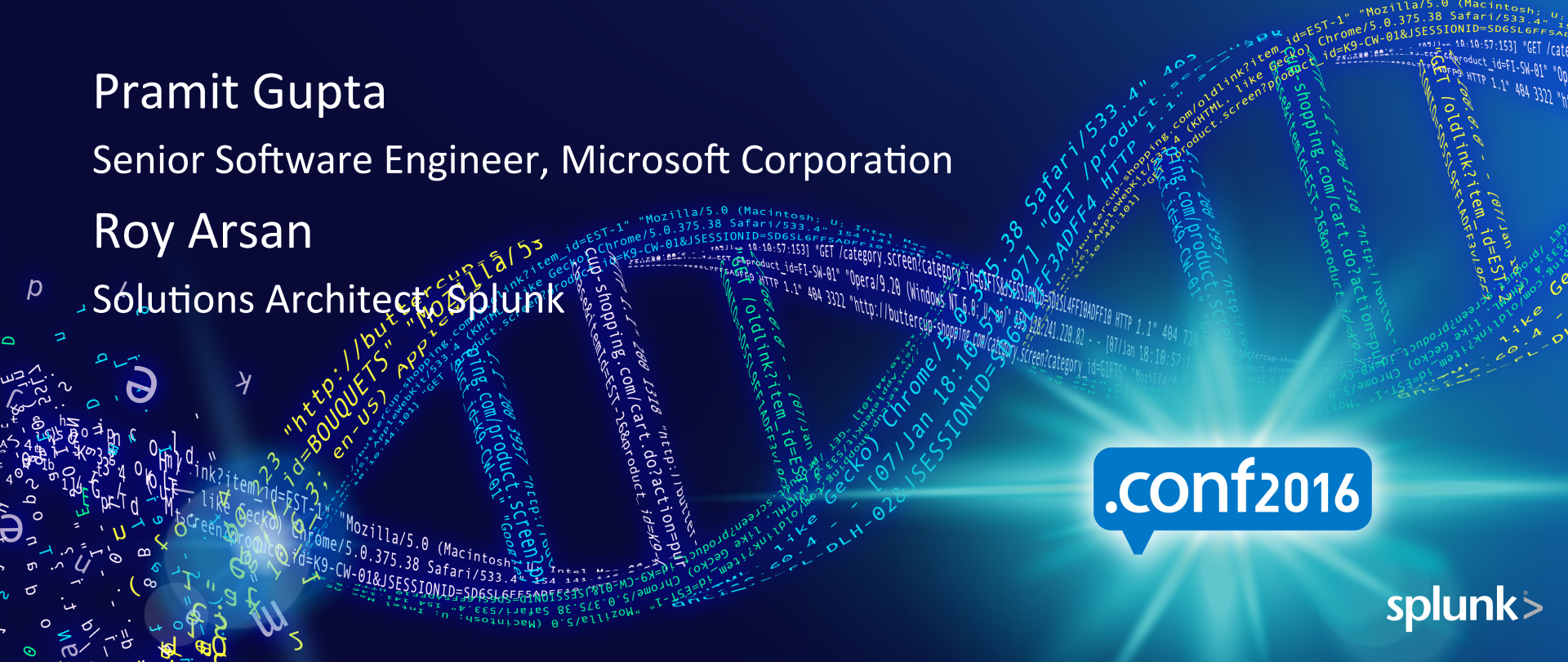
Deploying Splunk Enterprise On Microsoft Azure

Pramit Gupta

Senior Software Engineer, Microsoft Corporation

Roy Arsan

Solutions Architect, Splunk



.conf2016

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

Bios

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

VM Size	Daily Indexing Volume (GB)	Performance
Standard_F8(s)	Up to 100	Good
Standard_F16(s)	100-150	Better
Standard_D(S)15_v2	150-250	Best

Search Heads

VM Size	Concurrent Users	Performance
Standard_F16(s)	Up to 8	Good
Standard_D(S)15_v2	Up to 16	Better

Deployment Server,
License or Cluster Master

VM Size	Performance
Standard_F4(s)	Good
Standard_F8(s)	Better

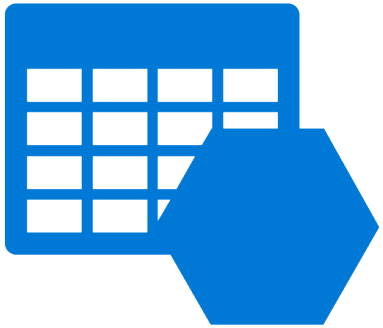
Disk Storage Selection

- 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

Disk Type	IOPS	Throughput
Standard	500 (8KB)	Up to 60MB/s
Premium	5,000 (256KB)	Up to 200MB/s



Technical Brief - Splunk On Azure

https://www.splunk.com/pdfs/technical-briefs/deploying-splunk-enterprise-on-microsoft-azure.pdf

splunk> TECH BRIEF

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 11,000 organizations use Splunk software to deepen business and customer understanding, mitigate cybersecurity risk, improve service performance and reduce costs. Splunk Enterprise indexes machine data in real time, enabling multiple roles across the organization—from system administrators to business analysts—to rapidly gain insight from the massive amounts of machine data generated by your environment.

Adopting a cloud strategy enables organizations to increase agility, reduce costs, decrease time to market and empower innovation. Splunk Enterprise is perfect for deploying in a cloud environment, offering enterprise-grade availability and scalability to support the collection of hundreds of terabytes of data per day from workloads residing on-premises, in the cloud or across hybrid environments. This document covers guidelines for deploying Splunk Enterprise on Microsoft Azure, an open and flexible cloud platform with a growing collection of integrated cloud services, including analytics, computing, database, mobile, networking, storage and web.

Splunk Deployment Components

A typical Splunk deployment includes Splunk forwarders, indexers and search heads. Splunk Enterprise is a single package that can perform one or many of the roles that each component would normally deliver, in addition to others. The software can be installed within minutes on your choice of hardware (physical, cloud or virtual) and operating system. The package is available publicly via the Azure Marketplace as a single-instance or a multi-instance Azure Resource Manager (ARM) solution template, in addition to downloadable packaged forms for most operating systems. While all major Splunk components can be run from a single installation on a single cloud instance, they can also run independently from within different cloud instances. Depending on the deployment infrastructure, considerations must also be taken to allocate the proper amount of resources per component type.

Forwarders perform data collection, data forwarding and data load balancing. Low amounts of resources are required to run a forwarder as they typically read and send data with minimal overhead. A Universal Forwarder is a lightweight package of the Splunk software that can perform most, if not all, of the forwarder functionality.

Indexers write the data to a storage device and perform searching on the data. These can be resource intense and require I/O and CPU allotment.

Search heads search for information across indexers and require CPU and memory allotment.

Budgeting system resources and bandwidth to enable search and index performance depend on the total volume of data being indexed and the number of active concurrent searches (scheduled or otherwise) at any time.

In addition to rapidly writing data to disk, indexers perform much of the work involved in running searches: reading data off disk, decompressing it, extracting knowledge and reporting. Since indexers incur most of the workload, increases in indexing volume should be tied to an increase in indexer instances. Deploying additional indexers will distribute the load of increased data volume, resulting in reduced contention for resources and improved search performance.

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

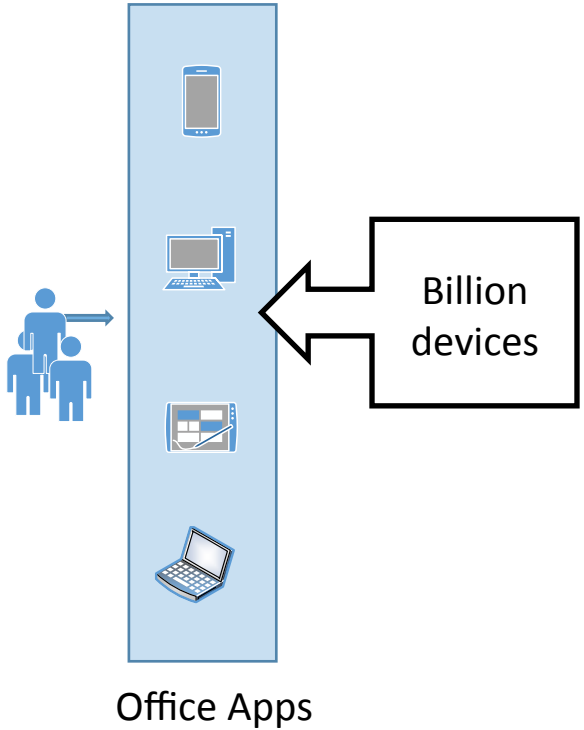
- Observe crash rates of an application
- Adoption of a new feature, a new button in the Office ribbon
- Dashboard to quickly monitor and alert on key metrics

Office Client Telemetry - Goals

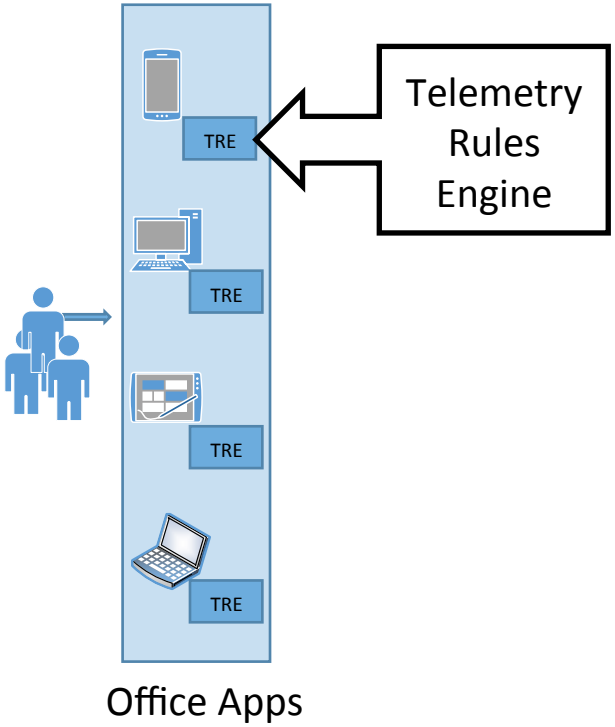
- Data access
  - 24hrs+ down to 30min
- MTTD
  - Several days down to 6hrs

Office Client Splunk - Architecture

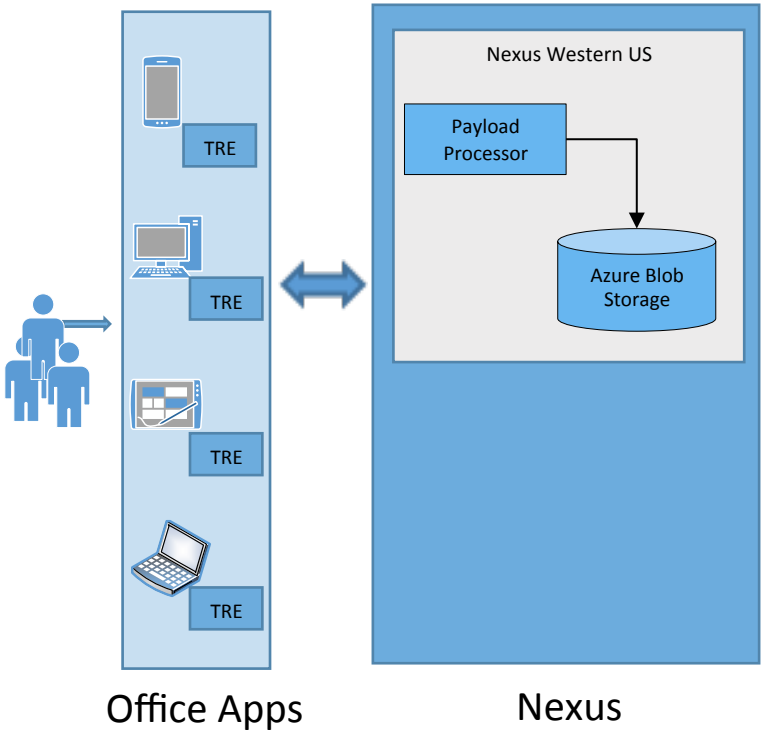
Office Client Splunk - Architecture



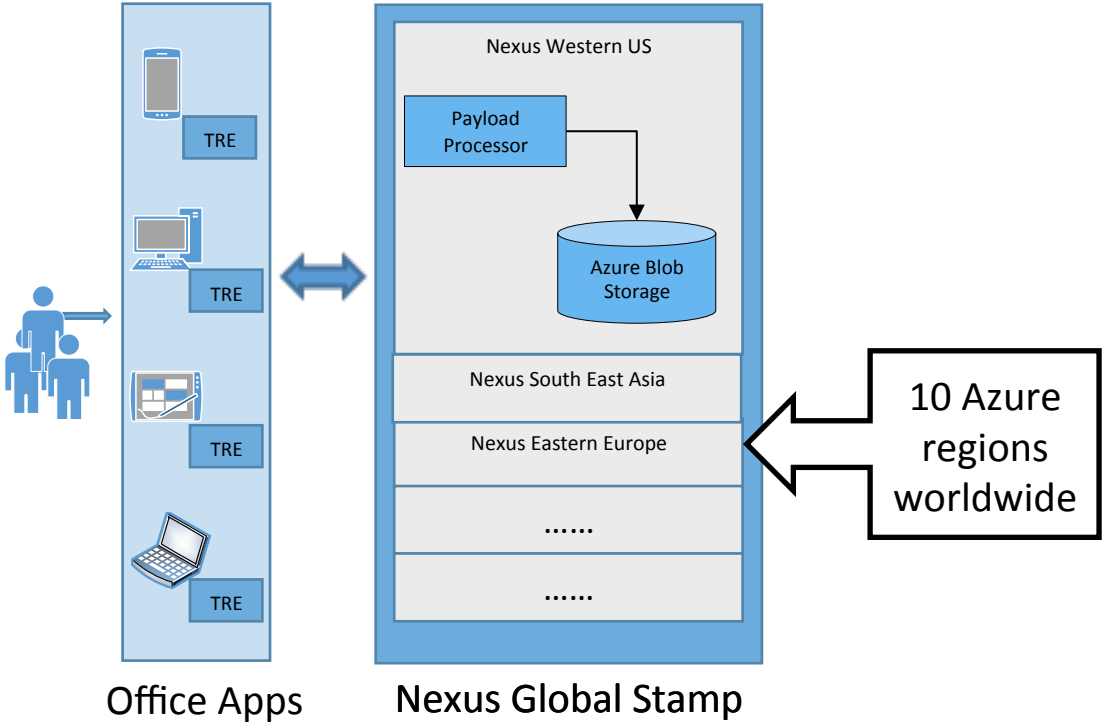
Office Client Splunk - Architecture



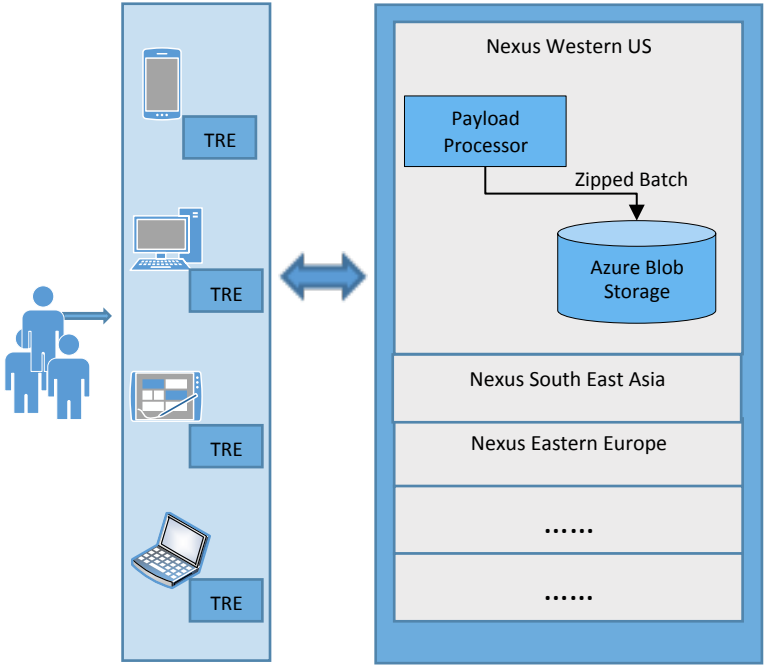
Office Client Splunk - Architecture



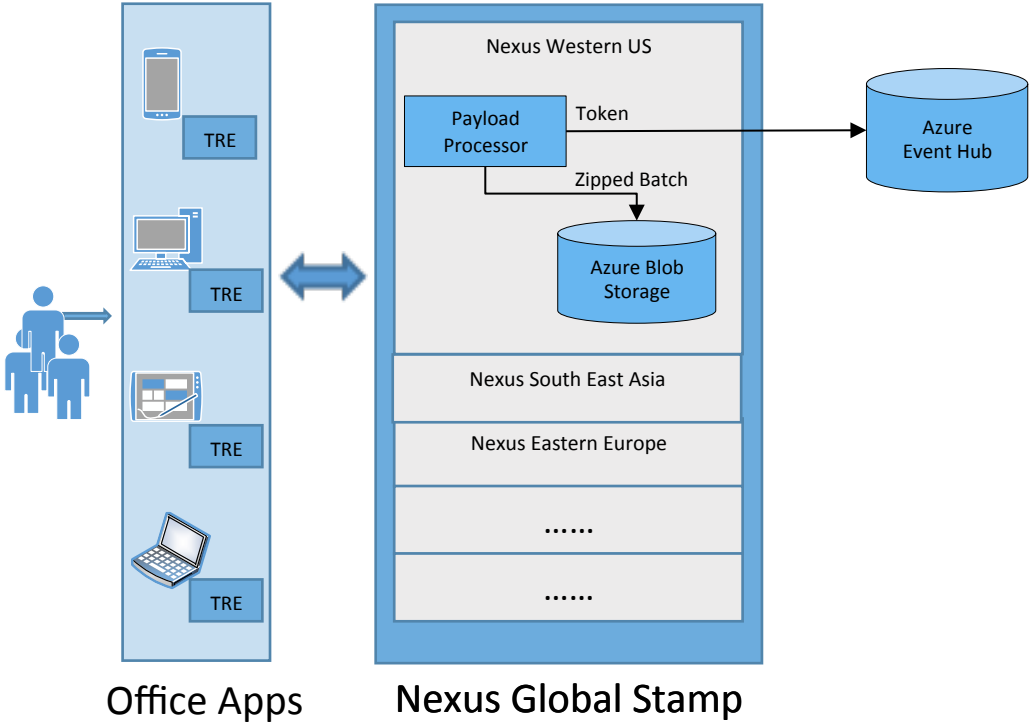
Office Client Splunk - Architecture



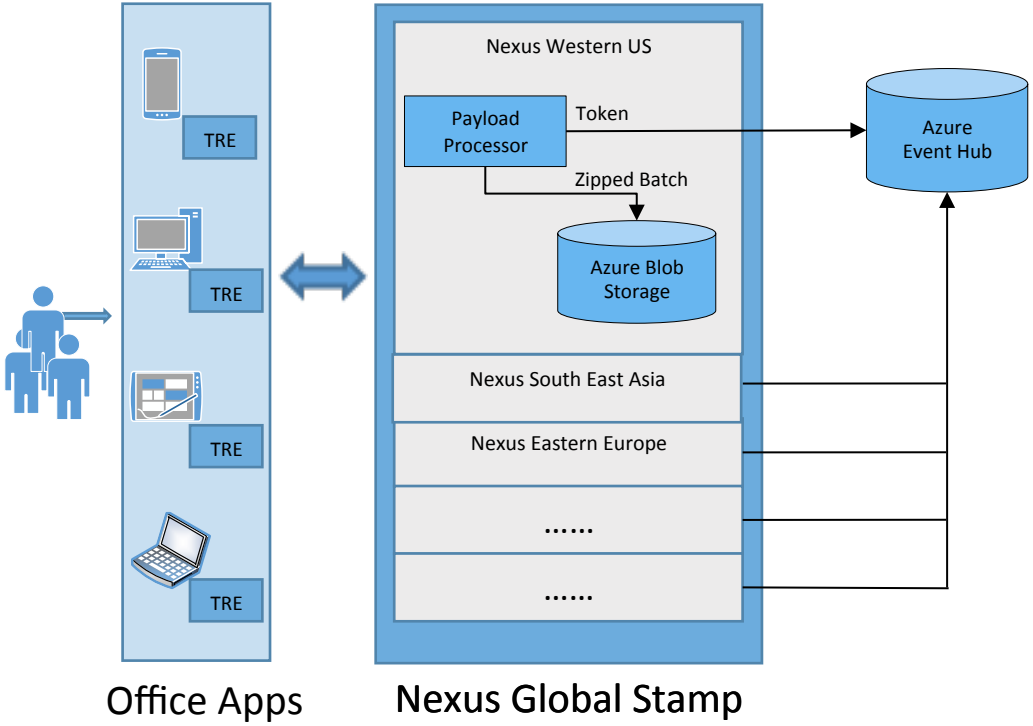
Office Client Splunk - Architecture



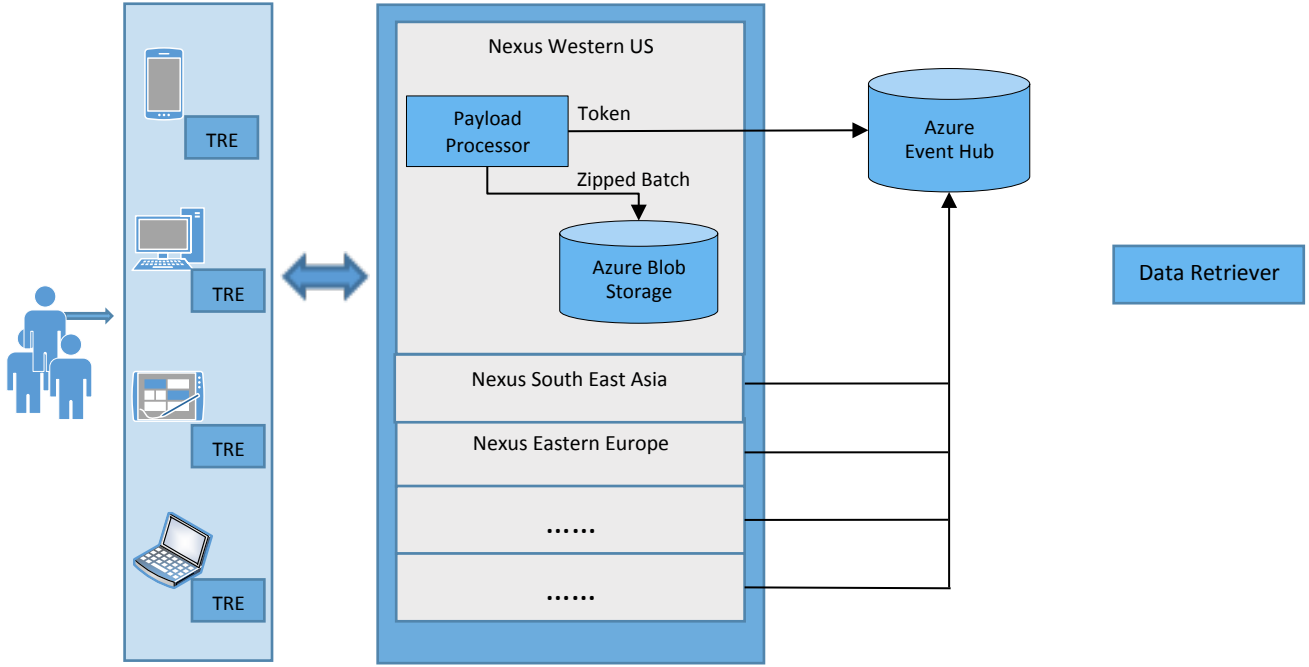
Office Client Splunk - Architecture



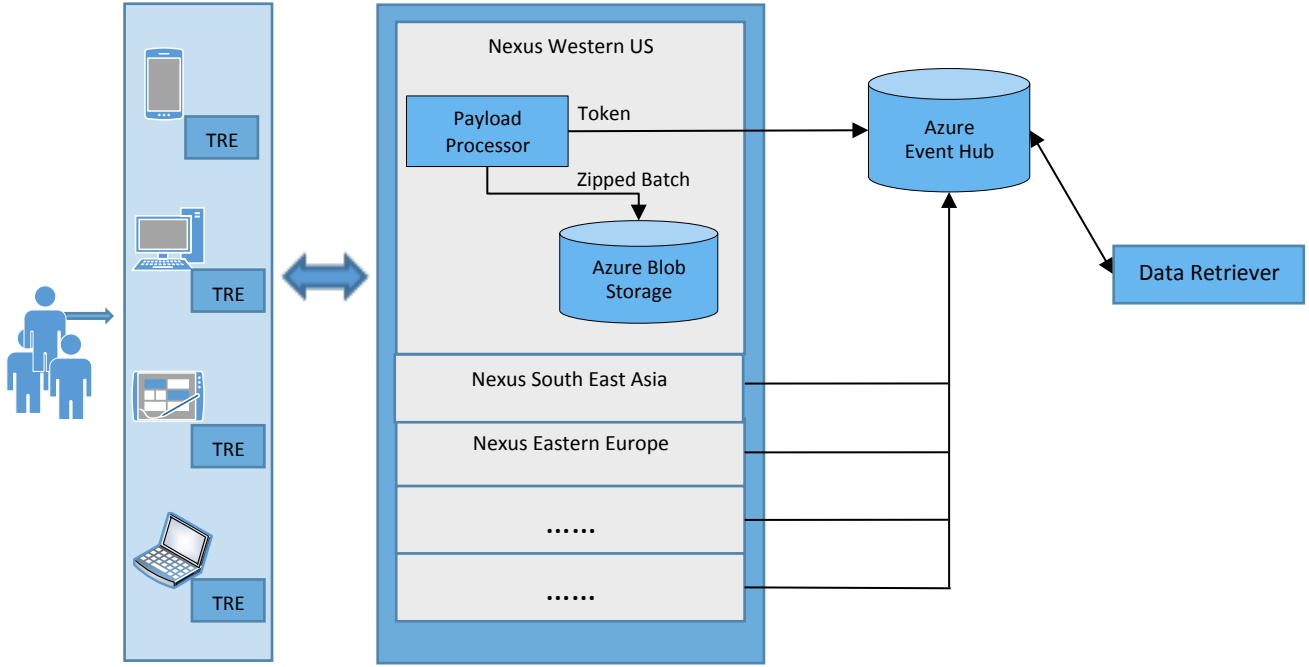
Office Client Splunk - Architecture



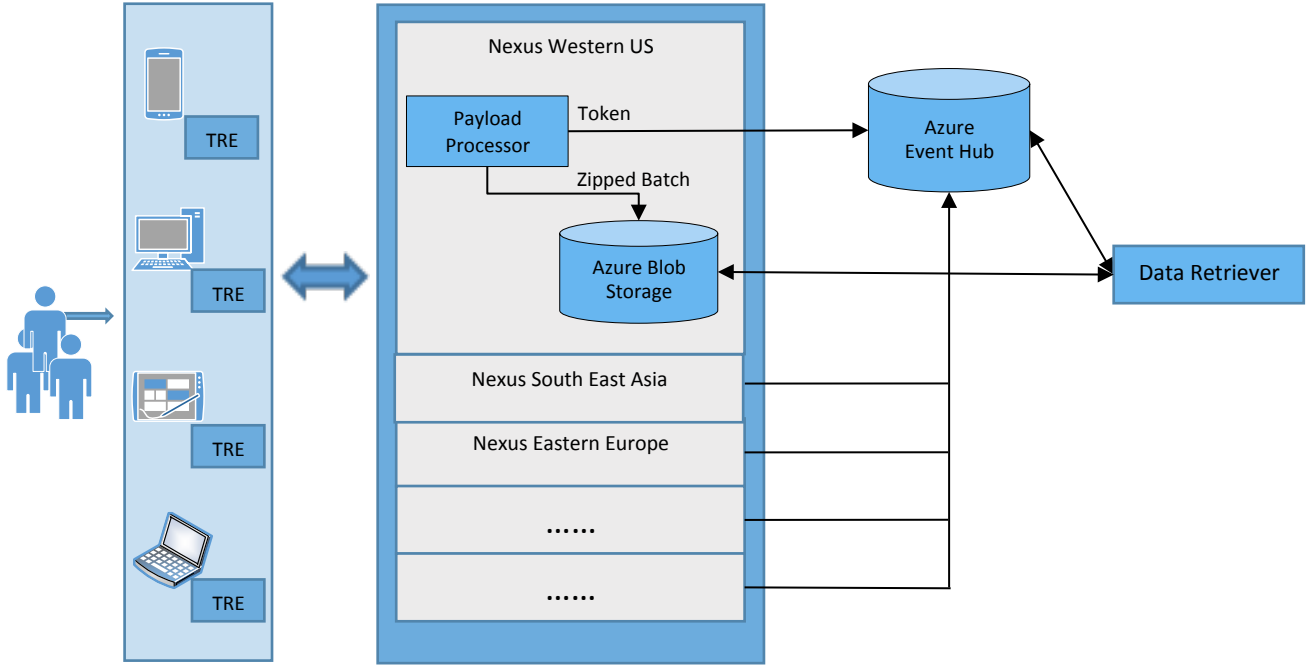
Office Client Splunk - Architecture



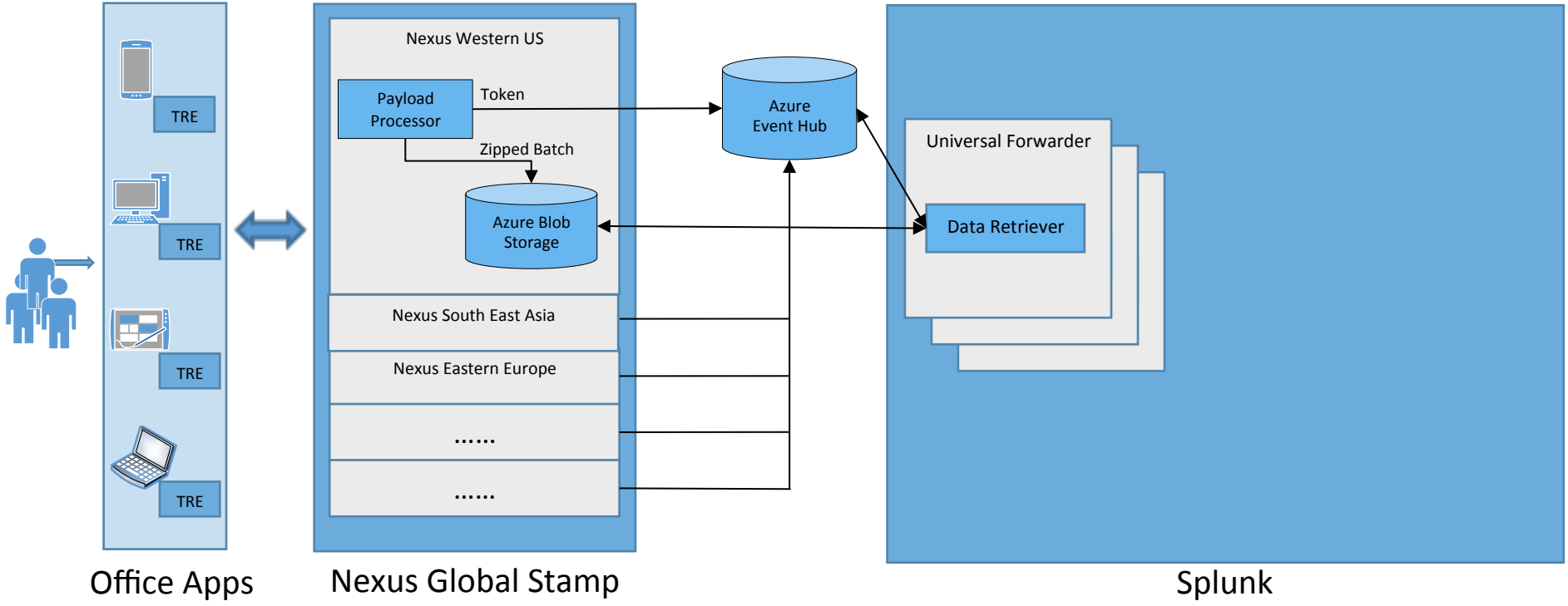
Office Client Splunk - Architecture



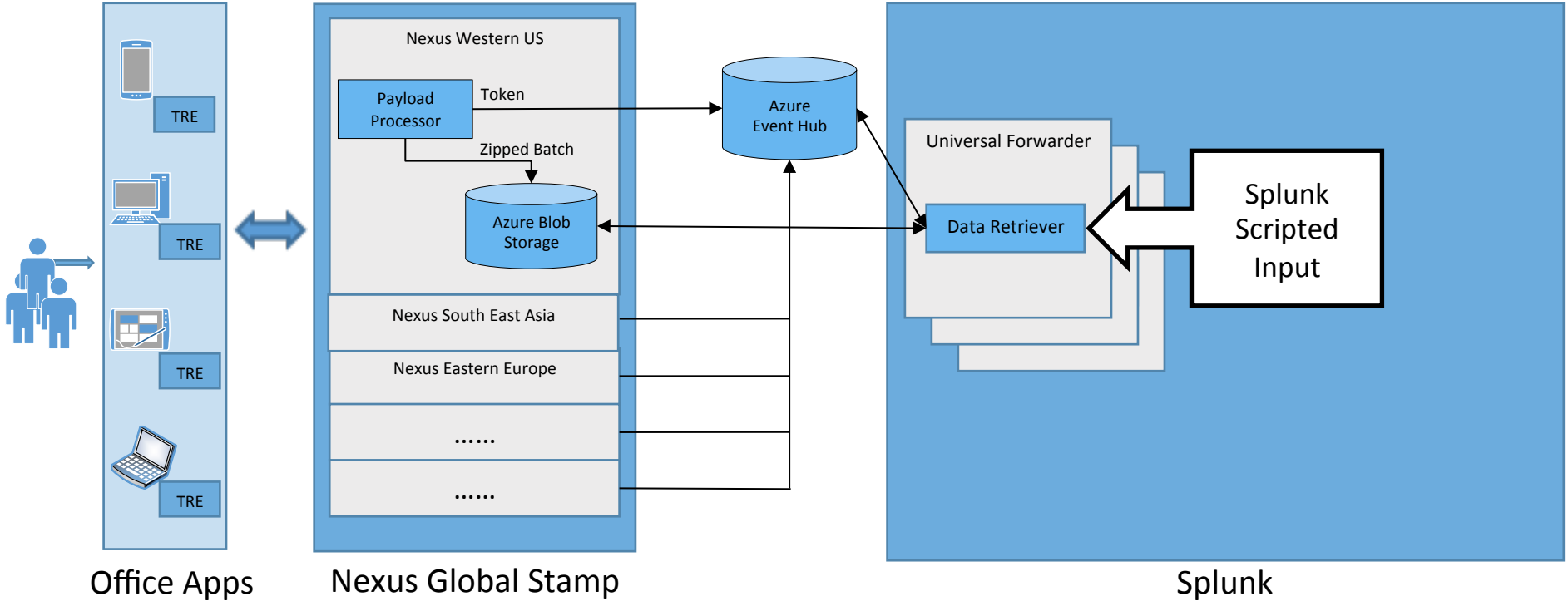
Office Client Splunk - Architecture



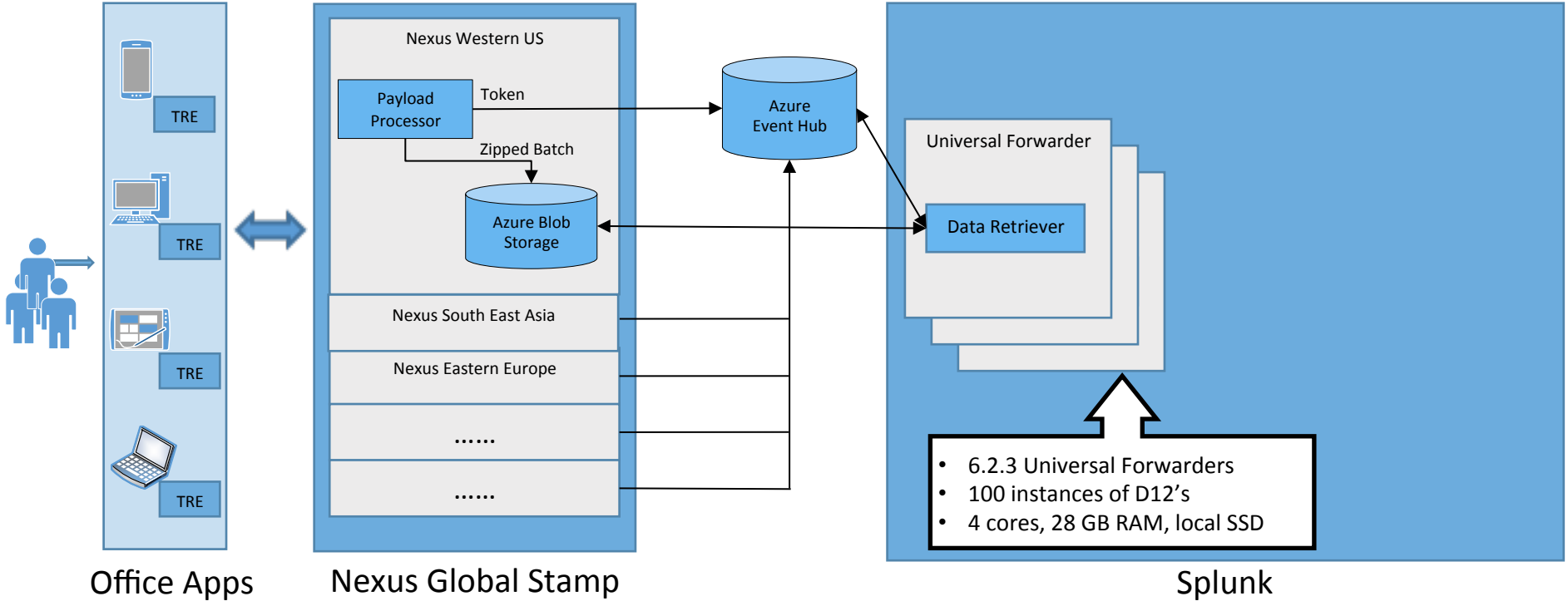
Office Client Splunk - Architecture



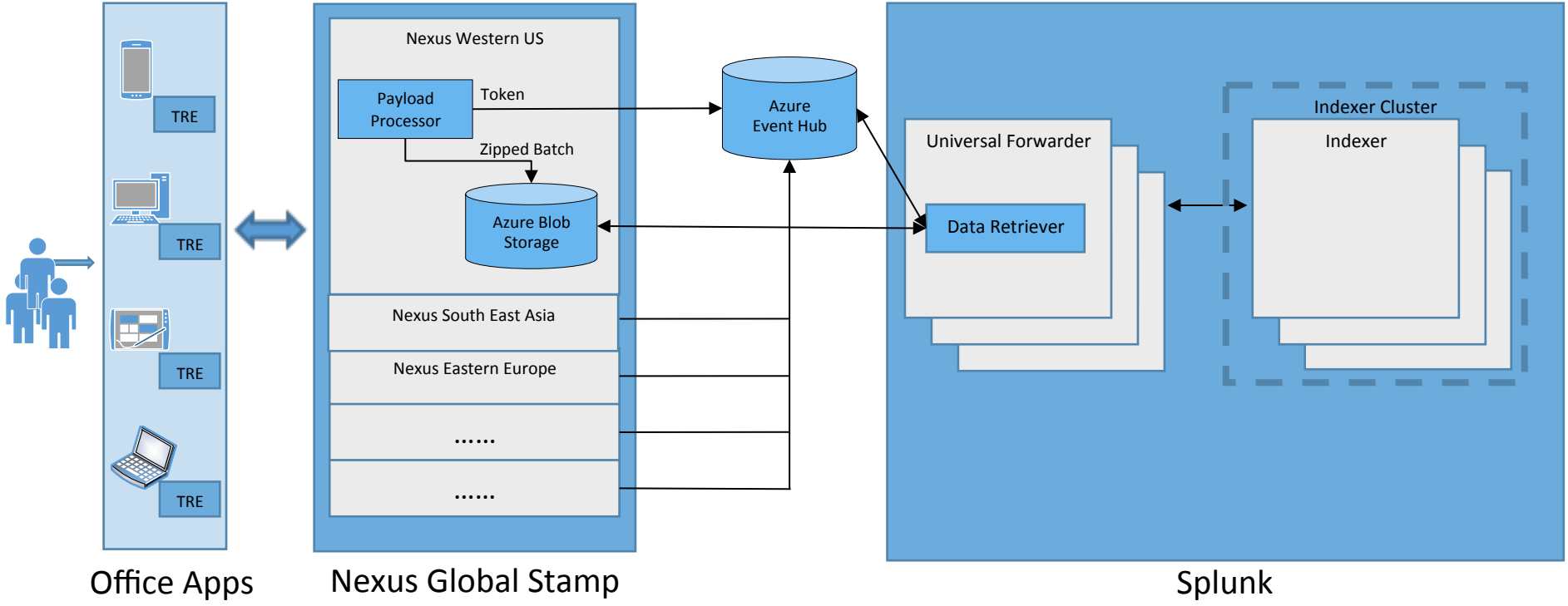
Office Client Splunk - Architecture



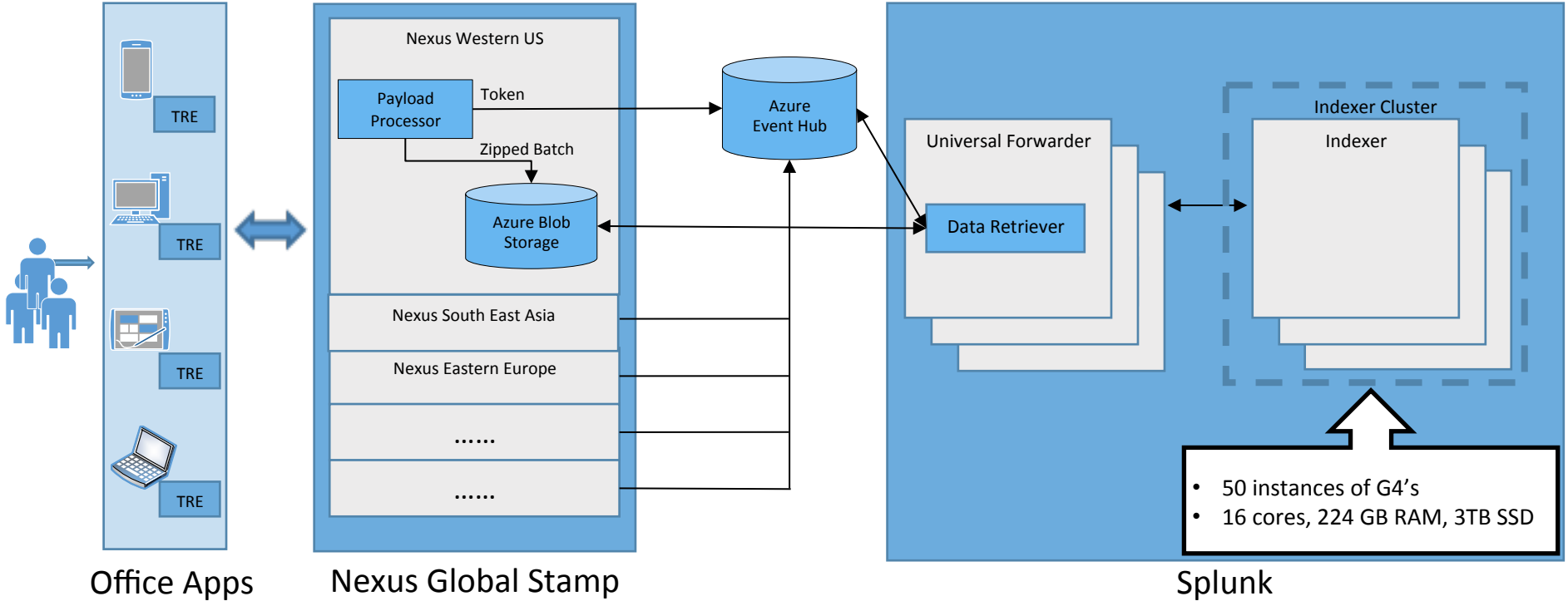
Office Client Splunk - Architecture



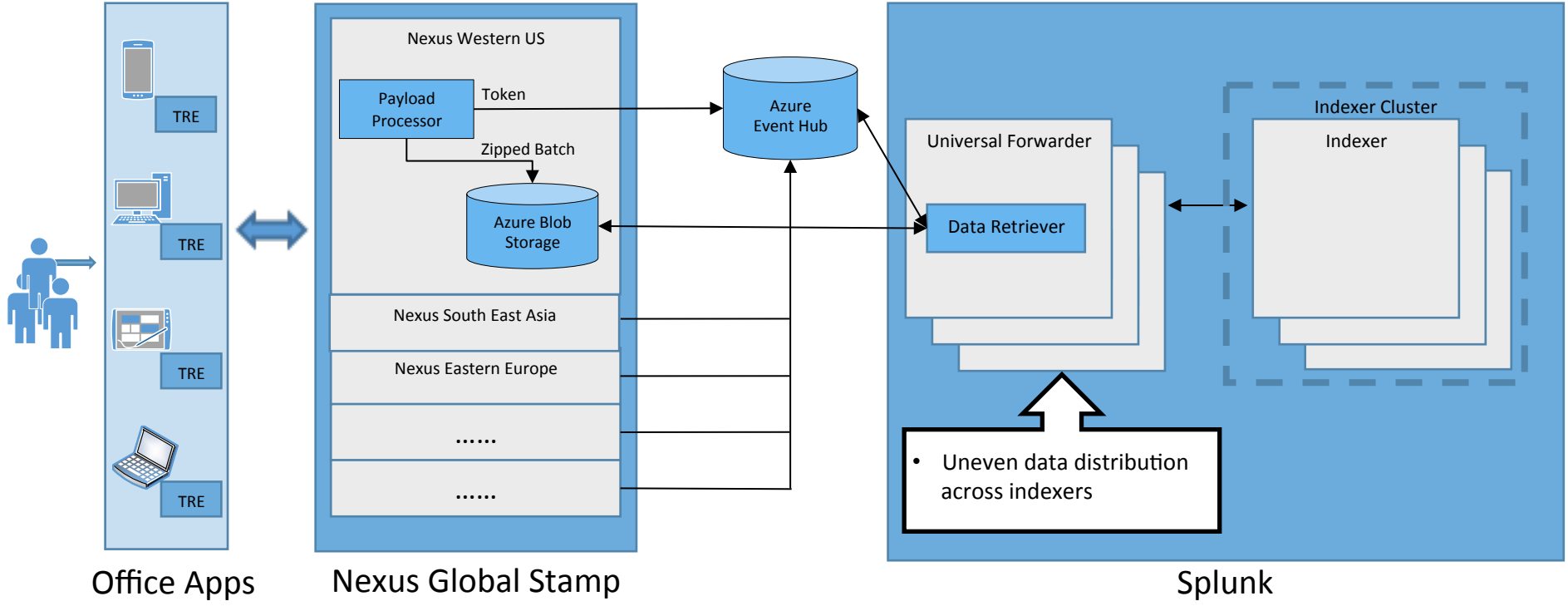
Office Client Splunk - Architecture



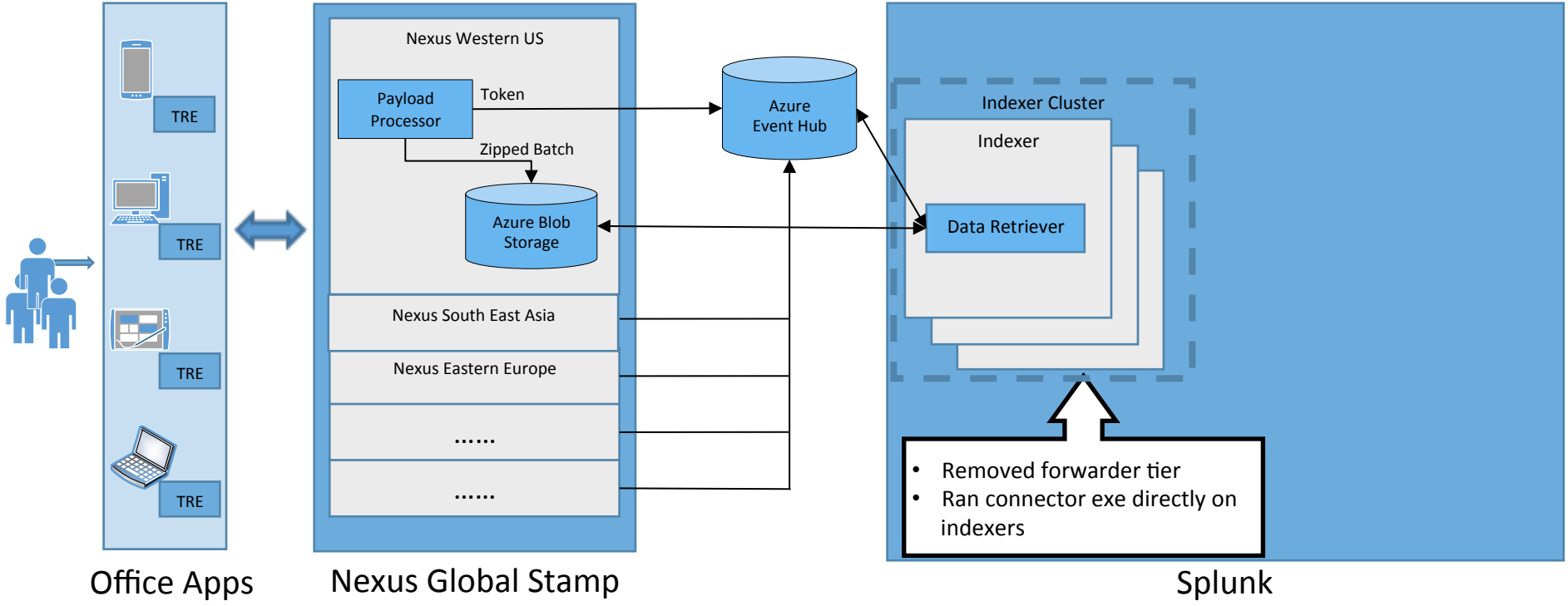
Office Client Splunk - Architecture



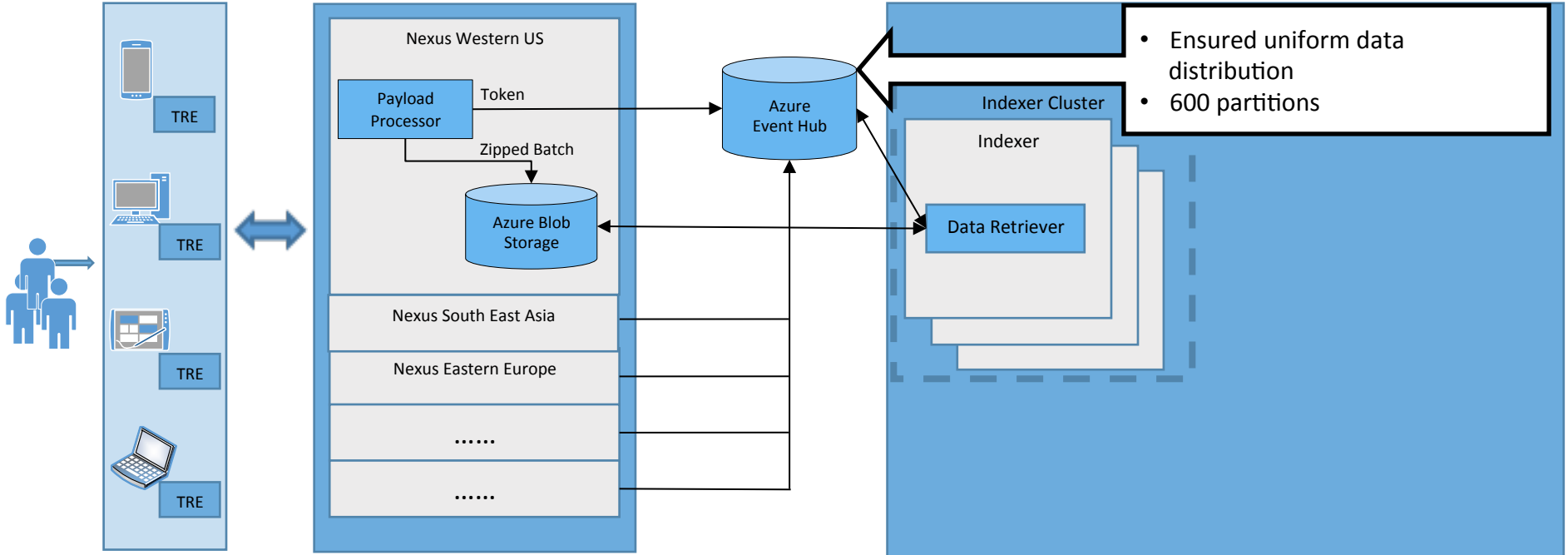
Office Client Splunk - Architecture



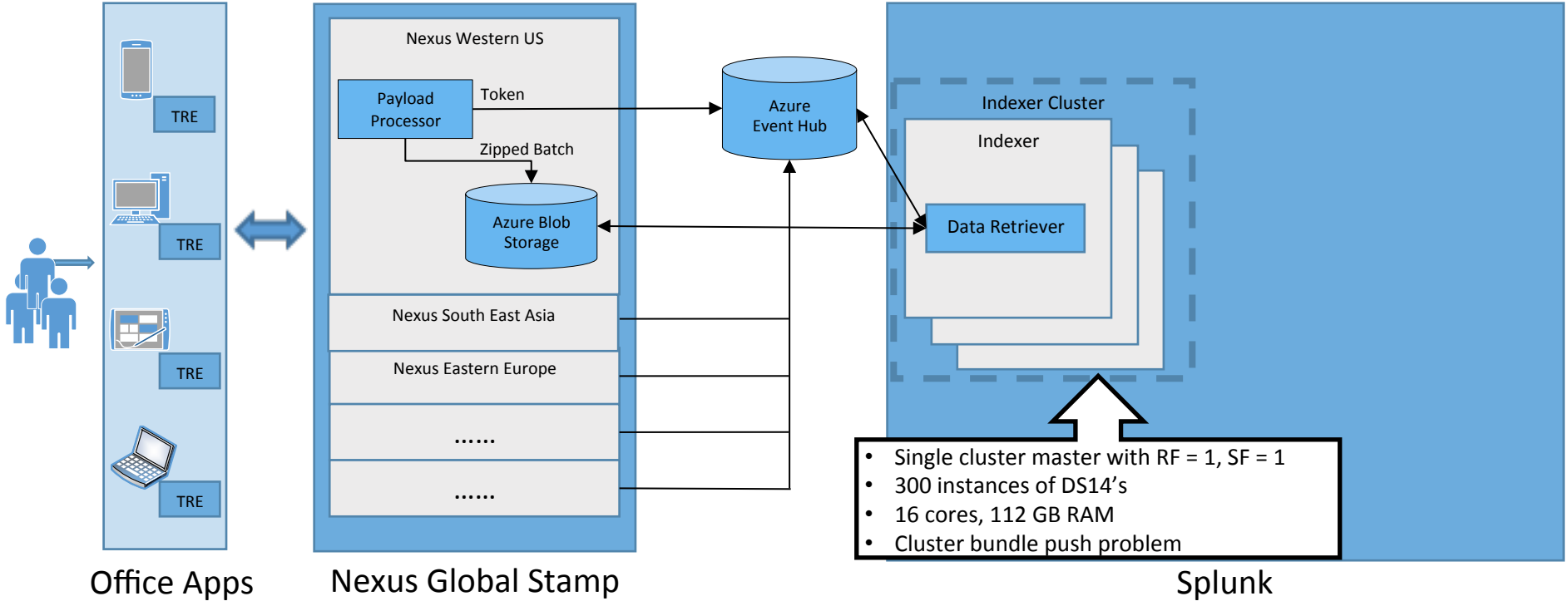
Office Client Splunk - Architecture



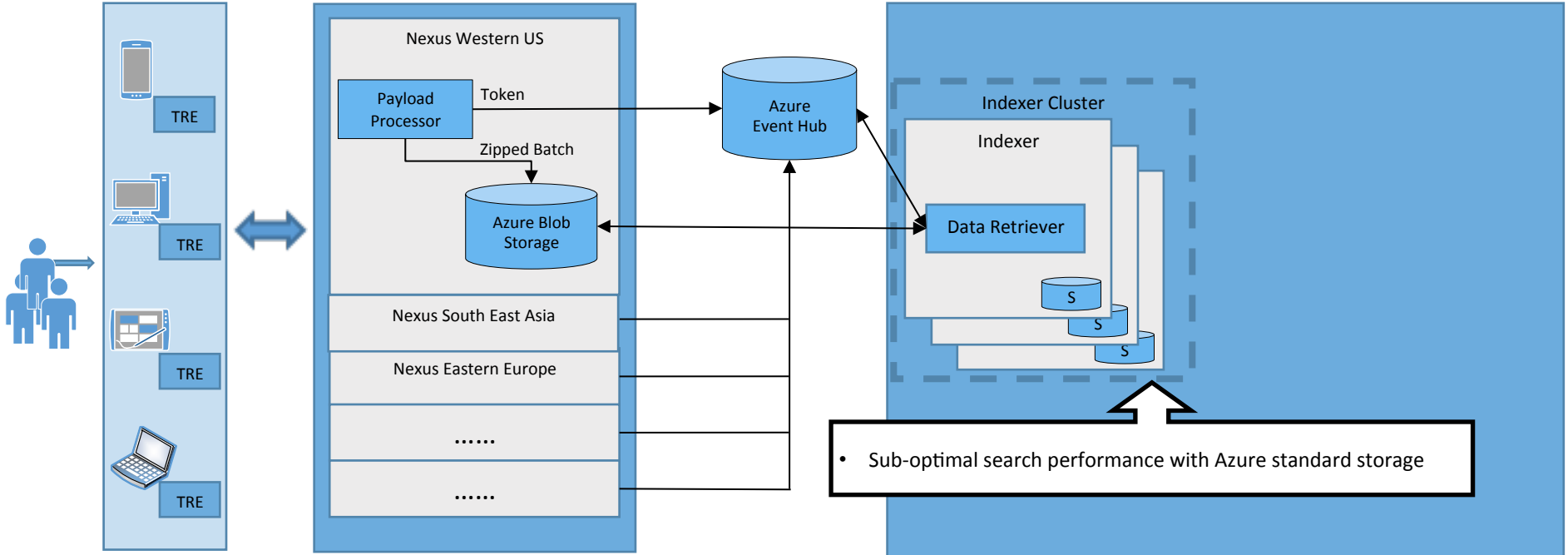
Office Client Splunk - Architecture



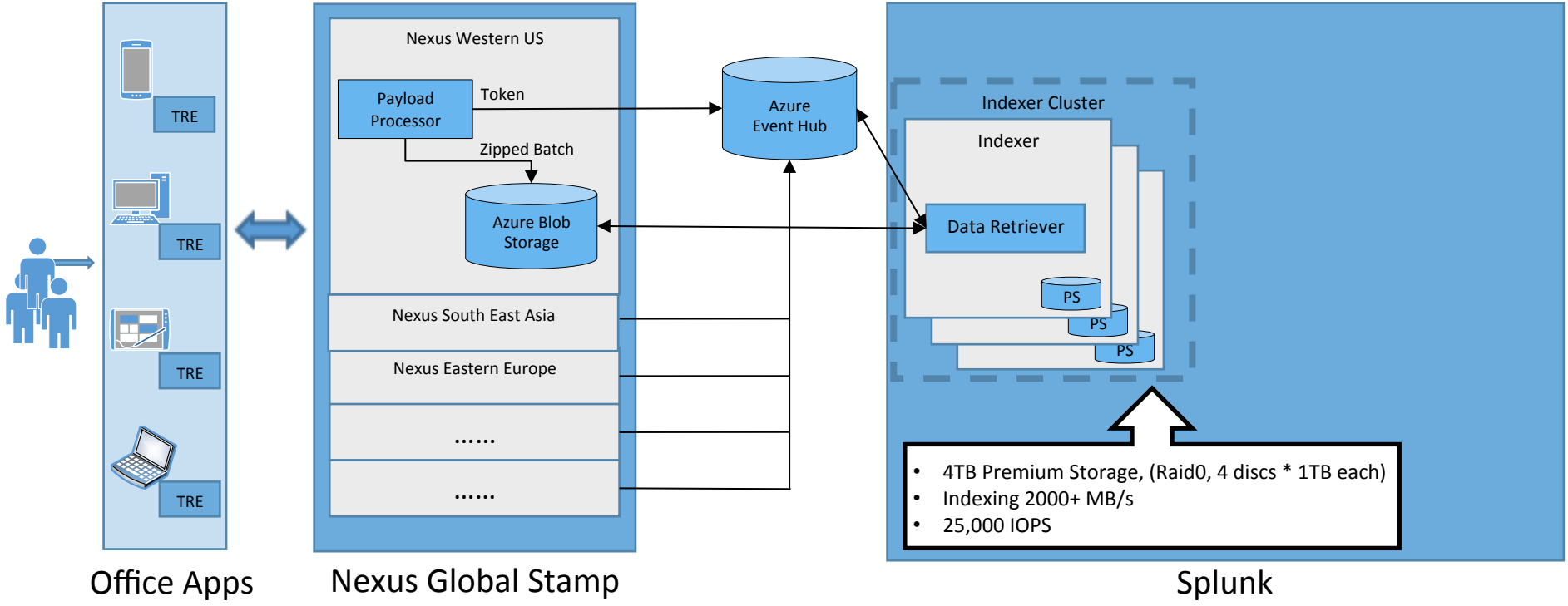
Office Client Splunk - Architecture



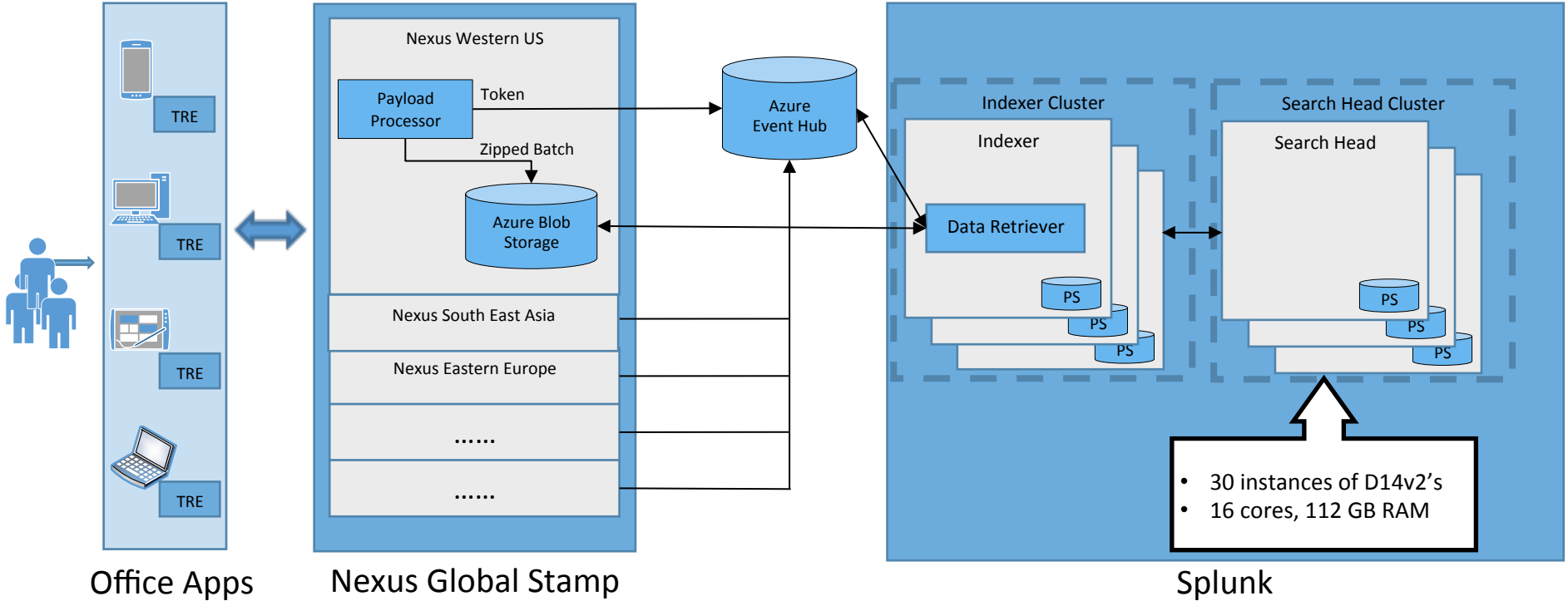
Office Client Splunk - Architecture



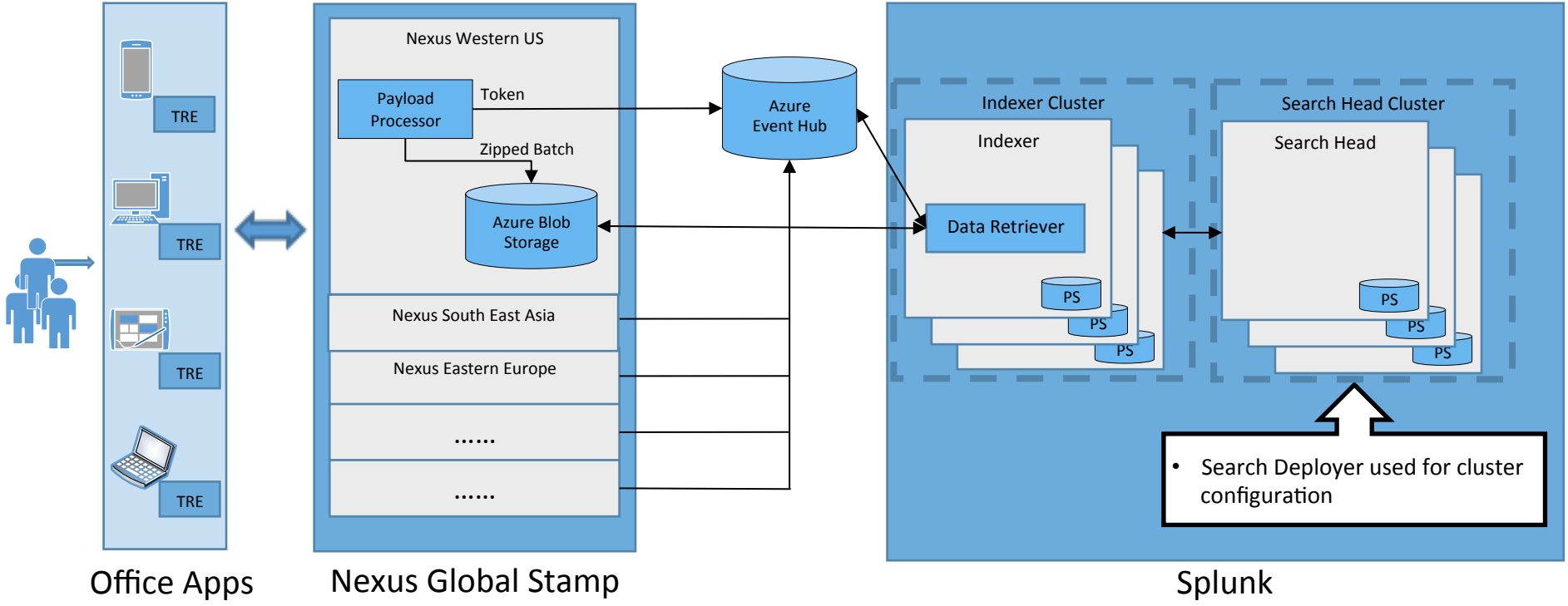
Office Client Splunk - Architecture



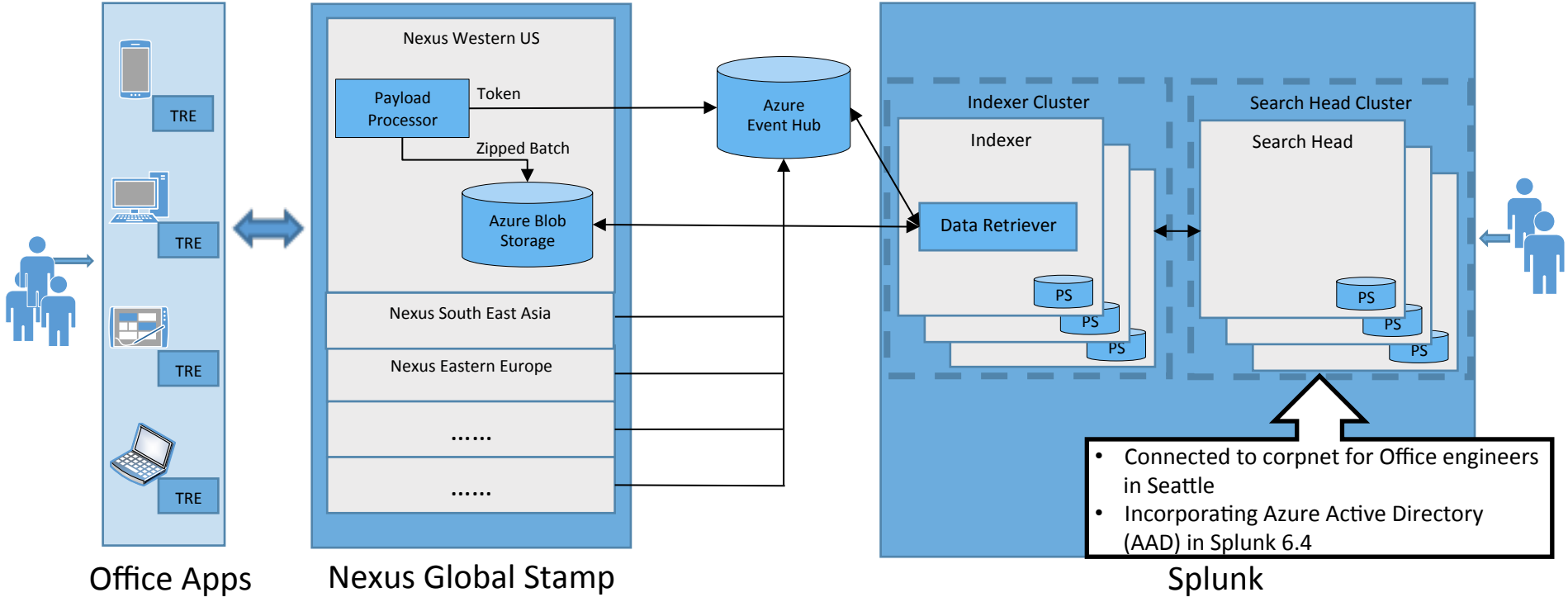
Office Client Splunk - Architecture



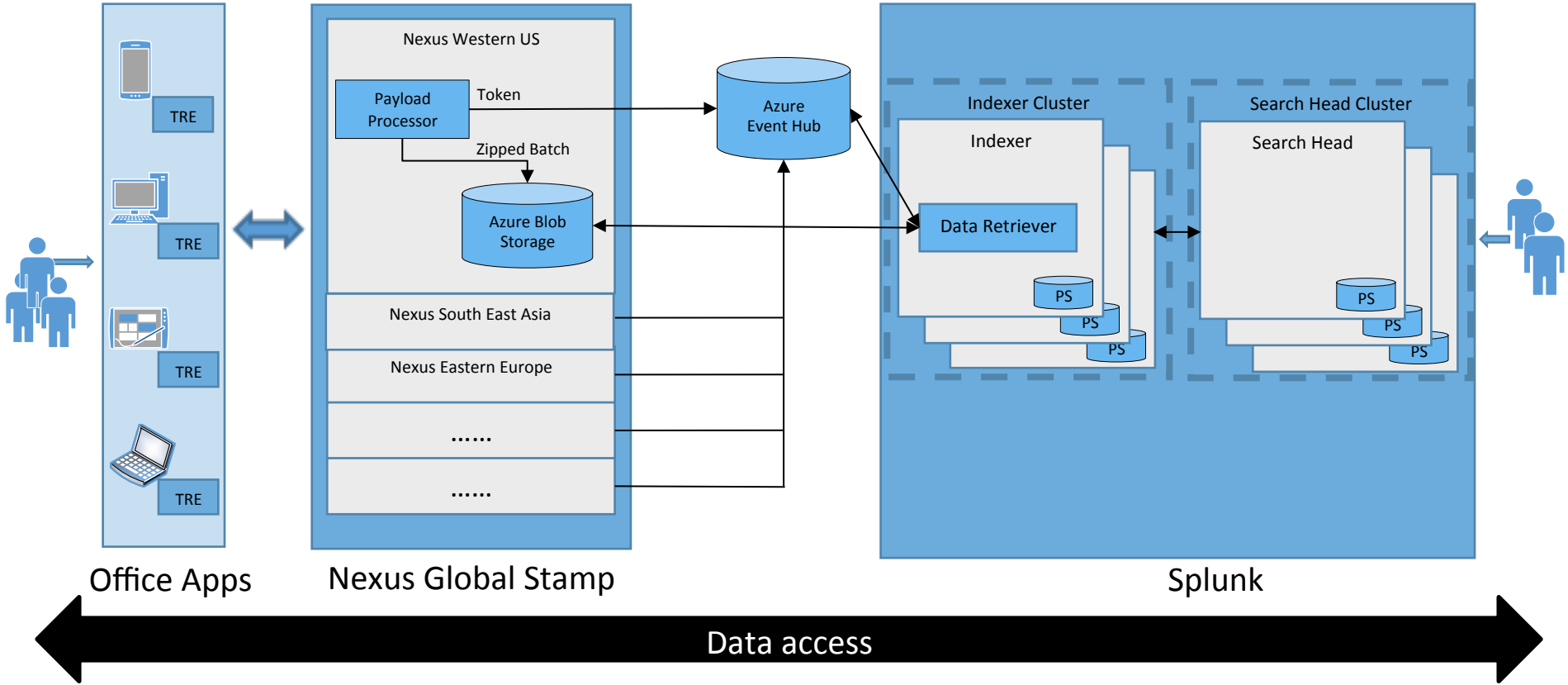
Office Client Splunk - Architecture



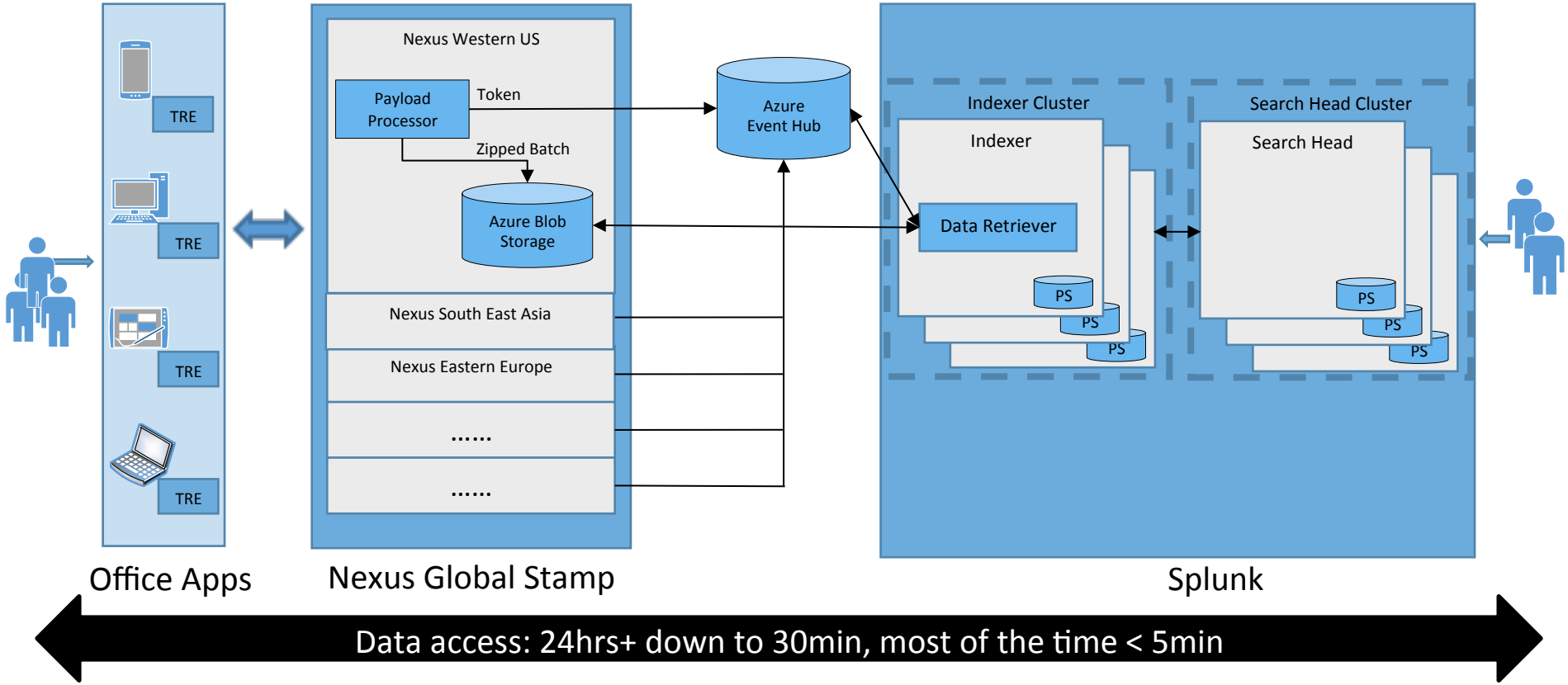
Office Client Splunk - Architecture



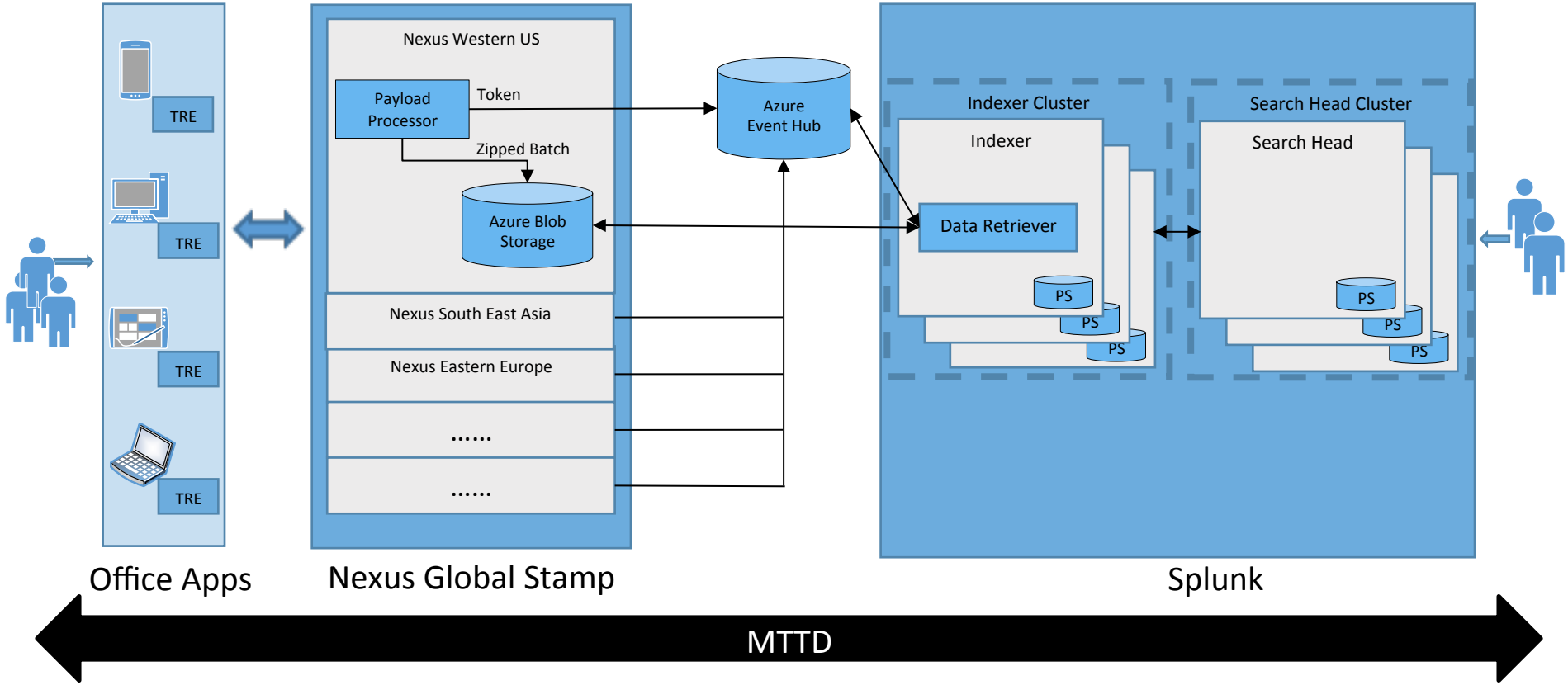
Office Client Splunk - Metrics



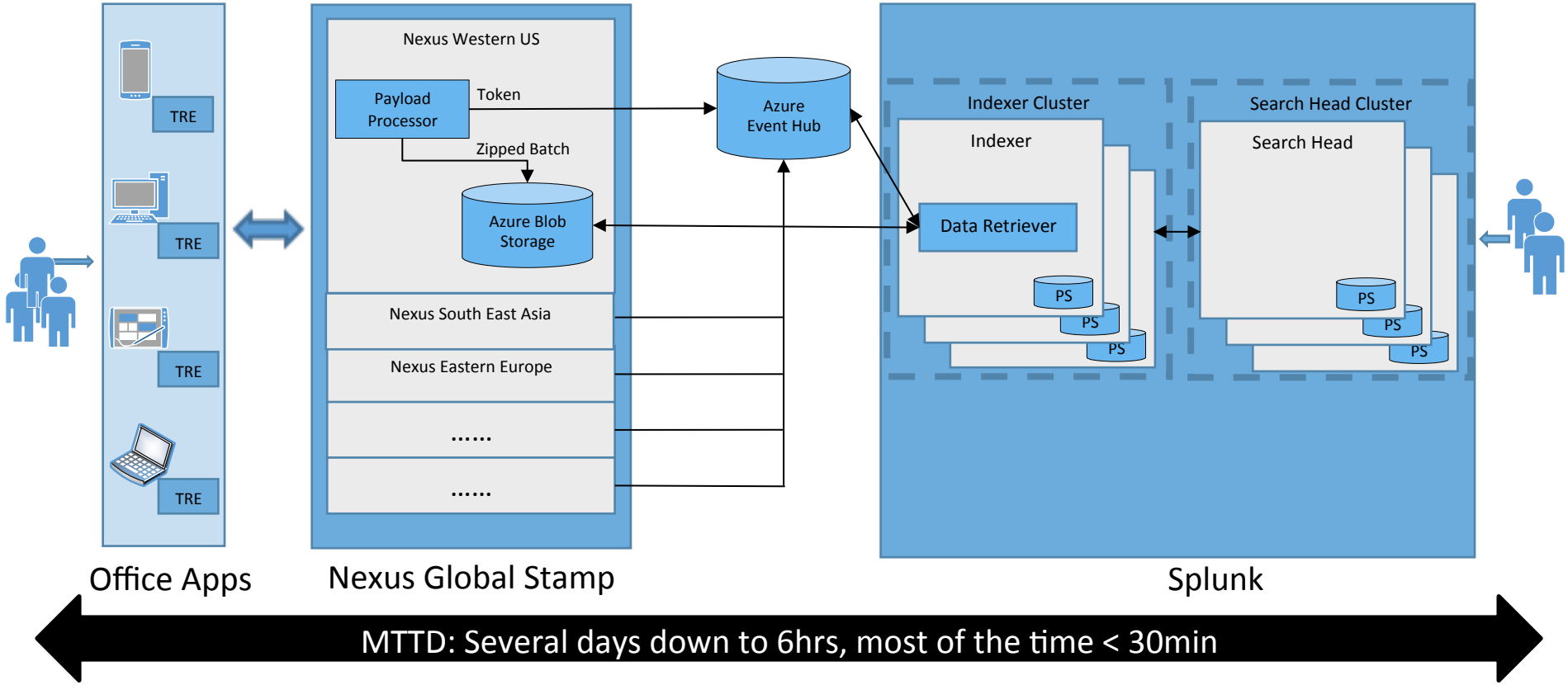
Office Client Splunk - Metrics



Office Client Splunk - Metrics



Office Client Splunk - Metrics



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

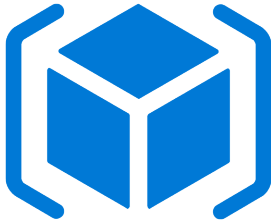
Attach premium storage disks, format & create single volume

- ```
  $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
  ```

- ```
  Format-Volume -DriveLetter H -FileSystem NTFS -NewFileSystemLabel
  "SplunkData" -AllocationUnitSize 65536 -Confirm:$false
  ```
  ```

Azure ARM Templates



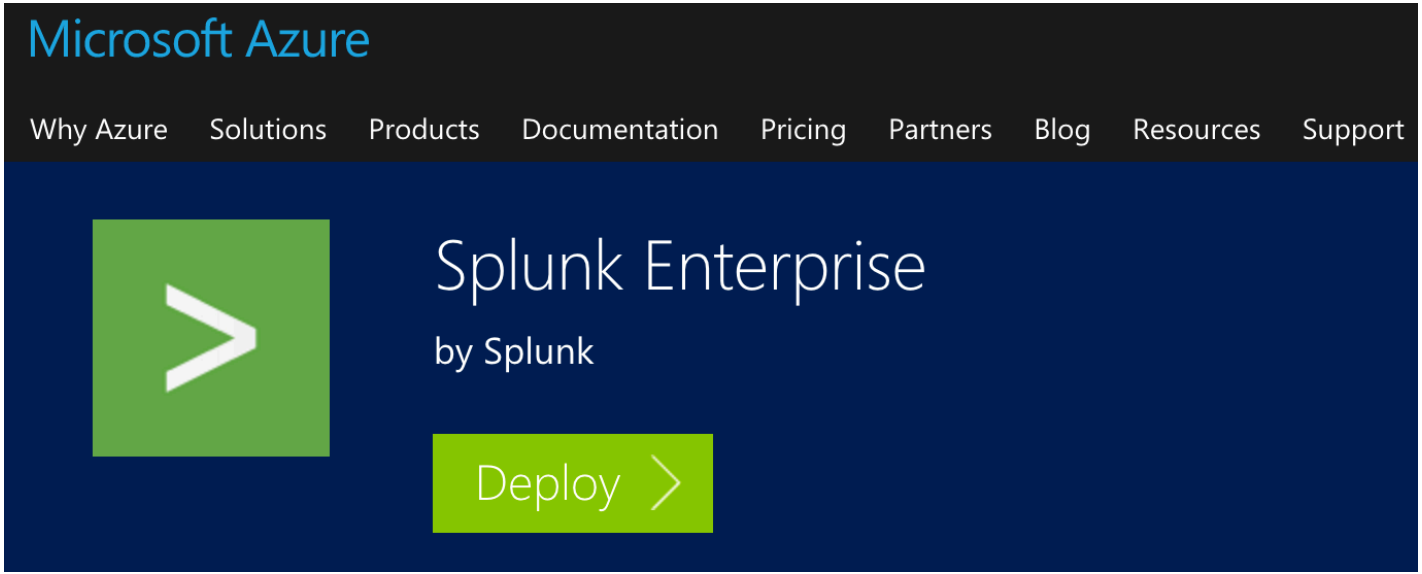
- Automated deployment provisioning via ARM templates
- Deploy via Azure portal, Azure PowerShell or Azure CLI:

```
$ 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



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

Technical Brief - Splunk on Azure

https://www.splunk.com/pdfs/technical-briefs/deploying-splunk-enterprise-on-microsoft-azure.pdf

splunk> TECH BRIEF

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 11,000 organizations use Splunk software to deepen business and customer understanding, mitigate cybersecurity risk, improve service performance and reduce costs. Splunk Enterprise indexes machine data in real time, enabling multiple roles across the organization—from system administrators to business analysts—to rapidly gain insight from the massive amounts of machine data generated by your environment.

Adopting a cloud strategy enables organizations to increase agility, reduce costs, decrease time to market and empower innovation. Splunk Enterprise is perfect for deploying in a cloud environment, offering enterprise-grade availability and scalability to support the collection of hundreds of terabytes of data per day from workloads residing on-premises, in the cloud or across hybrid environments. This document covers guidelines for deploying Splunk Enterprise on Microsoft Azure, an open and flexible cloud platform with a growing collection of integrated cloud services, including analytics, computing, database, mobile, networking, storage and web.

Splunk Deployment Components

A typical Splunk deployment includes Splunk forwarders, indexers and search heads. Splunk Enterprise is a single package that can perform one or many of the roles that each component would normally deliver, in addition to others. The software can be installed within minutes on your choice of hardware (physical, cloud or virtual) and operating system. The package is available publicly via the Azure Marketplace as a single-instance or a multi-instance Azure Resource Manager (ARM) solution template, in addition to downloadable packaged forms for most operating systems. While all major Splunk components can be run from a single installation on a single cloud instance, they can also run independently from within different cloud instances. Depending on the deployment infrastructure, considerations must also be taken to allocate the proper amount of resources per component type.

Forwarders perform data collection, data forwarding and data load balancing. Low amounts of resources are required to run a forwarder as they typically read and send data with minimal overhead. A Universal Forwarder is a lightweight package of the Splunk software that can perform most, if not all, of the forwarder functionality.

Indexers write the data to a storage device and perform searching on the data. These can be resource intense and require I/O and CPU allotment.

Search heads search for information across indexers and require CPU and memory allotment.

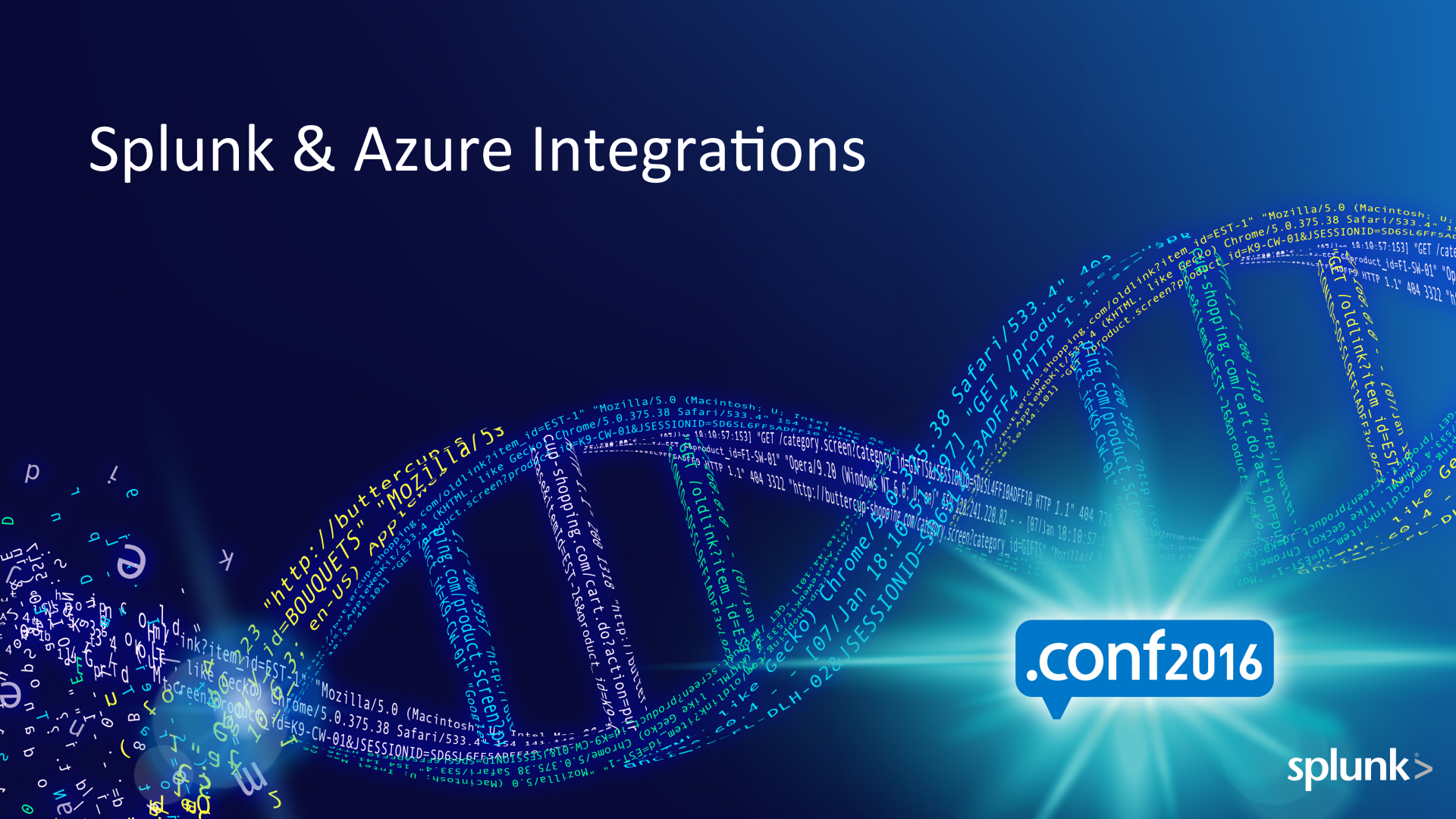
Budgeting system resources and bandwidth to enable search and index performance depend on the total volume of data being indexed and the number of active concurrent searches (scheduled or otherwise) at any time.

In addition to rapidly writing data to disk, indexers perform much of the work involved in running searches: reading data off disk, decompressing it, extracting knowledge and reporting. Since indexers incur most of the workload, increases in indexing volume should be tied to an increase in indexer instances. Deploying additional indexers will distribute the load of increased data volume, resulting in reduced contention for resources and improved search performance.

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)

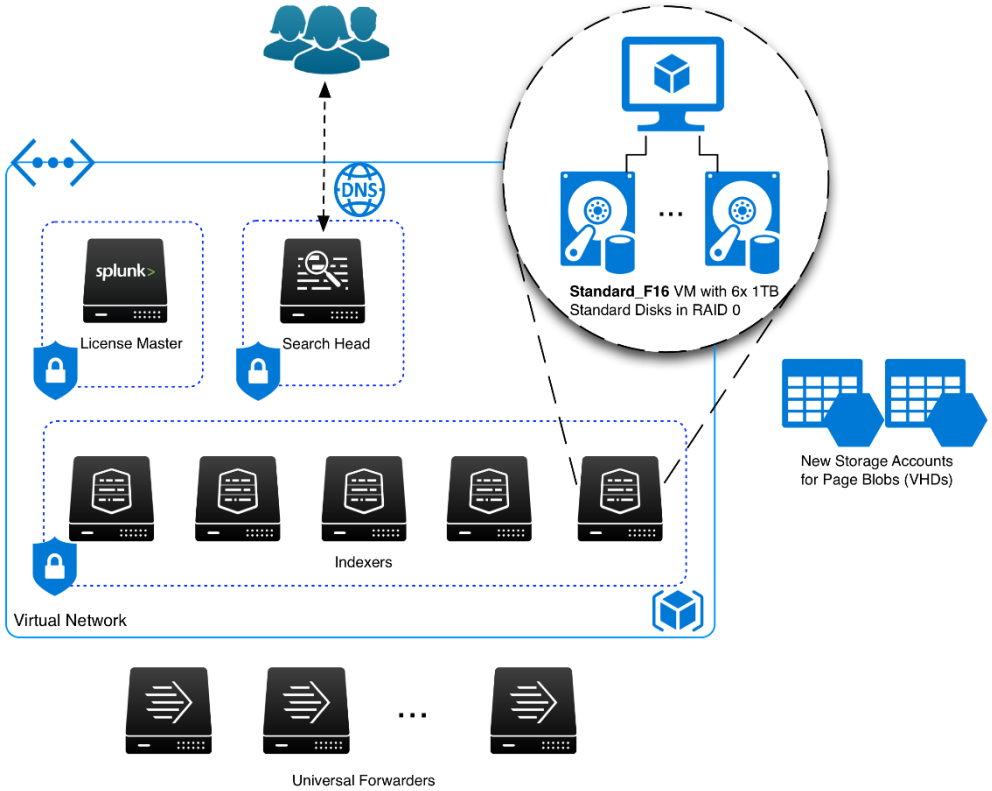
THANK YOU

splunk>

Example Deployment

16 concurrent users

1 TB/day
4 months retention



Example Clustered Deployment

8 concurrent users

500 GB/day
60 days retention