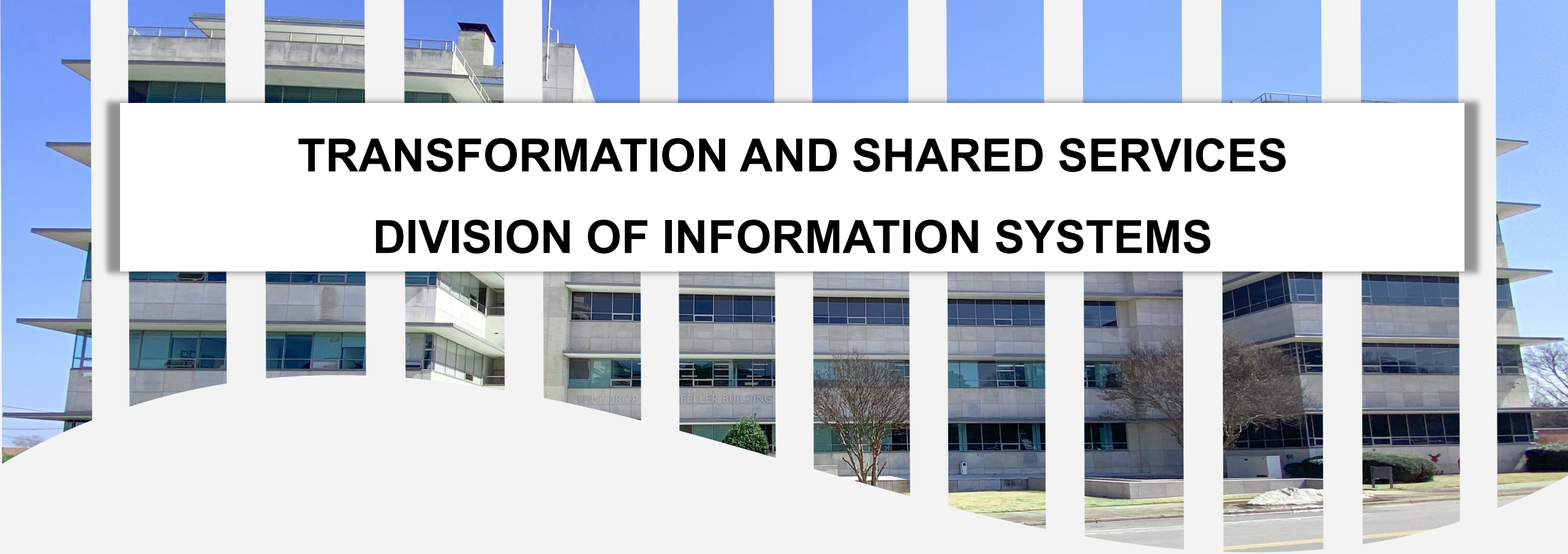


TRANSFORMATION AND SHARED SERVICES

DIVISION OF INFORMATION SYSTEMS

Arkansas Data Hub

Arkansas Data and Transparency Panel
Thursday, September 19, 2024

Agenda

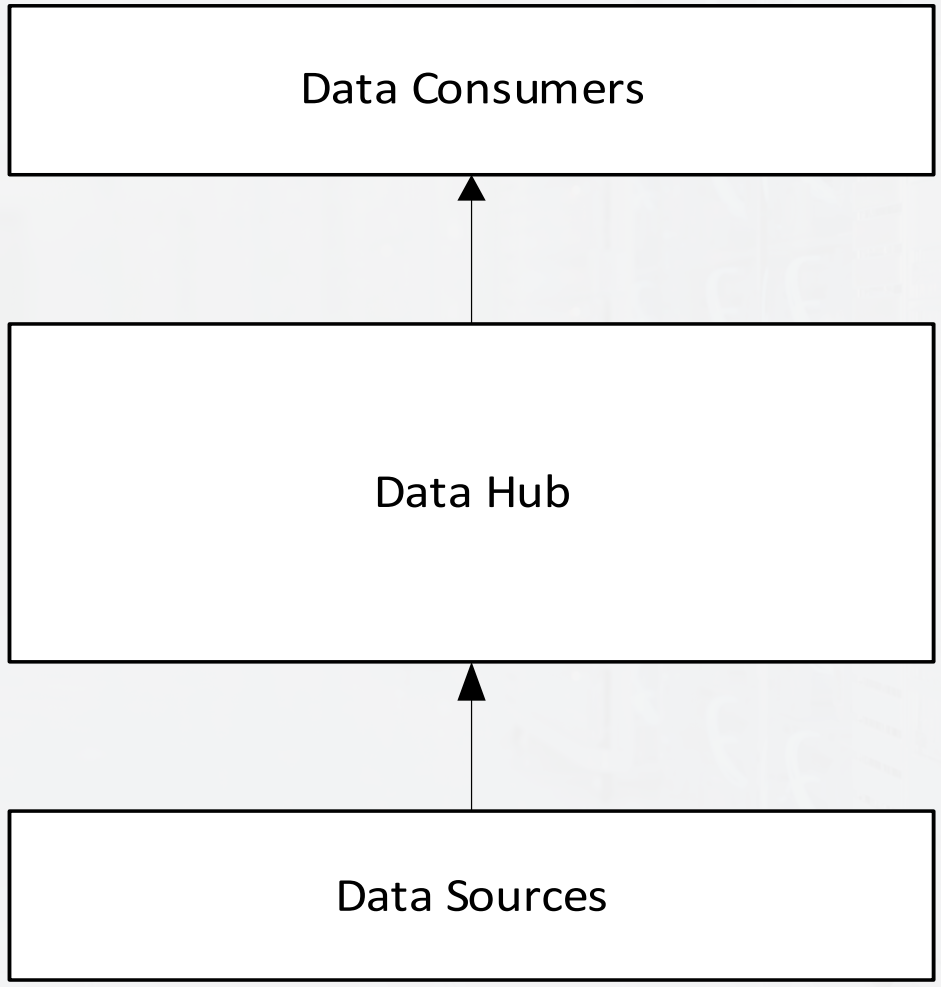
1. Data Hub Overview
2. Cross-Agency Data Sharing
3. Data Virtualization
4. Master Data Management
5. Data Catalog
6. Unlocking the Toolbox

Defining the Arkansas Data Hub

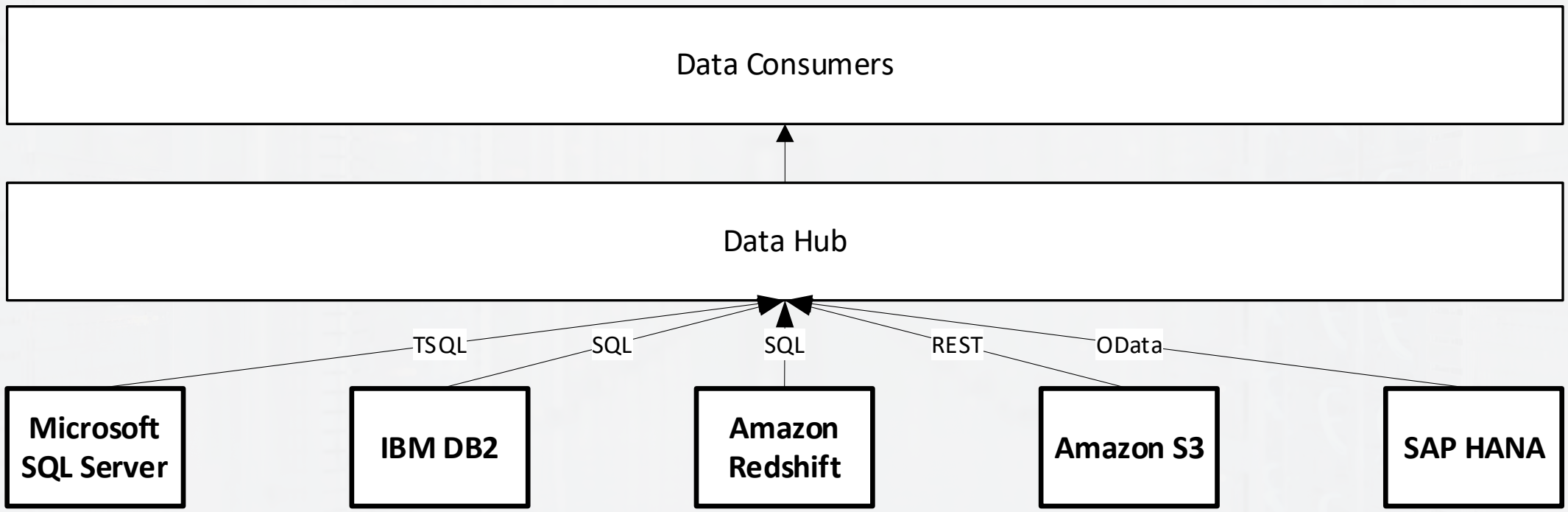
A statewide shared service data infrastructure required by Act 634 of 2023 for facilitating the sharing, integration, accessibility, quality, standardization, and governance of data.

High Level Architecture

The Arkansas Data Hub facilitates secure, governed, sharing, and integration of data across multiple sources for use by a wide variety of operational and analytical data consumers and use cases.

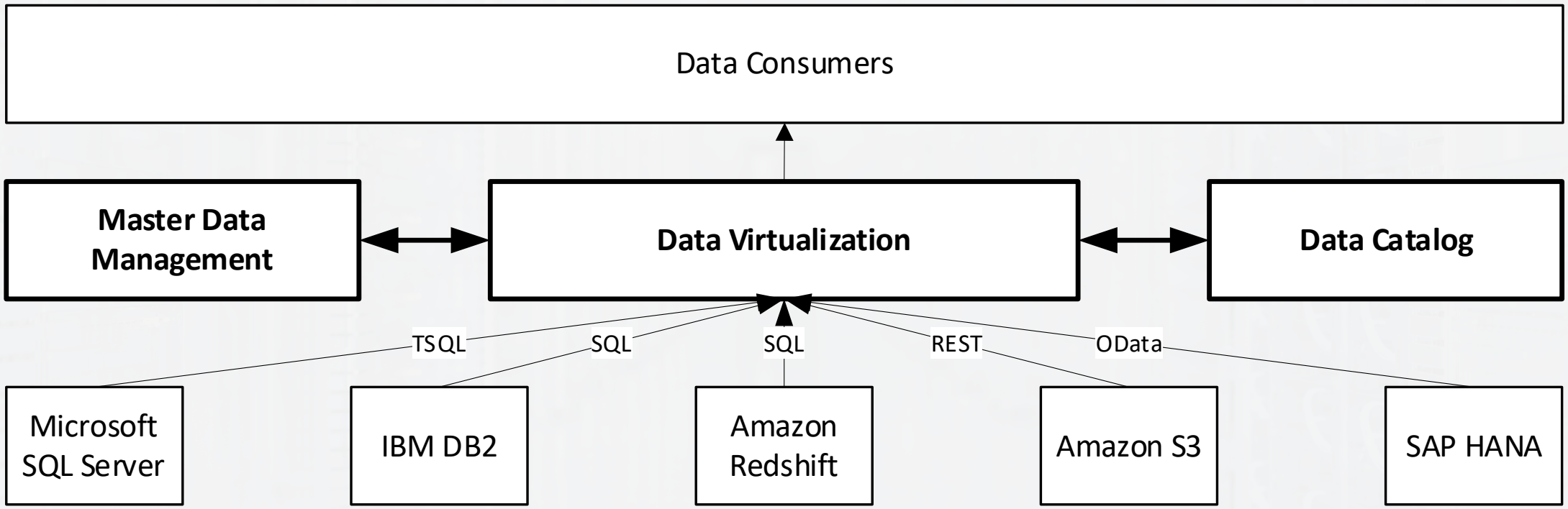


Data Sources



State data are distributed across a wide variety of sources with different stewards, locations, formats, and representations.

Key Data Hub Components



Data Virtualization, Master Data Management, and a Data Catalog work together as key components of a “data fabric” to provide a comprehensive, real-time view across disparate data sources.

Defining a Data Fabric

- Connects all the state's business data, no matter where it is stored, making it easily accessible, secure, and ready for real-time use.
- Simplifies data management, enabling faster decisions, better insights, and smoother operations across departments.
- Secures data by controlling who can access it, encrypting it and ensuring compliance.

Arkansas Data Hub - Security

Advantages:

- On-premise environment hosted by the state.
- Authentication uses our internal Active Directory.
- By default, inbound and outbound connections are denied unless specified to allow.
- Data is secured by using industry-supported encryption.
- Security best practices include:
  - Ongoing vulnerability management
  - Penetration testing
  - Privacy assessments
  - Security testing
  - Patch management
- Multiple layers of physical, network, and application security.

Arkansas Data Hub – Data Sharing

- The Arkansas Data Hub allows fast and secure data sharing through many integration points and views.
  - The data remains in the source system but allows a snapshot view of data that the agency data steward controls.
  - 90+ source connectors built in to allow for flexibility.
- Features allow the user to see what data is available in a protected view for the user then to contact the agency-specific data steward to arrange data sharing.

Data Hub Example Use Case

Corrections and Arkansas Court Connect

- Incarcerated individual has upcoming court date.
- Court has no way of knowing the individual is incarcerated.
- Individual is issued a warrant for missing their court date, so they recidivate upon release.

Data Hub Example Use Case

Offender Intake Process:

Gathering High School Diploma or GED

- Corrections must verify a high school diploma or equivalency (GED) upon intake.
- Corrections staff calls the school district to search for and confirm high school diploma or equivalency attainment, which involves staff time on both Corrections and the school district.

Arkansas Data Hub – Data Virtualization

Real-time data integration tool that allows for a unified view of data from multiple sources.

Key benefits:

- Link heterogeneous data sources into a common language.
  - Cloud, On-Prem, Database, API
- Fast and easy data sharing reduces workload on staff. (Improves timeliness)
- Data protection and governance
  - Role-based access controls
  - Data privacy rules (masking, redaction, obfuscation, etc.)
  - Federal Information Processing Standards (FIPS) compliant

Data Virtualization Demonstration – Dummy Data

Microsoft SQL Server

Table: EMPLOYEE | Schema: WHATEVER1

Find columns

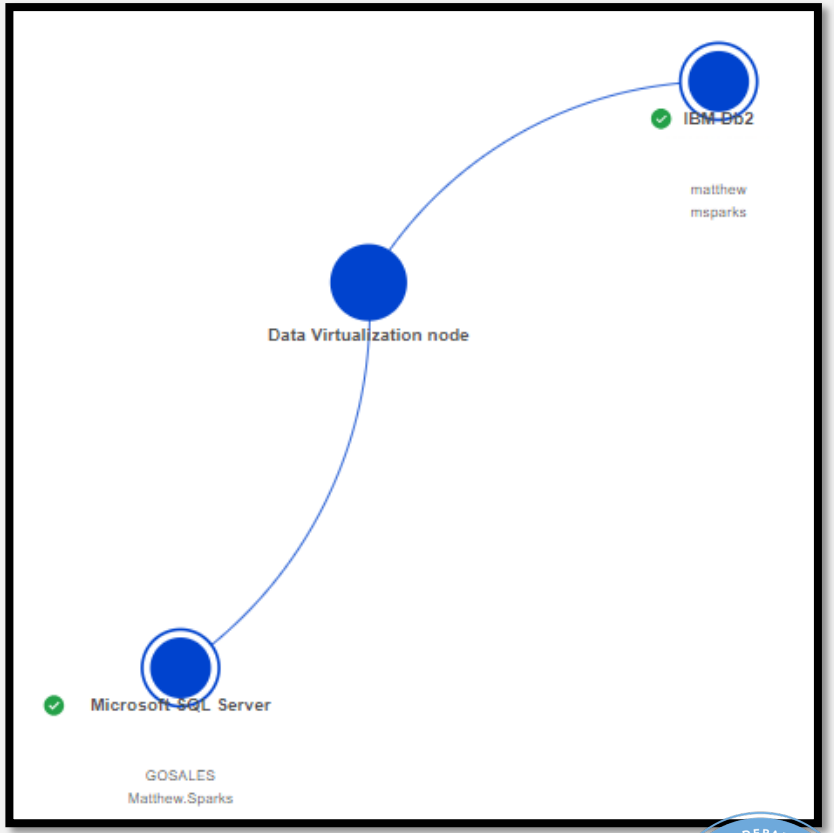
Column name	Data type
EMPLOYEE_CODE	INTEGER
FIRST_NAME	VARCHAR
FIRST_NAME_MB	VARCHAR
LAST_NAME	VARCHAR
LAST_NAME_MB	VARCHAR
DATE_HIRED	TIMESTMP
TERMINATION_DATE	TIMESTMP
TERMINATION_CODE	INTEGER
BIRTH_DATE	TIMESTMP
GENDER_CODE	SMALLINT
WORK_PHONE	VARCHAR
EXTENSION	VARCHAR
FAX	VARCHAR
EMAIL	VARCHAR

IBM DB2

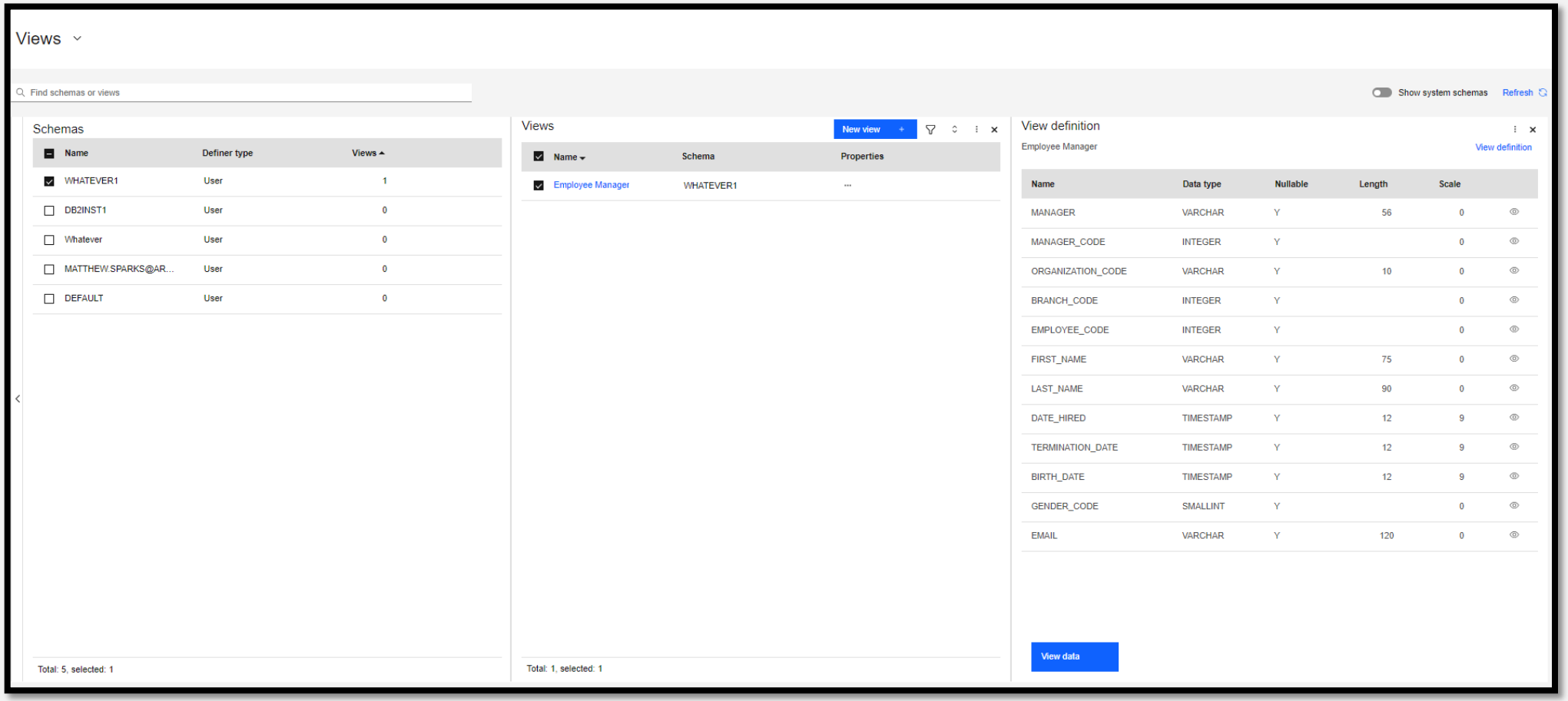
Table: EMPLOYEE_HISTORY | Schema: WHATEVER1

Find columns

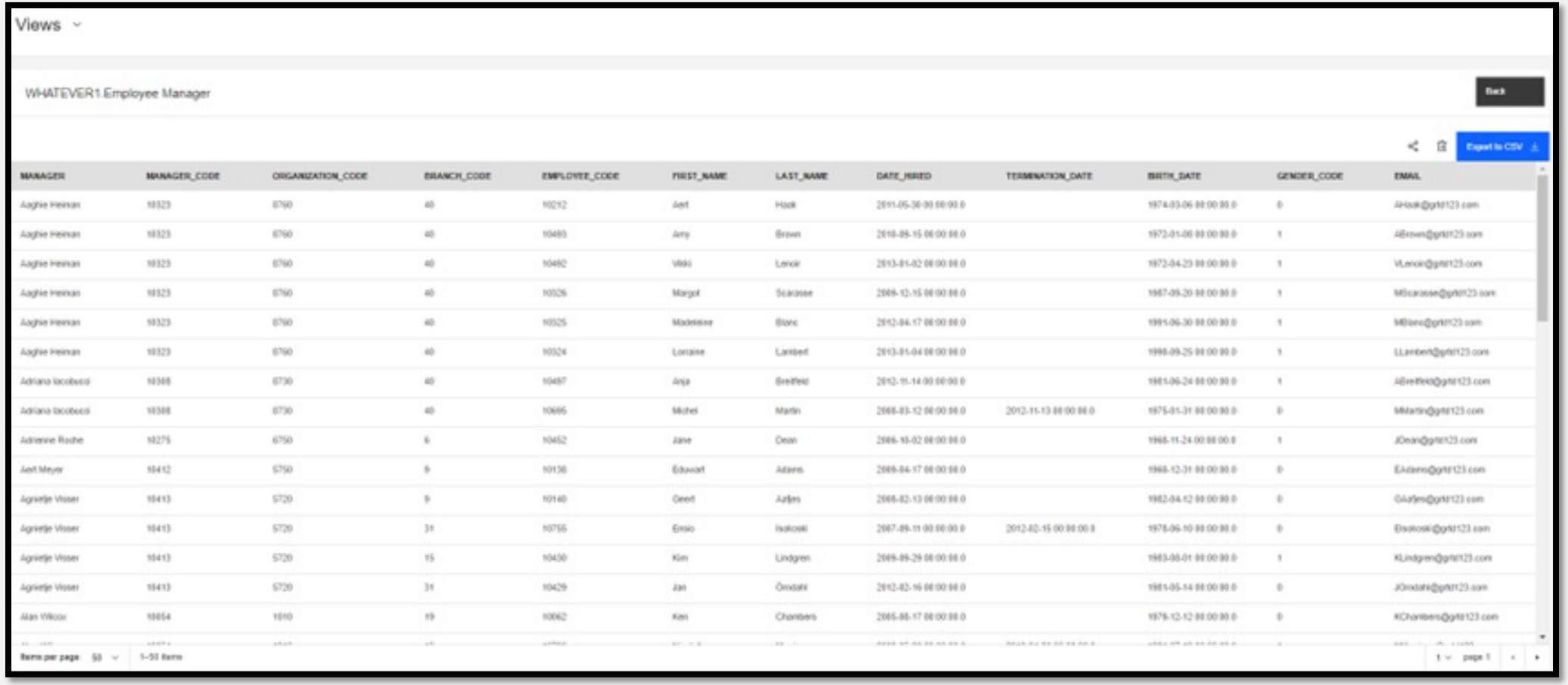
Column name	Data type
EMPLOYEE_HISTORY_CODE	INTEGER
EMPLOYEE_HISTORY_PARENT	INTEGER
EMPLOYEE_CODE	INTEGER
RECORD_START_DATE	TIMESTMP
RECORD_END_DATE	TIMESTMP
POSITION_CODE	INTEGER
POSITION_START_DATE	TIMESTMP
MANAGER_CODE	INTEGER
MANAGER	VARCHAR
MANAGER_MB	VARCHAR
BRANCH_CODE	INTEGER
ORGANIZATION_CODE	VARCHAR



Data Virtualization Demonstration – Dummy Data



Data Virtualization Demonstration – Dummy Data



Arkansas Data Hub – Master Data Management (MDM)

Master Data Management (MDM) resolves differences in semantics across different references across one or more sources to the same real-world entity. (person, household, organization, etc.)

A statewide MDM service facilitates:

- Entity resolution
- Clerical review
- Data consolidation (golden record survivorship)
- Consent management

Arkansas Data Hub – Data Catalog

Defining a Data Catalog

- Cross-agency **data discovery and data lineage**
  - Self-service access
  - Ability to see where data has come from and what has been done to the asset along the way.
  - Supports data change management
- Provides **data quality** insights to enhance data
  - Identifies data issues at the source
  - Data profiling gives a quick analysis of completeness, representational consistency and other relevant profiling metrics
- Protected user view through **data privacy** and **protection** practices

Arkansas Data Hub – Data Catalog

Data Discovery

- Empowers users to find domain-specific data for various agency projects and/or use cases.
- All users can be granted access to view metadata.

Data Lineage

- Complete history of your data, from its source, through each transformation, to its final endpoint.
- Identify causes of data errors.
- Track changes that have happened to the data asset.
- Assess the impact of changes by identifying where the asset is used.

Arkansas Data Hub – Data Catalog

Data Quality

How does a data catalog help with data quality?

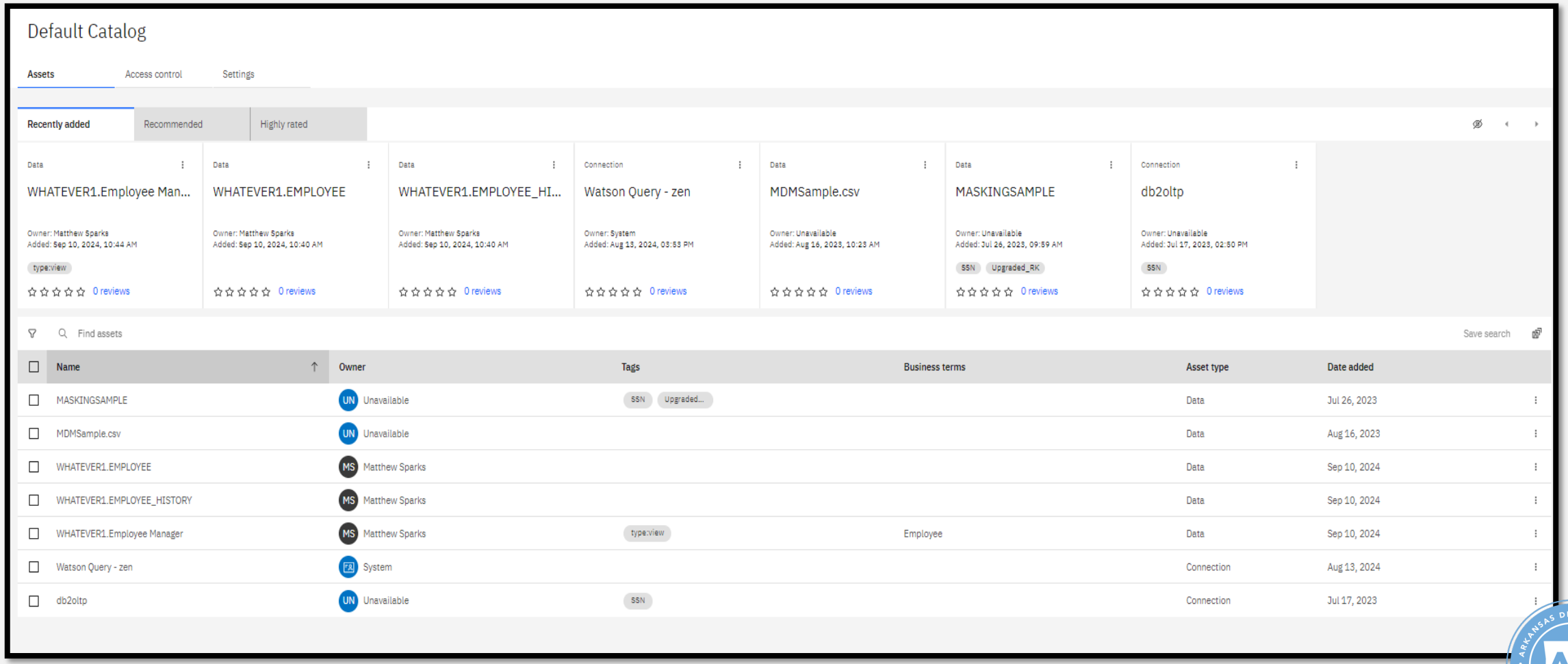
- Enhances data quality by providing structure, visibility, and governance over data assets
- Ensures that the information is trustworthy and suitable for:
  - Analysis
  - Decision-making
  - Reporting
  - Other data-driven activities

Arkansas Data Hub – Data Catalog

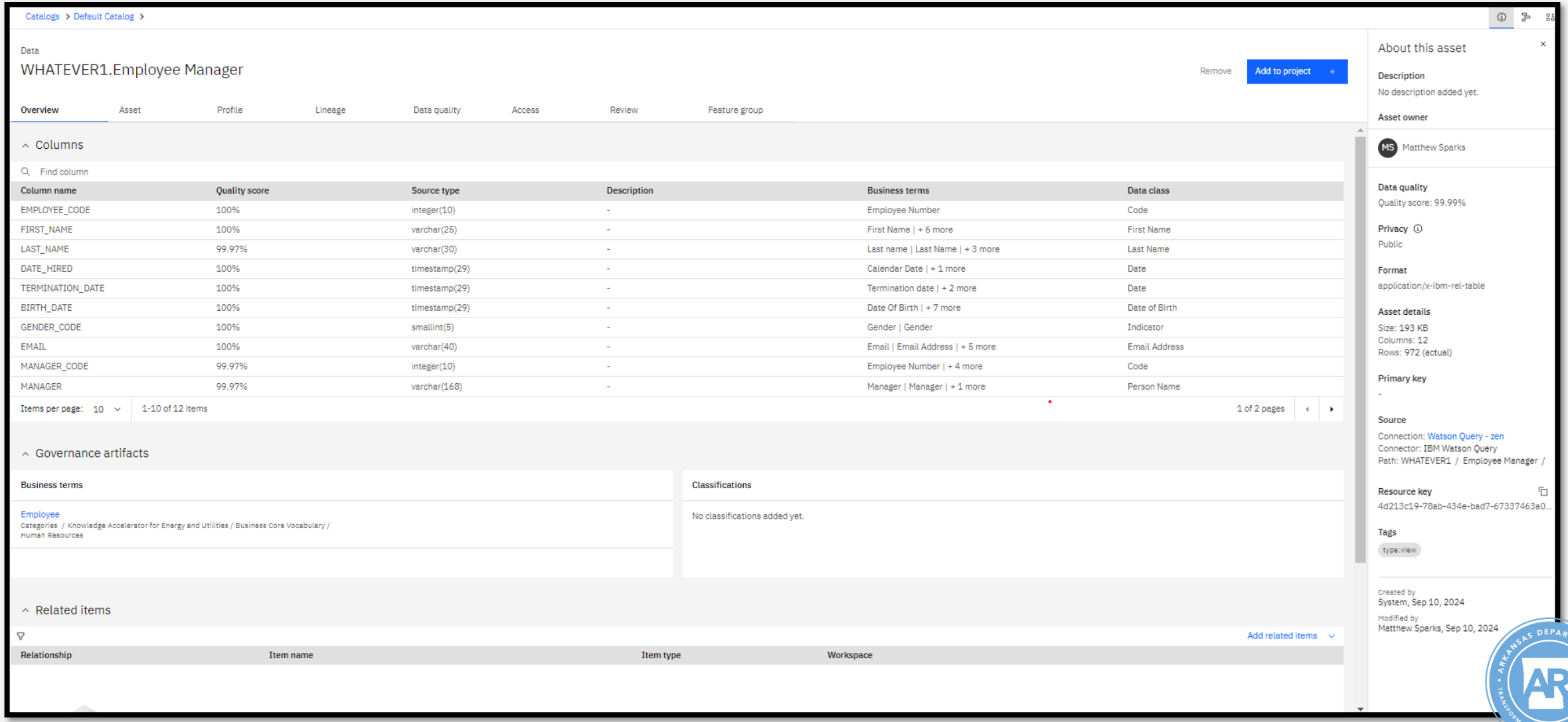
Data Protection

- Data governance policies will define how data should be accessed, shared, and protected (under the direction of the State Cybersecurity Office).
- Data privacy
  - Data can be masked, obfuscated, filtered, redacted or denied access to.
  - The platform provides encryption at rest and in transit.
  - Helps adhere to privacy regulations.
- Role-based access control
  - Roles include Admin, Viewer, and Editor.
  - Data protection rules apply to all users. (Exception: catalog and asset owners)

Data Catalog Demonstration – Dummy Data

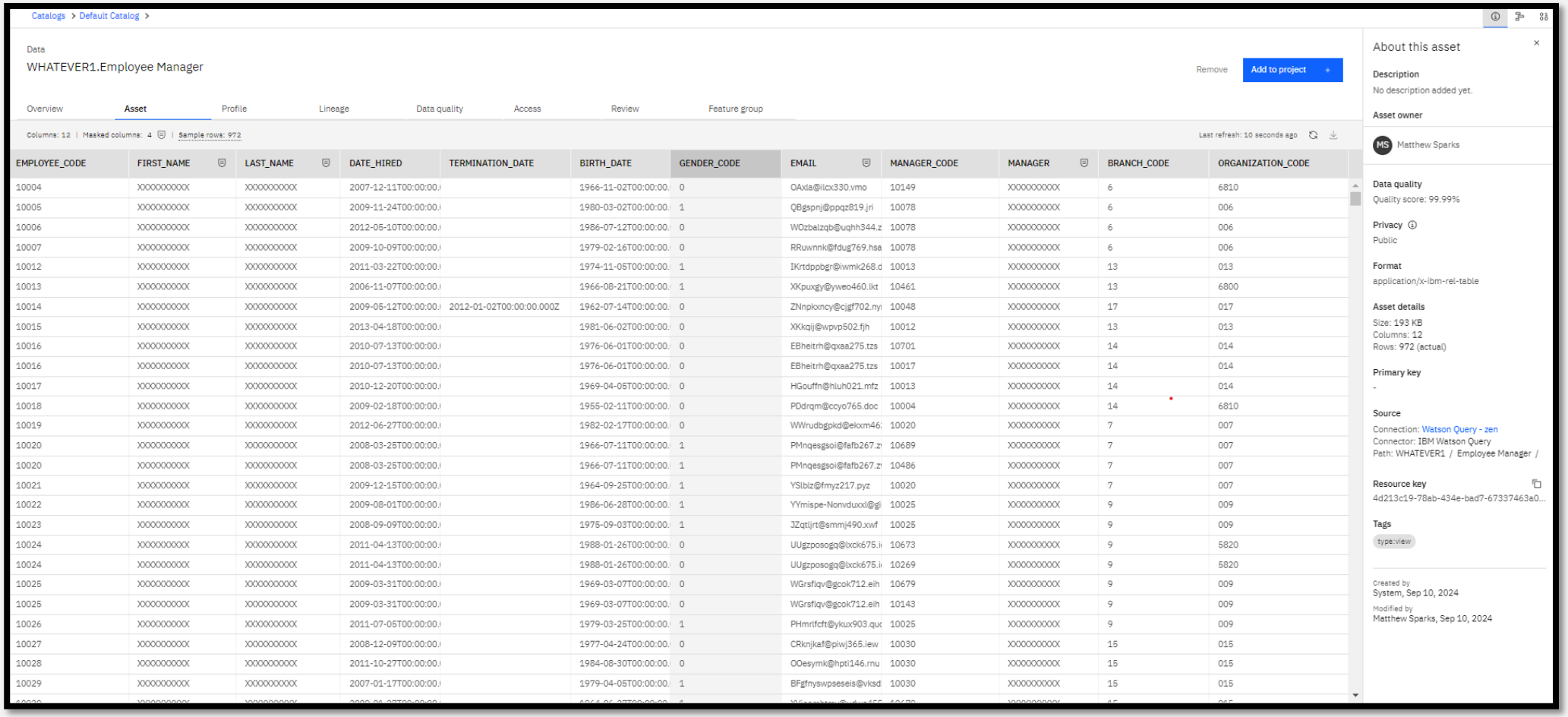


Data Catalog Demonstration – Dummy Data

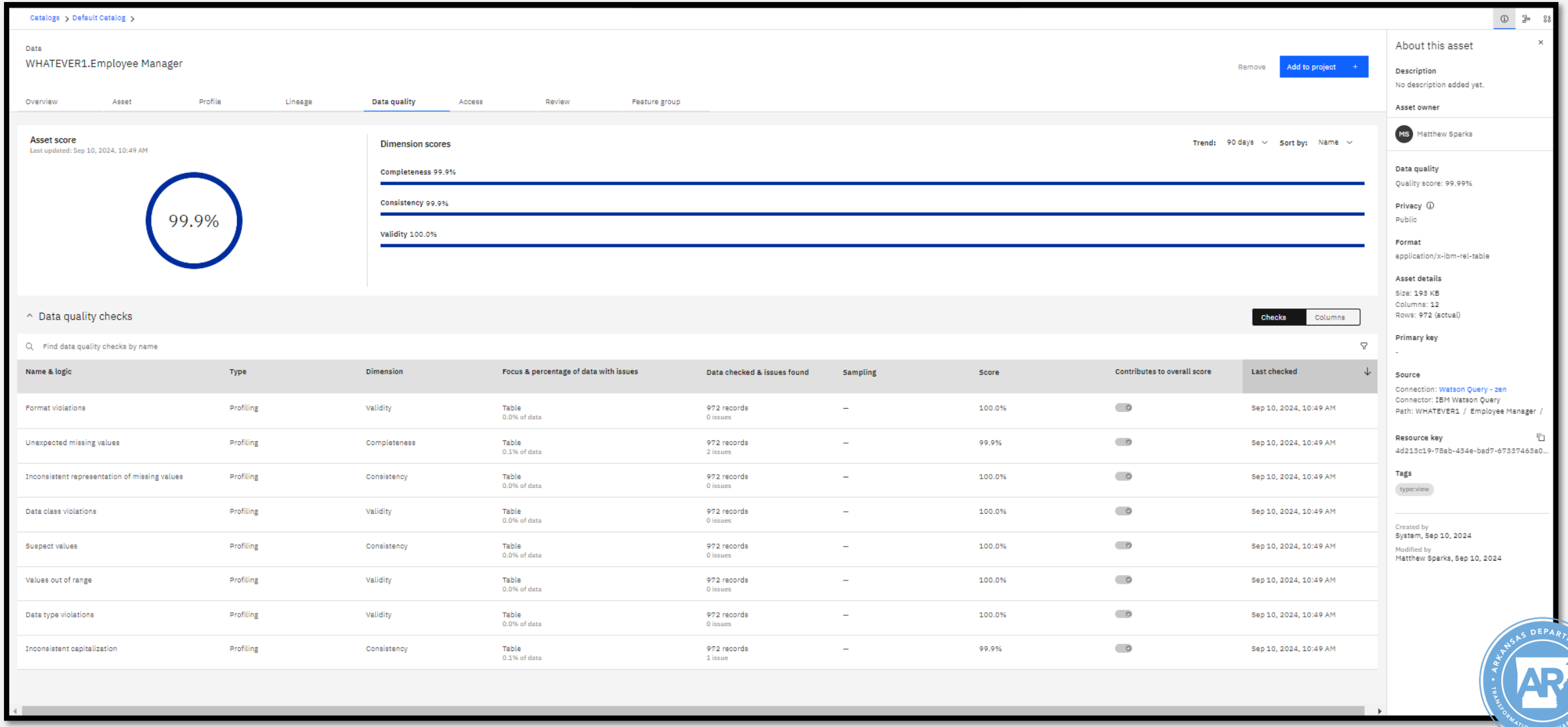


Data Catalog Demonstration – Dummy Data

Redacted data and obfuscated email



Data Catalog Demonstration – Dummy Data



Unlocking the Toolset

- **Machine Learning**
  - Deep learning platform that data scientists can use to build, train, and deploy deep learning models.
- **AI Model Management**
  - Integrate Development Environment (IDE) to build, run and manage AI models.
- **Reference Data Management (RDM)**
  - Centrally manage and distribute reference data across the enterprise.
- **Decision Optimization**
  - Simulation, optimization, and operations research for large-scale dynamic systems.

Unlocking the Toolset

Business Automation Tools

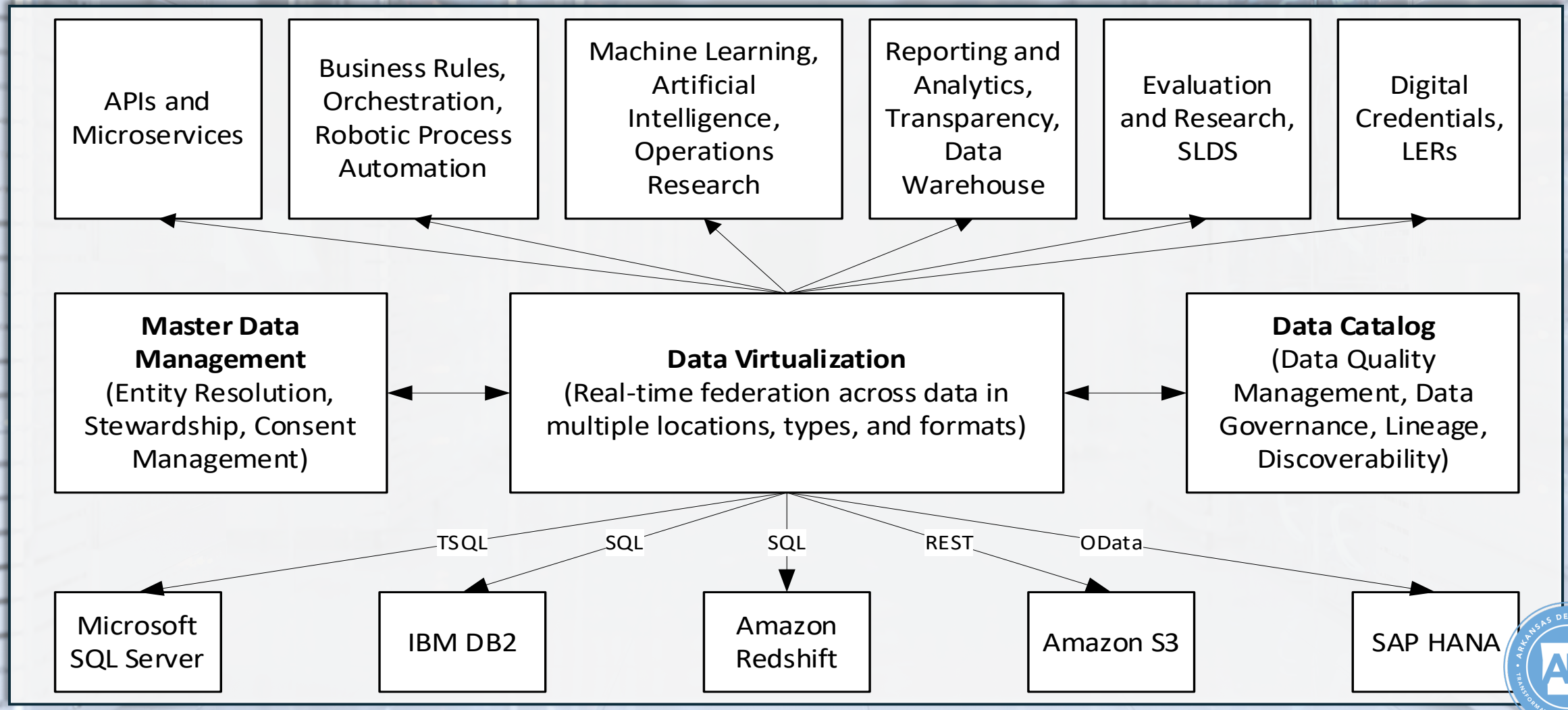
- **Workflow Builder**
  - Automate case and process workflows to reduce or eliminate time-consuming tasks.
- **Decision Management**
  - Make customized decisions by creating and managing business rules independently from applications and processes.
- **Process Mining**
  - Model, analyze, monitor, and optimize overall business processes and proactively identify cases in need of attention.
- **Robotic Process Automation (RPA)**
  - Automate processes at scale with software robots that act on AI insights to complete tasks with no lag time.

Unlocking the Toolset

Integration Tools

- **Enterprise API Management**
  - Create, manage, secure and socialize APIs.
- **Message Queue**
  - Supports the exchange of information between applications, systems, services and files by sending and receiving message data via messaging queues.
- **Big Data Transfer Tool**
  - Data transfer software for moving amounts of data quickly and securely, regardless of network conditions.

Unlocking the Tool Set



Questions and Open Discussion

Objectives of the Arkansas Data Hub

The core objectives identified in the feasibility study informed by Act 1282 of 2015, the gap analysis of agency needs, and the open data and transparency task force include:

1. Enable openness, transparency, and pervasive, **self-service data access and delivery**.
2. **Share data** to enhance its value while enforcing privacy and security (Sharing)
3. Support **data-driven decision making** and analytic maturity through development and support of analytic skills and shared services.
4. **Integrate data** for improved cross-agency analysis and **reduced duplication of data and efforts**
5. Enable **real-time assessment, support, alignment, and automation** of decisions, programs, and resources.
6. **Manage enterprise data** as a state asset.