

Data Democratization

Training Material





Introduction to Data Democratization

- Data Democratization in an organization makes data available to all stakeholders and provides easy to use tools to access data
- Data is now everyone's business and everyone in organization should have access to data to generate meaningful content, insights, etc.
- How do we empower users with equal access to data:
 - Data Discovery
 - Data exploration
 - Data Visualization
 - Data Experimentation

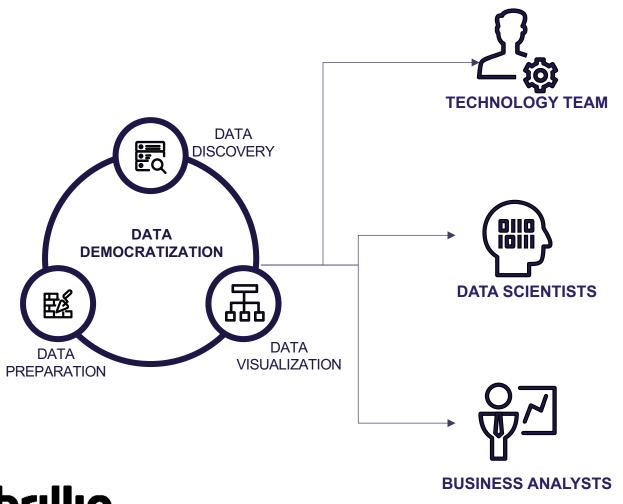


Why is Data Democratization important for an Organization?

- Ability to Harness actionable insights
- Foster the culture of informed decision making
- Move faster on initiatives within the organization
- Drive data driven innovations
- Reduction of duplicate work from data analyst



How is Data Democratization leading Organizations towards Data Literacy?



SINGLE VERSION OF TRUTH

- Create and maintain standardized business reports
- Maintain Infra, applications and data



MULTI-VIEW CONSUMPTION

- Quick analysis and hypothesis testing
- Zero waiting for 'perfect' version of data elements



REPRESENTATION OF INTELLIGENCE

- On the fly data consumption and modification
- Dashboards for analysis and decision making





Enablers of Data Democratization

CORE CAPABILITIES



Data **Foundation**



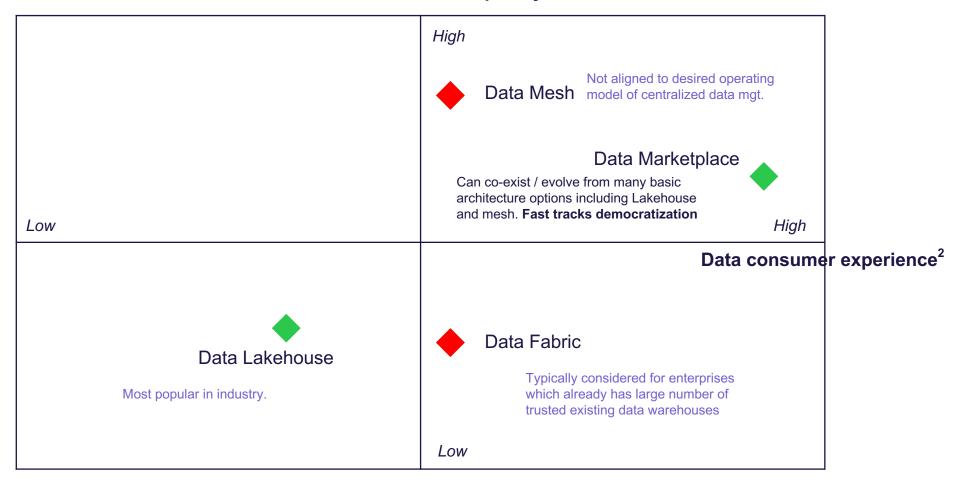
Human Centric Design



Right speed governance



Transformation complexity ¹





- Complexity determined by technology, business process and operating model changes
- Data consumers include business users, analysts, digital product owners, data scientists and engineers

Success and Adoption of Data Democratization is dependent on providing a Human-centric Layer for experience

KEY COMPONENTS

Lake House

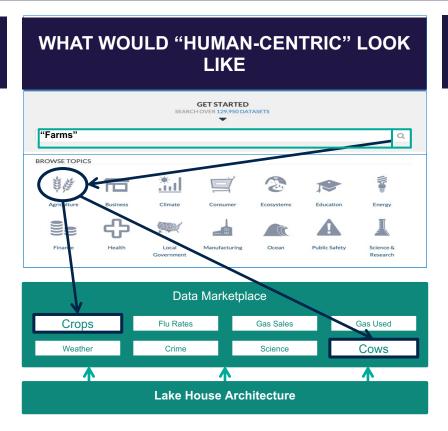
The collection of data sets within the Enterprise to be searched and browsed

Data Marketplace

A Data Marketplace would provide a transparent. collaborative visual medium to drive data consumption

Data Catalog

Collected business and technical information that describes the data sets in the Marketplace



WHAT WOULD IT PROVIDE

Search & Browse Data Keywords Fuzzy Matching Categorized Data Sets Exposed API

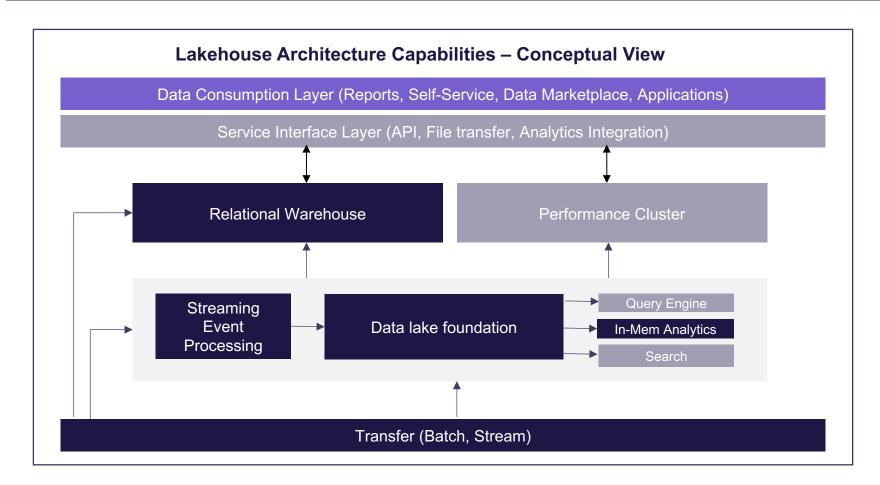
Information Available

Business Meaning Technical Metadata Recommended use Quality

User Comments & Tags Related Data Sets



Data Lakehouse Architecture provides the right springboard for Data Democratization



Key outcomes

Best in class architecture

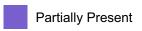
Reduced time to market

Least incremental/startup cost



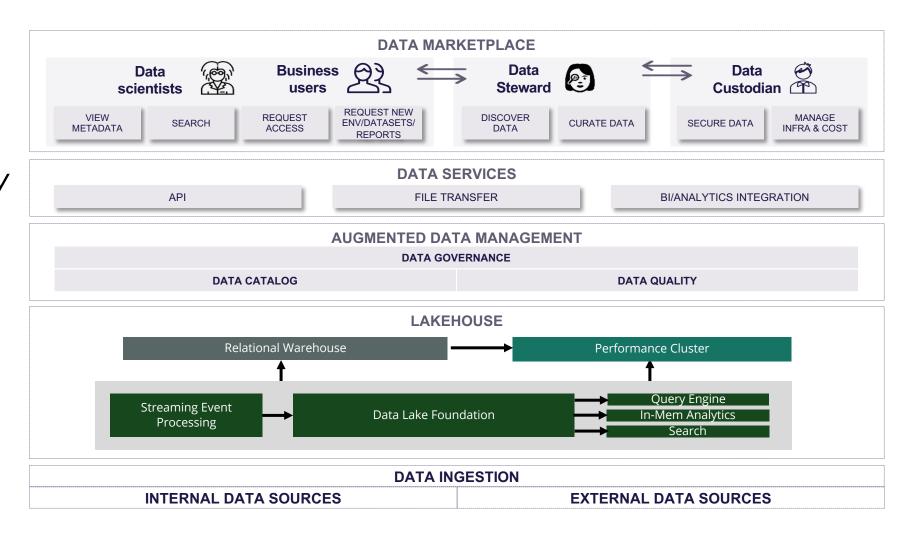






Data Marketplace

Brillio envisions a Data Marketplace on top of the Lakehouse architecture, enabled by an Enterprise Data Catalog to drive Transparent, Collaborative Data Democratization with right-speed Governance









Administration



User Management, Customize User profiles and Access management & group assignments



Back End

Federation Engine – Used to Harmonize data coming from different source systems.

User Management Engine – intelligently updates users from LDAP and assigns appropriate user access levels



Collaboration



User News Feed – tailored towards user 's demographics (role, location,, interests, etc.) within the company.



Share & Follow datasets – notification if changes are made



Recommendation Engine – Explore data. Driven from Federation Engine as well as user behavior.



Business Process Workflow



Telemetry

User Analytics – Analyze user behaviors including usage, demographics, user activity, etc.



Data set Analytics – Analyze datasets including access frequency, popularity, etc.



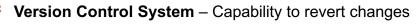
System Analytics – Analyze system usage including peak time usage analysis, etc.



Combined Analytics – This will provide a granular view of user access for data sets.



Customer Experience





Application Program Interface (API) – API which allows the generation of data feeds



Search & Browse – keyword search, fuzzy logic



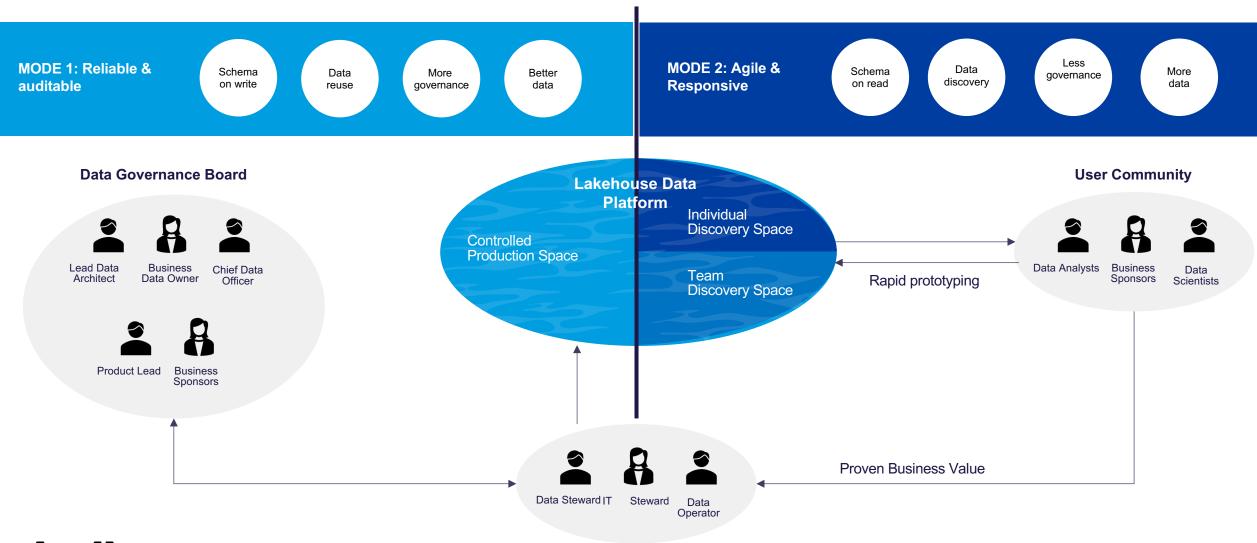
Data Exploration - Visualization & Drilldown features



Export data to multiple formats – CSV, PDF,



Right-Speed Governance would drive success and adoption





Right-Speed Governance starts with defining the key roles and responsibilities

Data Librarian



Ensures appropriate Metadata information is available for all datasets on the lake.

Data Owner

Create spec and provide necessary information for ingesting.



Creates necessary metadata policies and tagging.

Key Data Governance Roles

Technical Architect



Reviews use case requests against platform capabilities.

Performs architecture review on building new platform capabilities to accommodate valuable use cases.

Data Governance Board

Agrees and monitors process for ingesting and granting access to data.



Used to ensure appropriate security & privacy risk assessment, and can fully approve, approve with restrictions or reject requests.



IT Steward

Performs technical assignment of users to groups.

Supports implementation of data lake tools.

Processes for Data Access



Scenario

Prior to this, the Data Consumer has checked the data catalogue and knows the data he/she wants resides on the data lake.



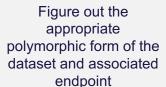
SANDBOX AREA **REQUEST**

Scenario

The Data Consumer requests data ingestion into his/her individual discovery area.

Arrive at the Data Catalog Interface with a business problem

Recommendation of other data sets which may be needed for similar business problem













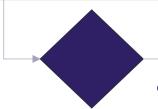


Search the Right Data from N no of instances possible in the data catalog

Select "right fit" data based on quality, ranking, rules of use

Data provisioning into personal work space





Tech Innovation that propels data democratization

- 1. Cloud Storage
- 2. Data Federation Software
- 3. Data Virtualization Software
- 4. Self-Service BI Application
- 5. Cloud-based ETL/ELT SaaS Platforms

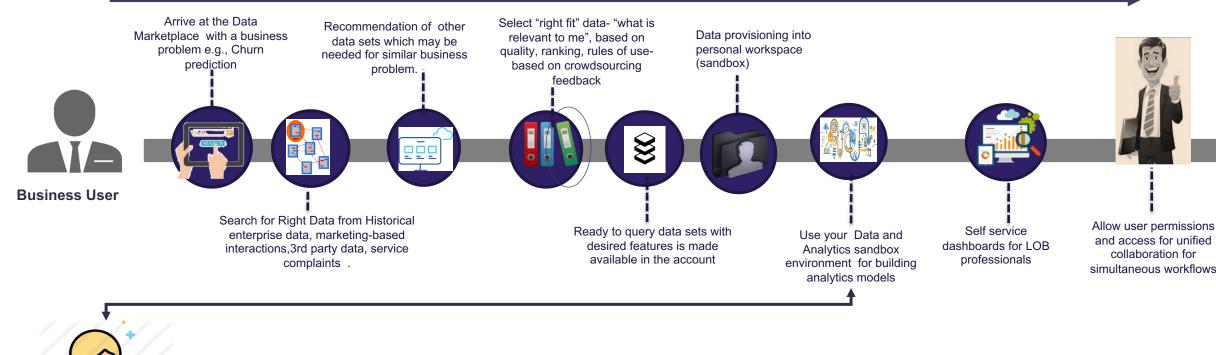


Best Practices of Data Democratization

- Start by establishing a culture of data literacy
- Embrace self-service analytics tools
- Implement user permissions and create policies for data accessibility
- Involve your employees
- Manage Data Quality Proactively
- Prioritize Security, Privacy, and Compliance



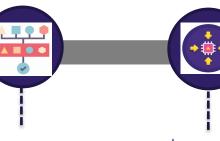
Case Study – User Journey within a Data Marketplace for Leading **Real-Estate Listings Provider**





Data scientist





Leverage prebuilt features from the feature store



Faster experimentation cycles: leveraging prebuilt models or validating the existing business model or new model with customizations



Framework for automated testing workflows. A/B testing with CI/CD tooling and workbenches



Containerization/ dockerisation of models of faster deployment, reusability and reproducibility

Thank you

