Achieving Data Maturity

Executive Summary

This playbook serves as a guideline for achieving and maintaining data maturity at Intuit, which is a critical strategic asset essential for informed decision-making, operational efficiency, compliance, and creating Done For You Experiences (DFY) for customers. Mature data fuels innovation and drives competitive advantage, enabling Intuit to deliver superior customer experiences with velocity and achieve its business objectives. It introduces the foundational L0 Data Capability that supports data-producing and consuming systems. Definitions for Data Assets, Data Products, and Hosting Locations are explicitly outlined to ensure a standard understanding across the organization. This initiative is guided by the Intuit Clean Data Maturity Standard ITAD 1.3, which establishes the foundational principles and standards for data management across the organization.

Objectives of Data Maturity

The objectives of achieving data maturity at Intuit include ensuring that data is accurate, complete, reliable, and consistently formatted. This maturity is not merely an operational goal, but a strategic imperative critical for powering AI systems, executing successful marketing campaigns, enhancing data analytics capabilities, and boosting overall business performance. The Data Maturity Star Scoring system (ranging from 0 to 5 stars) is a key component in measuring progress, with a set target for all data assets to achieve at least a 3-star (Minimally Mature) scoring by FY25Q4.

Utilizing Data Discovery and Scoring Systems

To facilitate the usage and management of data maturity, the playbook emphasizes the implementation of Data Discovery tools for locating and accessing mature data. Additionally, the L3 Data Scoring System is advocated for maintaining data quality and ensuring transparency and accountability in data handling processes, all of which are essential for leveraging data strategically.

Monitoring and Metrics

Important metrics such as the Clean Data Adherence Rate (CDaR) and Schema Management Efficiency (SME) are described to track progress towards improved data maturity. These metrics help monitor the effectiveness of the data management strategies implemented and assess the impact on reducing operational costs and boosting efficiency, ultimately contributing to Intuit's strategic goals.

Getting Started

This section covers the basics of getting started with achieving data maturity.

Introduction

The objective of this playbook is to provide a comprehensive guide for leaders and teams to understand, adopt, and maintain data maturity practices. By following this guide, teams will learn the importance of data maturity, the steps required to achieve data maturity sets, and how to maintain these standards consistently. This guide will walk users through the L0 Data capability, the foundational data platform supporting data-producing and consuming systems. It will define key terminology like Data Assets, Data Products, and Hosting Locations, ensuring a standard understanding across the organization.

The playbook emphasizes the Intuit Clean Data Maturity Standard ITAD 1.3, targeting a minimum 3-star scoring for all data assets by FY25Q4. It details the data maturity achievement process, including assessment, defining standards, cleansing, and integration. Users will learn about utilizing Data Discovery tools and the L3 Data Scoring System to locate, manage, and maintain data quality. Metrics such as Clean Data Adherence Rate (CDaR) and Schema Management Efficiency (SME) will be discussed to track progress. The guide further explains what data maturity is, why it is crucial for AI development, marketing, analytics, compliance, and customer satisfaction, and the specific steps needed to produce, promote, and consume data maturity within Intuit.

Overview of the L0 Data Capability

This section serves to outline the main concepts within the **L0 Data Capability**, a data platform developed to support the creation and management of data-producing and data-consuming systems that is thoroughly documented in the L0 Concepts doc. Based on Domain-Driven Design (DDD) principles, it provides a scalable and standardized approach to data management within Intuit. The main concepts are described in the following sections.

Context

The L0 Data Capability provides a robust, standardized foundation for data management at Intuit, promoting operational excellence, secure data exchange, and efficient data-driven decision-making. From the L0 Concepts doc, in FY25, Intuit envisions an ideal data-driven state characterized by:

- Data maturity production available in real-time with low latency
- Capabilities that enable fast and data maturity consumption and production across multiple clouds
- Mandatory adherence to data maturity policies and security measures (e.g., fine-grained access control and encryption)

- High availability, performance, and reliability reflecting operational excellence

Data Strategies

This section provides detailed information about L1 Capabilities, data definitions, quality, and persistence. For a comprehensive understanding, refer to the following related documents:

- Data Powered Intuit FY26 strategy document
- Clean and Secure Data FY25 strategy document

Architecture

Key architectural elements include:

- **Data Recipes**: Blueprints for data transformation and movement
- **Hosting Locations and Data Ports**: Interfaces for controlled data exchange
- **Credential Management**: Policies for data access
- **Data Asset and Data Product Integration**: Organizing data for discoverability and usability

Key Performance Metrics

Key performance indicators include:

- **Clean Data Adherence Rate (CDaR)**: Measure of compliance with data quality standards
- **Schema Management Efficiency (SME)**: Evaluation of schema management effectiveness
- **Principle of Least Privileges (PLaP)**: Assessment of minimal access permissions
- **Hosting Location Lockdown Score (HLLS)**: Enforcement of access control policies
- **Developer Velocity (DpV)**: Speed and efficiency in data management
- **Data Access Request Time (DaRT)**: Efficiency of the data access process

Data Terminology Review

Understanding the following terms is crucial for developing a foundational knowledge of data terminology at Intuit. These terms appear throughout several key documents:

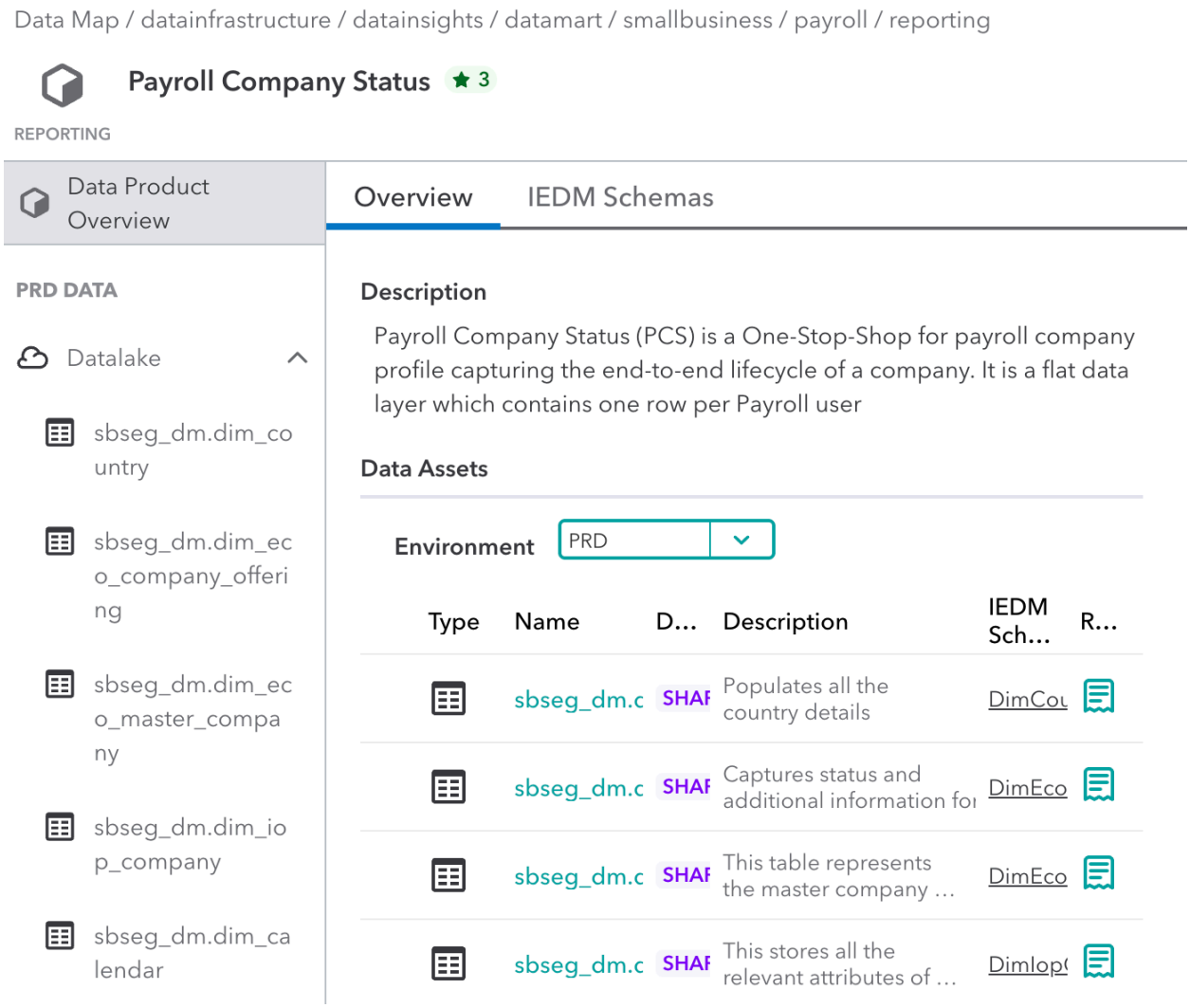
- Intuit Clean and Secure Data Strategy
- L0 Data Capabilities
- Clean Data Maturity Model
- Various paved path experiences found in Data Discovery

More detailed and comprehensive descriptions of these and other terms can be found in the L0 Data Capability Concepts and the Entity Relationship Diagram (ERD). Key terms are highlighted below for quick reference.

- **Hosting Location** - A system for persisting and serving data. Hosting locations must adhere to the hosting location requirements decision, ensuring proper cataloging of Intuit's entire data estate and proper control and exchange of data with other teams and systems. Properly controlled hosting locations promote efficient and compliant data collaboration and management across the company. Examples include, but are not limited to, the Intuit Data Lake, CDC, Event Bus, Credit Karma (CK) Data Lake, and Mailchimp (MC) Data Lake.
- **Data** - An item of information distinguishable from other items through its organization in a collection of items in a Data Asset. When used as a plural noun, it is often synonymous with Data Asset. Examples include a row in a table, an event message on a topic in an eventing system, a response from a service API, or a document in a Google Drive folder.
- **Data Asset** - A resource at Intuit that contains data. It aggregates essential information such as ownership, intent, lifecycle status, and environment. Data Assets are also referenced for data lineage and data access control management. Examples include Event Bus topics, Data Lake tables, and CDC attribute sets.

  The following image displays some of the Data Assets associated with various

resources:



- Data Product - A collection of consumable data assets across one or more hosting locations, designed to represent the state of an important business domain in order to solve a well-defined business problem. Data products have ownership, data entities defined in IEDM semantic definitions and schema, service level agreements (SLAs), star scorings, security and compliance tags, and access policies.The **Intuit Enterprise Data Model (IEDM) semantic** refers to a key concept in database management and software engineering in that it describes the fundamental operations that can be performed on data within a database or information system. For detailed information on available data operations in the IEDM, refer to the IEDM Github repo.
- **Data Asset Intent** - The categorization scheme that classifies data based on its intended use, access level, and potential consumers. The scheme is meant to be complete and exhaustive, meaning that all data assets at the company will exclusively fall into one of the following three categories:
    - **Internal** - A data asset generated, managed, and consumed within a specific team's or system's business scope. The business scope boundary is defined by

a **Dev Portal project**, which itself is intended to have a reasonably well-defined business problem and scope of responsibility with a reasonable number of people and systems that belong to it.
    - **Temporary** - A data asset not meant to be long-living. It is typically generated during data exploration, for internal decision-making and analysis. It is often data that does not have a regularly scheduled pipeline associated and is not monitored.
    - **Consumable** - A data asset made available for cross-dev_portal project use. It should be easy to understand, durable, and usable by various consumers.

Understanding Data Maturity

This section helps you understand the different levels of data maturity.

What is Data Maturity?

Data Maturity is defined by the Intuit Clean Data Maturity Standard ITAD 1.3 as data that is discoverable, trustworthy, and usable, establishing a clear understanding between data producers and consumers within Intuit. It involves defining maturity levels and criteria for data assets to ensure they are accurate, consistent, and reliable.

Mature data is crucial for several reasons:

- **AI Development:** Essential for training effective and reliable AI systems
- **Marketing:** Enables precise targeting and customized campaigns, improving engagement and ROI
- **Data Analytics and Insights:** Prerequisite for accurate analytical models and valuable insights
- **DFY Experiences:** Enhances service quality and personalization, leading to increased satisfaction and Done For You (DFY) experiences for customers
- **Compliance and Risk Management:** Ensures adherence to regulations and reduces legal risks
- **Data Integrity and Trust:** Maintains data integrity, building trust between producers and consumers
- **Efficiency:** Reduces time and resources spent on correcting data errors
- **Cost Savings:** Minimizes errors and optimizes processes, leading to cost reductions
- **Decision Making:** Supports informed and strategic decisions, minimizing analysis errors

In summary, data maturity is a vital asset that supports various aspects of business performance, enhancing decision-making, operational efficiency, compliance, customer satisfaction, creating DFY AI-native experiences, data integrity, cost management, and analytical capabilities. Maintaining data maturity is a priority for sustained growth and success.

Accountability and Strategic Alignment

The executive-level Data Maturity Dashboard provides leaders with a comprehensive view of their data maturity, and has been integrated into TOS (Technology Super Scrum), as well as all CTO+1 operating mechanisms, ensuring integrity, ownership, and accountability for data quality. TOS is a strategic meeting where leaders from different technology departments collaborate to align strategies, roadmaps, and address operational challenges, highlighting and addressing the most important operational items and raising awareness about what is happening across Technology.

Note: For detailed information on metrics, refer to Data Maturity - Accountability & Impact.

Understanding the Data Maturity Strategy

The data maturity initiative has a clear objective for fiscal year 2025: to elevate all data assets to a minimum maturity level of 3-star by the end of Q4. This foundational step ensures a baseline of data quality, reliability, and usability across the organization. Building upon this, plans are in place to further enhance data maturity, with requirements targeting 4-star or higher to be implemented within fiscal year 2026 on path to the 5-star target state levels. The Mature Data Product Strategy document provides a comprehensive overview of Intuit's approach to enhancing data quality.

Specific High-Impact Use Cases

This section summarizes high-impact use cases from FY20-FY25. For detailed information for use cases by fiscal year, refer to this document.

The Data Maturity Dashboard enables granular monitoring of data assets against standards across all business units, ensuring leadership accountability particularly in remediating highly used assets to help limit what can affect customers and revenue.

High Impact Use Case for Enabling Granular Monitoring of Data Assets

Mature data underpins the development of efficient and trustworthy AI models, leading to significant operational improvements. This aligns with the goals outlined in Intuit's Clean & Secure Data FY26 Strategy. For instance, utilizing higher-quality data has led to 50% faster data discovery processes and 27% quicker launches of marketing campaigns.

- **Data Quality and Accountability:** The Data Maturity Dashboard provides a detailed view of data quality across Intuit business units, enabling accountability at the PD leadership level. This aims to establish a culture of clean data management to prevent significant data incidents.

- **Support for Strategic Initiatives:** The dashboard supports the "Clean & Secure Data FY26 Strategy" and fulfills leadership commitments related to data quality.
- **Foundation for AI Models:** High-quality data is crucial for developing effective and trustworthy AI models.
- **Performance Improvements with High-Quality Data:**
    - Data discovery is 50% faster
    - Marketing campaigns launch 27% quicker
    - Time spent on data quality issues by Machine Learning Engineers (MLEs) decreases from 20% to 5%
    - Root Cause Analyses (RCAs) decreased from 92% to 20%
    - MLEs can consume data from the event bus in minutes instead of days

High Impact Use Cases from FY20-FY24

- **Reduced model processing time**: The Stolen Identity Refund Fraud Model has a 12X reduction
- **Reduced manual work hours**: A reduction of 700hr/year
- **Increased lending capacity**: Lending capacity refreshed more frequently (from 5 to 1 times/week)
- **Increased funding for small businesses**: $1.2B in funding for over 30,000+ small businesses
- **Improved customer verification**: Ability to verify customers calling in for support increased from 40% to 72%
- **Reduced onboarding time**: Onboarding time in QuickBooks Payroll reduced by 60%
- **CG Marketing Data Get Well (MDGW)**: Contributes to an expected business impact of approximately $700M of TurboTax Revenue across marketing channels
- **SBSEG Marketing Data Get Well (MDGW)**: Contributes to a projected incremental $50M annually for Ecosystem Lifecycle Marketing revenue

High Impact Use Cases from FY25

- **Intuit Assist (IA)**: Clean data for IA is projected to drive down costs by approximately $26.4M by the end of FY25
- **AI Cortex Matcha Team**: Launch of a High Quality Document Foundation leads to a 30x cost reduction in document data sourcing and human annotation, with estimated savings in Q2 and Q3
- **Automated expert note-taking**: Projected to bring in over $4M in savings
- **Generative Assist**: (Done for you Q&A and Chat) Projected to result in approximately $3M in annual savings each
- **IDX Account Ownership Verification/Account Ownership Details (AOD)**: Estimated to have a $32M loss saving by reducing unauthorized returns

- **CG Marketing Campaigns**: Clean data enables the successful launch of critical campaigns with significant revenue targets (e.g., $70.7M for the LAB Welcome campaign, $95M generated by form campaigns)
- **Incremental Clickstream Pipeline for AI Powered Personalization**: Implementing incremental data processing results in cost savings of $15,000 per month and $180,000 per year

Adopting Data Maturity Scoring

The **Data Maturity Star Scoring system** evaluates data assets through a detailed framework that rates their security, management, and readiness based on a 0 to 5-star scale. This section discusses the specific criteria for achieving 3-star and 4-star data maturity at Intuit, emphasizing the distinction and enhancements required for each level in data quality and reliability.

Understanding the Data Maturity Star Scoring System

Data Maturity Star Scoring refers to a scoring system for data assets as officially described in the Clean Data Maturity Standard decision. The Data Maturity Star Scoring system scores the degree to which a data asset is secure, stewarded, documented, modeled, high quality, reliable, and usable. Each successive star in the scoring system builds on the attributes added by the one before it. Mature Star scores are:

- **0 Star** - *Uncontrolled* - Represents a security/compliance risk
- **1 Star** - *Compliant* - Owner identified; known security/compliance disposition
- **2 Star** - *Controlled* - Documented; appropriately limited access and minimum scoring requirement; appropriately encrypted
- **3 Star** - *Minimally Mature* - Stewarded, registered, interoperable, data movement paved paths available, proven operational monitoring; At this level, the data is consumable and can be shared across dev_portal projects
- **4 Star** - *Advanced Mature* - Stewardship team measured on responsiveness; higher documentation bar; semantic relationships defined; exhaustive data quality checks; lineage available
- **5 Star** - *Target State Mature* - Highly functioning stewardship team with well established consumer-feedback loop via Scaled Stewardship on data model, documentation and exhaustive data quality checks

At Intuit, the difference between 3-star data and 4-star data maturity involves higher standards of data refinement, processing, and enhancement to elevate the level of data quality, reliability, and usability. Mature data is ranked based on its reliability and usability, as follows:

- 2-star data is discoverable but restricted to internal use within DevPortal project teams, offering the least flexibility.
- 3-star data meets minimum standards but may have gaps affecting its reliability, thus usage is not advised for high priority cases where 4-star or higher is preferable.

- Both 4-star and 5-star data are consumable, with 4-star data prioritized in search results and 5-star data being highly reliable for all use cases. Higher star ratings indicate better quality and broader usability.

For more information, refer to the Intuit Clean Data Maturity Standard ITAD 1.3.

Understanding the L3 Data Scoring System

This section describes the target state architecture for the **L3 Data Scoring System** that aims to provide clarity to data consumers about the data they are using by assigning a score to Data Assets and Data Products based on the Intuit Clean Data Maturity Standard. It also provides data producers with a clear framework to improve the quality of their Data Assets and Data Products according to the Intuit Clean Data Maturity Standard ITAD 1.3. It helps producers identify improvement areas, optimize data quality, and meet higher standards, increasing the value and usability of the data for consumers.

Key aspects of the L3 Data Scoring System include:

- **Data Asset/Product Scoring**: Every data asset and product will receive a score from zero to five, reflecting its compliance with the clean data maturity standards
- **Scoring Categories**: Assets and products will be scored on multiple aspects, such as data modeling and data quality
- **Score Calculation**: Scores can be calculated on demand, upon state changes in related systems, or periodically
- **Score Change Notification**: Changes in scores will be notified to producers and data stewards and updated in the data registration service
- **API Routes**: The system will provide APIs to calculate and fetch scores for data products and data assets

The preceding list also outlines the aspects used for data asset and data product scores, as well as domain events that will be published for score changes.

The L3 Data Scoring System runs and rescores data assets as follows:
:

- Runs/rescores data assets on a weekly basis
- Reruns in a push system whenever edits are made

Tip: For questions on the scoring systems, reach out to `#data-maturity-score-support`.

Criteria for 3-Star Data Maturity

At Intuit, achieving 3-star data maturity signifies a robust foundation in our data management practices, ensuring that our datasets are accurate, consistent, and reliable. According to the

Intuit Clean Data Maturity Standard ITAD 1.3, here are some of the critical dimensions for achieving 3-star data quality:

- **Stewardship**: Satisfactory Stewardship with a stewardship grade of >= 3 of 5
- **Documentation**: Satisfactory documentation with a documentation grade of >= 3 of 5
- **Data Model**: Satisfactory model defined in IEDM with modeling grade >= 3 of 5
- **Data Quality Monitoring Coverage**: Satisfactory data quality coverage defined with a grade >= 3 of 5
- **Operational Readiness**: Satisfactory operational grade of 3 of 4

For 3-star data maturity, consumable data only meets the minimum standards, leaving open the possibility of gaps in testing and monitoring that impact reliability. Your goal should be to push for 4+ stars, aligned with the target data maturity tier of your use cases. Again, refer to the Intuit Clean Data Maturity Standard ITAD 1.3 for detailed information about the data maturity tiers.

3-star data is machine-readable and typically stored in non-proprietary formats such as CSV or JSON. It is recommended for use cases that require a higher level of data interoperability and accessibility.

Criteria for 4-Star Data Maturity

Refers to a set of data quality standards that must be met to achieve a high level of data integrity and reliability. These standards build on the 3-star data maturity standards and typically focus on the following critical dimensions:

- **Stewardship**: Good Stewardship with a stewardship grade of >= 4 of 5
- **Documentation**: Good documentation with a documentation grade of >= 4 of 5
- **Data Model**: Good model defined in IEDM with modeling grade >= 4 of 5
- **Data Quality Monitoring Coverage**: Good data quality coverage defined with a grade >= 4 of 5
- **Operational Readiness**: Good operational grade >= 4 of 4

The move from 3-star to 4-star involves tightening the requirements on all aspects of data management to ensure that data is not only reliable and accurate but is also readily actionable and based on the most timely and well-documented records possible.

Producing Mature Data

This section details the processes and practices for creating and ensuring data meets established maturity standards. It covers adopting data maturity practices, producing new data maturity, and upgrading existing data assets to 3-star maturity, including using My Data and Paved Path resources.

Adopting Data Maturity Practices

When it comes to data maturity, **adopting** refers to the process of integrating and implementing practices and standards that ensure data is accurate, complete, consistent, and reliable. This includes establishing data quality checks, data validation, data maturity, and data governance procedures to ensure the data is free of errors, discrepancies, or inconsistencies, making it suitable for analysis, decision-making, and other business or research purposes. It means treating data as a product or P0 service within your organization, responsible to downstream, cross dev_portal project teams consuming it for Done For You (DFY) customer experiences, marketing, executive insights, etc.

Producing New Mature Data

There are different Paved Paths available for producing new data at Intuit to ensure it meets the outlined maturity standards. Achieving 3-star maturity for data assets at Intuit involves using the Paved Paths framework to systematically improve data management, focusing on data cleansing to rectify inaccuracies, duplicates, and inconsistencies. By following these structured paths, teams can ensure their data meets the 3-star+ maturity compliance, enabling more accurate decision-making and increased efficiency.

Improving Data Quality for Paved Path Data

To improve data quality for Paved Path data, you must go through the interface used to create the data product in order to make relevant updates to increase data score. Here are links to the Paved Path resources for promoting data asset maturity:

- Produce Clean Data using IPS
- Produce Clean Data using Realtime Source Paved Paths
- Produce Clean Data using Event Bus Paved Path
- Produce Clean Data using Unified Ingestion Paved Path

Also, be sure to review Data Stewardship Onboarding, which is a degreed pathway designed for Data Stewards who produce data. This pathway course will help you understand your role and responsibilities as a Data Steward, and the impact clean data has on the company.

Promoting Existing Data Assets to a Higher Level of Maturity

Data stewards have the ability to upgrade data assets post-publication. The specific changes to be made are based upon the reason why the asset is not presently meeting the current standard. The Data Maturity Dashboard can help teams identify both the pockets of data in greatest need of improvement and the scoring components that currently do not meet the data maturity standards described in the next section.

You can improve data maturity for non-Paved Path data pipelines using My Data, and for Paved Path data within Data Discovery and Exploration, as described later in this section.

Tip: Regardless of which type of data you need to mature, be sure to follow the guidelines for data scoring components listed in Producing New Mature Data.

Understanding the Data Scoring Components

As you produce your new data, keep in mind the following data scoring components, as listed in the Data Maturity Dashboard, which reflect the scoring system results to hold teams accountable for producing data maturity that is rated 3-star and higher:

- **Classification**
  - **Purpose**: The presence of security classifications on data attributes ensures that each piece of data is categorized according to its sensitivity and the security requirements tied to its handling and storage.
  - **Impact**: This helps in applying appropriate access controls and protection measures, preventing unauthorized access and leaks of sensitive information. Data classification supports compliance with data protection standards and enhances the overall security posture of the organization.
- **Compliance**
  - **Purpose**: Compliance scoring involves tagging data assets with relevant regulatory identifiers (e.g., 7216 which refers to IRS rules on disclosure or use of tax information).
  - **Impact**: This ensures that all data handling practices are in line with specific legal and regulatory requirements, reducing the risk of legal penalties and supporting ethical data practices. Compliance tags help in quickly identifying the rules applicable to each data asset, simplifying legal audits and ensuring transparency.
- **Data Model**
  - **Purpose**: The data model component refers to the schema definition stored in a central IEDM (Intuit Enterprise Data Model) repository.
  - **Impact**: Storing schema definitions in a centralized repository ensures consistency across the organization, facilitates easier integration of data sources, and enhances the accuracy of data analytics. A well-defined schema aids in data normalization and helps in maintaining the integrity of data across different systems.
- **Documentation**
  - **Purpose**: Documentation includes providing detailed descriptions at both the attribute and asset levels.
  - **Impact**: Robust documentation contributes to better understanding and usability of the data by all stakeholders. It reduces ambiguity and potential misuse of data, enhances collaboration across departments, and supports training and onboarding processes.
- **Quality**

    - **Purpose**: Quality scoring involves the implementation of a Data Quality Observer mechanism that assesses various dimensions of data quality, with a passing score being over 90%.
    - **Impact**: High-quality data is critical for accurate analytics and decision-making. Implementing a Data Quality Observer ensures continuous monitoring and evaluation of data quality, leading to timely identification and rectification of quality issues. It enhances the reliability and credibility of data-driven insights.
- **SLO (Service Level Objectives)**
    - **Purpose**: SLOs relate to specific commitments regarding the delivery timelines, availability, and content quality of data assets.
    - **Impact**: Establishing clear SLOs ensures that data services meet the performance expectations of end-users and internal stakeholders. This leads to better customer satisfaction, improved operational efficiency, and helps in setting realistic expectations for data consumption.

Promoting with My Data

You can find the steps for promoting data assets to 3-star maturity in My Data 3-Star Promotion Steps. The document provides instructions on how to promote data assets to a 3-star score using My Data within Data Discovery and Exploration. It outlines the steps required, focusing on areas like documentation, compliance, data quality, operational readiness, schema, and ownership, specifically for hive_tables. You can also find more detailed information for promoting data assets to 3-star maturity in this demo recording.

Promoting for Paved Path Data Within Data Discovery and Exploration (DDE)

To promote your data assets to a higher level of maturity, select the Paved Path that was used to produce your data asset and follow the instructions. Refer to Improving Data Quality for Paved Path Data for a listing of available Paved Paths.

Consuming Mature Data

Consuming mature data refers to the process of using data that has been thoroughly verified via the L3 Scoring System to ensure it is accurate, consistent, complete, and reliable. Before you consume any data, you should first ensure that the data meets your data requirements by using a Data Map Search in Data Discovery.

Note: This section covers the basic concepts of how to consume data maturity with Data Discovery. For more detailed information, refer to Consuming Clean Data in Data Discovery.

Performing a Data Map Search

In the context of data maturity and data management, using a Data Map Search to find data products, attributes, and topics can be explained as follows:

- **Data Products**: These are collections of data, such as datasets or data models, used for different purposes like analysis or reporting
- **Attributes**: Attributes are specific features or properties of data, like columns in a table or fields in a dataset. Mature attributes are well-defined, correctly formatted, and error-free.
- **Topics**: Topics categorize data into different subject areas, like finance or customer information. Mature data here means that the data is relevant and properly categorized under these topics.

This section covers how to access and understand data maturity within Intuit's Data Discovery application. It explains that users can find data products and assets through search or browsing, and then view key information such as documentation, ownership, usage, and lineage. It also describes navigating within Data Discovery, particularly the Overview and IEDM Schemas tabs in Data Product view, and the details available in Data Asset view, including lineage and usage statistics. Additionally, it covers checking/requesting access to data, exploring data using available functions, and understanding the difference between flattened and unflattened data structures.

The view of the data that you find allows you to understand various aspects of the data maturity, including:

- Ownership of the data
- Any documentation for the data provided by the data owner, or you can add your own as a community member if you want to enhance the existing description
- Who is using the data, and how popular it is
- The lineage of the data in terms of how it is built or derived, and where it is used downstream

Finding Mature Data

You can find data maturity using either the search functionality in the Data Discovery application or by browsing through the Data Map, which is organized by City Map taxonomy.

Using the Data Discovery Application

The starting point for finding mature data at Intuit is the Data Discovery application. This is an in-house application designed to help Intuit data workers discover data efficiently. It is intended to support anyone at Intuit who needs to find and understand data. Key features include the ability to:

- Search the data catalog and Data Map
- View schema/table/column/entity/feature metadata
- Find out who owns data
- Request access to data

For more information on Data Discovery, refer to the Data Discovery documentation.

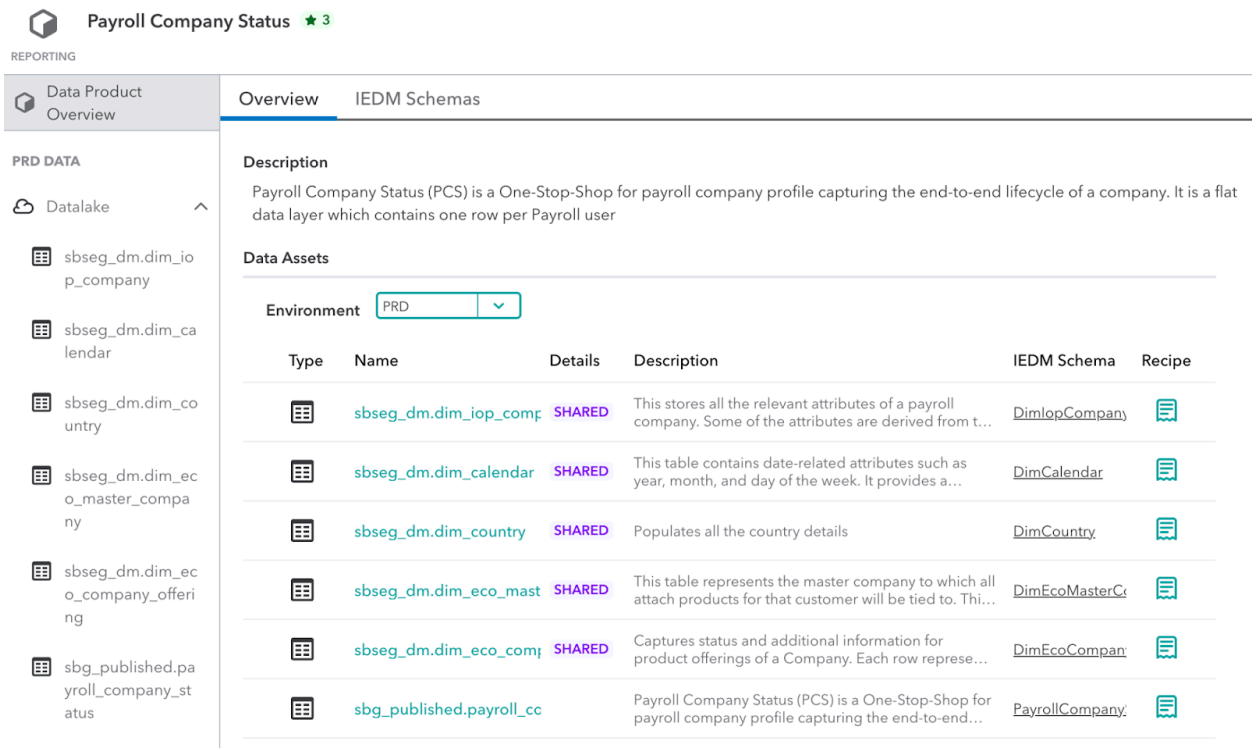
A search bar appears at the top of every page in the Data Discovery application and can be used for keyword search through much of the Intuit data ecosystem. This has also recently been enhanced to support AI assisted search. For more information see the Data Discovery page on Search.

Sometimes it is difficult to find the instance of the data product that you are interested in. For example, if you search for “Product” in Data Discovery you will see that there are many “Product” Data Products (15 at the time of writing). You can add further information to your search to narrow the results. For example, using “Product PPM” will rapidly narrow your search down to the version of Product that comes from the PPM catalog.

The following steps demonstrates how to search for data in the Payroll Company Status data product using the Data Discovery search bar:

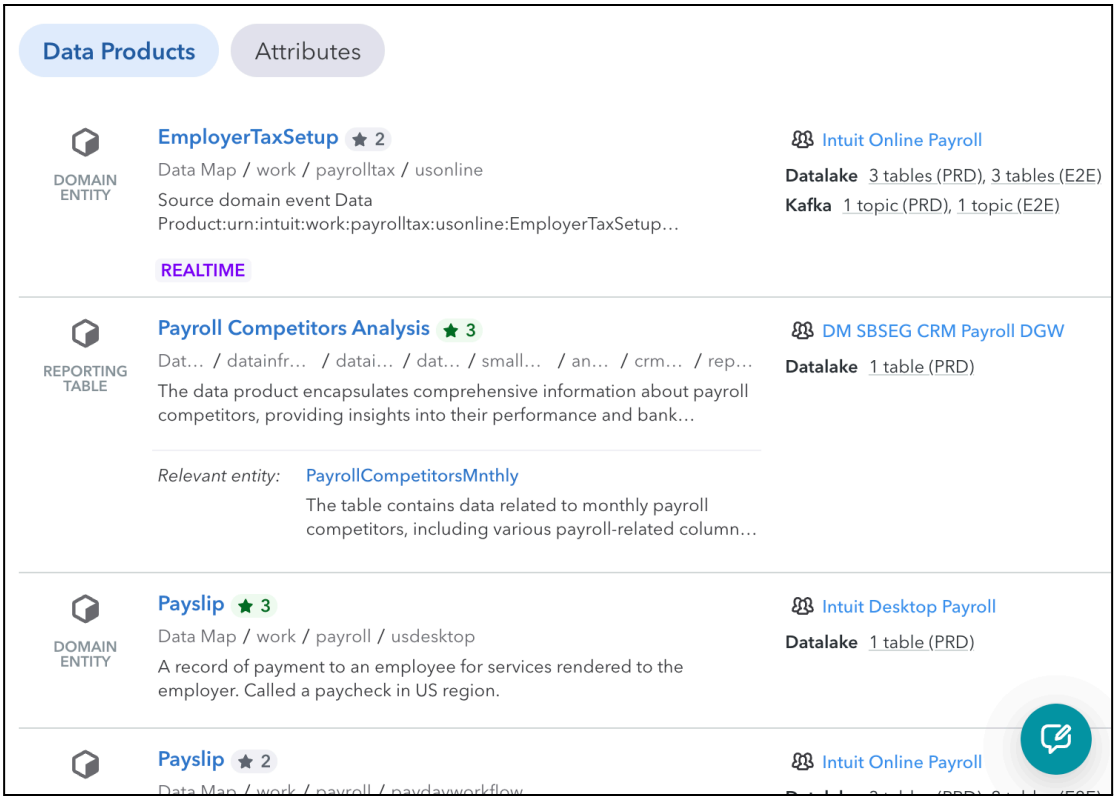
1. Search for “**Payroll Company Status**” using the search bar at the top of the Data Discovery application.
2. Click **Data Products** on the returned results.
3. Click on the **Payroll Company Status reporting table**.

The following image displays the data assets returned for the Payroll Company Status reporting table:



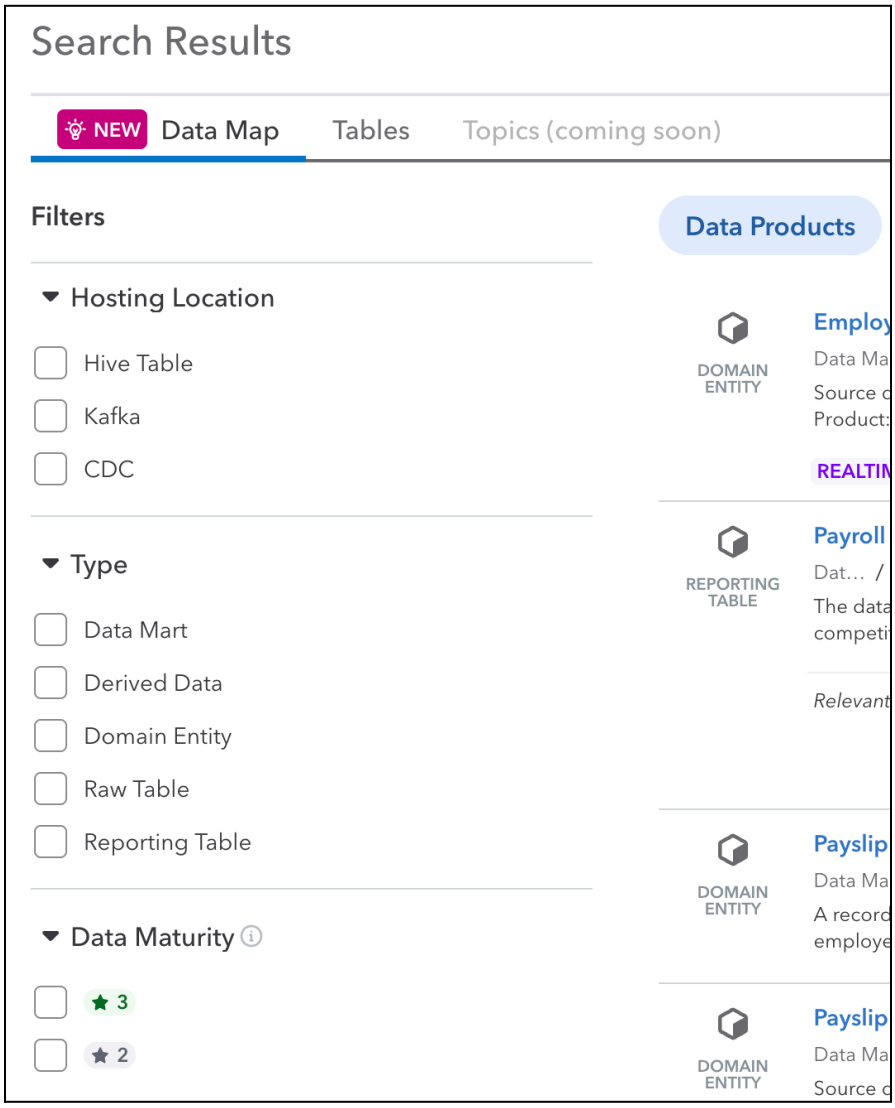
As another simple example of how to find data products in Data Discovery and verify they meet your data maturity requirements:

1. Open **Data Discovery** and enter "**payroll us**" in the **Data Map Search** field
2. Press **Enter** and then select **Data Products**. A list of matching data products displays in the results pane on the right, as shown in the following image:



3. Notice that next to the title of each data product are the number of data maturity stars evaluated by the data scoring system. This is how you determine the maturity of each data product.
4. If you want to filter the returned results by their data scores, scroll down in the Filter pane on the left until you see the Data Maturity score options, as shown in the following

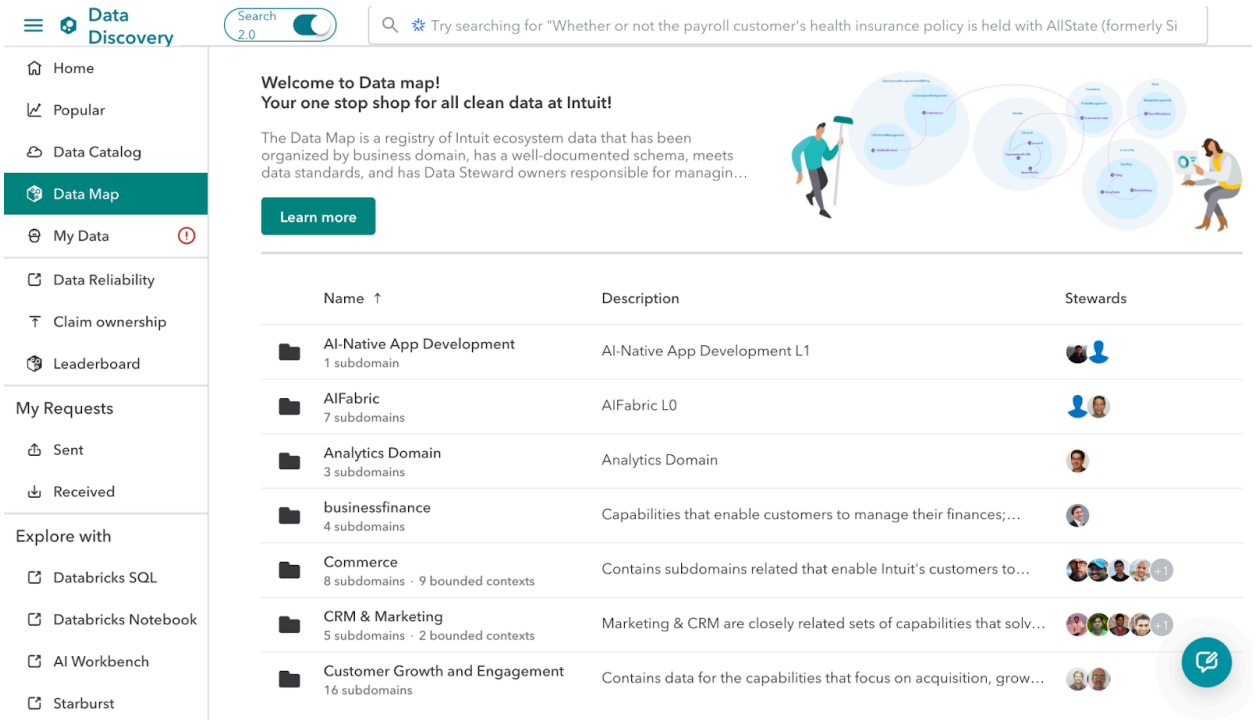
image:



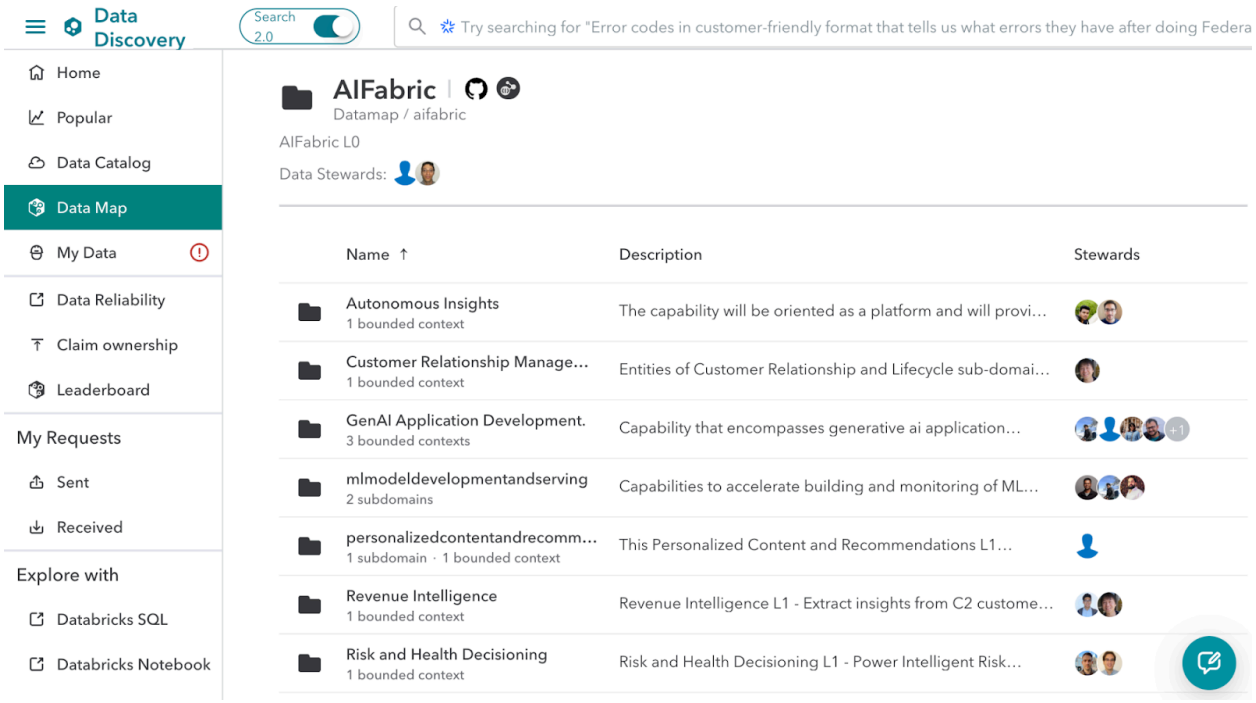
5. In the **Data Maturity** section, select the data maturity star scoring for the data products you are interested in order to narrow or expand the returned results.

Using Browse to Find Mature Data

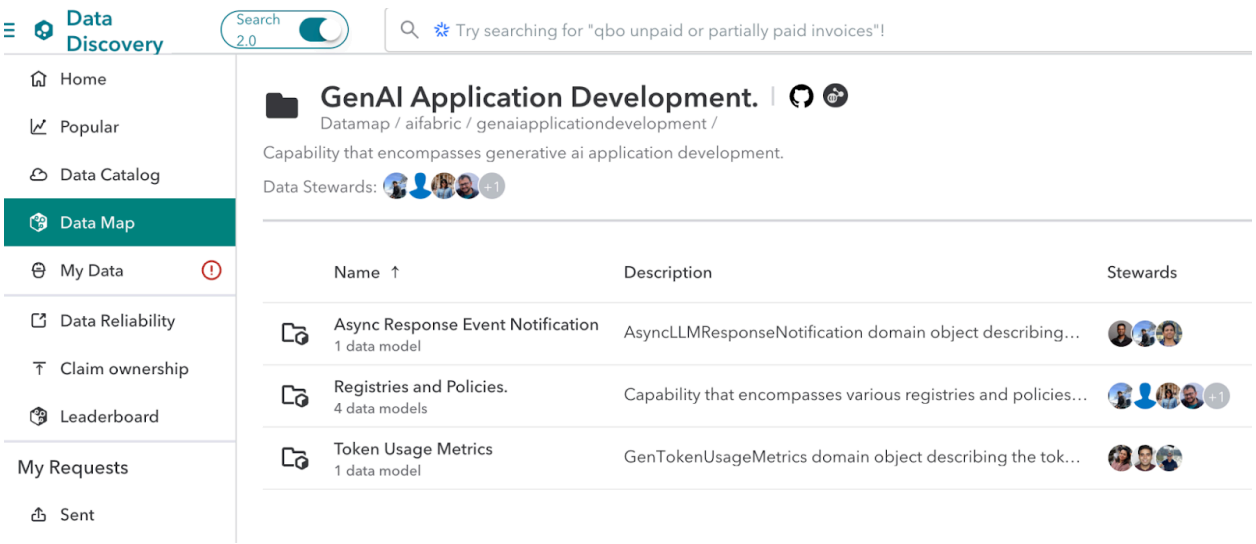
To access a Data Product, start by going to the Data Discovery section of the platform. On the left side of the screen, you will see a navigation sidebar. From this sidebar, click on the option labeled "Data Map". You can also go directly to the Data Map by clicking Data Map view. Here's how the initial Data Map view appears:



Once you are in the Data Map view, you will see a visual representation of data assets organized by domains. Here, you can browse through the available domains or use the search function to locate the specific Data Product you are looking for. When you find the Data Product, you can simply click on it to access more detailed information or to start working with the data. For example, if you select the AIFabric category in the Data Map view, you will see the following resulting folders:



If you click into a folder in the AIFabric domain, such as the GenAI Application Development folder, you will see the data models shown in the following image:



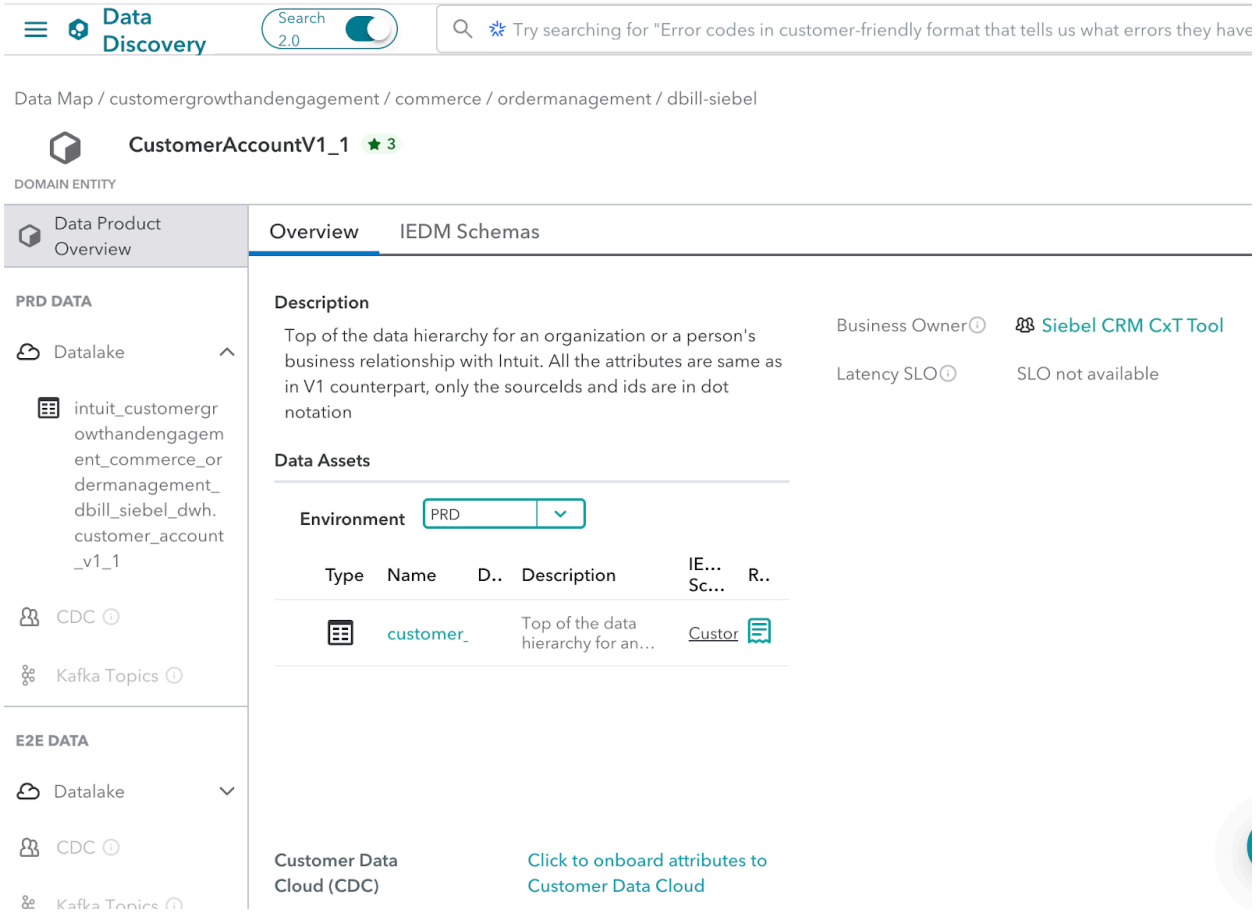
To navigate through a data model, continue clicking into nested folders to explore the associated data assets. Begin by selecting an initial domain, which will reveal any subdomains or bounded contexts linked to it. Keep drilling down through these subdomains until you reach a bounded context. Within each bounded context, you will find specific data products available for exploration.

Mature Data in Data Product View

Now that you've found the Data Product (or table) you are interested in, how do you navigate the views in Data Discovery? Select the CustomerAccountV1_1 Data Product from dBill Data Product as an example. The link opens in Data Discovery, where you will see two tabs:

- The **Overview tab**: Provides information about the Data Assets associated with this Data Product, as well as a description of the Data Product (which in this case explains why this version is different from the original Customer Account Data Product)
- The **IEDM Schemas** tab**:** Displays the current IEDM schema for the Data Product, as well as any past or draft versions of the schema, and allows you to:
    - **Link directly to the schema definition in Github**: Useful if you want or need to request that the producer make any changes to the IEDM definition.
    - **Navigate from the Data Product to associated Data Assets**: From within Data discovery you can view information on the Data Assets associated with the Data Product in either production or E2E.

Here's how the Overview tab appears in Data Discovery:



Mature Data in Data Asset View

If you navigate to the Data Asset view or access it directly, you will see information that is specific to the individual Data Asset. A Data Asset is a resource at Intuit that contains data. Each Data Product can be associated with multiple Data Assets. A Data Product is a collection of consumable data assets across one or more hosting locations, designed to represent the state of an important business domain in order to solve a well-defined business problem.

A real-time Paved Path Source Data Product would typically be associated with:

- A **Snapshot table** containing only the latest status of each asset
- An **Event History table** that logs all the events from the Event Bus topic for this Data Product
- A **Kafka Topic**, which refers to the topic on which the Domain Events are published

Review the Data Discovery documentation on Data for general information on the Data Asset view.

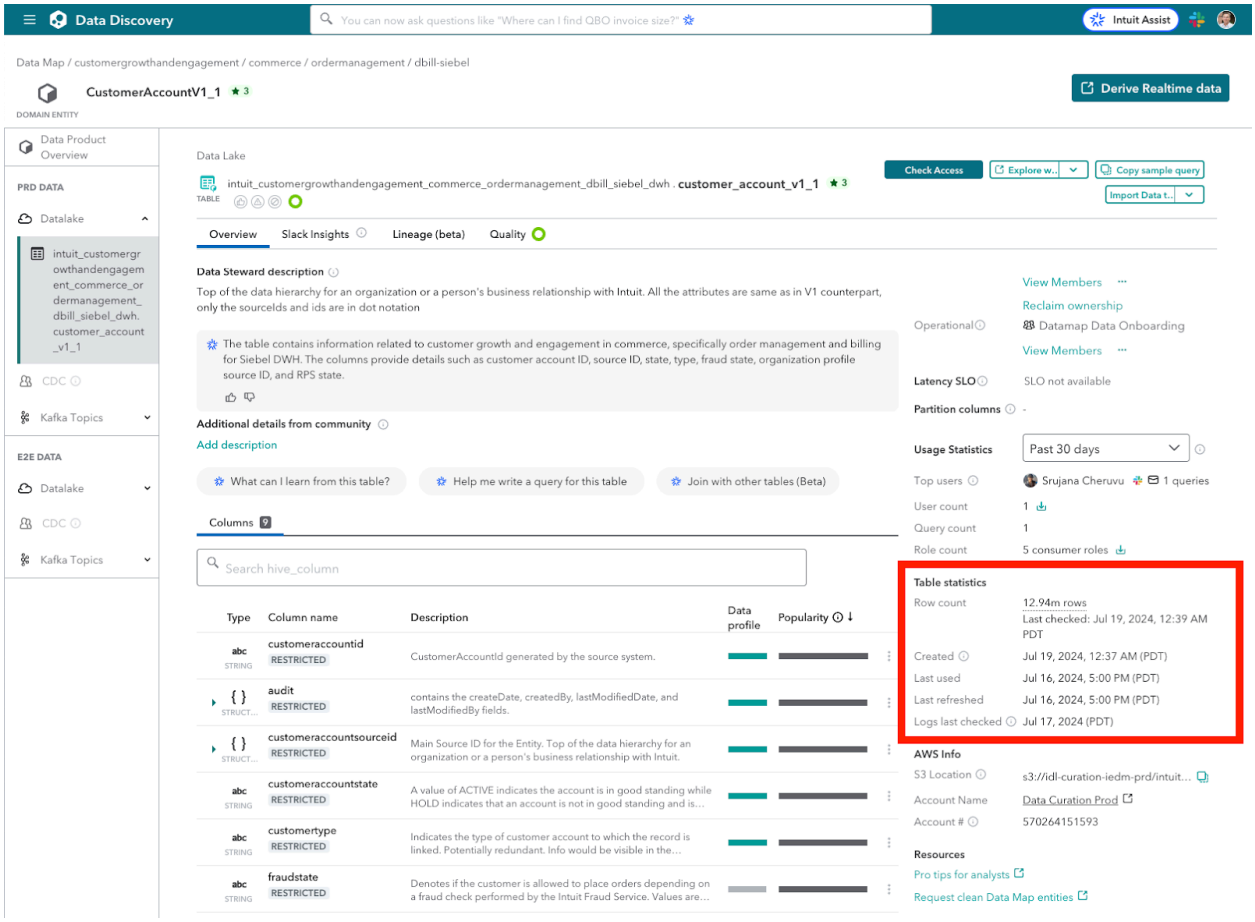
Note: To navigate from the Data Asset to the associated Data Product, from within the Data Asset view, you can click on Data Product Overview to return to or access the Data Product view.

Working with Data Asset View

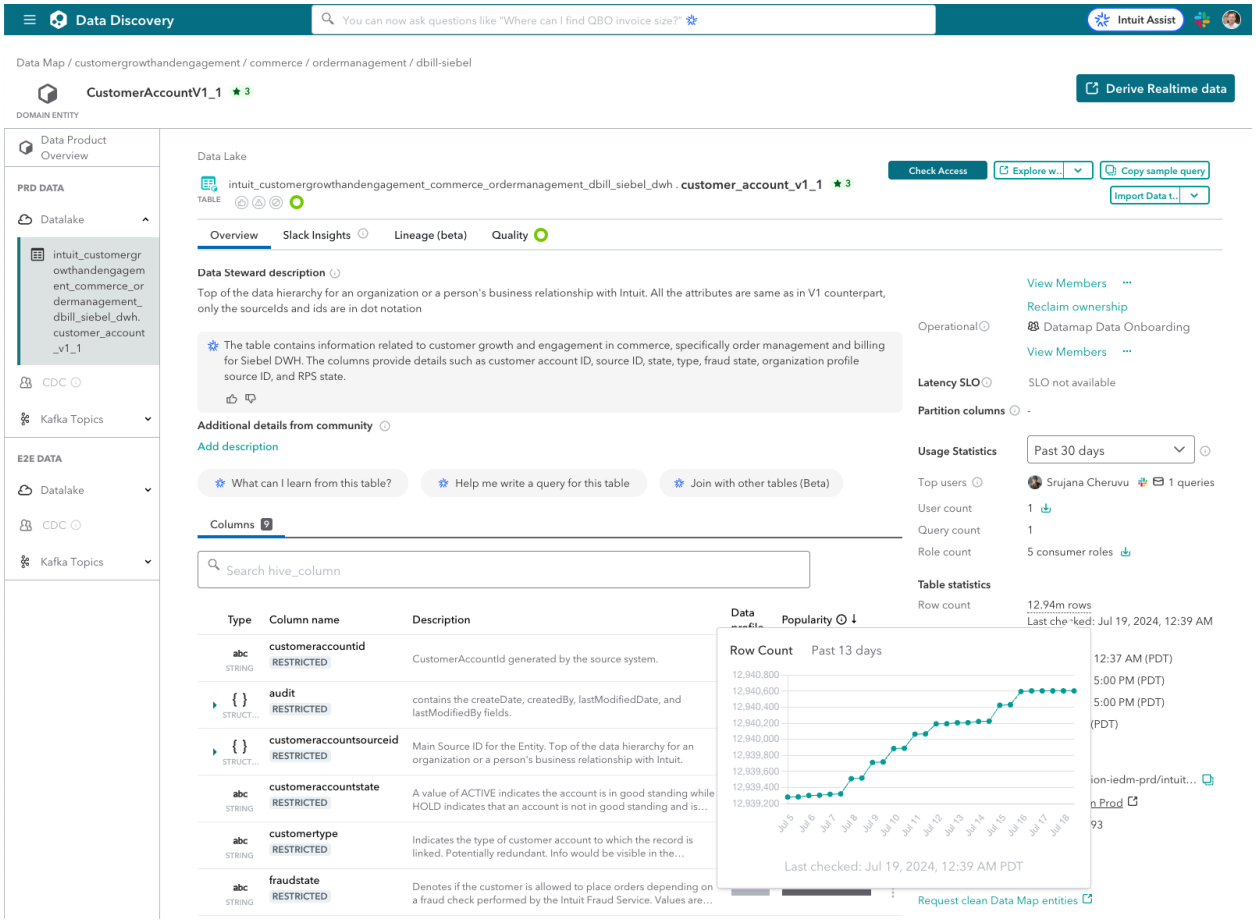
From within the Data Asset view, you can perform the following tasks:

- From within the Data Asset view, click on the Lineage tab to:
    - View both upstream and downstream lineage for the Data Asset
- Use the information in the Usage Statistics section to:
    - Understand who is running queries against the data asset
    - Determine if there are other roles that are accessing the Data Asset

The following image displays the Data Asset view:


Data Discovery
You can now ask questions like "Where can I find QBO invoice size?"
Intuit Assist
Data Map / customergrowthandengagement / commerce / ordermanagement / dbill-siebel
CustomerAccountV1_1
DOMAIN ENTITY
Derive Realtime data
Data Product Overview
PRD DATA
Datalake
intuit_customergrowthandengagement_commerce_ordermanagement_dbill_siebel_dwh.customer_account_v1_1
CDC
Kafka Topics
E2E DATA
Datalake
CDC
Kafka Topics
Data Lake
intuit_customergrowthandengagement_commerce_ordermanagement_dbill_siebel_dwh . customer_account_v1_1
TABLE
Check Access
Explore w..
Copy sample query
Import Data t..
Overview
Slack Insights
Lineage (beta)
Quality
Data Steward description
Top of the data hierarchy for an organization or a person's business relationship with Intuit. All the attributes are same as in V1 counterpart, only the sourceIds and ids are in dot notation
The table contains information related to customer growth and engagement in commerce, specifically order management and billing for Siebel DWH. The columns provide details such as customer account ID, source ID, state, type, fraud state, organization profile source ID, and RPS state.
Additional details from community
Add description
What can I learn from this table?
Help me write a query for this table
Join with other tables (Beta)
Columns
Search hive_column
Type
Column name
Description
Data profile
Popularity
abc STRING customeraccountid RESTRICTED CustomerAccountId generated by the source system.
{ } STRUCT... audit RESTRICTED contains the createDate, createdBy, lastModifiedDate, and lastModifiedBy fields.
{ } STRUCT... customeraccountsourceid RESTRICTED Main Source ID for the Entity. Top of the data hierarchy for an organization or a person's business relationship with Intuit.
abc STRING customeraccountstate RESTRICTED A value of ACTIVE indicates the account is in good standing while HOLD indicates that an account is not in good standing and is...
abc STRING customertype RESTRICTED Indicates the type of customer account to which the record is linked. Potentially redundant. Info would be visible in the...
abc STRING fraudstate RESTRICTED Denotes if the customer is allowed to place orders depending on a fraud check performed by the Intuit Fraud Service. Values are...
View Members
Reclaim ownership
Operational
Datamap Data Onboarding
View Members
Latency SLO
SLO not available
Partition columns
-
Usage Statistics
Past 30 days
Top users
Srujana Cheruvu
1 queries
User count
1
Query count
1
Role count
5 consumer roles
Table statistics
Row count
12.94m rows
Last checked: Jul 19, 2024, 12:39 AM PDT
Created
Jul 19, 2024, 12:37 AM (PDT)
Last used
Jul 16, 2024, 5:00 PM (PDT)
Last refreshed
Jul 16, 2024, 5:00 PM (PDT)
Logs last checked
Jul 17, 2024 (PDT)
AWS Info
S3 Location
s3://idl-curation-iedm-prd/intuit...
Account Name
Data Curation Prod
Account #
570264151593
Resources
Pro tips for analysts
Request clean Data Map entities

By mousing over the row count you can see how the size of the data has changed over time, as follows:

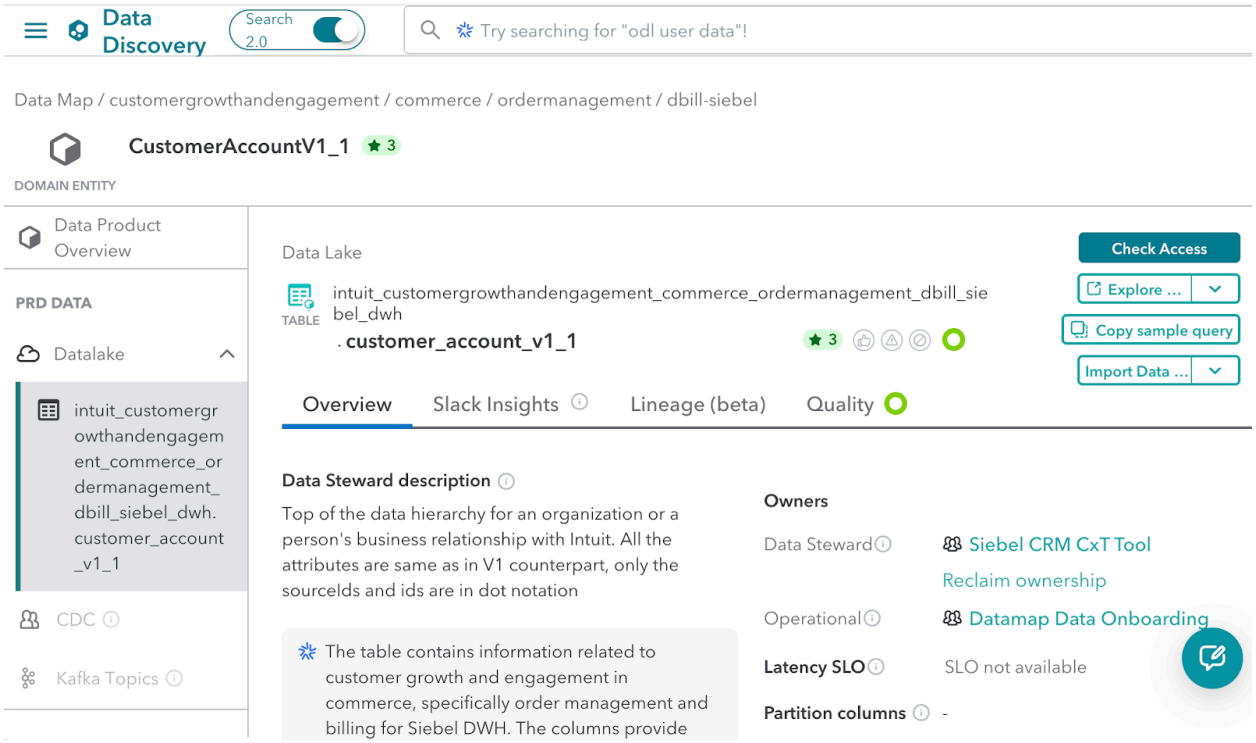


Accessing and Using Mature Data

After identifying the data you need, you can further access and understand it using the Exploration functions within the Data Discovery application. For a detailed overview of these functions, consult the Data Exploration documentation. Below are some helpful tips to get you started.

Checking/Requesting Access to Mature Data

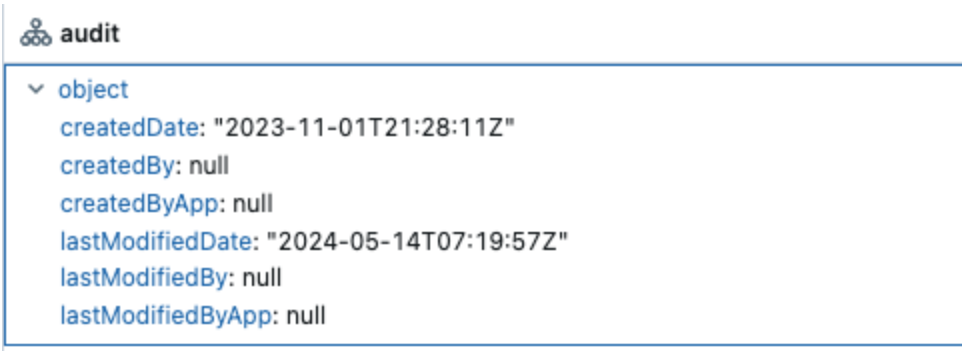
When you are looking at a Data Asset in the Data Discovery application, you will see a Check Access button in the top right corner, as shown in the following image, that you can use to check your access to a table or request access if you do not already have it. Click the button to get started.



Working with Flattened and Unflattened Data

When working with data maturity tables, you might encounter the terms "Flattened" and "Unflattened Data." These terms describe how the structures defined in the Intuit Enterprise Data Model (IEDM) Schema are represented as columns in the data lake. In essence, **unflattened data** stores each structure as a JSON object within a single column, while **flattened data** breaks down each element of the structure into its own individual column. The underlying data remains the same regardless of whether it's flattened or unflattened. However, as a data consumer, the key difference lies in how you query the data. With flattened data, you can directly reference the column, such as `` `SELECT audit_createdBy FROM ...` ``. In contrast, unflattened data requires you to qualify each level of the data structure, like `` `SELECT audit.createdBy FROM ...` ``. It's important to note that Data Discovery does not explicitly indicate whether a table is flattened or unflattened; you will need to review the table's schema to determine this. To accommodate both formats for the same data entity, you would need to create two separate storage structures or tables in the data lake: one for the flattened data and another for the unflattened data.

Here is an example of the Audit column from an unflattened table:



And here is an example of the same Audit column from a flattened table:

audit_createdBy	audit_createdByApp	audit_createdDate	audit_lastModifiedBy	audit_lastModifiedByApp	audit_lastModifiedDate
2-8F6S0E5	null	2012-05-05T00:00:00Z	1-4YPI	null	2024-06-03T00:00:00Z

Requesting Mature Data with the Accountability Dashboard

This section details the procedure for sending Jira tickets to data owners who possess data below the 3-star standard. The goal is to ensure all data assets meet Intuit's required data maturity standards. This process, vital for data integrity and reliability, will be monitored using the Accountability Dashboard and reviewed by leaders at TOS. The Data Success team will handle the distribution of tickets and monitor their progress.

Here is the process for requesting mature data with the Accountability Dashboard:

1. **Identification:** Jira tickets will be created for all data assets below the 3-star standard.
2. **Assignment:** Jira tickets will be assigned to the Dev Portal Manager responsible for the specific data table. This individual is responsible for taking necessary actions to upgrade the asset to meet the 3-star standard.
3. **Tracking:** Progress will be monitored via a Jira Dashboard. The dashboard should track:
    - # of tickets sent out, categorized by Level 0 (L0) organization
    - # of tickets resolved
    - Time taken against the Service Level Agreement (SLA)
4. **Service Level Agreement (SLA):**
    - FY25: SLA will be two (2) weeks
    - FY26 onwards: SLA will be one (1) business week

When creating Jira tickets to request mature data, keep in mind the following requirements:

- **Labels:** Labels such as "clean-data-score-remediation" and "capability name (L0 and L1)" will be used to track and categorize Jira tickets by organization.
- **Template:**
    - **Title:** Data Upgrade Required to Meet Intuit's 3-Star Standard

- **Labels:** (See above)
- **Description:** "The Accountability Dashboard has identified the following data asset that you own as not meeting the 3-star standard per the Intuit Data Maturity Standards ITAD. <asset details> The process for upgrading is straightforward through the promoted paved paths. You can refer to the attached document for detailed instructions on how to elevate your data to meet our required standards. Please take the necessary steps to upgrade your data asset to a minimum of 3 stars by <date>."