Project Alexandria Architecture

Kerstin Kleese van Dam (BNL), John Darrington (INL), Mark Myshatyn (LANL), Tim Bender (LLNL), Dan McFarland (NETL), Sean Story (PNNL), & Ed Caroll (SNL)

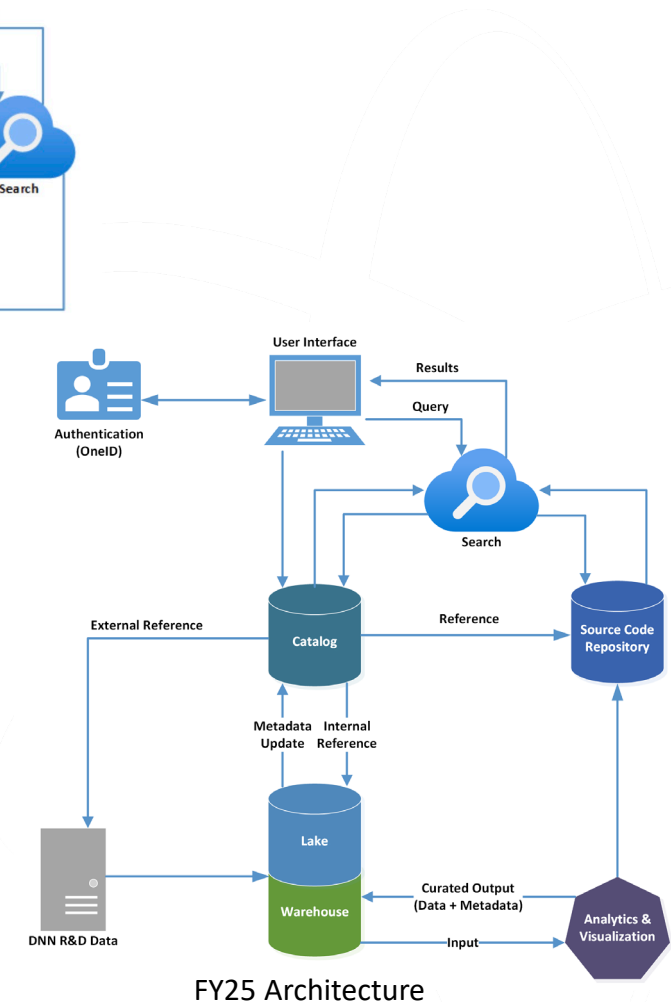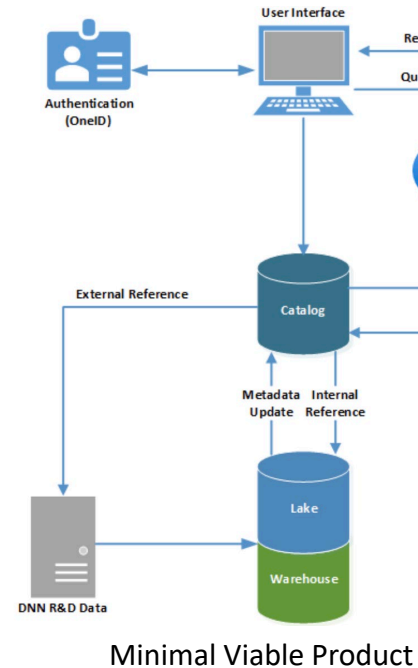Federal Program Manager: Paul Adamson

Technical Advisor: Lakshman Prasad

Principal Investigator: Christopher Ritter (INL)

PROJECT ALEXANDRIA

Brookhaven National Laboratory

INL Idaho National Laboratory

Lawrence Livermore National Laboratory

Los Alamos NATIONAL LABORATORY

OAK RIDGE National Laboratory

NETL NATIONAL ENERGY TECHNOLOGY LABORATORY

Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

Sandia National Laboratories

SRNL SAVANNAH RIVER NATIONAL LABORATORY

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

NNSA National Nuclear Security Administration

Defense Nuclear Nonproliferation Research & Development Program

Alexandria Architecture

- Authentication: Central Login
- Catalog: Index of Data
- Lake: Unstructured Data
- Warehouse: Structured Data
- Search: Centralized Queries
- Source Repo: Code Storage
- Analytics: Integrated Canned Analytics
- Compute: Custom Analytics (under study)

Minimal Viable Product

FY25 Architecture

Defense Nuclear Nonproliferation R&D

Common Services

INNOVATE. COLLABORATE. DELIVER.

Authentication

- OneID is multi-laboratory authentication solution already deployed by LLNL.

- Proven through Atlassian JIRA and Atlassian Confluence deployment for Alexandria development.

Cloud Environment

- Reuse of existing DNN Ion infrastructure and cloud services full authority-to-operate granted from 2020.

- Reuse of INL hub-spoke infrastructure for real-time cyber security controls.

Source Repository

- Reuse of existing multi-laboratory GitHub Enterprise.

Defense Nuclear Nonproliferation R&D

Search & Analytics

Search

- Search allows users to **quickly find** and **discover** information within both the data catalog and data warehouse.

- The search functionality within Alexandria will utilize a highly scalable, deep learning algorithm for fast user queries. (Not within MVP)

- This will leverage lessons learned and tools originally developed for the DOE-funded NETL EDX platform.
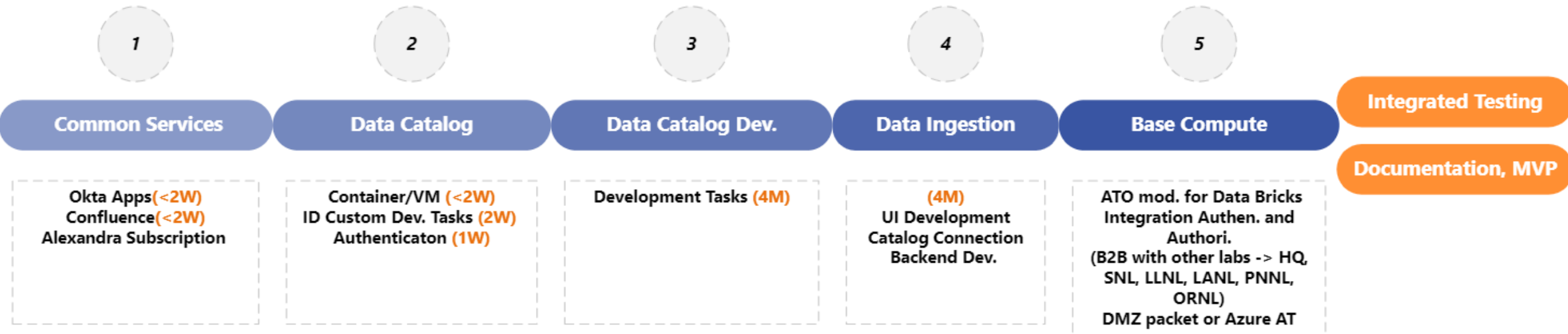
Analytics:

- A set of **optional** integrated analytics for basic data science activities that use the **optional** centralized lake and warehouse storage.

- Provide quick answers to typical questions a venture or project team may seek to answer.

- Can provide **immediate** scientific benefits to projects which integrate with Alexandria's architecture.

Defense Nuclear Nonproliferation R&D

Data Lake

- Computing infrastructure to store **non-proliferation data files**. Can use a combination of the following infrastructures
  - Cloud
  - On-premises (edge)
  - High-performance infrastructure
- **Persist** data ingested into Alexandria in its raw native format
- Data providers have the option to **choose** the model that fits their project best.
- Will accommodate both large numbers of **small** files and **large** files through a **federated approach** using a combination of:
  - DOE provided data center storage
  - Cloud storage
  - Testbed localized storage approaches.

Defense Nuclear Nonproliferation R&D

Data Warehouse

- Store data from Alexandria ventures that need a **standardized ontological** format to enable:
  - Digital twins
  - Digital threads
  - Fast ad hoc queries across the non-proliferation community.

- This optional service can be used for **novel data integration** for projects which need new data insights.

- Accelerates deployment **of scientific machine learning models** to mission, the team will leverage recent work in **adapting verification** and **validation** methodologies to quantity **model credibility** and **end-user trust**.

- NNSA has invested in world-, one-of-a-kind test beds and facilities to support non-proliferation research.

Defense Nuclear Nonproliferation R&D

Architecture and Development FY24 Schedule

PROJECT ALEXANDRIA

1

Common Services

Okta Apps(<2W)
Confluence(<2W)
Alexandra Subscription

2

Data Catalog

Container/VM (<2W)
ID Custom Dev. Tasks (2W)
Authenticaton (1W)

3

Data Catalog Dev.

Development Tasks (4M)

4

Data Ingestion

(4M)
UI Development
Catalog Connection
Backend Dev.

5

Base Compute

ATO mod. for Data Bricks
Integration Authen. and Authori.
(B2B with other labs -> HQ, SNL, LLNL, LANL, PNNL, ORNL)
DMZ packet or Azure AT

Integrated Testing

Documentation, MVP

Defense Nuclear Nonproliferation R&D

Work Breakdown Structure

Defense Nuclear Nonproliferation R&D