure Marketplace

Demo Time

Microsoft Azure

Why Azure Solutions Products Documentation Pricing Partners Blog Resources Support

Splunk Enterprise
by Splunk

Deploy
Agenda

- Azure IaaS
- Splunk Azure Deployment @ Microsoft Office
- Provisioning & Automation
- Azure Best Practices
- Splunk & Azure Integrations
Azure Best Practices
Best Practices - Scalability

- Multiple Storage Accounts
  - Standard storage: 20,000 IOPS limit per account
    - No more than 40 disks per standard storage account
  - Premium storage: 50 Gbps limit per account

- Tiered Storage
  - Use both standard & premium for hot/cold data tiering
    - Optimal performance & cost tradeoff
Best Practices - Availability

- Azure Availability Sets
  - VMs on different update & fault domains

- Backup
  - VHD Snapshots
  - Azure Blob storage

- Archive
  - Archive indexes with Hunk into HDFS-compatible Azure Blob Storage
DEPLOYING SPLUNK® ENTERPRISE ON MICROSOFT® AZURE

Splunk provides the leading platform for Operational Intelligence. Splunk software searches, monitors, analyzes and visualizes machine-generated big data from websites, applications, servers, networks, sensors and mobile devices. More than 7,000 organizations around the world use Splunk to make sense of their critical data and to gain insight into the operation of their business. Splunk's Simple, Secure, and Scalable deployment allows you to deploy and scale Splunk to meet your business needs through a variety of deployment options.

DEPLOYING SPLUNK® ENTERPRISE ON MICROSOFT® AZURE

Splunk provides the leading platform for Operational Intelligence. Splunk software searches, monitors, analyzes and visualizes machine-generated big data from websites, applications, servers, networks, sensors and mobile devices. More than 7,000 organizations around the world use Splunk to make sense of their critical data and to gain insight into the operation of their business. Splunk's Simple, Secure, and Scalable deployment allows you to deploy and scale Splunk to meet your business needs through a variety of deployment options.

Technical Brief - Splunk on Azure

Agenda

- Azure IaaS
- Splunk Azure Deployment @ Microsoft Office
- Provisioning & Automation
- Azure Best Practices
- Splunk & Azure Integrations
Splunk & Azure Integrations
Splunk & Azure Integrations

- Splunk Enterprise SSO support for Azure AD as of 6.4
- Splunk Add-on for Microsoft Cloud Services
- Splunk Add-on for Azure
What Now?

Related breakout sessions and activities...

- **Splunking Azure: Gain Insights into your Microsoft Azure Data using Splunk** by Jason Conger & Cory Fowler (Wed Sep 28, 4:35-5:20pm)

- **Splunks of War: Creating a better game development process through data analytics** by Phil Cousins (Tue Sep 27, 3:15-4:00pm)
THANK YOU
Example Deployment

16 concurrent users

1 TB/day

4 months retention
Example Clustered Deployment

8 concurrent users

500 GB/day
60 days retention