Speedy Server Deployment on AWS

Anoop Kunjuraman
Software Engineering Manager
Capital One

Michael Foree
Software Engineer
Capital One

#data19
Tableau and Capital One
Capital One

10th largest bank in US
Fortune 100 company

3rd largest credit card issuer

~50,000 Associates

Banking Reimagined
Tableau and Capital One

Tableau is used by EVERYONE

- Sale Managers
- Call Center Agents
- Data Analysts
- Accountants

Tableau is used for EVERYTHING

- Tracking originations
- Monitoring delinquencies
- Risk model monitoring
- Employee volunteer engagement
Our AWS Tableau Journey at Capital One
Agenda

- 5 Point Check List
- Key Drivers for Successful Deployment
- Common Mistakes in Deployment
- Monitoring & Managing Tableau
5 Point Tableau Deployment

Your Tableau Deployment should meet these 5 criteria

- Change Controlled Codebase
  - Infrastructure is defined as code
  - Binaries are stored in the environment

- Infrastructure Dependencies
  - Clearly identified and loosely coupled from each other

- Environment Isolation and Parity
  - Production, Disaster Recovery, and Test are similar but separate

- Disposable Infrastructure
  - Fast Startup
  - Scalable
  - Graceful Failures

- Thoughtful Design
  - Designed to meet RTO and RPO timelines
  - Minimizes manual interaction
<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flexible</td>
<td>Blue-Green Deployment provides fast failover</td>
</tr>
<tr>
<td>Well Documented</td>
<td>Consolidated Documents Simplifies Version Control</td>
</tr>
<tr>
<td>Organized</td>
<td>Solid Organization Structure Facilitates Upgrades</td>
</tr>
<tr>
<td>Robust</td>
<td>Regional Migration is a Breeze</td>
</tr>
<tr>
<td>Optimized</td>
<td>Optimized Configurations reduces manual clicks From &gt;1000 clicks to 10</td>
</tr>
<tr>
<td></td>
<td>Prod and QA differences are clearly delineated</td>
</tr>
<tr>
<td></td>
<td>Test clusters are completely isolated from Prod cluster</td>
</tr>
<tr>
<td></td>
<td>Move anywhere in the country in 1 hour – without prep work</td>
</tr>
</tbody>
</table>
Typical Terminology

- **S3 Bucket**: Stores data in the cloud
- **EC2**: On-Demand computer
- **Route 53**: AWS branded DNS server
- **Region**: Geographically isolated collection of cloud services
- **Availability Zone**: Small section within a region
- **Cloud Formation Template**: Text document describing your infrastructure
Capital One’s AWS Infrastructure

Main Region

- Tableau Production Cluster
- Tableau Staging Node
- Availability Zone

Availability Zone

Upload backups
Download settings

DR Region

- Tableau Production Cluster
- Tableau Staging Node
- Availability Zone

Availability Zone

Upload backups
Download settings

Cross Region Replication

Monitor KPIs

Send Alerts to Administrators

Tableau Production Cluster

Tableau Production Cluster

Tableau Production Cluster

Tableau Production Cluster

Tableau Production Cluster

Tableau Production Cluster

Tableau Production Cluster

Tableau Production Cluster

Tableau Production Cluster
Key Drivers for Successful Deployment on AWS
Parameterized Scripts Connect Each Component

- S3 Bucket
- CloudFormation
- Initial Node
- Additional Nodes
- Amazon CloudWatch
- S3 Bucket
- Tableau Backup

Deployment Checklist
- Changed Controlled Codebase
- Infrastructure Dependencies - Loosely Coupled Components
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
Cloud Formation Template Initiates Creation

- Manually Assigned
  - Region
  - Production/Test Deployment
  - Host URL

- Dynamically Identified
  - Security Groups
  - AMI
  - S3 Bucket Name
  - Account
  - Initial Tableau Node
  - Additional Tableau Nodes

Deployment Checklist
- Changed Controlled Codebase
- Infrastructure Dependencies – Infrastructure as code
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
S3 Buckets Reliably Store All Data

Structure Files For Consistency
• Shell Scripts
  • Install Instructions, Configuration Files,
  • Topology Files, Node Configuration File
• Drivers
• Binary Files
• Cloud-Formation-Template JSON Files
• Tableau Backup Files
  • Latest Backup File
  • Historical Backup for Prod Region and DR Region

Deployment Checklist
✓ Changed Controlled Codebase
☐ Infrastructure Dependencies – Infrastructure as code
✓ Environment Isolation & Parity
☐ Disposable Infrastructure
☐ Thoughtful Design
EC2s Run Automated Shell Scripts

Shell Scripting is Initiated by CFT

- **Input Parameters:**
  - S3 Bucket, Node Identifier, Prod/Test Deployment, Host URL

- **Dynamically Calculate:**
  - Trusted Host, SAML file, Email Configurations, Backup File Name, Node Configuration File

- **Download all files from the S3 bucket**

- **Uploads:**
  - Tableau Backup Files, Topology File,
  - Node Configuration File is downloaded by additional nodes

---

**Deployment Checklist**
- Changed Controlled Codebase
- Infrastructure Dependencies – Loosely Coupled Components
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
EC2s Scripting is bifurcated to allow for testing

Stage 1 Scripting
• Activates Tableau as trial license
• Disables Subscriptions
• Configures with a test Host URL
• Restores from latest production Tableau backup file

Stage 2 Scripting
• Initiated after testing is completed
• Activates Tableau with license
• Enable Subscriptions
• Configures with production URL
• Restores Tableau repository

Deployment Checklist
☑ Changed Controlled Codebase
☑ Infrastructure Dependencies – Loosely Coupled Components
☑ Environment Isolation & Parity
☐ Disposable Infrastructure
☑ Thoughtful Design
Cloud Watch and SNS Topics Alert Admins

- **Use Cloud Formation Template to create Alarms**
  - EC2 CPU Usage
  - EC2 Network utilization
  - EC2 Status unavailable

- **Subscribe to SNS topics for alerts**
  - Tableau Admins
  - Text Messages
  - 3rd Party monitoring Systems

---

**Deployment Checklist**
- Changed Controlled Codebase
- Infrastructure Dependencies – Loosely Coupled Components
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
Common Mistakes In Deployment
Common Mistakes to Avoid When Deploying Tableau

- **Reducing common Errors**
  - Logging in to TSM
  - Log error codes
  - CFTs are sensitive to formatting
  - Install drivers last
  - Long polling times
  - Coordination service cleanup
  - Recommended order of applying settings:
    - login, activate, register, other settings, repository access, apply *but don’t start Tableau*
    - Restore from backup, Generate Node Configuration File, start Tableau

---

**Deployment Checklist**
- Changed Controlled Codebase
- Infrastructure Dependencies – Loosely Coupled Components
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
Common Mistakes to Avoid When Deploying Tableau

- **Restarting Tableau takes a long time**
  - Minimize restarts
    - Stop Tableau as soon as it is installed
  - Changes that require a restart
    - Apply pending changes
    - Repository access
    - Restore from backup
    - Adding or removing a node
    - Change Cluster Controller
    - Moving File Store

---

**Deployment Checklist**

- Changed Controlled Codebase
- Infrastructure Dependencies – Loosely Coupled Components
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
Let’s See This In Action
Monitoring Tableau Infrastructure
We monitor our Tableau cluster with a suite of tools

- CloudWatch
- Prometheus
- Pagerduty

**Deployment Checklist**
- Changed Controlled Codebase
- Infrastructure Dependencies – Loosely Coupled Components
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
CloudWatch is Native to AWS

- CPU monitoring
- Status Checks
- Network load
Prometheus Integration Displays Trending of Many Metrics

- CPU usage
- Memory consumption
- Network load
- Hard drive load
- Error Packets
- Hard drive space
Pagerduty Notifies on-call Personnel When Critical Metrics are Tripped

- CPU usage
- Memory consumption
- Network load
Managing Tableau Well
Managing Tableau properly is critical for a healthy environment

- Automated backups
- Workbook refresh details are sent to Capital One Streaming Data Platform
- SRE team monitors
  - Workbook refresh failures
  - Time for data source refreshes
  - Tableau Process statuses

Deployment Checklist
- Changed Controlled Codebase
- Infrastructure Dependencies – Loosely Coupled Components
- Environment Isolation & Parity
  - Disposable Infrastructure
- Thoughtful Design
Governing Tableau
Governing Tableau ensures a solid platform going forward

- **Two clusters**
  - Staging and Production
  - SLA enforced on migration to Production

- **Auto delete users who leave Capital One**

- **Auto delete abandoned workbooks**

---

**Deployment Checklist**

- Changed Controlled Codebase
- Infrastructure Dependencies – Loosely Coupled Components
- Environment Isolation & Parity
- Disposable Infrastructure
- Thoughtful Design
Upcoming Development

- Triggered Builds of our DR Region Cluster
- Dynamic Tableau Topology
- Programmatic Datasource Creation
- Enhance Existing Email Notification Service
Please complete the session survey in the mobile app

View ‘My Evaluations’ in the menu or find your session under ‘Schedule’
Thank You

@anoopengineer
@Mforee741