EVA Data Integration

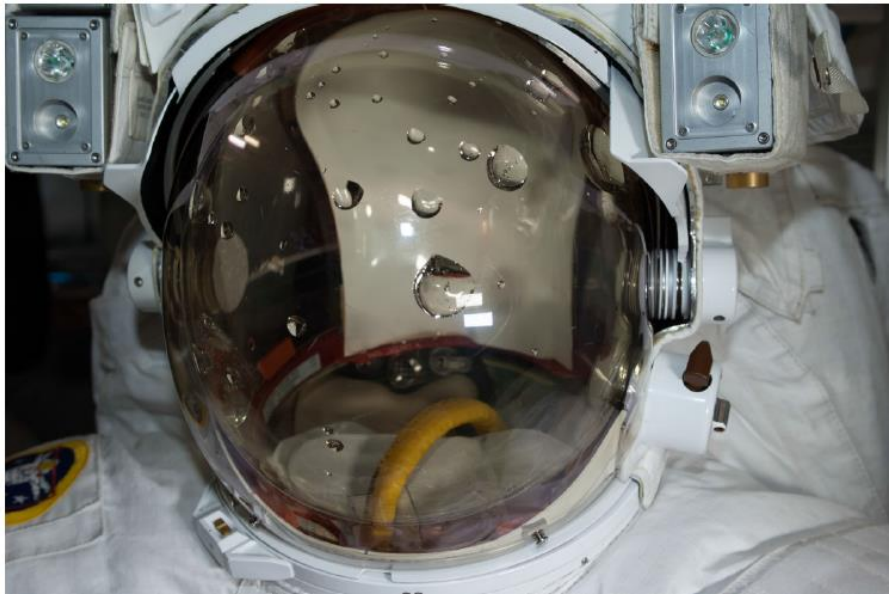
Technology Collaboration Center
Data Analytics Workshop
Rice University

Cuong Nguyen

10/03/2017

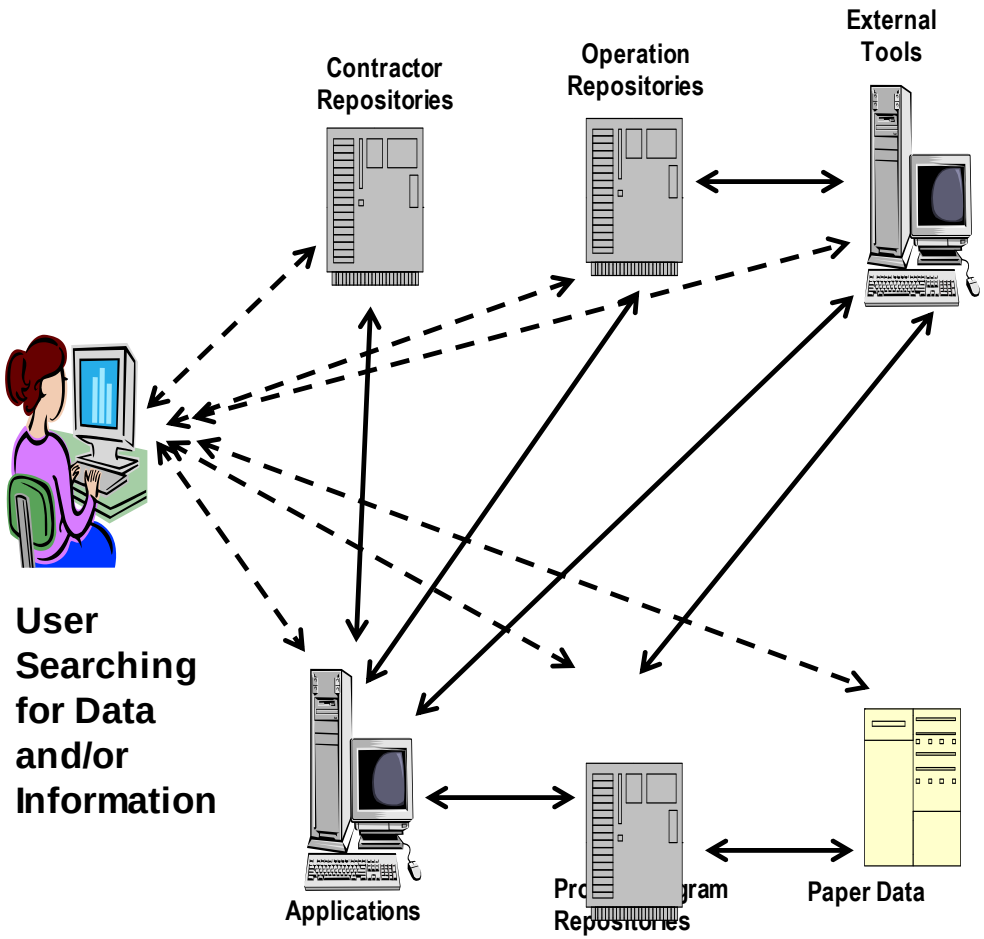
Executive Summary

- **Background**
  - On July 16, 2013, two US crew members exited the International Space Station (ISS) US Airlock to begin U.S. Extravehicular Activity (EVA) 23. Roughly 44 minutes into EVA 23, EV2 reported water inside his helmet on the back of his head.
  - Mishap Investigation Board Recommendation to combine all EVA knowledge databases into a set of databases that are easily accessible to the entire EVA Community



Challenges

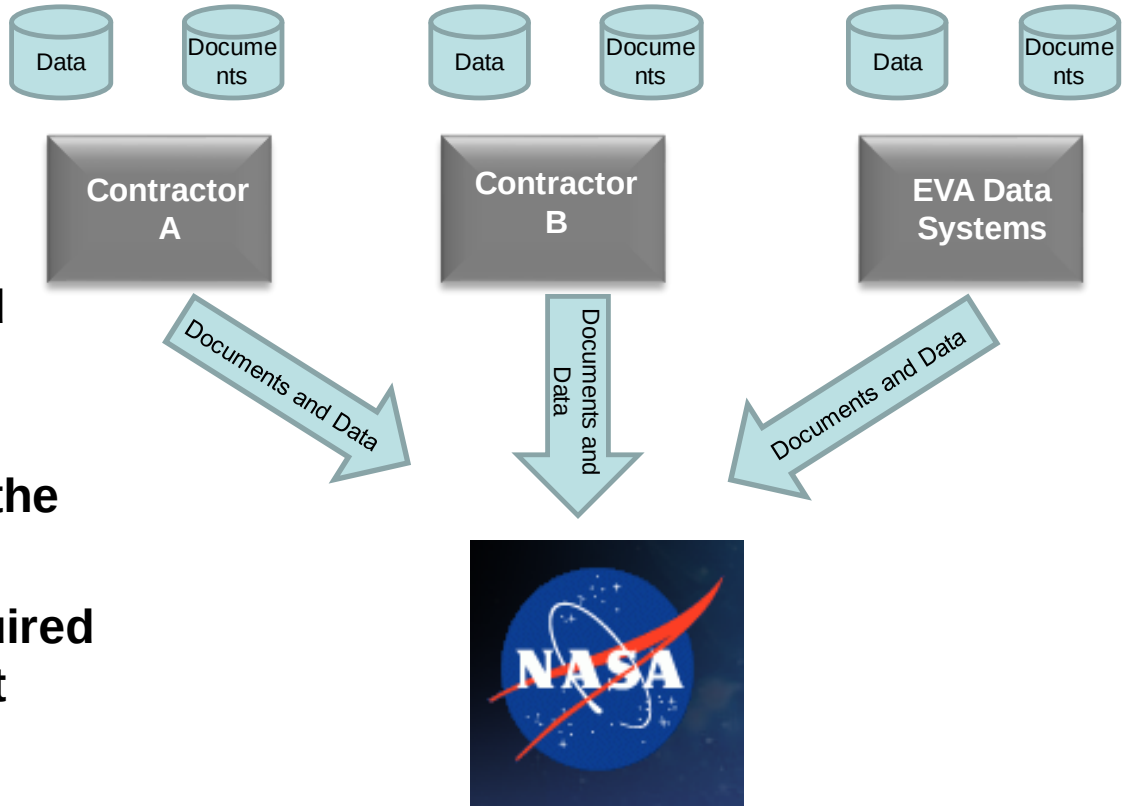
- Data are scattered and disconnected
- Users don't have access to the data based on uniform security
- Data and documents are copied and duplicated
- No standards for sharing and exchanging data
- Lack of Data Interoperability
- Limited time for data analysis; time wasted on data gathering
- Project-exclusive approach results in disparate data definitions
- Incomplete and Inconsistent Information; various formats



Traditionally, focus has been on solutions based on Projects and Applications resulting in Data Silos

Problems

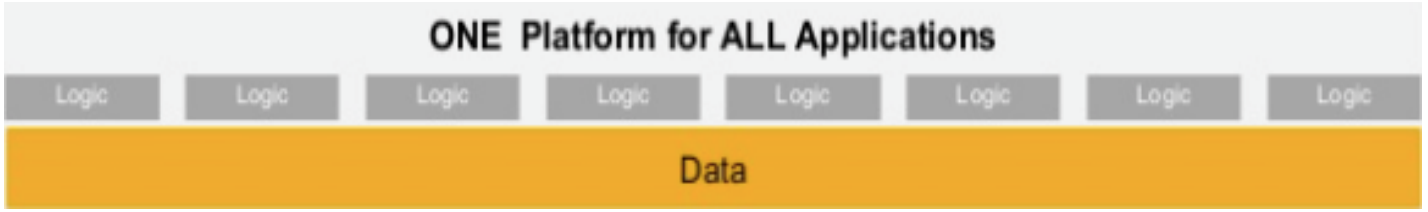
- Data comes in multiple sources and in multiple formats
- Data is dispersed into multiples non-integrated databases
- No process in place for vesting and integrating the data for end users
- Nearly 8 TB of data acquired from a contract closeout
- Data are being captured every day into EVA data systems
- Need to bring in data for new suit development



Proposed Solution (Conceptual)

Solution: Make All Data Readily Available to All Applications

Reduce Data Movement, Latency, Errors, and Manual Work



No waiting for Data Access and Processing → All Applications Processed in one System → All Data Types Processed in one System → **Speed Simplicity Cost Effective**

Advantages to availability of data to users: Mission/Crew Safety

- Turn Data into real-time Information
- No Delays in searching and accessing Data/Information

Enterprise Data/Information Framework

- **Establish a Framework to support changing Data and IT Landscape**
- **NASA must own all its DATA**

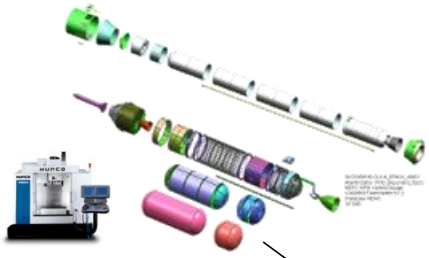
Key Considerations

- ❑ Data Architecture & Management
- ❑ Data Integration
- ❑ Business Intelligence and Data Analytics
- ❑ Agile Methodology
- ❑ Data Governance
- ❑ IT Working Group
- ❑ Data Competency Center

How can we Leverage ALL our Data?

Support Mission Lifecycle

Each domain of practice uses different data formats, conventions, representations, and tools making Interoperability and reuse challenging

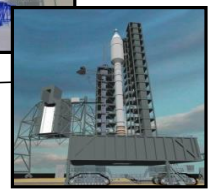


Manufacture

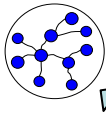
Test & Eval

Operate

Maintain



Upgrade



Lessons Learned

How do both computers and humans do this?

Learn

How do the data and IT help us really learn?

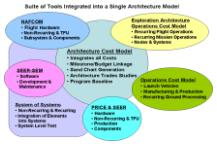
System Lifecycle

Manufacture

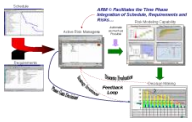
Upgrade

Design

Perf



Cost



Risk

Simulation

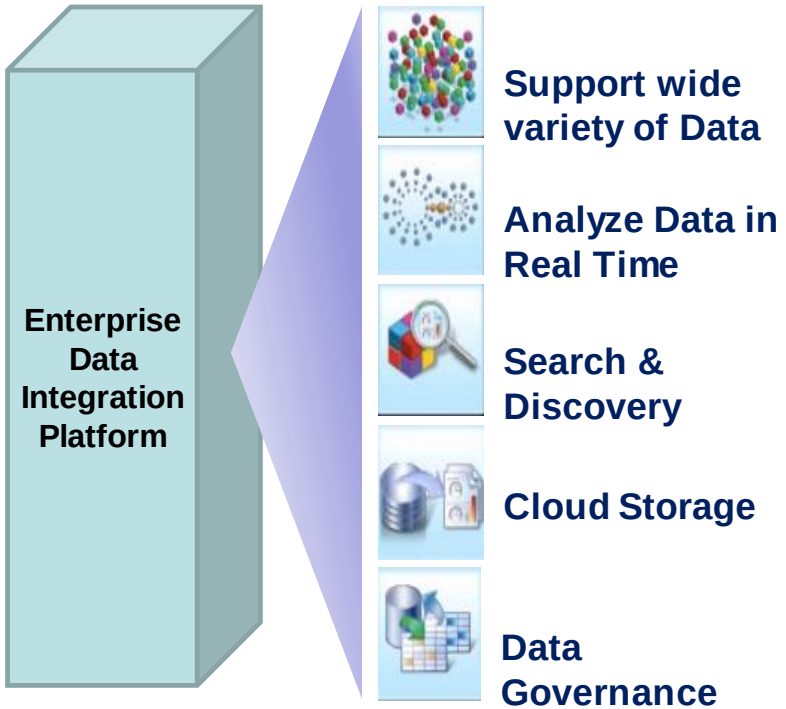
Adding to the issue we note that Information evolves as it is used by each domain

System Concept

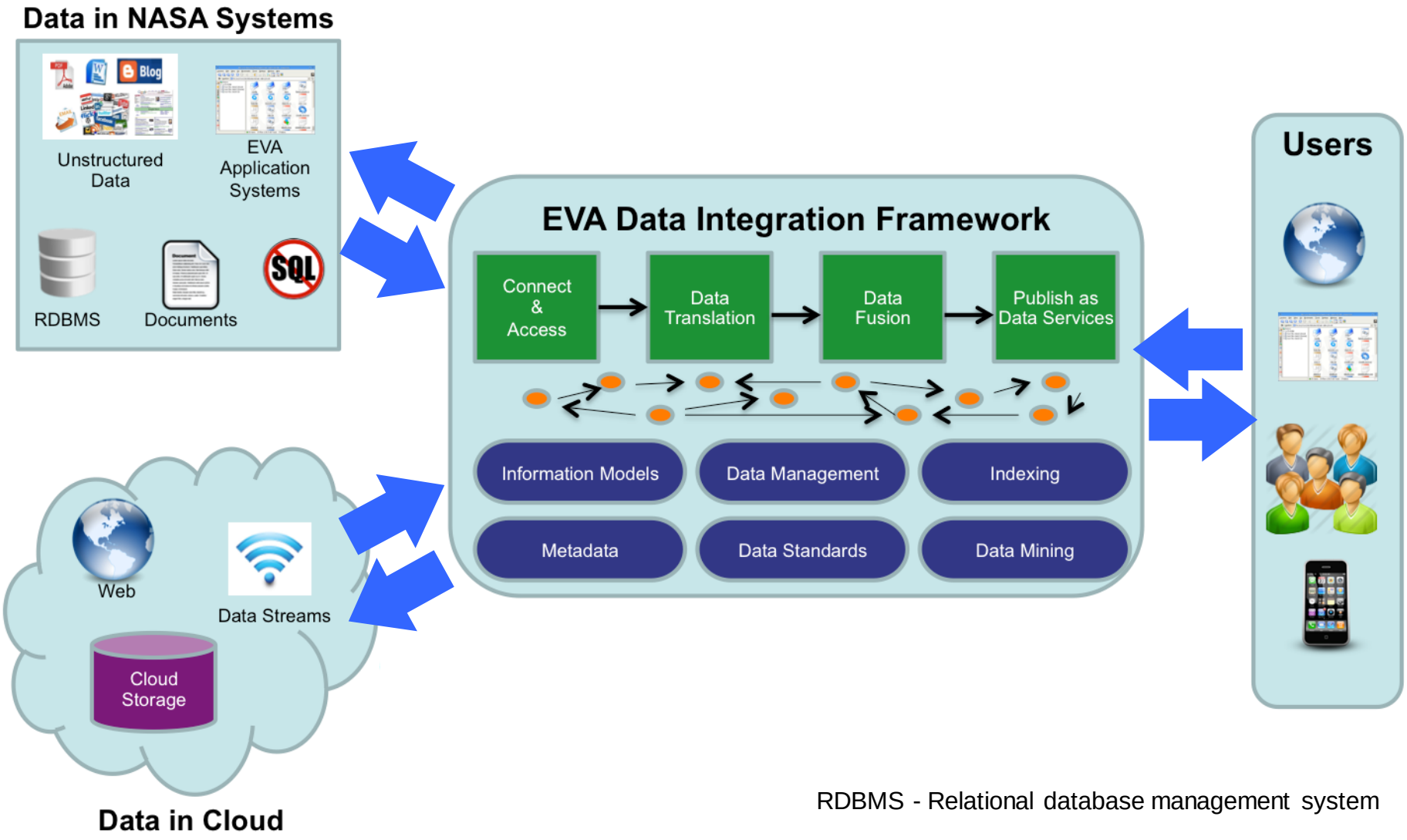
The overall system concept is to provide a set of core shared services for EVA Data Integration (EDI), with some of the core services having end user application interfaces including an integrated search application and a document management application (e.g., for current suit data).

Goals include:

- *Enable easy secure access and integration of EVA data & applications for authorized users in the EVA community.*
- *Assurance that EVA data is complete, accurate and up to date.*
- *Enable rapid low-cost development and operation of EVA applications through shared services*.

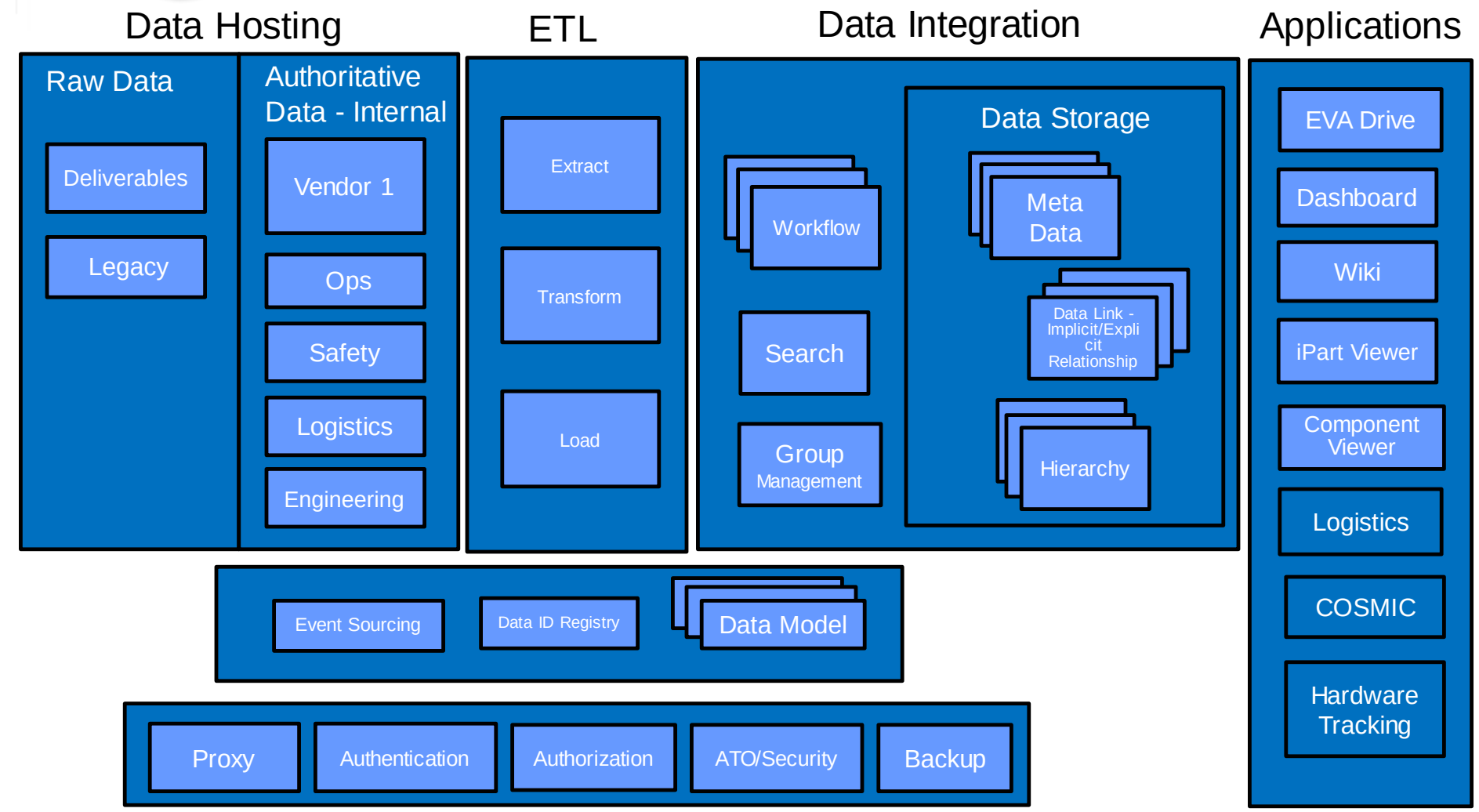


Proposed High Level Data Integration Architecture

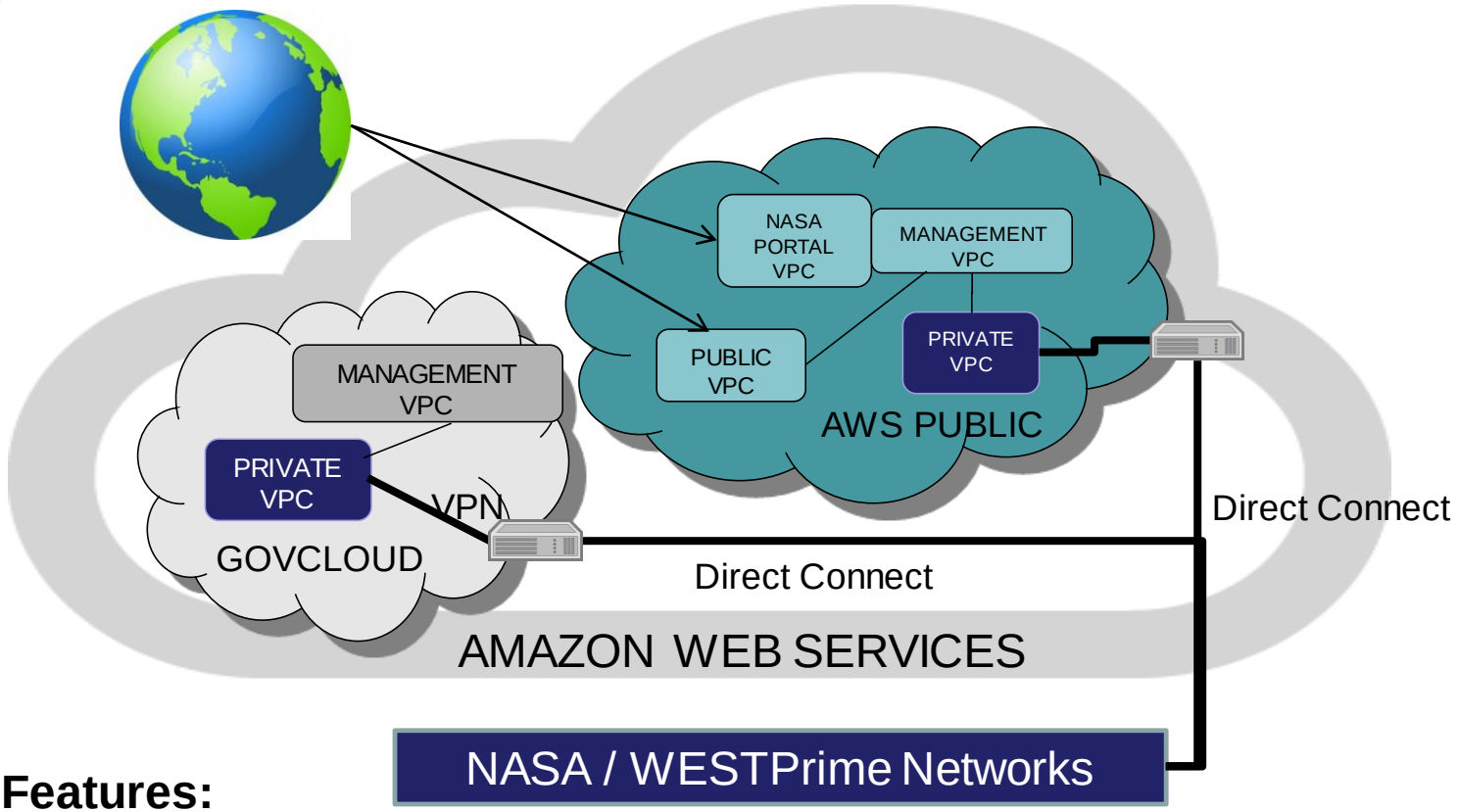


RDBMS - Relational database management system

What is EVA Data Integration – Logical Architecture



NASA Cloud Architecture



Required Features:

- **Complete Secured Solution**
- **Optimized for Purpose**
- **Extensible**
- **Faster Deployments**
- **Easy Operations Support**
- **Low Cost**

VPC – Virtual Private Cloud
AWS – Amazon Web Services

Data Integration Requirements

- **Single login access across numerous data sources and types**
  - EVA Data Portal
- **Simplified unified access management of internal datasets**
  - Data protection/security
- **Cross platform compatibility (Mobile devices, desktop)**
- **Uber Search Capabilities**
  - Google like keyword search
  - Graphical navigation search
  - Follow-the-link capability
  - Intelligent linking (text-to-text, text-to-graphic-hotspot, e-mail-to-mediafile, person-to-part, etc.)
  - Generalized data aggregation and extraction
- **Confidence in Data integrity regardless of where data is located**
- **Open standards deployed**
- **Flexibility in architecture to allow system to evolve**

EVA Data Integration: Benefits

- Eases Integration of Systems and Applications across the lifecycle
- Improves discovery of relevant information
- Tool / Application Independence; avoiding Vendor lock-in of systems with proprietary schemas and formats through neutral models
- Lower the barriers for collaboration and facilitates Communities of Practice through actionable, model-based knowledge capture and reuse
- Helps make working knowledge (tacit) explicit
- Provides a query-able resource of who produces and uses what, when and how
- Serves as a backplane for Information Sharing
- Provides a foundation for Linking Data elements – navigation, hierarchies, etc.
- Increases confidence in data interoperability through consistency of data types, structures and taxonomy

Lessons Learned – EVA Data Integration

- Data Integration is hard
- Data stove-pipes are major hurdle to overcome – data sharing policy needed
- Data sets quickly become very large when including sub-assembly components and all “build” and “change” artifacts.
- Required deliverables only accounted for a small part of data
- Differences in procedure/process resulted in many formats for the same deliverable.
- Majority of the data are in boxes of paper or scanned PDF – work needed to make them searchable.
- When processing data deliverables from a legacy source, the data received may be unorganized, in paper form, and/or delivered without context .
- Company special processes or sensitive, proprietary, or ITAR data complicates the solution.
- New technology has made the data integration task achievable as long as the scope of old data is kept at a manageable level.
- Major cost savings over time due to easy access to data.
- Data Integration has the possibility to the safety margins of a system as failures can be predicted before they happen due to trending analysis.