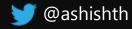# Data Architects guide for successful Open Source analytics workloads with Azure HDInsight
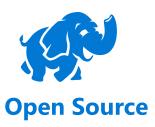
Ashish Thapliyal
Principal Product Manager
Azure HDInsight
Microsoft Corp.

@ashishth

Azure HDInsight

A secure and managed Apache Hadoop and Spark platform for building data lakes in the Cloud

Open Source

- 100% Apache Open Source
- The most popular open source frameworks
- Part of the Hortonworks HDP distribution

Managed

- 99.9% availability SLA
- Cluster Health Monitoring
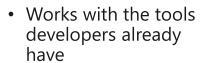- Integration with Azure Log Analytics
- Highly optimized for Azure

Secure & Compliant

- Role based access control
- Azure AD & Kerberos based authentication
- Strong VNET and service endpoint support
- The most trusted and compliant platform

Productive

- Works with the tools developers already have
- Special extensions for advanced debugging and diagnostics

Lift & Shift

- Move workloads from on-prem or other clouds without code changes
- Curated application platform for wide variety of use cases

HDInsight Solution Architecture: Keys topics for today

OSS Framework Choices	**Storage**	**HA & DR**
Security	**Monitoring**	**Cost Optimization**

@ashishth

HDInsight Technology choices

@ashishth

HDInsight Technology choices

@ashishth
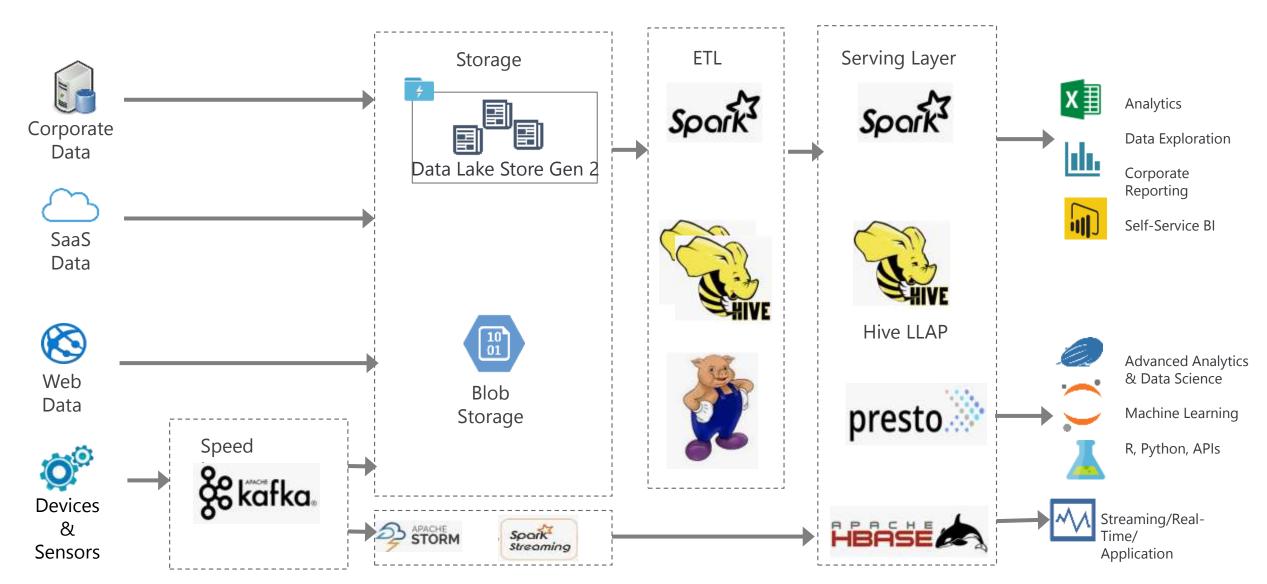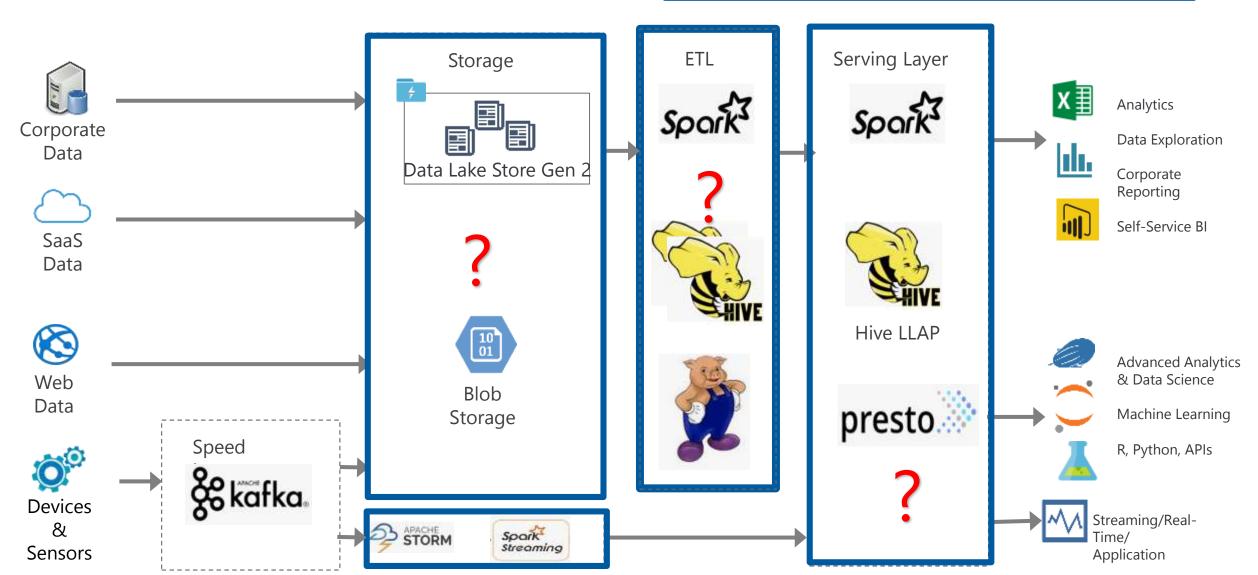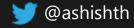
Scenario/ Technology mapping

1. ETL technology choices

	Spark	Pig	Hive
Designed for	ETL	ETL	Data warehousing
Adoption	High, increasing	Low, decreasing	Stable
Number of connectors	Highest	High	High
Languages	Python, R, Scala, Java, SQL	Pig	SQL
Performance	High	Medium	Medium

@ashishth

2.Streaming engine technology choices

	Spark Structured Streaming	Storm
Adoption	High, increasing	Decreasing
Event processing guarantee	Exactly once	At least once
Throughput	High	Low
Processing Model	Micro Batch	Real-Time
Latency	High	Low
Event time support	Yes	Yes
Languages	Python, R, Scala, Java, SQL	Java

@ashishth
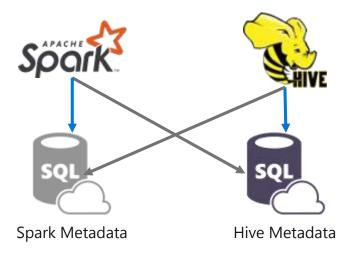
3.Interactive Query technology choices

Capability	Hive LLAP	Spark SQL	Presto
Interactive Query Speed	High	High	Medium
Scale	High	High	Low
Caching	Yes	Yes	Early Support
Result Caching	Yes	No	No
Intelligent Cache Eviction	Yes	No	No
Materialized Views	Yes	No	No
Complex Fact to Fact Joins	Yes	Yes	No
Transactions	Yes	No	No
Query Concurrency	High	Low	Low
Row , Column level security	Yes [Apache Ranger+ AAD]	Medium	Medium
Rich end user Tools	Yes	Yes	Yes
Language Support	SQL, UDF	SQL, Scala, Python	SQL
Data Source Connector Support	Storage Handlers	Data Sources	High number of connectors

@ashishth
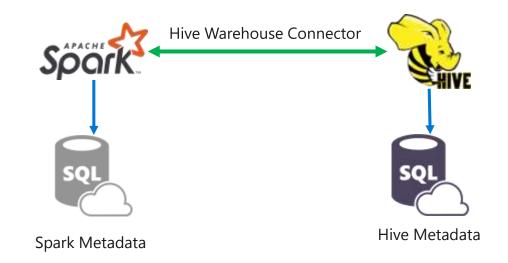
Tip: Spark & Hive Metastore

Azure HDInsight 3.6 with Hadoop 2.6

Spark Metadata   Hive Metadata

- Spark executors talk directly to Hive Metastore
- Reliability and compatibility issues
- Cannot take advantage of the native query engine

Azure HDInsight 4.0 with Hadoop 3.x

Hive Warehouse Connector

Spark Metadata   Hive Metadata

- **New Hive Warehouse Connector**
- Apache Arrow based communication between Spark executors and Hive LLAP
- Smart predicate pushdown
- Transactional access to Hive tables from Spark

Hive Metastore migration tool: https://azure.microsoft.com/en-us/blog/hdinsight-metastore-migration-tool-open-source-release-now-available/

4.Data pipeline orchestration technology choices

	ADF	Airflow	Oozie
Service management	Azure PaaS	IaaS VM	HDInsight
Code	JSON	Python	Java
GUI	ADF V2 has great UX	Good UX	Below Average UX
Community	Microsoft	Growing (10893 Stars)	Declining (454 Stars)
On-demand clusters	Yes	No, but extensible	No
Extensibility	Custom action-only	Full, graph + actions	Custom action-only
Pipeline definition	JSON/UX	Python/ UX	XML/UX
Devops-first design	Yes	Yes	Yes
Pipeline monitoring	Yes	Yes	Yes
Scheduling	Event, Time	Event	Event, Time

@ashishth

Migrating to Azure HDInsight Guide!

[Motivation and benefits](#) covers the benefits of migrating on-premises Hadoop ecosystem components to HDInsight and how to plan for the migration.

[Architecture best practices](#) provides best practices for the architecture of HDInsight systems and addresses different types of workloads.

[Infrastructure best practices](#) goes into detailed recommendations for managing the infrastructure of HDInsight clusters.

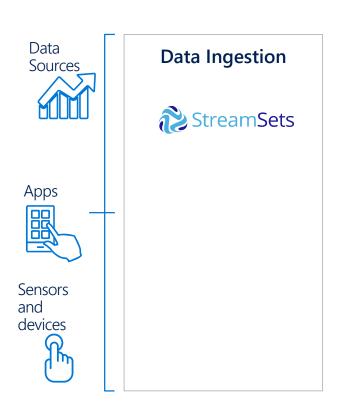[Storage best practices](#) gives recommendations for data storage in HDInsight systems.

[Data migration best practices](#) provides recommendations for data migration to HDInsight.

[Security and DevOps best practices](#) gives recommendations for security and DevOps in HDInsight systems.

https://azure.microsoft.com/en-us/blog/migrating-on-premises-hadoop-infrastructure-to-azure-hdinsight/
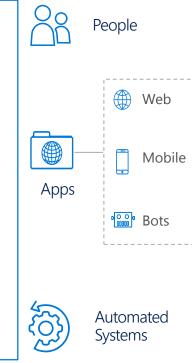
HDInsight Applications: Something for everyone

Data Sources

Apps

Sensors and devices

Data Ingestion	Data Prep/ Management	Advanced Analytics	BI/ Visualization
StreamSets	Datameer	data iku	Datameer
	TRIFACTA	H₂O.ai	ATSCALE
	Paxata	R Studio	tableau
	Informatica		MicroStrategy
	CASK		Qlik
			DOMO

People

Apps
- Web
- Mobile
- Bots

Automated Systems

Productivity Tools

Data catalog/ Governance/ Lineage — Waterline Data

Connectors: JDBC, ODBC — Simba technologies A MAGNITUDE SOFTWARE COMPANY

Enterprise grade add-ons (hybrid, backup, DR, security, performance) — IMANIS DATA — wanDISCO — @ashishth
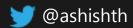
Storage

HDInsight Storage topics

Data movement

Storage options and tradeoffs

Caching

@ashishth

Data Qty	Network Bandwidth		
	45 Mbps (T3)	100 Mbps	1 Gbps
1 TB	2 days	1 day	2 hours
10 TB	22 days	10 days	1 day
35 TB	76 days	34 days	3 days
80 TB	173 days	78 days	8 days
100 TB	216 days	97 days	10 days
200 TB	1 year	194 days	19 days
500 TB	3 years	1 year	49 days
1 PB	**6 years**	**3 years**	**97 days**
2 PB	12 years	5 years	194 days

Storage Transfer options

Network Transfer with TLS

- Over Internet

- Express Route

- Databox online Transfer

Shipping data offline

- Import / Export service

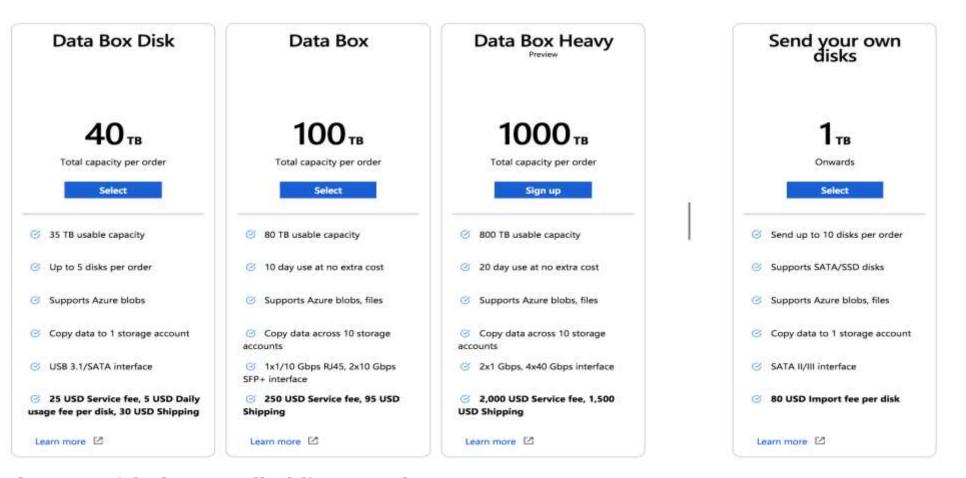- Data Box offline data transfer

Data Box Transfer Options

Data Box Disk	Data Box	Data Box Heavy *Preview*	Send your own disks
40 TB	**100** TB	**1000** TB	**1** TB
Total capacity per order	Total capacity per order	Total capacity per order	Onwards
Select	**Select**	**Sign up**	**Select**
35 TB usable capacity	80 TB usable capacity	800 TB usable capacity	Send up to 10 disks per order
Up to 5 disks per order	10 day use at no extra cost	20 day use at no extra cost	Supports SATA/SSD disks
Supports Azure blobs	Supports Azure blobs, files	Supports Azure blobs, files	Supports Azure blobs, files
Copy data to 1 storage account	Copy data across 10 storage accounts	Copy data across 10 storage accounts	Copy data to 1 storage account
USB 3.1/SATA interface	1x1/10 Gbps RJ45, 2x10 Gbps SFP+ interface	2x1 Gbps, 4x40 Gbps interface	SATA II/III interface
25 USD Service fee, 5 USD Daily usage fee per disk, 30 USD Shipping	**250 USD Service fee, 95 USD Shipping**	**2,000 USD Service fee, 1,500 USD Shipping**	**80 USD Import fee per disk**
Learn more	Learn more	Learn more	Learn more

https://github.com/alkohli/azure-docs-pr/blob/4023eb52cc6ed103e0fa7e794e039c143b6d2a6a/articles/storage/blobs/data-lake-storage-migrate-on-prem-HDFS-cluster.md

@ashishth

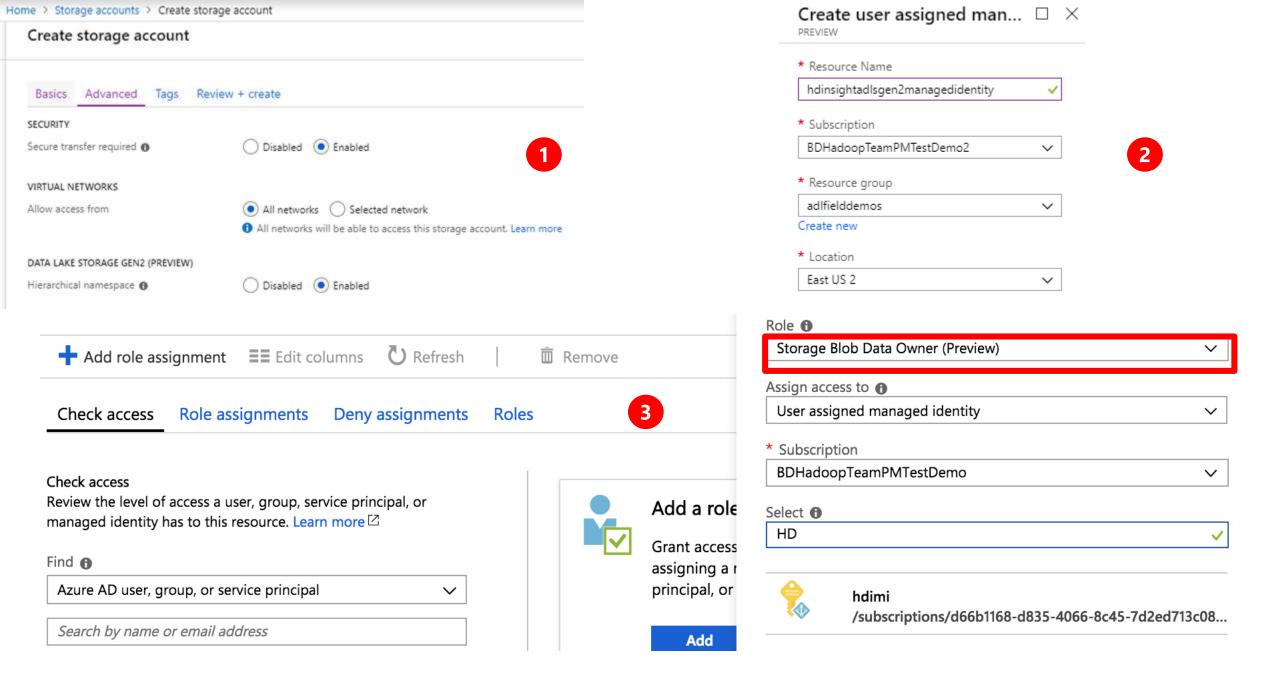# Storage options and tradeoffs

Storage Options with HDInsight

|	Type	Latency	Consistency	Workloads	Bandwidth	Key Benefits
ADLS Gen 1	Hierarchical	10-100ms	Low	HDInsight 3.6(No HBase)	High	Atomic Rename, File Folder level ACL's
ADLS Gen 2	Hierarchical	10-50ms	Medium	HDInsight 3.6 & 4.0	Unconstrained	Atomic Rename, File Folder level ACL's
Standard BLOB	Object Store	10-50ms	Medium	HDInsight 3.6 & 4.0	Unconstrained	Mature
Premium BLOB	Object Store	~5ms	High	HBase in Preview	Unconstrained	Fast
Premium Managed Disks	Hierarchical	~5ms	High	Kafka, HBase in preview	Based on disk	Consistent

ADLS Gen 1 and Gen 2 & HDInsight Compatibility
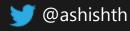
Scenario	Supported	Workaround
HDInsight 3.6 & 4.0 with **Standard Blob as Primary and/ or secondary**	Yes	
HDInsight 3.6 & 4.0 with **ADLS Gen2 as primary**	Yes	
HDInsight 3.6 & 4.0 with **ADLS Gen2 as primary & Blob as additional**	Yes	
HDInsight 3.6 & 4.0 with **Blob as primary & ADLS Gen2 as additional**	No	
HDInsight 3.6 with multiple **ADLS Gen2 accounts**	Yes	
HDInsight 3.6 & 4.0 with **ADLS Gen1 and ADLS Gen 2**	No	Distcp across two clusters
HDInsight 4.0 **with ADLS Gen 1**	No	Distcp across two clusters

@ashishth

Create storage account

Basics | **Advanced** | Tags | Review + create

SECURITY

Secure transfer required ⓘ — ◯ Disabled ⬤ Enabled

VIRTUAL NETWORKS

Allow access from — ⬤ All networks ◯ Selected network

ⓘ All networks will be able to access this storage account. Learn more

DATA LAKE STORAGE GEN2 (PREVIEW)

Hierarchical namespace ⓘ — ◯ Disabled ⬤ Enabled

①

Create user assigned man...
PREVIEW ☐ ✕

②

* Resource Name
hdinsightadlsgen2managedidentity ✓

* Subscription
BDHadoopTeamPMTestDemo2 ⌄

* Resource group
adlfielddemos ⌄

Create new

* Location
East US 2 ⌄

Role ⓘ
[Storage Blob Data Owner (Preview) ⌄]

Assign access to ⓘ
User assigned managed identity ⌄

* Subscription
BDHadoopTeamPMTestDemo ⌄

Select ⓘ
HD ✓

🔑 **hdimi**
/subscriptions/d66b1168-d835-4066-8c45-7d2ed713c08...

---

③

➕ Add role assignment   ▤ Edit columns   ↻ Refresh | 🗑 Remove

Check access | Role assignments | Deny assignments | Roles

Check access

Review the level of access a user, group, service principal, or managed identity has to this resource. Learn more ↗

Find ⓘ
Azure AD user, group, or service principal ⌄

Search by name or email address

Add a role
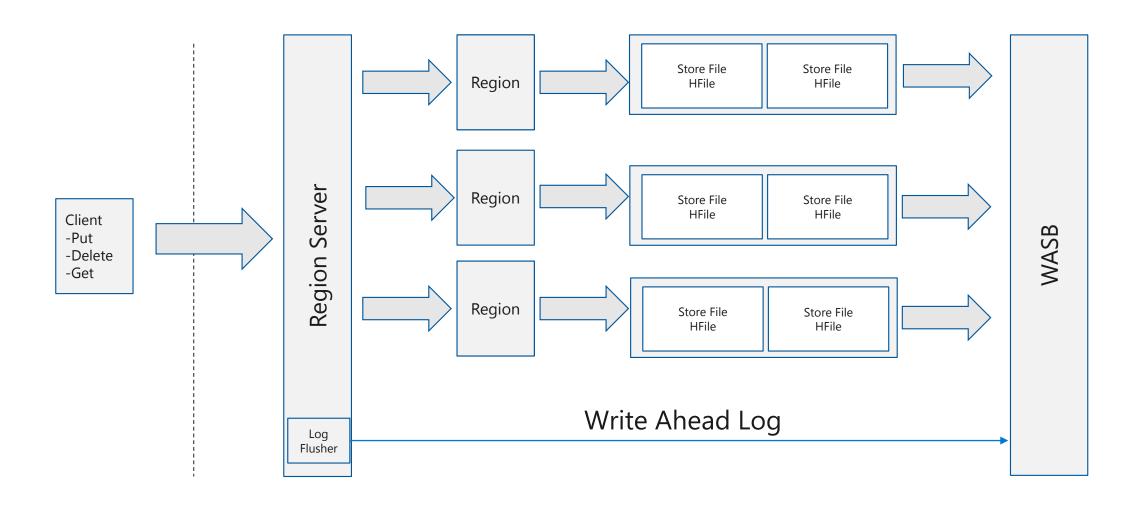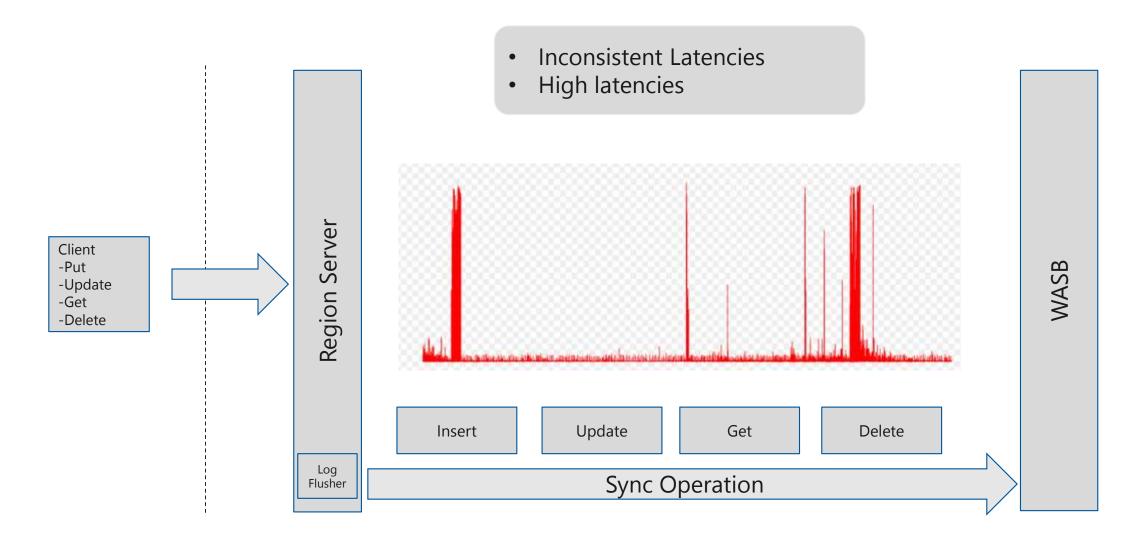Grant access
assigning a
principal, or

Add

🐦 @ashishth

New Storage Improvements: Enhanced reads and writes for low latency workloads (Private Preview)
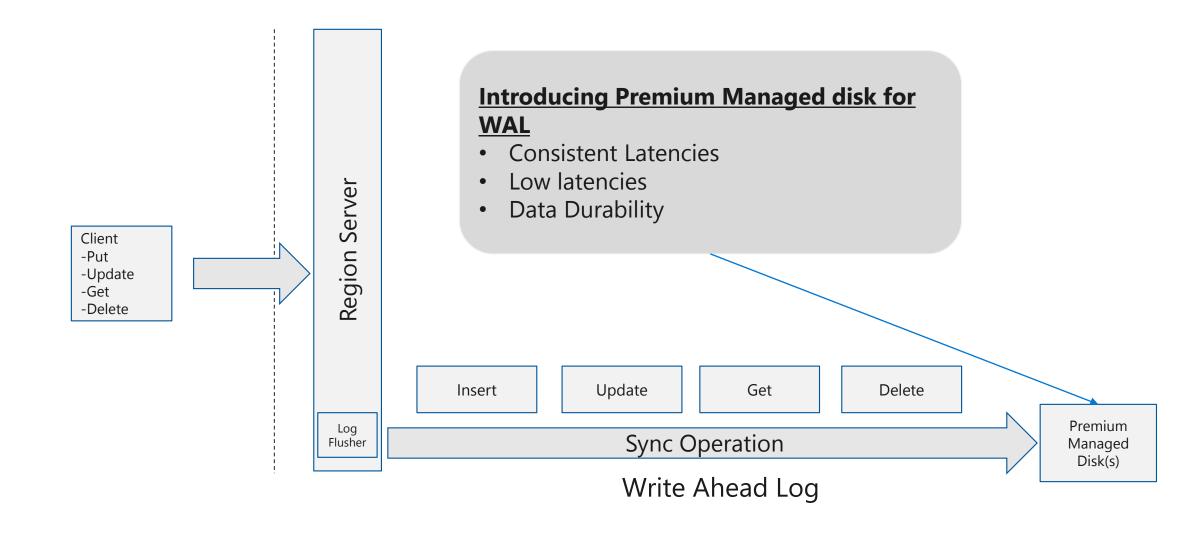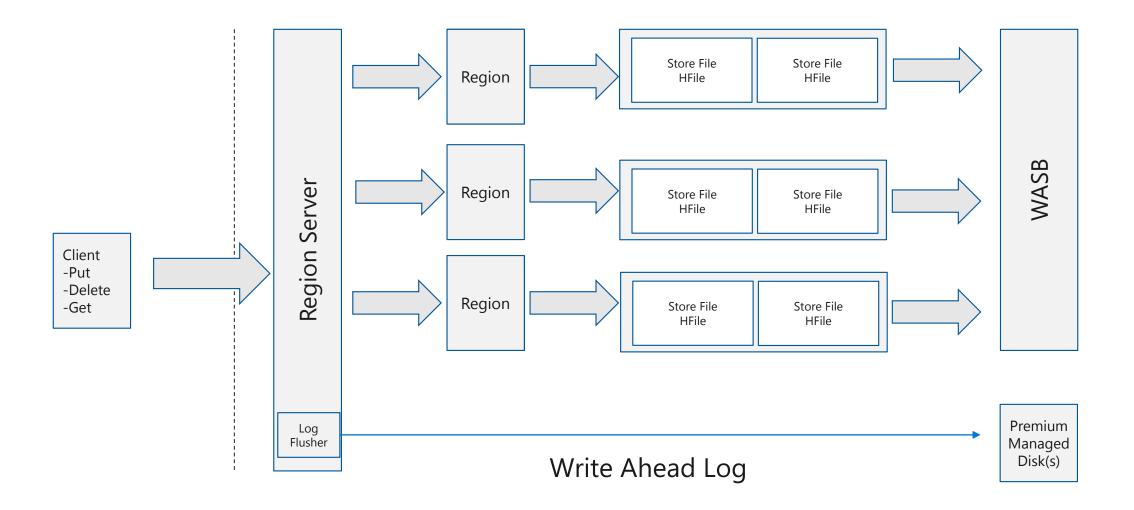
Example: Low latency workload HBase/ Small write

Write path challenges with Write Ahead Log
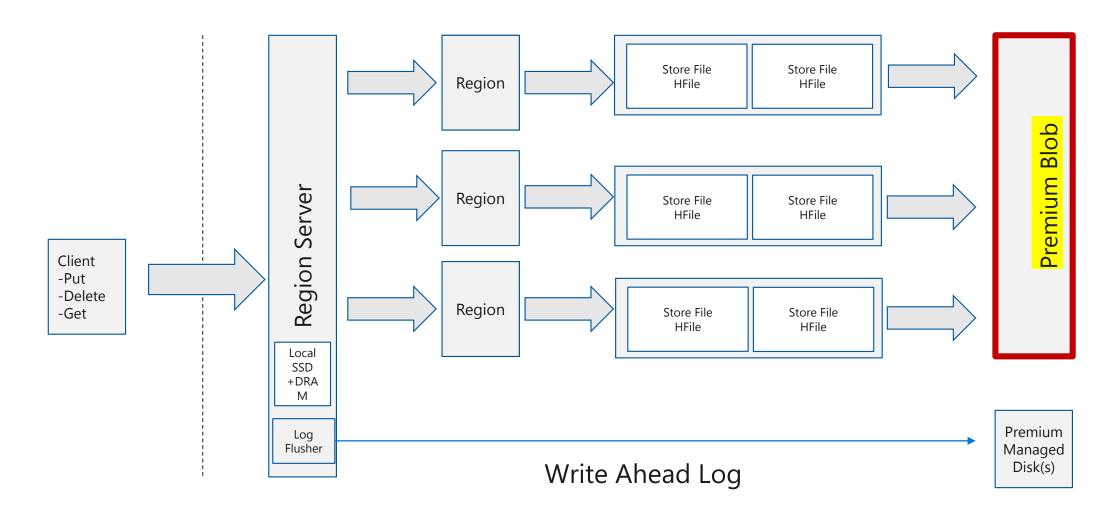
@ashishth

@ashishth

How about Reads?

Introducing support for Premium Blob

@ashishth

Performance (YCSB)

Cluster Type	Operation	Row Size	# ops	#Region Servers	Region Server Node Size	#Clients	Throughput	Avg Latency (ms)	Run Time (min)
Standard	Write	1KB	107,374,182	4	Standard_D4_V2	2	37,958	**0.417**	47
Premium WAL	Write	1KB	107,374,182	**4**	Standard_D4_V2	2	57,812	0..271	31
Standard	Small Write	100 Bytes	1,073,741,824	**4**	Standard_DS4_V2	2	**84,910**	0..186	210
Premium WAL	Small Write	100 Bytes	1,073,741,824	**4**	Standard_DS4_V2	2	**701,234**	0.016	25
Standard	Read	100 Bytes	925,075	4	Standard_D4_V2	2	256	**62**	60
Premium WAL & Premium Blob	Read	100 Bytes	33,503,676	4	Standard_D4_V2	2	9,306	1.7	60
Standard	Large Read	1K	945,682	4	Standard_D4_V2	2	262	**61**	60
Premium WAL & Premium Blob	Large Read	1K	24,846,209	4	Standard_D4_V2	2	6901	**2.3**	60
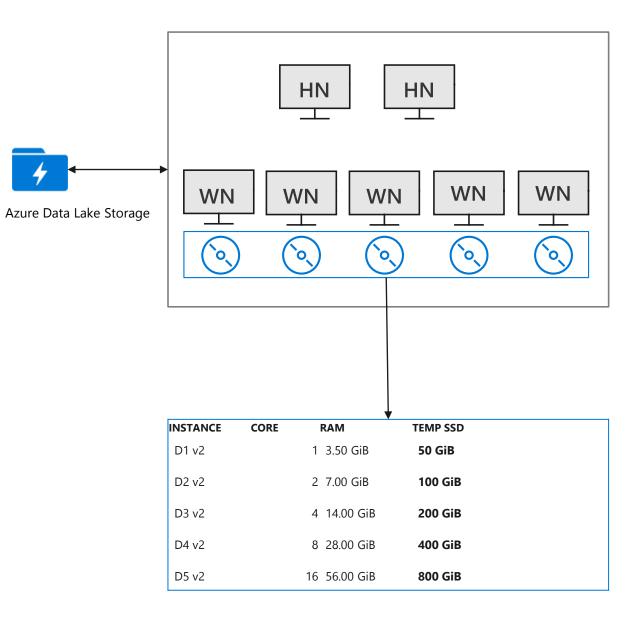
Caching options in HDInsight
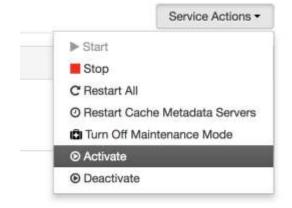
Caching Options by workload
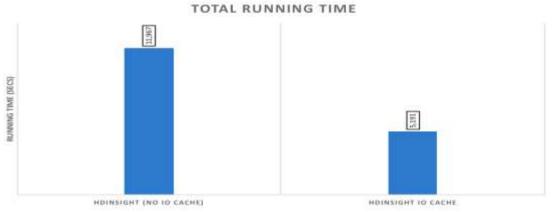
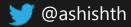| Workload	Caching Options	Key benefits
Spark	Spark IO Cache	Up to ~8 to 10x perf improvements
HBase & Phoenix	Bucket cache	Up 5-10x perf gains on recently read or written data
Hive + LLAP	LLAP Intelligent cache/Result Cache	Up to ~4-100X gain on cached data

New HDInsight IO Cache

- Significant Spark performance speed up with IO cache (up to 9X perf gains)
- Automatic cache resource management
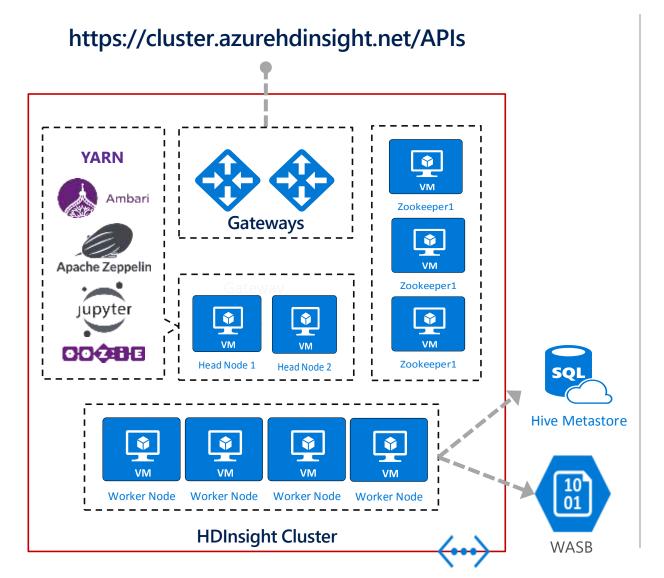- DRAM + Temp SSD makes large cache

Azure Data Lake Storage

HN   HN

WN   WN   WN   WN   WN

INSTANCE	CORE	RAM	TEMP SSD
D1 v2	1	3.50 GiB	50 GiB
D2 v2	2	7.00 GiB	100 GiB
D3 v2	4	14.00 GiB	200 GiB
D4 v2	8	28.00 GiB	400 GiB
D5 v2	16	56.00 GiB	800 GiB

Service Actions ▾

▶ Start
■ Stop
C Restart All
⊙ Restart Cache Metadata Servers
▣ Turn Off Maintenance Mode
⊙ Activate
⊙ Deactivate

TOTAL RUNNING TIME

RUNNING TIME [SECS]

HDINSIGHT (NO IO CACHE)   HDINSIGHT IO CACHE

HA & DR

Azure HDInsight: Highly Available End-points

https://cluster.azurehdinsight.net/APIs

YARN
Ambari
Apache Zeppelin
Jupyter
OOZIE

Gateways

Gateway

Head Node 1   Head Node 2

Zookeeper1
Zookeeper1
Zookeeper1

Worker Node   Worker Node   Worker Node   Worker Node

HDInsight Cluster

SQL
Hive Metastore

WASB

Highly Available APIs:

Livy – Spark job submission and interactive session management

Yarn – cluster resource management, Yarn job submission

Ambari – cluster management

Oozie – Oozie workflow scheduling and coordination (legacy APIs, ADF is recommended as a replacement)
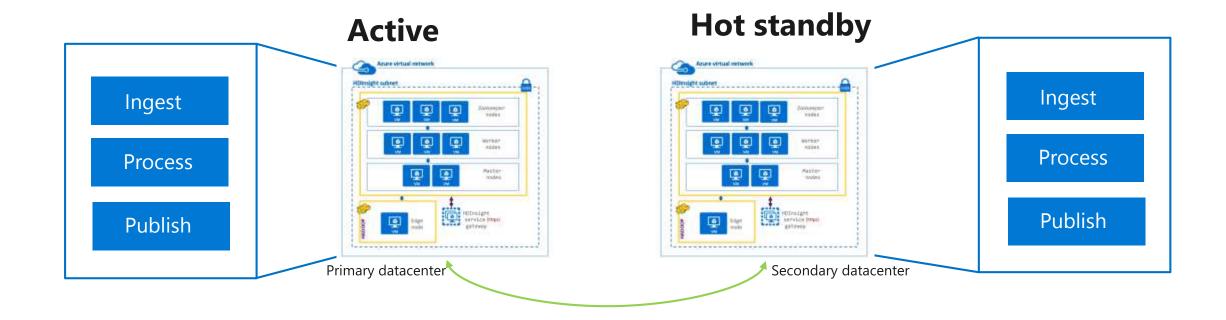
@ashishth

DR options by workloads

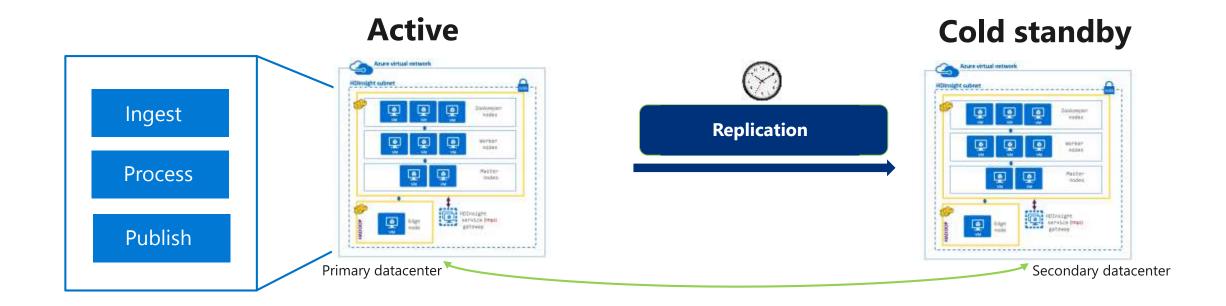| Workload	DR Option
Spark / Hive	Manual, Partner solution
HBase	HBase replication, Snapshot export, Import Export, Copy Tables
Kafka	Mirror Maker

Option 1: Active-hot standby with dual ingest

Active

Hot standby

Ingest

Process

Publish

Primary datacenter

Secondary datacenter

Ingest

Process

Publish

RPO	**Low**
RTO	**None**
Cost	**High**

- Applications/integration processes write to both clusters
- Both clusters run identical batch jobs
- Standby cluster is <u>offline</u> for reads by applications and end users
- Synchronization tasks need to be run to ensure clusters are in sync

Its also not uncommon to have the hot standby used for non-critical workloads like adhoc exploratory analytics by data analysts, for reporting
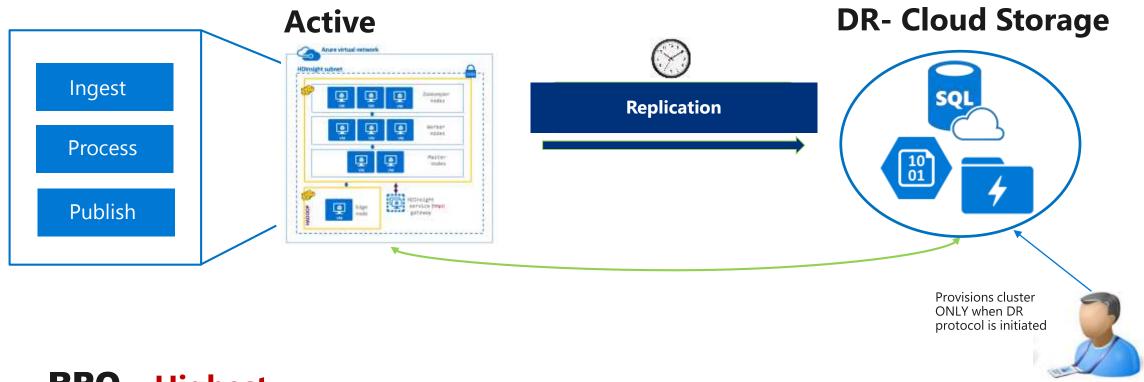
Option 2: Active- cold standby with batch replication

Active

Cold standby

Ingest

Process

Publish

Primary datacenter

Replication

Secondary datacenter

RPO	**Medium**	
RTO	**Medium**	
Cost	**High**	

- Applications write to active-primary ONLY
- Replication to DR cluster is incremental, batch, scheduled
- Synchronization tasks need to be run to ensure clusters are in sync
- Its is not uncommon to have a storage dense, compute light DR cluster for cost optimization
- Only curated data is replicated
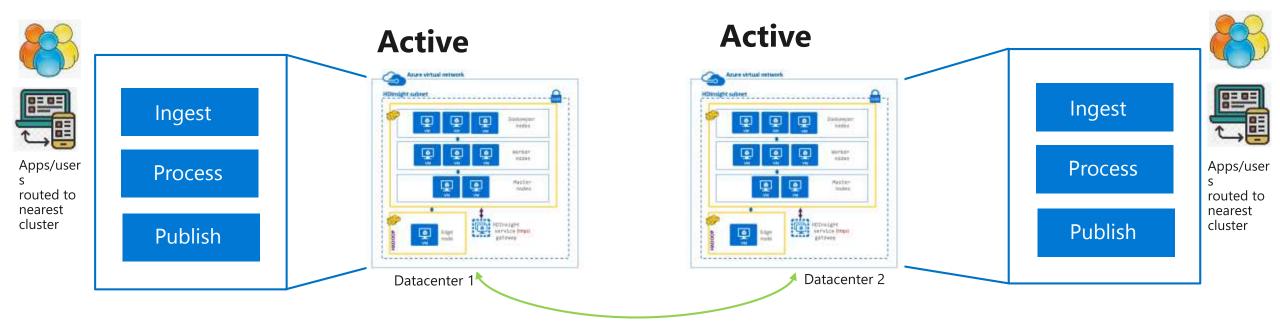- No batch processes are run

@ashishth

Option 3: Active-DR on demand

Active

DR- Cloud Storage

Replication

Provisions cluster ONLY when DR protocol is initiated

| Ingest |
| Process |
| Publish |

RPO **Highest**
RTO **Highest**
Cost **Lowest**
Effort **Highest**

- Applications write to active-primary cluster, no DR cluster provisioned
- Replication utilities synchronize data and metastore to cloud storage in DR datacenter
- Administrator provisions cluster on-demand in DR datacenter
- Synchronization tasks need to be run to ensure clusters are in sync

@ashishth

Option 4: Active-Active with multi-master (WANdisco)

Active

Active

Apps/users routed to nearest cluster

Ingest

Process

Publish

Datacenter 1

Datacenter 2

Ingest

Process

Publish

Apps/users routed to nearest cluster

RPO **Lowest**
RTO **None**
Cost **Highest**

- Applications/integration processes write to/read from nearest cluster clusters
- Both clusters run identical batch jobs
- Synchronization tasks need to be run to ensure clusters are in sync
- WANdisco fusion supports automated metadata replication (sync) and data replication (async), scheduling, monitoring, alerting, bandwidth throttling and more
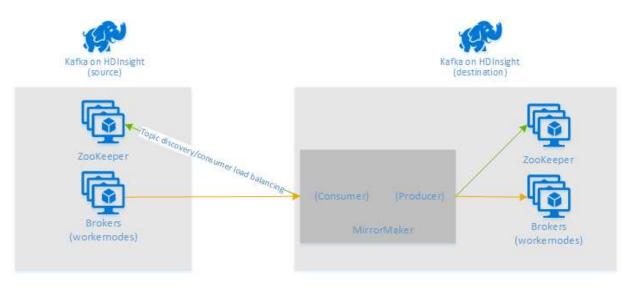
@ashishth

Kafka Replication with MirrorMaker

- Kafka is often deployed in multiple environments for Disaster Recovery, high availability, and on-prem to cloud hybrid scenarios.

- These require replication of data from one Kafka to the other. HDInsight has worked closely with enterprise customers to understand this need, and provides support for data replication scenarios through Apache MirrorMaker

https://docs.microsoft.com/en-us/azure/hdinsight/hdinsight-apache-kafka-mirroring

More Resources

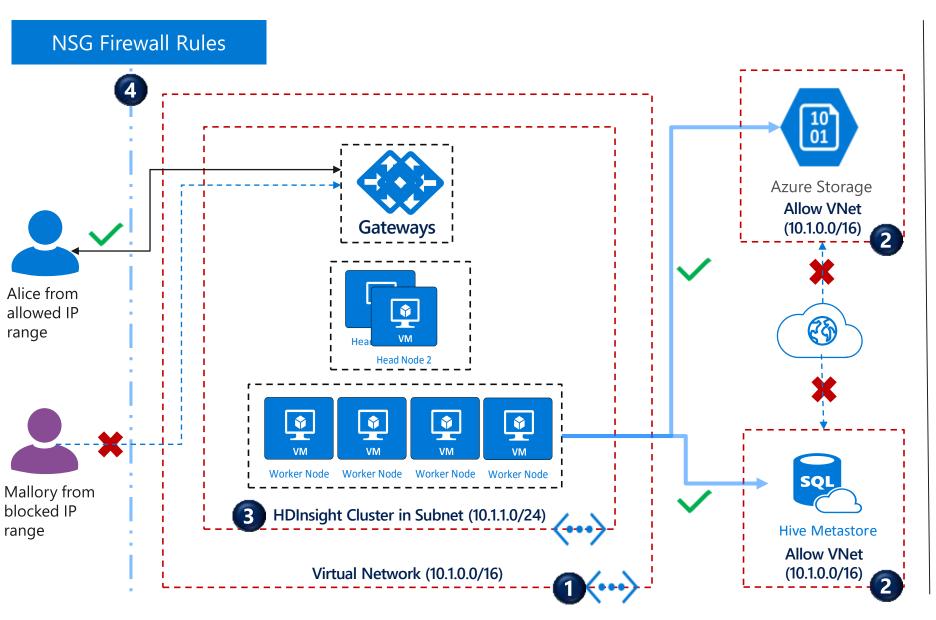Spark/ Hive HA & DR https://github.com/anagha-microsoft/hdi-spark-dr

Kafka HA & DR https://github.com/anagha-microsoft/hdi-kafka-dr

HBase Backup, Replication https://docs.microsoft.com/en-us/azure/hdinsight/hbase/apache-hbase-backup-replication
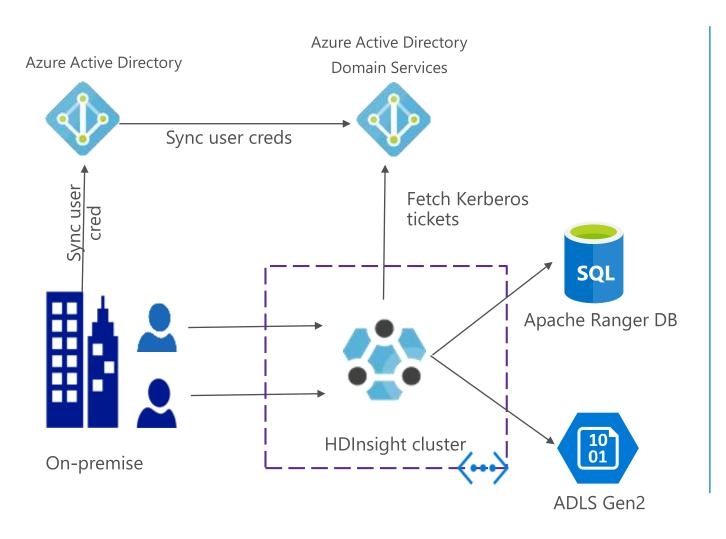
Security

Azure HDInsight Network Security

Securing Data sources with Virtual Network Service Endpoints

NSG Firewall Rules

Alice from allowed IP range

Mallory from blocked IP range

Gateways

Head Node 2

Worker Node · Worker Node · Worker Node · Worker Node

3 HDInsight Cluster in Subnet (10.1.1.0/24)

Virtual Network (10.1.0.0/16)

Azure Storage
Allow VNet
(10.1.0.0/16) **2**

Hive Metastore
Allow VNet
(10.1.0.0/16) **2**

1 Create VNet, a subnet and enable service endpoint

2 Restrict network access to Storage & SQL

3 Create HDInsight cluster within subnet

4 Create NSG rules to control inbound access to HDInsight cluster

@ashishth

Azure HDInsight: Authentication & Access Control

Azure Active Directory

Azure Active Directory Domain Services

Sync user creds

Sync user cred

Fetch Kerberos tickets

SQL

Apache Ranger DB

On-premise

HDInsight cluster

10 01
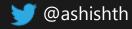
ADLS Gen2

Authentication:
- Supports identities managed in **Azure Active Directory (AAD)**
- Clusters are joined to **Active Directory Domain Services (ADDS)** based Kerberos Domain Controllers.
- On-premise corporate identities are synced to AAD and ADDS via AD Federation Services.

Access Control:
- Apache Ranger based access control and auditing
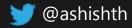- Ranger plugins for Hive, Spark, Kafka and HBase.

@ashishth

Ranger + ADLS Gen 2 Auth Scenarios in HDInsight

| Scenario	Authorizing Component
Yarn: Submit-App	Apache Ranger: Yarn Plugin
Hive Operations: Create, Select, Update, Drop, index, Lock, Read, Write, Masking, Row level filter on Hive Database, Table & Columns	Apache Ranger: Hive Plugin
Create/ Alter Table with storage location reference	Apache Ranger + ADLS Gen 2 ACL's
Spark SQL access with Hive Metastore	Apache Ranger: Hive Plugin
HBase Access Policies	Apache Ranger/ HBase plugin
Kafka Access Policies	Apache ranger/ Kafka Plugin
Access Azure Data Lake Storage Gen2 using the Spark DataFrame API	ADLS Gen 2 ACLs
Access Azure Data Lake Storage Gen2 using the RDD API	ADLS Gen 2 ACLs
HDFS operations: Mkdir, ls, put, copyFromLocal, get, cat, mv, cp etc	ADLS Gen 2 ACLs
Running Map Reduce jobs	ADLS Gen 2 ACLs

Monitoring
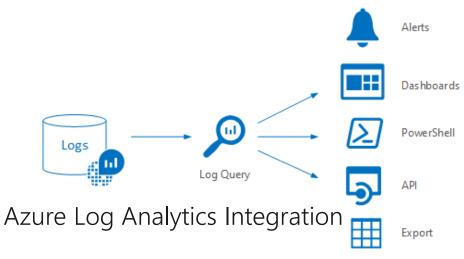
HDInsight Monitoring options

Apache Ambari

- View cluster metrics like CPU, memory, and disk usage at a glance in real time
- Identify malfunctioning components with Ambari alerts
- Monitor queue capacities, jobs, and view associated OSS logs

HDInsight Cluster Metrics

- See gateway requests to monitor cluster stress and cluster size to monitor costs
- Apply filters and chart splitting to extract important data
- Set up alert rules to receive notifications and trigger actions for key metrics

Azure Log Analytics Integration

Logs → Log Query → Alerts, Dashboards, PowerShell, API, Export

- Organizes cluster metrics and OSS log records into queryable tables
- Create custom dashboards to surface all the metrics you need from multiple clusters on a single pane of glass

Cost Optimization

Lower Price

Different VM Pricing Options: HDInsight pricing model is VM price + HDInsight markup

Different VM SKU Options

INSTANCE	VCPU	RAM	TEMPORARY STORAGE	PAYG
D14 v2	16	112.00 GiB	800 GiB	$1.196/hour
E16 v3	16	128.00 GiB	400 GiB	$1.064/hour

11%

Different Region Options
- A8 V2 West US 2 region, HDI price: $0.469/hour
- A8 V2 East US 2 region, HDI price: $0.536/hour

12.5%

@ashishth

Thank You!