**Concept Note**

# **Introduction**

The Directorate General of Hydrocarbons (DGH) was established in 1993 under the administrative control of the Ministry of Petroleum & Natural Gas through a Government of India Resolution. Its primary role is to regulate and oversee upstream activities in the petroleum and natural gas sector, while also providing advice to the Government in these areas. The objectives of DGH include promoting the responsible management of oil and natural gas resources, with careful consideration for environmental, safety, technological, and economic aspects of petroleum activities.

DGH is responsible for implementing, monitoring, and regulating operations related to oil and gas blocks throughout their entire lifecycle, encompassing the stages of bidding, exploration, production, site restoration, and abandonment.

# **Context:**

To fulfil its responsibilities, DGH requires various data, reports, and information from operators as needed. Throughout the E&P operation's lifecycle, operators also require periodic approvals, clearances, and guidance from DGH.

These interactions between DGH and the operators generate numerous documents and communications, which are crucial for compliance, regulatory, and decision-making purposes.

In the current system, stakeholders within DGH typically store this digital information on their individual machines, leading to the creation of multiple data silos. This fragmentation increases inefficiencies and delays in the approval process.

Furthermore, localized storage of sensitive data presents significant risks, such as unauthorized access and data exfiltration due to external attacks or insider threats, including IT misconfigurations, human error, or other security vulnerabilities.

# **Objectives**

To address the gaps and shortcomings in the current system, a comprehensive IT solution is required. This solution should include centralized storage for critical data, with regular backups aligned with industry best practices. Additionally, a robust mechanism is needed to secure sensitive information on end-user devices, preventing accidental or malicious leakage, loss, misuse, or destruction.

The primary objectives of the comprehensive IT solution are as follows:

1. **Enhanced Data Security:** Monitoring and controlling the transmission, access, and sharing of data within and outside the organization to protect sensitive information.

2. **Centralized Management and Control:** Enforcing IT policies, managing access controls, and monitoring data usage from a single, centralized point.

3**. Improved Compliance and Regulatory Adherence:** Ensuring that data retention and archival processes meet compliance and regulatory standards.

4. **Cost Efficiency:** Reducing the need for multiple storage systems across departments, thereby minimizing redundancy and lowering costs.

5. **Improved Data Visibility and Analytics:** Enhancing the ability to identify potential risks and mitigate them, leading to better decision-making and more strategic use of information across the organization.

6. **Enhanced Collaboration and Productivity:** Facilitating secure information sharing within the organization, reducing the risk of data breaches, and boosting overall productivity.

# **Proposed Solution**

A comprehensive IT solution shall be procured through competitive bidding on the GeM platform. The solution will consist of three key components to achieve the desired objectives:

***Centralized Network Attached Storage (NAS):*** To provide a unified storage system for critical data, ensuring easy access, management, and scalability across the organization.

***Backup Storage and Software:*** To regularly back up data according to industry best practices, ensuring data integrity, availability, and quick recovery in case of data loss or system failures.

***Data Loss Prevention (DLP) Solution:*** To safeguard sensitive information by monitoring, detecting, and preventing unauthorized access, transmission, or leakage of data, both within and outside the organization.

The required features of each component are broadly listed below.

1. ***Centralized Network Attached Storage (NAS)***
2. **Capacity**:
   * Initial usable capacity: 200 TB
   * Scalability: Up to 500 TB
3. **Performance**:
   * At least 10% of capacity on NVMe storage for high-performance tier
   * Auto-tiering: Automatically place data on appropriate storage class based on access frequency
4. **High Availability**:
   * No Single Point of Failure (NSPOF) for both front-end and back-end connectivity
5. **Data Optimization**:
   * De-duplication and compression functionality before data is written to actual data drives
6. **Data Protection**:
   * Native support for point-in-time copies for specific paths or directories
   * User, user group, and directory quotas
   * Quotas definable per directory and per user
7. **Security**:
   * Encryption support with compatibility for internal and external key management systems
   * WORM (Write Once, Read Many) capabilities to prevent data modification or erasure
8. ***Backup Storage and software***
9. **Backup Storage:**
   * Provide a low-cost storage solution with a capacity of at least 30% of the production storage
   * Ensure redundancy by storing backups in a separate rack from the production storage
10. **Backup Software:**
    * Policy-driven automation: Automate backup processes based on defined policies
    * Monitoring and reporting: Provide real-time monitoring and detailed reporting of backup operations
    * Deduplication: Ensure efficient storage by deduplicating data
    * Customized retention policies: Allow for tailored retention policies for recovery and compliance purposes
    * Cloud support: Support backup, recovery, and migration to, in, and from the cloud
11. ***Data Loss Prevention (DLP) solution***
12. **Device Control**
    * Restrict USB and auxiliary devices from accessing and copying sensitive data
    * Allow printing of relevant documents with a watermark for user convenience
13. **Content Inspection**
    * Automated in-depth content inspection to detect structured and unstructured sensitive data
    * Pinpoint locations of sensitive data
14. **Policy Creation**
    * Automatic, behavior-based DLP policy creation and extension
15. **Content Aware Protection**
    * Scan data in motion to prevent sensitive data from leaving the network
    * Monitor data movement and control exit points (web browsers, instant messaging, cloud storage uploads, etc.)
16. **Data Categorization**
    * Consolidate and categorize sensitive data using:
      1. Predefined templates
      2. Custom mechanisms (keyword search, fingerprinting, document matching, RegEx pattern)
17. **Data Shadowing**
    * Mirror data copied to external devices (serial, parallel, network ports) in violation of DLP policies
18. **Clipboard Control**
    * Selectively control user and group access to clipboard (copying files, text, images, etc.)

# **Implementation Strategy**

The deployment of the comprehensive IT solution will be divided into three phases to ensure a structured and efficient rollout, minimizing disruptions and ensuring a smooth transition for users:

**Phase 1: User Onboarding and Directory Setup**

* Onboard users to Microsoft Active Directory platform
* Ensure hierarchical segregation of users based on departments and functional reporting

**Phase 2: Storage and Backup Deployment**

* Deploy and configure Centralized NAS storage in DGH premises
* Allocate storage pools and enable auto-availability of mapped directories to respective desktop PCs
* Deploy and configure backup solution to ensure redundancy and data availability

**Phase 3: DLP Solution Deployment**

* Install and configure DLP solution on end-user desktops as per organization's policies
* Ensure compliance with data protection and security requirements