BEYOND.2019
DATA ANALYTICS CONFERENCE

ThoughtSpot

# Making Cloud Analytics Work for You at Scale

**Sameer Satyam**
Director of Product Management, ThoughtSpot

**Sangeetanjali Panda**
Senior Systems Reliability Engineer



## Chaz Bademan

**Financial Advisor**
Services Technology Specialist

## AM Grobelny

**Startup Partner Solutions Architect**

# ThoughtSpot in the Cloud




**Deployable as VM instances in your tenancy**

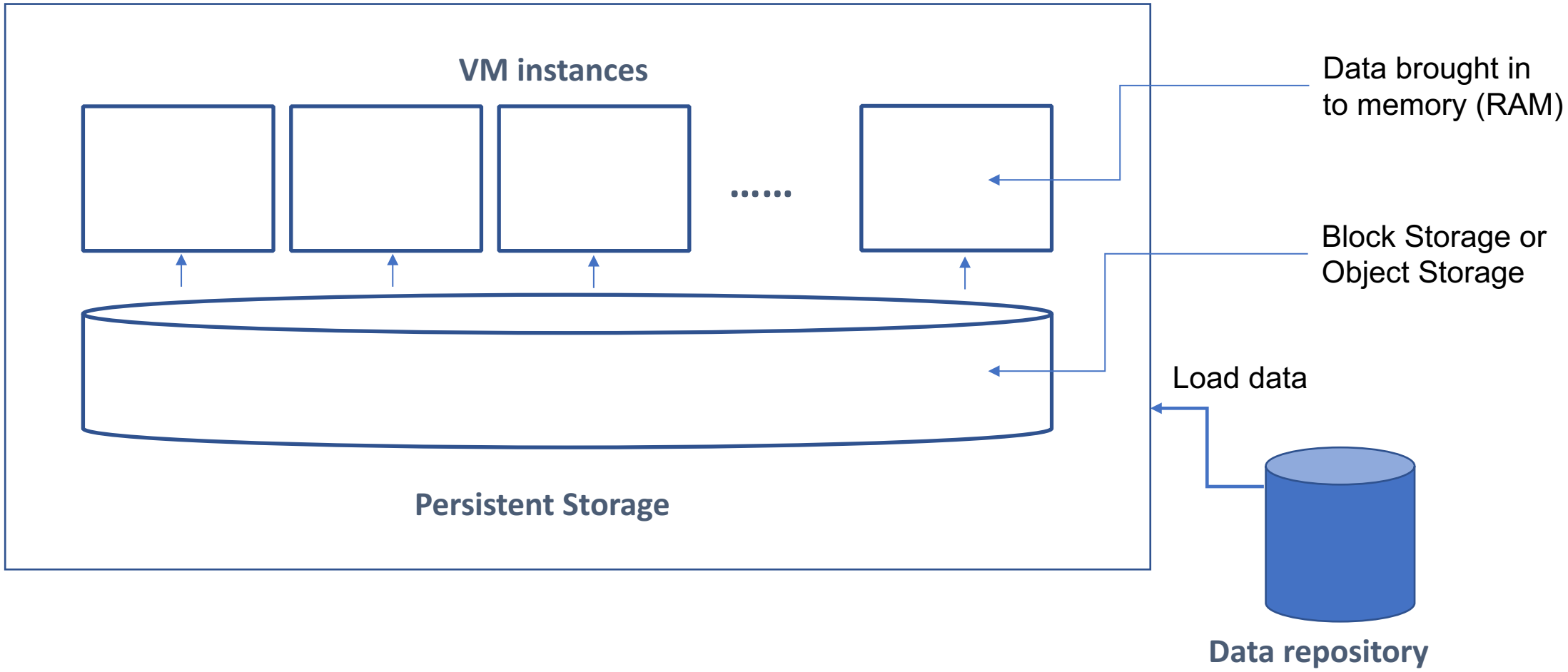
# Cost Reduction Focus

- Multiple instance types
- Changes to persistent storage layer
- Heterogeneous clusters
- Better In-memory Data compression

# Architecture

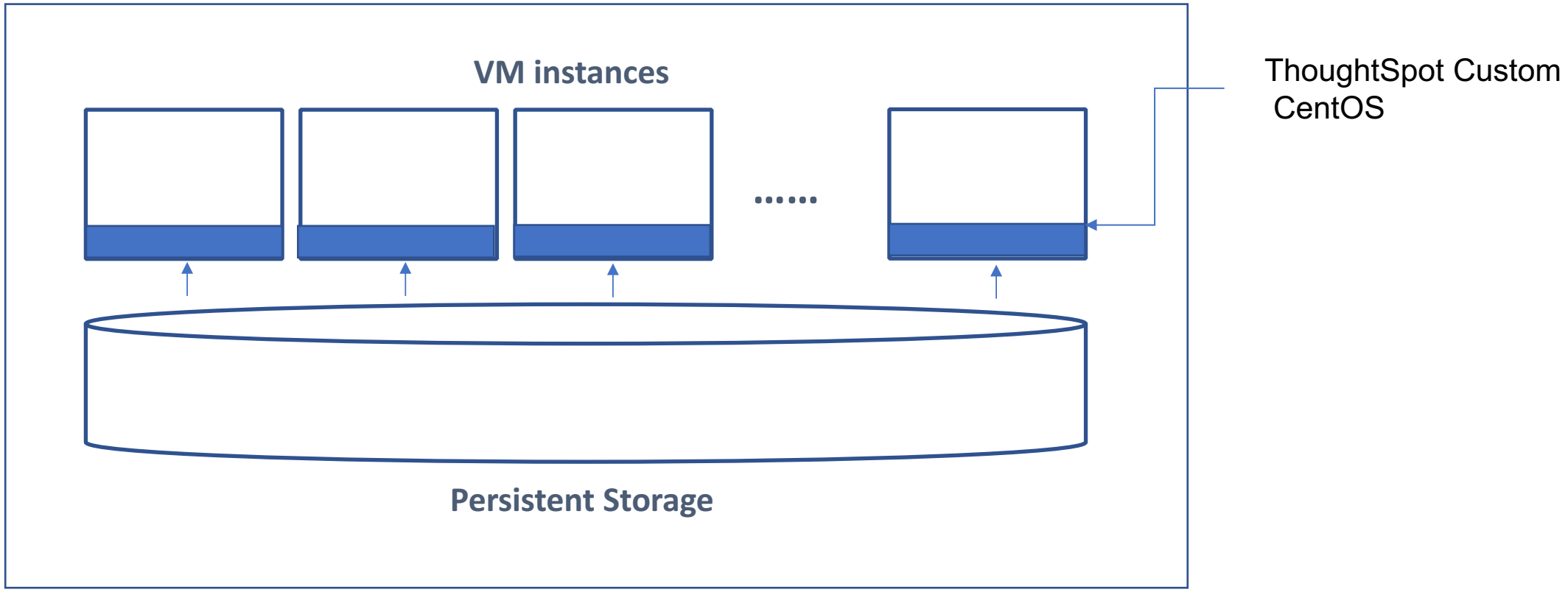
# ThoughtSpot Architecture

## Infrastructure



# ThoughtSpot Architecture

## Operating System



# ThoughtSpot VM instances


- Select instance types from the (memory-optimized) r4, r5 class and the m5 class
- Support for EBS and S3 as persistent storage

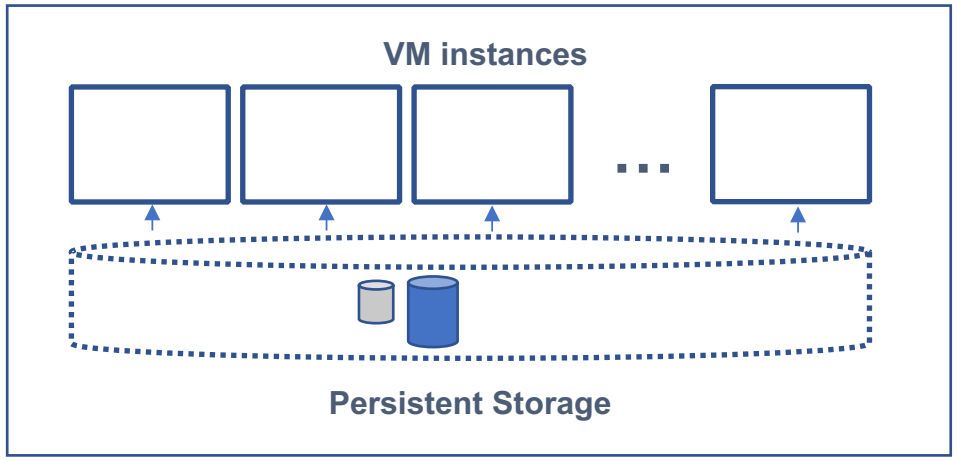

- Select D-series instances (general-purpose) and E-series instances optimized for in-memory applications
- Premium SSD managed disks for persistent storage


- Select instances from the general-purpose high-memory N1 family
- Zonal persistent SSD disks and GCS (Object Storage) for persistent storage

https://docs.thoughtspot.com/5.3/appliance/cloud.html

# Persistent Storage Deployment Options

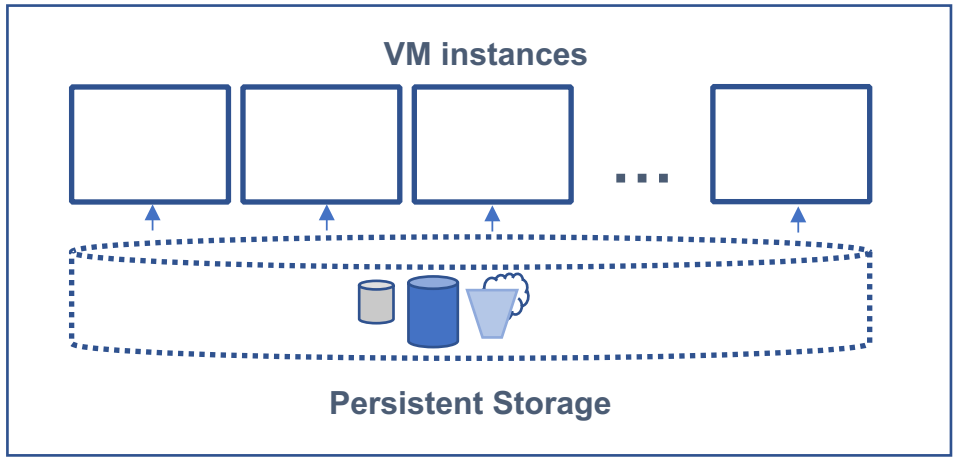
## Block Storage only




250GB boot disk per VM

2TB to 3TB data disk per VM

## Block Storage + Object Storage




**Up to 10% cheaper**

250GB boot volume per VM

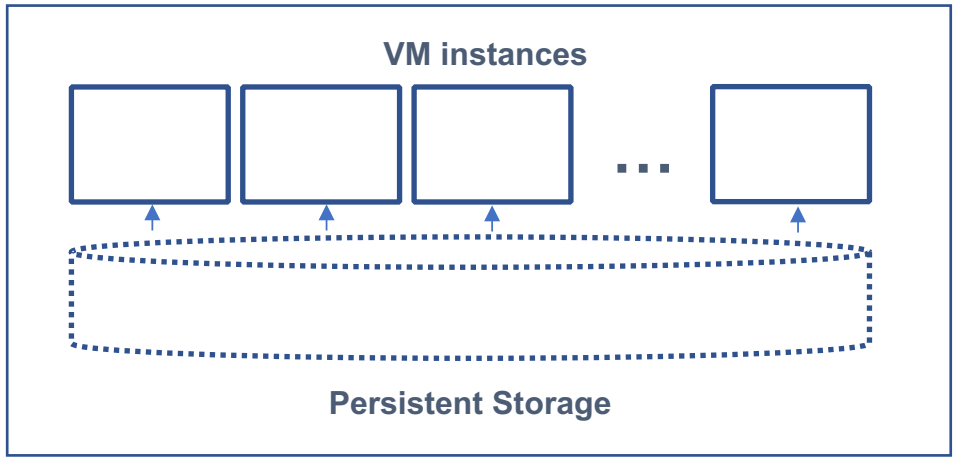
500 GB data volume per VM

Object Storage size approximately equal to size of user data

# Coming soon..

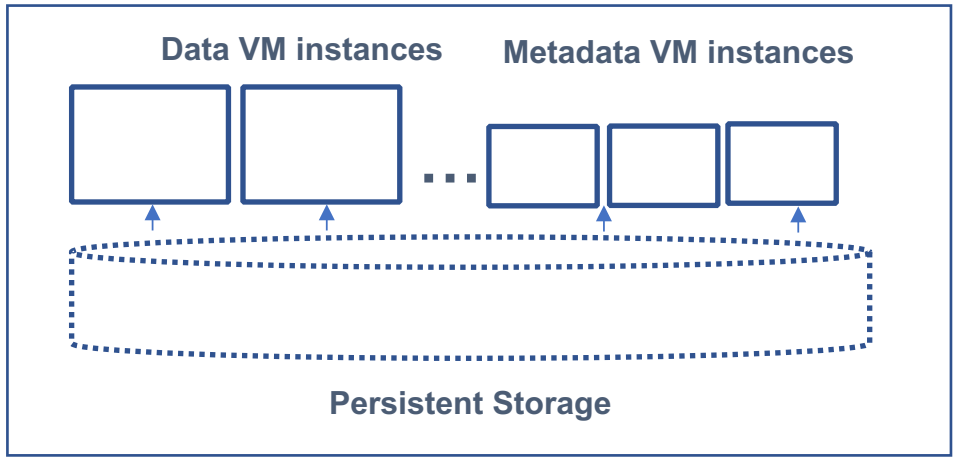
# Heterogeneous Clusters

## Current (Homogeneous clusters)




- All instances of the same type

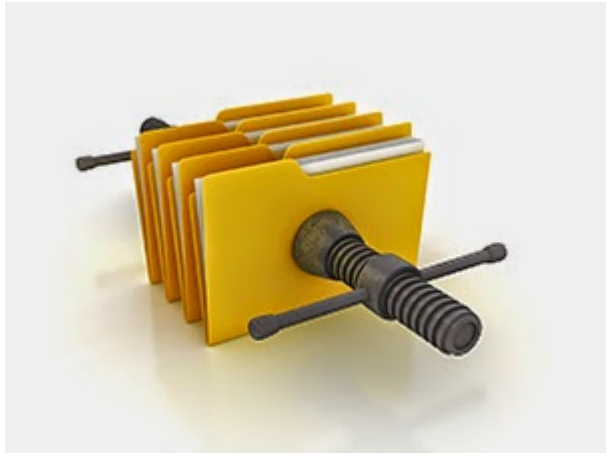
## Coming (Heterogeneous clusters)




**Up to 10% cheaper**

- Separation of instances into data and metadata (smaller) nodes
- Applicable at scale (4TB and beyond)

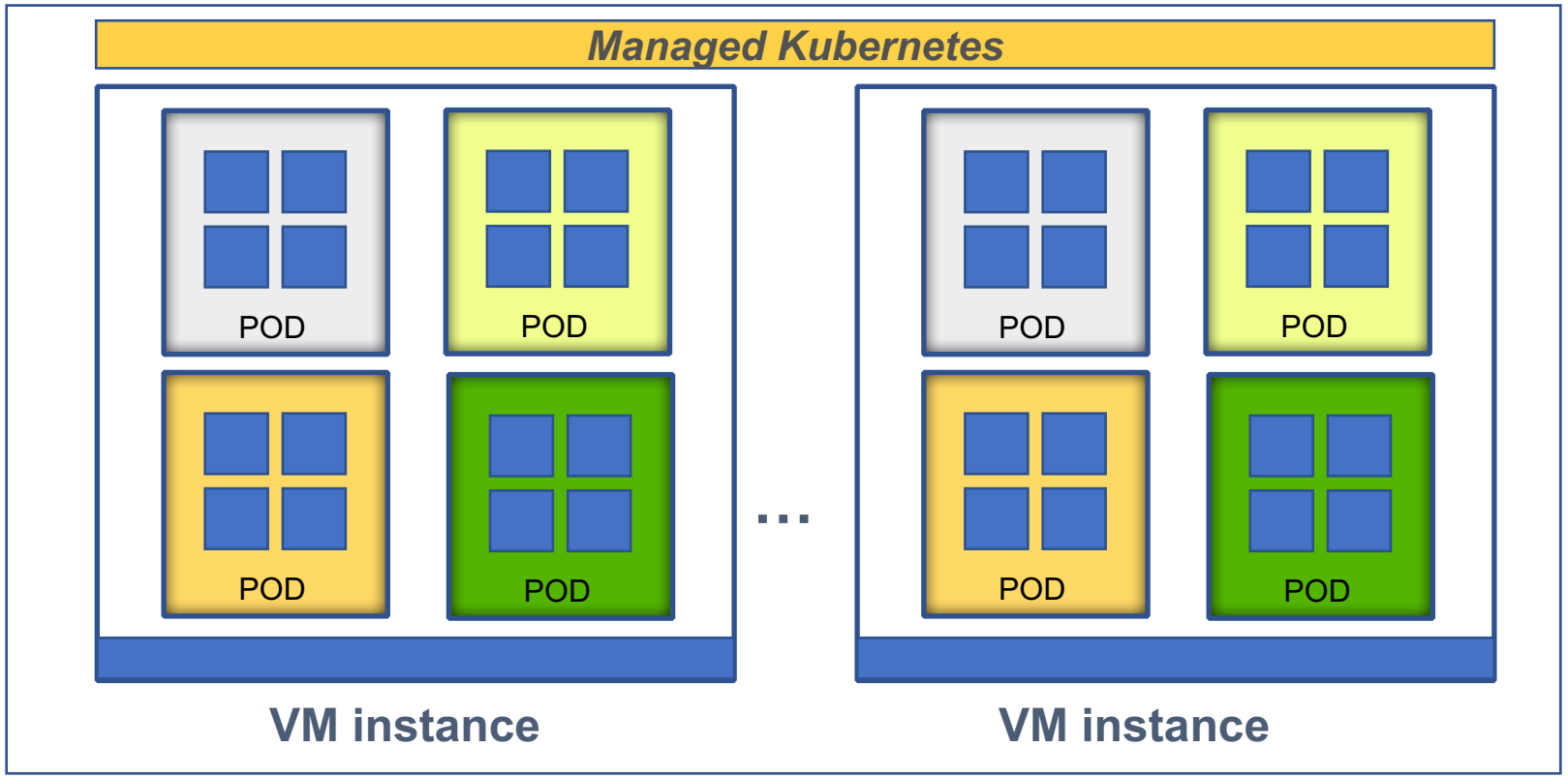
# Better In-memory Data Compression



- Two new compression algorithms (LZ4 and RLE) in addition to the already supported Dictionary compression
- Cost reduction 20% – 40%
- No significant impact to query responses

# ThoughtSpot Platform Evolution

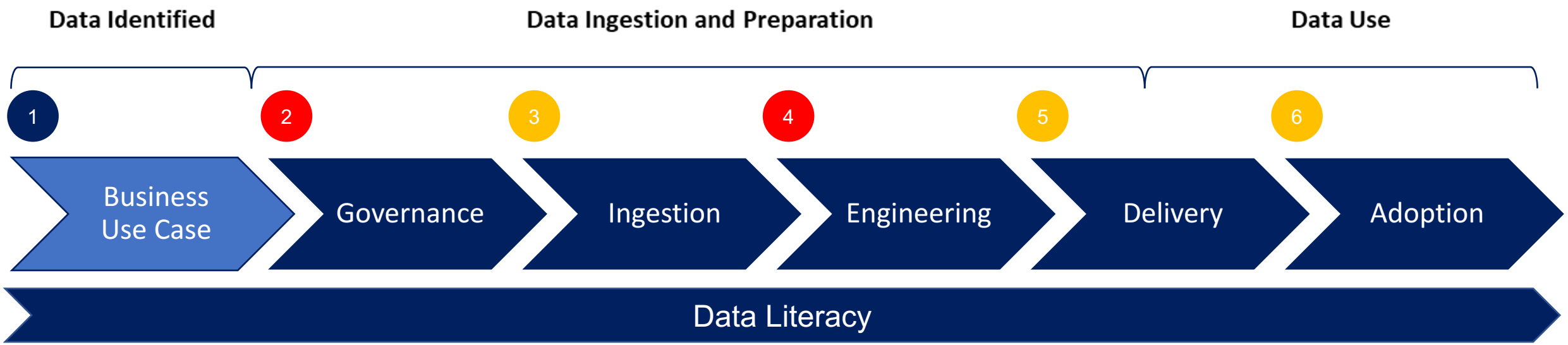
## Bring Your Own OS




- Choice of OS
- Improved Security compliance
- Managed K8s: Standardized cluster management

# Vanguard Use Cases

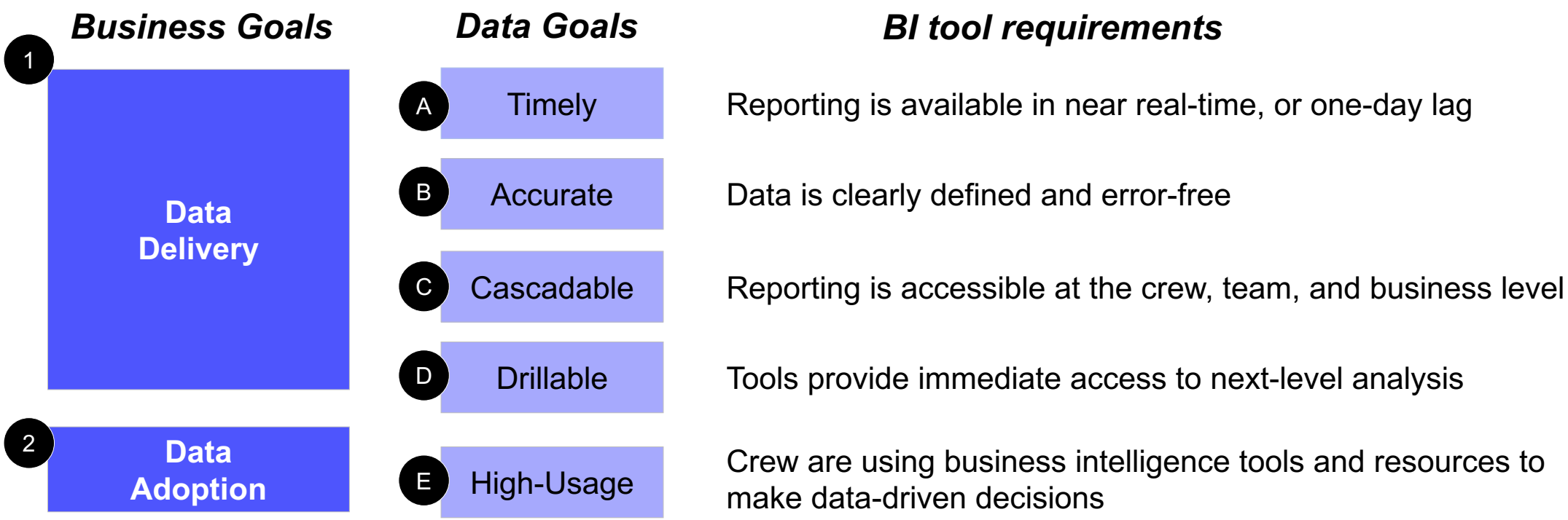
# Vanguard Financial Advisor Services: Data & Analytics




*Provide modern technology for automating cleansing, prepping, visualizing and searching clean and trusted data*

# Why ThoughtSpot ? The Vanguard Vision

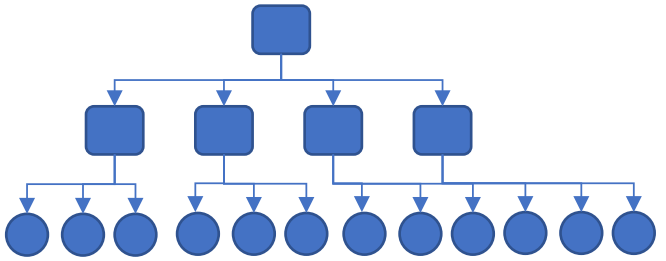
ThoughtSpot unlocks potential to improve business user experience, scale, manage risk and quality, and improve engagement.



# Vanguard Use Case: FINRA compliance

## Call quality

**Consumers**: Business leaders who need insights into call quality




### *Questions*

- Are associates helping with trade execution
- Are calls resulting in resolution

### *Data*

- Sources: Phone call audit records
- Frequency: Daily

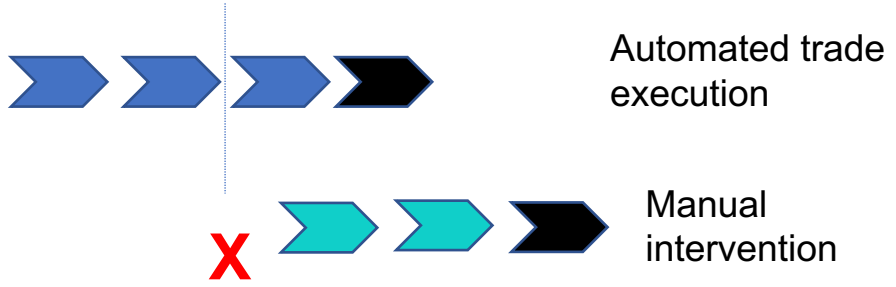
### *Insights*

- Identify outlier associates
- Trend analysis for sub-groups
- Weekly and Monthly tracking

# Vanguard Use Case: Automated trading analysis

## Rejected trade analysis

**Consumers:** Automated Trade Processing team and senior leadership




### *Questions*

- What is the volume of automated trading channels
- Why was a trade rejected (needing manual intervention)

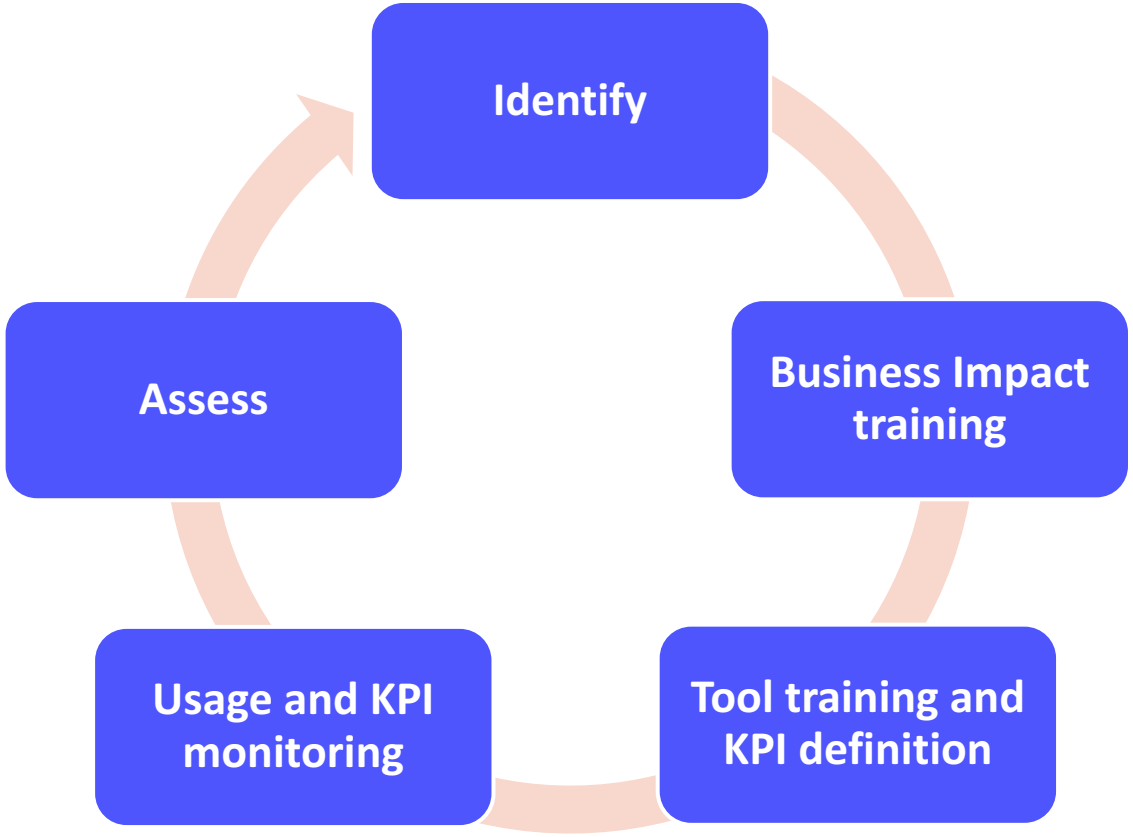
### *Data*

- Sources: Trade data
- Frequency: Daily

### *Insights*

- Track volumes by time
- Understand volumes by clients
- Rejected trade analysis (Manual intervention trades)

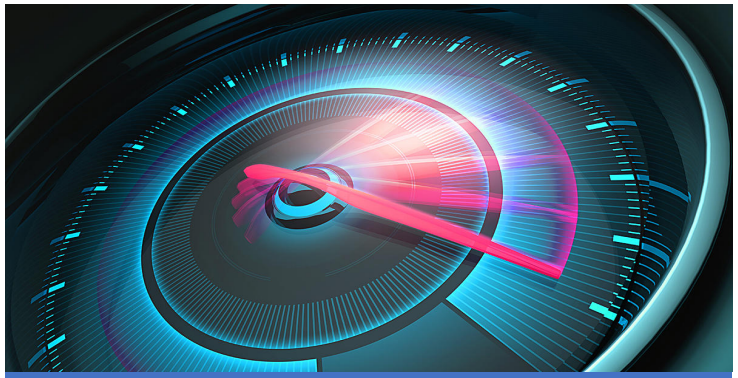
# ThoughtSpot User Adoption Lifecycle at Vanguard




## Success is measured through

- Usage rates
- Reduction of ad-hoc data requests
- Leadership feedback.

# Vanguard: Cloud deployment benefits



Speed and simplicity

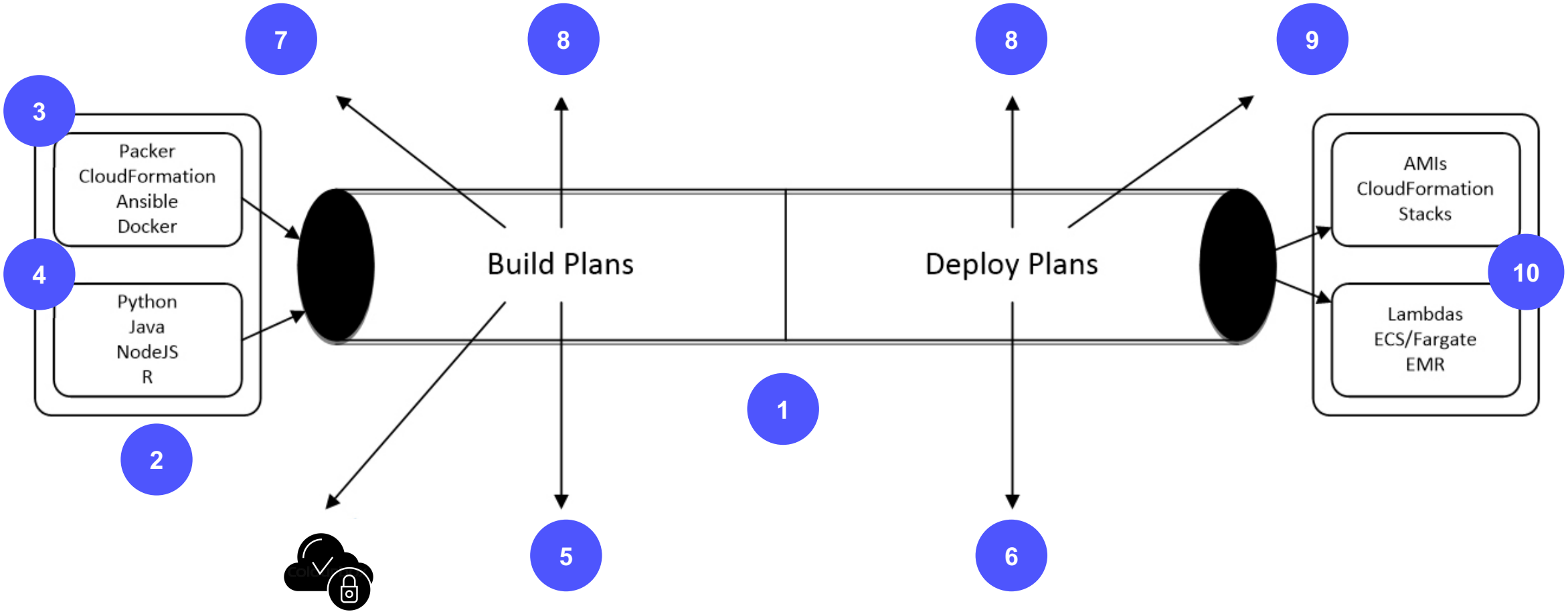
Standardized security

Risk minimization

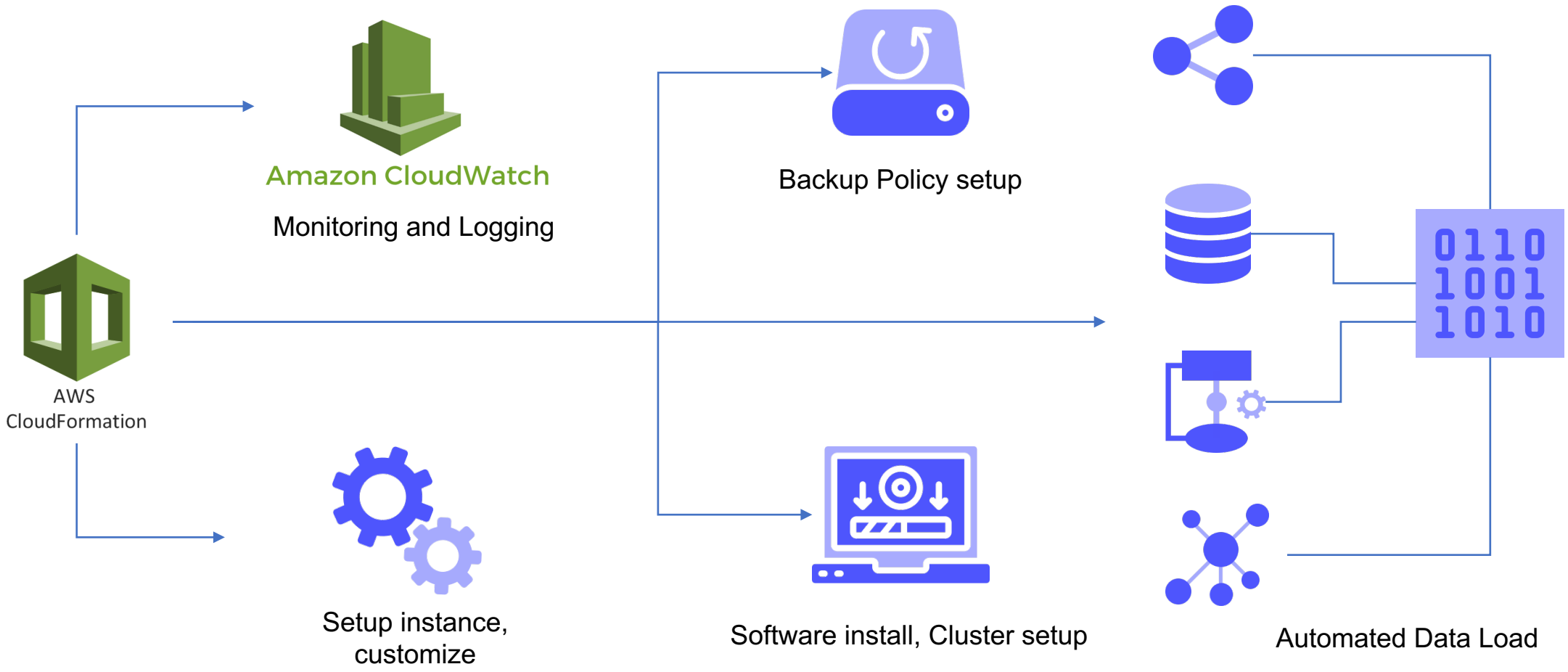
Configuration consistency


Cost savings

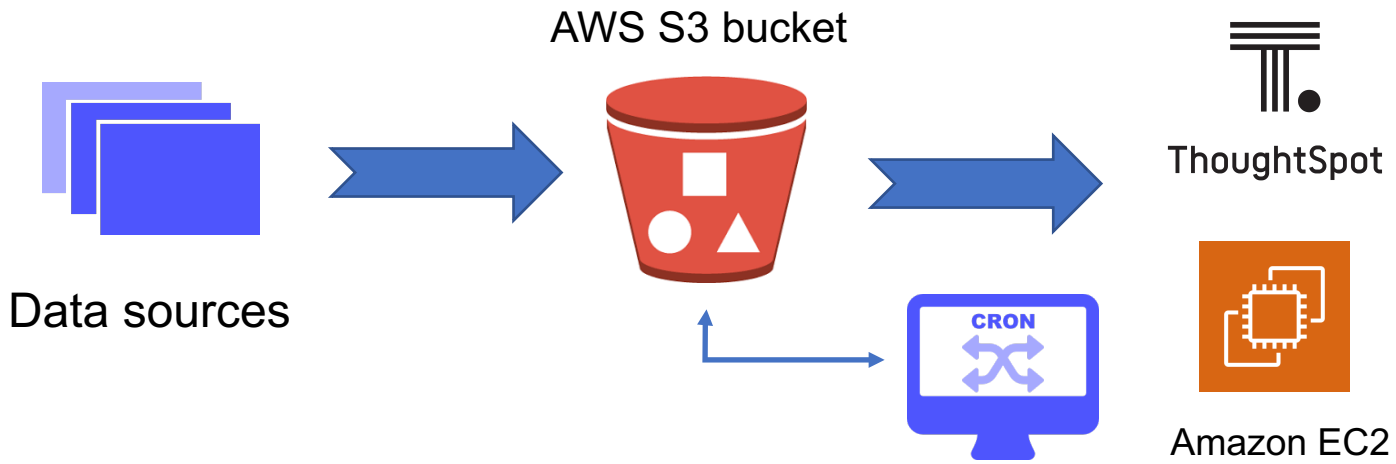
# Automation Framework



# Automated Deployment Lifecycle



# Vanguard ThoughtSpot Data loading



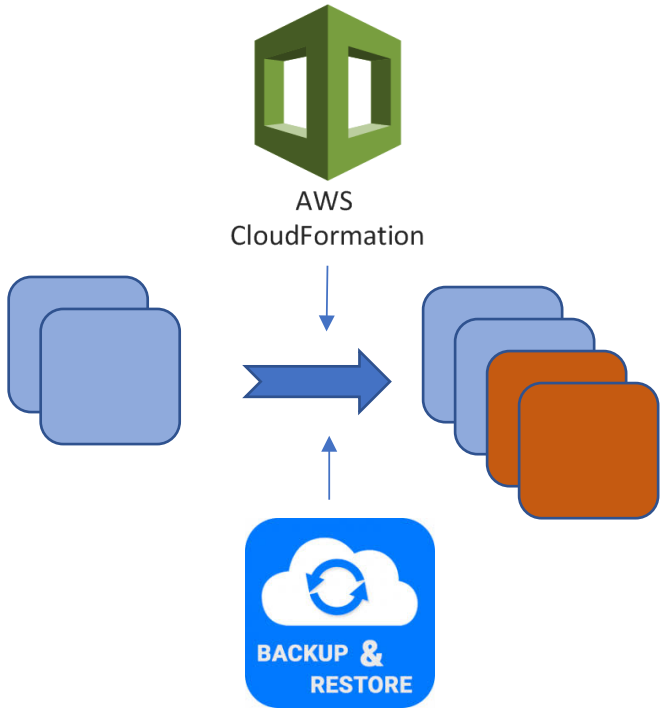

TS_Ingestion_Schedule

Columns | Schema | Data | Profile | Dependents | Row Security | SpotIQ Insights | SearchIQ

| FREQUENCY | CRON_SCHEDULE | ACTIVITY_TYPE | SOURCE_FILE_BUCKET | SOURCE_FILE_PREFIX | SOURCE_FILE_NAME | SOURCE_FILE_DELIMITER | HEADER_ROW_FLAG | TARGET_DB | TARGET_SCHEMA |
|---|---|---|---|---|---|---|---|---|---|
| Daily | 15 13 * * * | I | vgi-fas-test-us-east-1-fdi-... | IT-Regression-testing | IT_Cleansed_Test_Data.txt | , | Y | FAS_DB | Test |
| Daily | 15 13 * * * | I | vgi-fas-test-us-east-1-fdi-... | IT_Regression_testing | IT_Prepped_Test_Data.txt | , | Y | FAS_DB | Test |
| Delete | 55 11 * * * | D | vgi-fas-test-us-east-1-fdi-... | IT-Regression-testing | IT_Cleansed_Test_Data.txt | , | Y | FAS_DB | Test |
| Delete | 55 11 * * * | D | vgi-fas-test-us-east-1-fdi-... | IT_Regression_testing | IT_Prepped_Test_Data.txt | , | Y | FAS_DB | Test |
| OneTime | (null) | O | vgi-fas-test-us-east-1-fdi-... | IT-Regression-testing | Cibulsky_CASS_Addres... | , | Y | FAS_DB | Test |

# Scale, Deploy, and Restore from Backup

- Deployment and configuration of additional ThoughtSpot nodes automated using AWS CFT
- Initial Deployment in region sets a restore flag
- Backup policy leverages ThoughtSpot Backup & restore functions
- Save backup to specified S3 location



# AWS Automation Tools

# Infrastructure as Code with AWS

## AWS CloudFormation

AWS CloudFormation offers declarative templates for AWS infrastructure in JSON and YAML

Additional features:

- cfn-init
- cloud-init
- Custom Resources


## AWS CDK

Use imperative programming languages to generate AWS CloudFormation templates

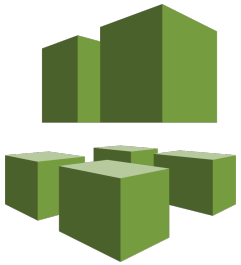

## HashiCorp Terraform

HashiCorp Terraform offers a cloud-agnostic infrastructure management option and is an AWS Partner.

# Configuration Management with AWS

## AWS OpsWorks

Provides managed Chef and Puppet instances to keep your running systems in a desired state



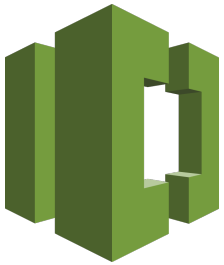
## AWS Systems Manager

Manage servers without using SSH or RDP, run automations against on-prem and AWS servers, keep desired state on your systems, and store configuration values.


## Ansible

A popular open source automation engine that has AWS playbooks available for automating deployments to AWS resources

# CI/CD with AWS



## AWS CodeCommit

A version control system to manage your code repositories. Allows for complete access controls and integrates with CI/CD pipelines

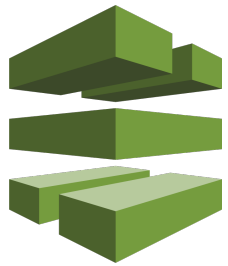
## AWS CodeBuild

Customize and preconfigure build environments with build tools like Gradle, npm, MSBuild, and more. Automates building code artifacts, Docker images, and more

# CI/CD with AWS

## AWS CodeDeploy

Automates deployment to Amazon EC2, AWS Fargate, AWS Lambda, and on-prem targets. Performs blue/green deployments, rolling updates, and more.



## AWS CodePipeline

Orchestrates the other AWS Code* tools to perform a full CI/CD process. Directly integrates with several AWS services, add Jenkins or GitHub with pre-built plugins, and even include manual approval steps into your pipeline

# CI/CD with AWS – An Easy Way to Get Started

## AWS CodeStar

AWS CodeStar creates a full CI/CD environment for your team in minutes

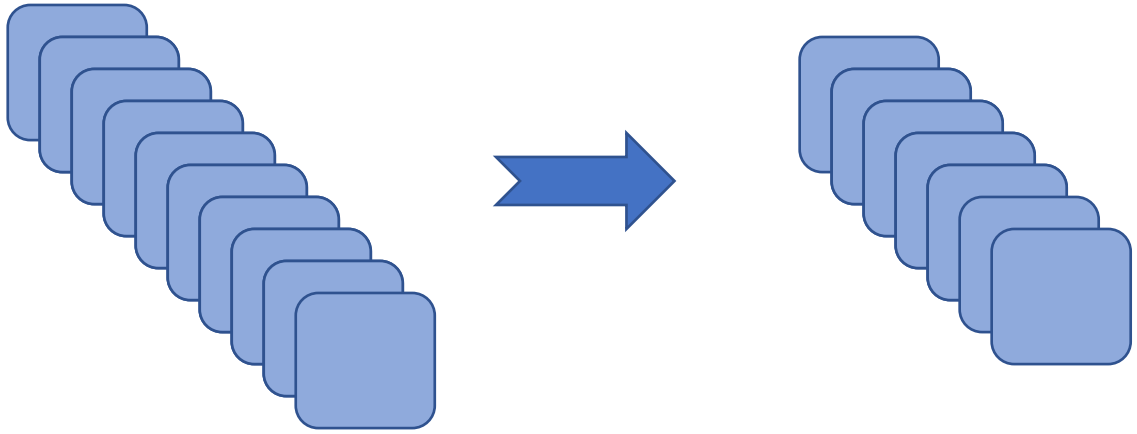
- JIRA integration
- Project templates
- Bundles all the AWS Code* services

# Demo

# In-memory Data Compression

- Demo of compression and RAM (cost) savings
- Sample synthetic product usage data
- Using Run Length encoding on all columns
- Query performance comparison with sample query

# In-memory Data Compression results



40% reduction in number of VMs needed

# Cloud Roadmap

| Delivered | November 2019 | Q1/Q2 2020 |
|---|---|---|
| • Multiple Instance types<br>• Persistent storage improvements on AWS<br>• Load data directly from AWS S3 buckets<br>• Data at rest and in-transit encryption | • Persistent storage improvements on GCP<br>• In-memory Data Compression<br>• Load data directly from Google Cloud Storage<br>• Heterogeneous clusters<br>• Backup to S3/GCS | • Bring Your Own OS (Kubernetes)<br>• Persistent storage improvements on Azure<br>• In-memory compression enhancements<br>• Load data /Backup to Azure Blob |

BEYOND.2019
DATA ANALYTICS CONFERENCE

# Thank You