DATA IN SCIENCE TECHNOLOGIES (DST)

Making sense of the world's data for research and modeling communities.

CATALOGING & ANALYTICS

DATALOGGER

EXECUTIVE SUMMARY

Data in Science Technologies (DST) makes sense of the world's data for analytics and modeling in government, healthcare and research organizations. We provide services and design solutions that empower informed decision-making, delivering a positive impact to your mission's success. Through our **Cataloging and Analytics**, we help solve common data management struggles in High Performance Computing (HPC) and Big Data. With chain-of custody and search across multiple data sources, our solutions lead to better analytics, efficiency and protection against data loss or misplacement.

SITUATION OVERVIEW

Organizations often view and search data using technology embedded in the ecosystem that created the data: internal file systems, databases, cloud services, e-mail folders, etc. Each ecosystem stores data as a separate media type with a variety of access methods. Data silos are created in disparate formats, platforms and applications. Global searching and correlation between these silos is often difficult, as current applications and toolsets are inadequately designed to authenticate, query and catalog results across different data silos.

Today, government, healthcare and research organizations have requirements for data auditing, reporting and access control. As well as, requiring Google-like search across all data and chain-of-custody validation. Additionally, recent federal mandates require research data to be made available to the public, where possible, to foster collaboration and economic growth with commercial entities. Compliance with these mandates in the absence of any data management tools will be a challenge.

DATALOGGER OVERVIEW

To solve these data management issues, DST developed DataLogger, a software solution that enables organizations to search, catalog their data and establish chain-of-custody in all environments. DataLogger establishes a singular view of the meaningful attributes of your data and identifies access rights to this data. Our complete approach establishes a common metadata repository, which benefits research and analytics in HPC environments.
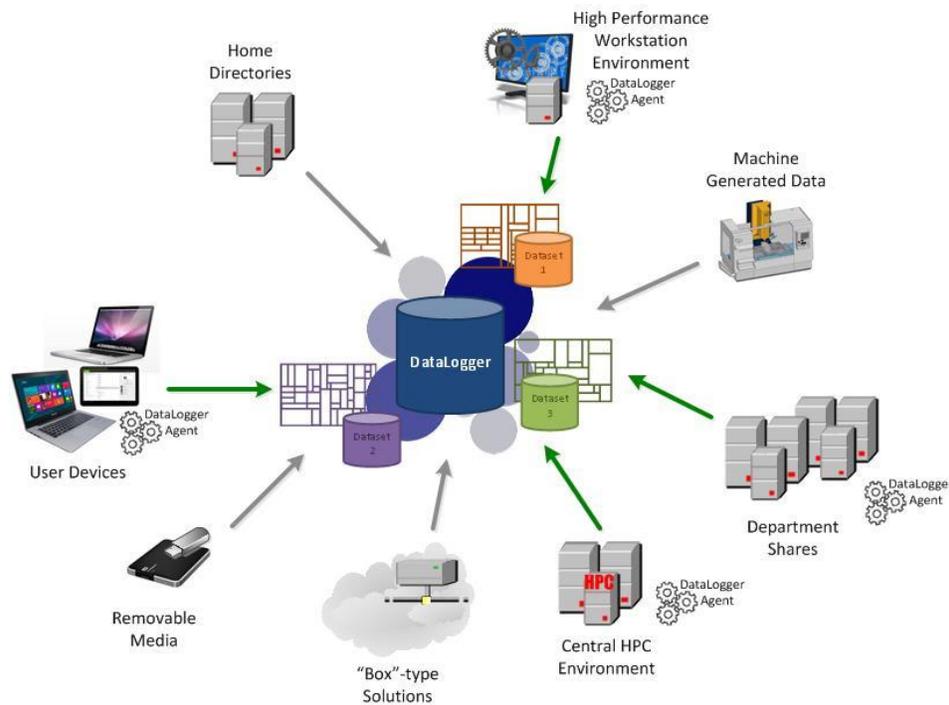
FEATURES

DataLogger augments your existing HPC and analytics systems to provide:

- ✓ Global, fast search and retrieval of all data
- ✓ Common dashboard for all data environments
- ✓ Advanced query capability
- ✓ Browser-based access to the data catalog with:
  - ∗ Predefined rules that govern how data sets are catalogued
  - ∗ Business logic that sets up default working parameters
  - ∗ An audit trail on all changes made to cataloged data
  - ∗ Access control is inherited by the data catalog and provides user authorization
  - ∗ A POSIX compliant file system so data can be managed in the same manner as done previously

✓ Capture of pre-defined metadata associated with the corresponding file system:

* Author, title and description

* Date created and/or modified

* File type, size and location

✓ A rules engine manager systematically reviews for compliance, security breaches or infrastructure expansion

DataLogger gives you complete control of your data management by providing:

✓ Insight into who has accessed data and how often

✓ Data preservation as your employee/s exit

✓ Research work quantifiable by individual or group

✓ Chain-of-custody validation

✓ Data auditing, reporting and alarming

✓ ISO Standard enforcement and documentation

Data visualization reports are also included with DataLogger. Reports include:

✓ Where data is located

✓ Who is using the data

✓ File type

✓ Credentials of researcher

DATALOGGER ARCHITECTURE

DataLogger is built on a scale-out, indexable architecture, similar to a relational database, yet with less rigidity in structure. The system allows applications to connect, store and extract data with a pre-defined set of libraries that support a wide variety of programming languages such as Java, PHP, Python and dotNet. A simple API performs the store and extract functions for the data and its associated metadata.