Enterprise Steam

Secure, Self-Service Artificial Intelligence Environments with Comprehensive IT Control

Enterprise Steam is a service to securely start and connect to H2O YARN jobs in a secure Hadoop environment. Encouraging adoption among Data Science users is also a key goal, so ease of use is paramount. Key user personas include Data Scientists, Hadoop Admins, Enterprise Architects, and IT Security Specialists. Enterprise Steam offers **security, resource control** and r**esource monitoring** out-of-the-box in a **multi-tenant architecture** so that organizations can focus on the core of their data science practice.

Enterprise Steam equips the stakeholders in AI practices with the capabilities required to perform their tasks without interfering with each other. Simply put, Enterprise Steam enables streamlined H2O adoption in a secure manner that complies with company policy. Administrators can easily control, monitor and measure H2O usage. This further enables use cases such as internal chargeback, internal cloud deployment, and H2O Platform as a Service (PaaS).

Enterprise Steam provides the following benefits to Artificial Intelligence (AI) practitioners:

Data Scientists

- Self-Service
  - Enterprise Steam provides easy R/Python APIs and a Web UI for starting H2O YARN jobs.
  - Without having to become Hadoop experts, Data Scientists can manage H2O clusters and connect to them using a stable service at a known IP address and port.
- Familiar Interface
  - Data Scientists can work in the comfort of familiar environments such as RStudio and Jupyter notebooks, without ever needing a terminal prompt.
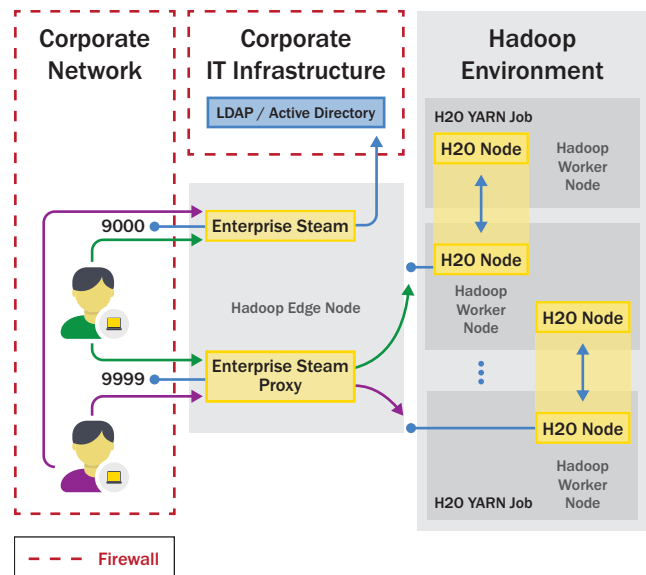

Figure 1: Enterprise Steam Deployment Diagram

- No SSH to Hadoop edge node required.
- Data Scientists can work directly from their laptops on the insecure side of a firewall.

Hadoop Admins, Enterprise Architects, IT Security Specialists

- Security
  - Enterprise Steam enforces full control of H2O YARN job security for administrators automatically without having to rely on the Data Scientist.
  - Enterprise Steam provides role-based Access Control for Admin Users and Standard Users.
  - Encrypted connections (SSL/TLS).
  - LDAP and Active Directory login authentication.
  - Kerberos authenticated YARN job submission.
  - Enterprise Steam offers the ability to put Hadoop clusters behind a firewall.
- Multi-Tenancy
  - Enterprise Steam is multi-tenant and prevents

Data Science users from accessing each other's H2O jobs (and data).

- Resource Control – Enterprise Steam allows Hadoop admins to:
  - Control which H2O versions are available.
  - Specify which YARN queue to use.
  - Cap the resources the Data Scientist can use.
  - Stop H2O jobs via a convenient Web UI.
- Resource Monitoring
  - Enterprise Steam provides Hadoop admins with monitoring capabilities to find dormant jobs tying up memory.
  - Enterprise Steam provides mechanisms for H2O usage measurement to enable chargeback and compliance use cases.

Key Benefits

- Next level operational efficiency in your AI environments with out-of-box **security, resource control,** and **resource monitoring**

- Data Scientists can freely and safely practice data science in their own H2O cluster since Enterprise Steam is **multi-tenant**

- Enterprise Steam enables new operation models such as internal chargeback, **internal cloud deployment,** and **H2O Platform as a Service (PaaS)**

License

- Annual Commercial License

System Requirements

- **OS:** Ubuntu 12.04 and up; RHEL 6.8 and up

- **Hadoop Distro:** CDH, HDP and MAPR distributions in last 2 years

- Databases: SQLite or Postgres SQL

About H2O.ai

H2O.ai is the open source leader in AI and automatic machine learning with a mission to democratize AI for everyone. H2O.ai is transforming the use of AI to empower every company to be an AI company in financial services, insurance, healthcare, telco, retail, pharmaceuticals and marketing. H2O.ai is driving an open AI movement with H2O, which is used by more than 18,000 companies and hundreds of thousands of data scientists. H2O Driverless AI, an award winning and industry leading automatic machine learning platform for the enterprise, is helping data scientists across the world in every industry be more productive and deploy models in a faster, easier and cheaper way. H2O.ai partners with leading technology companies such as NVIDIA, IBM, AWS, Intel, Microsoft Azure and Google Cloud Platform and is proud of its growing customer base which includes Capital One, Nationwide Insurance, Walgreens and MarketAxess. H2O.ai believes in AI4Good with support for wildlife conservation and AI for academics. Learn more at www.H2O.ai