Necessity for Fine-Grained Access Controls

Overview

This document seeks to describe the rationale and need for fine-grained access controls. This document provides problem context, sample use cases, and scale.

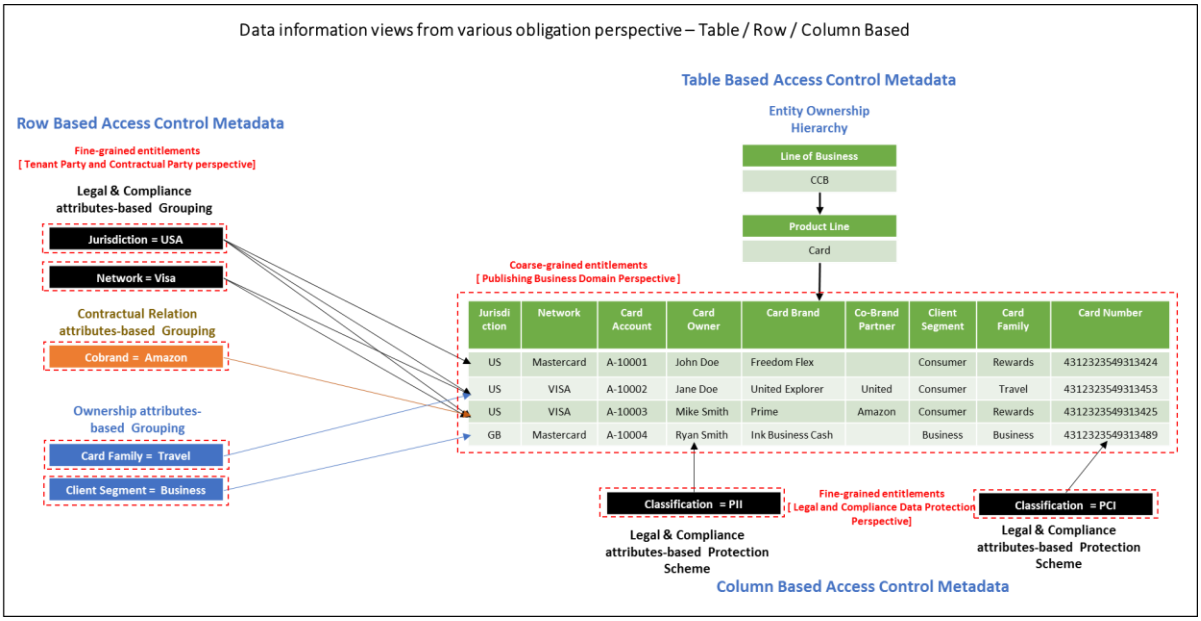
Problem Statement

One of the goals within CCB Data Technology is to provide a platform that supports a Lakehouse architecture, enabling multiple business functions and personas to operate against the same data. One of the challenges faced in achieving this goal has been the need to satisfy the data protection obligations associated with storing, processing, and consuming this data within the platform and across different interfaces. The origination of data from various sources, including internal products and external sources, brings with it data protection obligations associated to the various parties represented in the data.

Access control is one of the most critical aspects to meet these data obligations. From a physical storage perspective, data is organized based on application and product logical data models. There are two approaches to access control: *Coarse-grained access controls* describe a way to arrange resources in broad categories with access policies protecting at the resource level (i.e. database and table) – which does not consider the elements of the resource. As an alternative approach, *Fine-grained access controls* describe a way to control access to the components of the resource (i.e. column and row), where access control policies are written for accessing specific elements.

Data protection obligations originate from multiple sources and associated parties represented in the data:

- First party interactions from clients
- Data sourcing from third-party product partners (co-brands)
- Data sourcing from vendors (external)
- Data sourcing from regulatory agencies (public)



** Data can have multiple party claims that simultaneously apply based on the data content and the mechanism by which it is sourced between parties. Access controls need to be supported to protect against various party claims perspective, which can include table, column, row, and their cell intersection.*

Obligation Types

Contractual obligations

As data is sourced from various parties, it is collected based on various *Terms & Conditions* between source party and firm party. The terms and conditions dictate how the data can be used:

- Sharing within firm party across business functions, for specific uses only.
- Sharing between firm parties.
- Sharing with external parties.

Legal & Compliance obligations

There are various legal and compliance authorities, which require regulatory audits, minimum standards of control, and monitoring based on data falling within their purview:

- Jurisdictional authorities
    - Regional regulatory bodies.
    - Federal regulatory bodies.
    - State regulatory bodies.
- Industry regulators
    - Payment Card Industry Compliance.

Service & Operation's contexts

Service and operations personnel get access to data based on the need to perform specific job functions, and no more. The driver of their access is referred to as least privilege (minimal access scoped to perform the associated function). These personnel are typically assigned to a function with a specific scope:

- Function
    - Client servicing, Product servicing, Analytics etc.
- Scope
    - Private client, Partner segment, Jurisdiction etc.

Example Use-Cases

Contractual obligations: Co-brand Reporting

Context

JPMC engages in partnerships with affiliate organizations to offer co-branded products. These co-branded product offerings (multi-partner) include auto OEMs, credit and debit card products, and travel channels where transactions are made through affiliates that may or may not include Chase customers. Examples of co-branded product partners include Amazon Prime, Southwest Airlines, Subaru, and various others.

Challenge

Data is sourced from various parties and each party has different claims on the data that impact what that specific partner data can be used for and when it can be accessed. Each contractual agreement between partners, while similar in nature, is unique to that partner contract and carries with it specific constraints. Within Chase, the individual records containing claims from different partners exist within the same tables.

- Identification of in-scope records requires having one or many elements physically present in the data to be referred for logical segregation.
- Some partner records must be separately entitled as data can only be accessed when specifically supporting the partner relationship.
- Context of the intended use of the data must be understood as access to the data must be managed against specific use cases. Misuse of partner data carries significant risk for breach of contract.

Approach

The current solution is implemented on Teradata as row level security using views to secure access to rows for certain co-branded products. This approach is currently being repeated with replication of specific partner views in Snowflake. However, the current approach cannot scale to the number of combinations of access that are required across all co-brand scenarios where separate authorization is required.

Industry compliance: Payment Card Industry (PCI)

Context

The Payment Card Industry Data Security Standard (PCI DSS) is a set of rules and guidelines that protect credit card information and payment account data.

- Mandatory for all businesses that process, store, or transmit cardholder data.
- PCI DSS restricts use of cardholder data to only people and processes that specifically require PAN (Primary Account Number)/CVV (Card Verification Values) for performing their job.

Challenge

PAN data is pervasive within Chase as the card number is used as a key for identifying accounts and customers well beyond the Card domain. Due to the pervasiveness of cardholder data, analytics data stores are subjected to PCI compliance. This introduces a challenge for all systems interacting with cardholder data environments (CDE) as any environment that could potentially consume cardholder data needs to be brought into scope for PCI compliance.

Approach

To prevent bringing all consumption into scope for PCI compliance, a broker is used to segment environments interacting with CDE's. The responsibility of the broker is to ensure that any data passing from the CDE to out-of-scope environments does not contain any cardholder data. Today, this is achieved by leveraging column level access control policies through an enforcement layer around a catalog on top of the CDE.

- All data elements are identified as containing cardholder data via metadata (tags). The broker must assume that all data contains cardholder data unless explicitly indicated it does not.
- Column level policies (tag based) ensure that out-of-scope environments cannot consume cardholder data.
- The broker that governs access to the CDE must ensure access control policies properly segment in and out-of-scope environments, ensuring cardholder data does not traverse the boundary.
- The broker acts with zero trust, authorizing each data access at runtime and only providing temporary scoped down access to out-of-scope consumers.

Legal Jurisdictional: Luxembourg

Context

J.P. Morgan Bank Luxembourg (JPMBL) operates in a regulated jurisdiction and it is therefore subject to the "professional secrecy" requirements as set out in Luxembourg law. CSSF (The Commission de surveillance du secteur financier) market supervisor in Luxembourg, requires that financial institutions can demonstrate an appropriate protection and control of data entrusted to it by its clients for professional secrecy.

- Luxembourg data is the data of clients (both Investor and Institutional) contracted to the legal entity JPMBL.
- It is important to note that even usage of JPMC platforms via Interaffiliate agreements (IAS) is considered outsourcing in the context of European and Luxembourg Regulations.

Challenge

JPMBL data must be segregated from data pertaining to other JPMC legal entities in globally hosted systems to ensure that the appropriate level of local oversight is in place. The solution must have a provision to identify and "ring-fence" Luxembourg data. In context to the structure of the data residing in the same tables, controlling access to specific rows containing Luxembourg in-scope data requires segmenting access logically for separate authorization of access.

- In line with the principles of "need to know" and "least privilege", access to JPMSE Luxembourg "Confidential Information" (i.e. JPMSE Luxembourg client data, prospective client data, and/or Confidential Supervisory Information) must only be granted to persons whose functions so require, i.e. with a specific purpose and with strictly limited permissions and must be controlled by JPMSE Luxembourg employees.
- Identification of in-scope records requires having one or many elements physically present in the data to be referred for logical segregation.

Approach

The current solution is implemented on Teradata as row level security using views to secure the access to the rows related to Luxembourg entities. All tables in scope for Luxembourg have a row identifier column which is used to identify data related to Lux from others. Lux data is identified based on Lux BranchID. View logic segregates querying user's entitlement to restrict access to Luxembourg specific views of the data.

Legal Jurisdictional: California Financial Information Privacy Act (CalFIPA)

Context

The California Financial Information Privacy Act (CalFIPA) was enacted in 2003 to require financial institutions to provide California consumers notice and meaningful choice about how consumers' nonpublic personal information is shared. CalFIPA primarily applies to private entities in the financial sector (financial institutions) doing business in California that provide services to consumers who are California residents.

CalFIPA restricts the sharing of "non-public personal information" (NPI) of consumers. CalFIPA requirements apply with respect to all types of NPI, regardless of sensitivity and does not include special rules for particularly sensitive types. CalFIPA extends protection to NPI which is any personally identifiable information that is financial in nature about a consumer.

Challenge

Opt-out of sharing data with Affiliated Companies: JPMorgan Chase & Co is a multinational banking and financial services holding company consisting of multiple legal entities. Each of these legal entities are considered affiliated companies and sharing data across affiliates will be impacted by California residents who opt-out of this internal sharing.

- Identification of in-scope records requires having one or many elements physically present in the data to be referred for logical segregation.
- Every Table must include a corresponding View that segregates customers who have elected to opt out, in conjunction with other access controls that might apply to the same records.
- Multiple sets of roles to be properly governed and granted to right set of users.

Approach

The suppression of the data within the Wealth Data Mart is achieved by creating separate views. One which include customers in shared data other views which exclude customers from affiliate sharing. Both the views are protected with different sets of RBAC roles.

Scale Overview

Physical Scale

There are millions of resources in the data landscape that require fine-grained access controls:

- Hundreds of data producers
- Petabytes of data
- Hundreds of thousands of tables
- Millions of Columns
- Thousands of data consumer functions

At this scale, any scheme for managing access requiring an enumeration of physical resources is destined to be inadequate.

Technology Scale

There are thousands of use cases from a data consumption perspective. These use cases can bring dependencies on different tools and frameworks as the industry landscape is constantly evolving and changing to support varying use cases of data consumers. Our challenge is to support these technologies simultaneously with less friction to onboard, transition, interoperate, and offboard.

- Industry technologies
    - Storage formats – Structured data formats, unstructured data
    - Catalogs – Glue, Iceberg Rest, Unity, Polaris, Snowflake …
    - Compute Engines – Spark, Trino, AWS (multiple), Snowflake, Databricks, Teradata, ...
- Corporate Technologies
    - Atlas, JADE, …

Business Model Scale

Overall JPMC as a business system has a large amount of information, in which data can be categorized based on many dimensions:

- Data ownership attributes
    - Line of Business, sub-Line of Business etc.
    - Product, Market, Client Segment etc.
    - Third Party Origin Sources etc.
- Data protection attributes
    - Jurisdictional compliance
    - Industry compliance
    - Information Controls
- Actor (Users & Applications) membership attributes
    - Line of Businesses / sub-Line of Business etc.
    - Business / Product Function
- Access Entitlements Grouping as a Role
    - Service and Operations Functional Groups

Industry Solution Approaches

There are numerous approaches to managing and implementing fine-grained access controls. In each approach, access control policies need to be managed against relevant metadata that support appropriate policy authoring. This can achieve scalable policy management as well as enforcement across large numbers of resources.

Arranging physical data by access classification (Coarse-grained access control)

In this approach, physical data resources are aligned to a single data access class and possible combinations of data access classes are kept manageable by a central governance function, relying on consistent and broadly applicable categories.

- Advantages of this approach access class dimensions are kept simpler, are smaller in number, prevent conflicts and duplicate definition, and are commonly understood by people and systems.
- Disadvantages of this approach are the lack of flexibility in supporting narrow use cases, scaling to the granularity of access class dimensions, and adapting to frequently expanding business needs. This approach is not viable beyond a very small number of dimensions and access constraints.

Examples of this approach would be the following:

- Segregate data in a separate data store (table) based on the protection classification. This can be achieved by physical segregation (separate tables) or logically (separate views). Any user needing full access would refer to multiple separate resources for the entire set of data.

Apply access policy controls at runtime (Fine-grained access control)

In this approach, physical data resources can contain data with varying access classes and each data element has separate access attributes. At the time of access, the attributes of each element are evaluated against the access policy of the consumer. Additional optimization techniques can be implemented by mapping access policy to a combination of access attributes. Policy governance is needed to keep the possible policy combinations manageable.

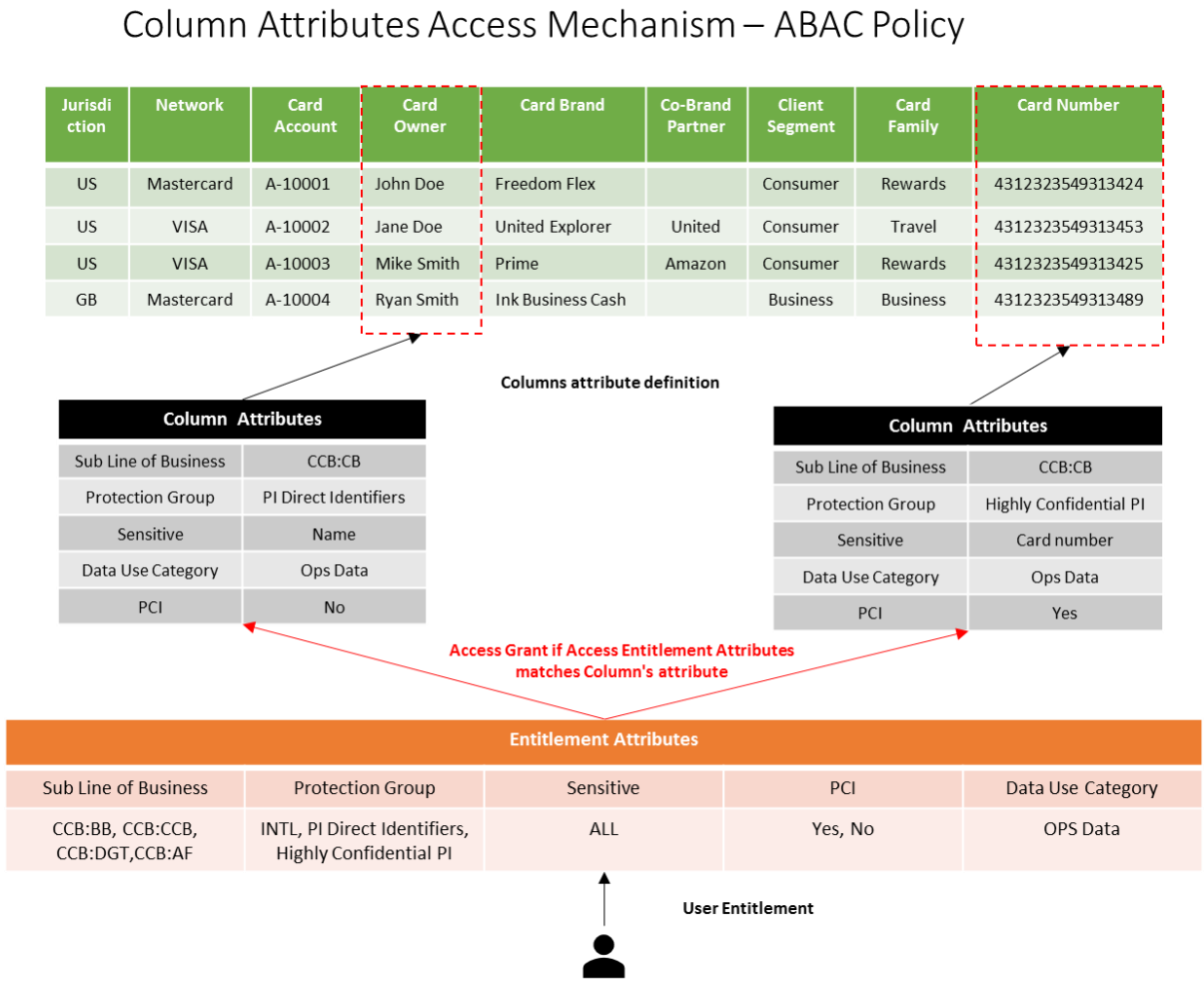
- Advantages of this approach are that access control policies can match the granularity of the obligation (accuracy). Policies can also be expressed using well defined attributes making policy authoring easier and is scalable.
- Disadvantages of this approach are the higher effort required to ensure that attributes, their attachment to resources, and the runtime evaluation by the engine can be coordinated. This approach is also challenging to ensure policies are consistent across different vendors and environments.

CCB Current Approach

These are high level feature highlights of CCB's current access control approach.

Fine-grained Access Control using Attribute Based policies

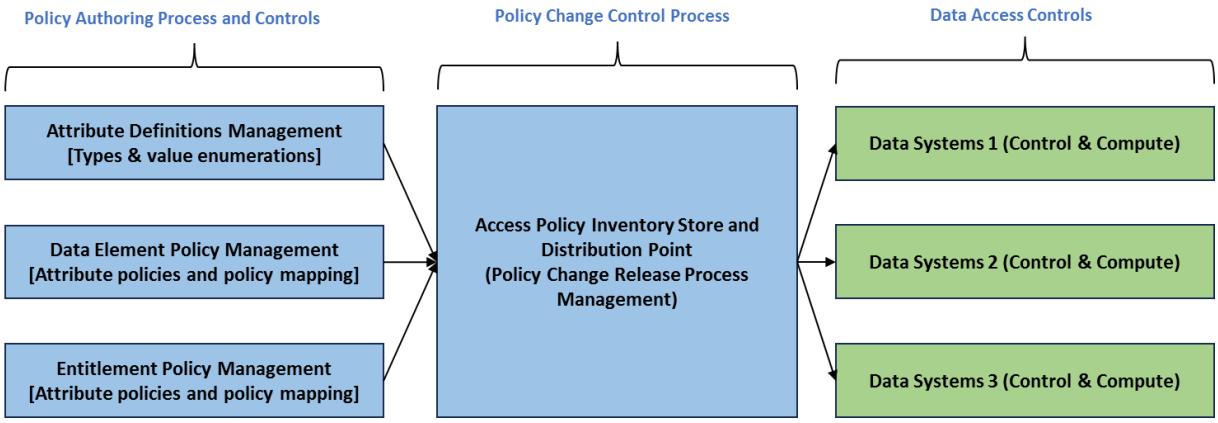
CCB's current approach is aligned to the fine-grained access control model for element level access using attribute-based access controls. Every data element has a definition based on standardized dimensions with subsequent attributes attached. Access control entitlements are also defined based on the same dimensions with each attribute being multi-valued. Access is granted to the accessor if the entitlement policy is satisfied by the attributes of the requested elements.



Decoupled policy authoring

With multiple data systems operating at any point of time and the evolution of firm choice within the general technology landscape, access policies should be externalized from enforcing data systems. This requires that policy authoring can be standardized to a format understood by multiple enforcement engines and subsequently enforcement engines are operating on standard policy constructs. This allows for polices to be leveraged across systems through machine interfaces with no expectation of human re-authoring or manual translation.

Policy Enforcement and Controls Systems



Access control policies carry information originating from different organizational units and systems. Across the policy authoring process, there are various pieces of policy information that need to be sourced, including:

- Attribute types and attribute value enumerations.
- Data element attributes and associated policies.
- Entitlement policies and policy mapping.

Prior to reaching the point of enforcement, data elements, their attributes, and resultant policies need to be managed through a policy change control process. The policy change control process, a system which manages the distribution of policies into data enforcement points, supports the following functions:

- Policy sourcing from authoring systems.
- Policy change impact assessment and notification.
- Policy review and release process.