DPP Data Extraction System: Services Overview

Nathan Lunceford

2025-03-06

Table of contents

Services Overview

This document outlines all the services required for our Distributor Data Extraction System, following the IDesign architecture pattern. Each service has a specific responsibility and

interacts with other services according to the defined communication rules.

Service Dependency Graph

Manager Layer

ExtractionManager

Responsibility: Orchestrates the complete data extraction workflow, from initiation to completion.

Key Functions:

- Coordinates the extraction process for both API and database sources
- Delegates specialized tasks to appropriate engines
- Manages the overall extraction state and progress
- Handles error scenarios and provides recovery mechanisms
- Ensures proper security context throughout the process

Engines Layer

ConfigurationEngine

Responsibility: Implements business logic for managing extraction configuration.

Key Functions:

- Determines which configuration parameters are needed for specific ERP types
- Processes raw configuration data into usable extraction settings
- Applies default values where needed
- Validates configuration completeness and correctness
- Manages configuration versioning and compatibility

ExtractorEngine

Responsibility: Implements business logic for extracting data from various sources.

Key Functions:

- Contains algorithms for different extraction methods
- Implements ERP-specific extraction strategies
- Manages data retrieval processes including pagination
- Handles extraction-specific error scenarios

- Optimizes extraction performance based on data characteristics

TransformerEngine

Responsibility: Implements business logic for transforming extracted data to standard formats.

Key Functions:

- Applies field mapping rules for standardization
- Converts data types to consistent formats
- Implements calculation rules for derived fields
- Manages data cleansing operations
- Handles special transformation cases by data type

ValidationEngine

Responsibility: Implements business logic for validating data quality.

Key Functions:

- Applies schema validation rules
- Implements business rule validations
- Checks data completeness and integrity
- Categorizes validation issues by severity
- Produces detailed validation reports

LineageEngine

Responsibility: Implements business logic for tracking data provenance.

Key Functions:

- Determines what metadata should be captured during extraction
- Implements rules for lineage record structure
- Processes lineage information for reporting
- Manages the relationship between extractions and data assets
- Provides data history and traceability

Accessors Layer

VaultAccessor

Responsibility: Provides secure access to credentials and secrets.

Key Functions:

- Retrieves credentials from HashiCorp Vault
- Manages secure credential handling
- Implements credential caching with appropriate expiration
- Applies least-privilege access principles
- Handles credential rotation and revocation

ConfigurationAccessor

Responsibility: Provides access to configuration storage.

Key Functions:

- Retrieves configuration data from FerretDB
- Stores updated configuration when needed
- Implements efficient querying for configuration data
- Manages configuration data caching
- Handles configuration database connection details

ERPApiAccessor

Responsibility: Handles communication with ERP APIs.

Key Functions:

- Manages HTTP connections to ERP API endpoints
- Implements authentication and authorization with APIs
- Formats requests according to ERP-specific requirements
- Processes API responses into standardized formats
- Handles API-specific error responses and recovery

ERPDatabaseAccessor

Responsibility: Manages direct database connections to ERP systems.

Key Functions:

- Establishes secure database connections
- Executes optimized queries for data extraction
- Manages connection pooling and resource cleanup
- Processes database result sets efficiently
- Handles database-specific error scenarios

StorageAccessor

Responsibility: Manages storage of extracted and processed data.

Key Functions:

- Uploads data to S3 storage in appropriate formats
- Organizes storage structure by client, ERP, and data type
- Implements efficient storage operations
- Manages storage permissions and security
- Handles storage-specific error scenarios

LineageAccessor

Responsibility: Manages persistence of lineage information.

Key Functions:

- Stores lineage records in the designated database
- Retrieves lineage information when needed
- Implements efficient lineage data querying
- Manages lineage database connections
- Handles lineage data relationships and integrity

Utilities Layer

SecurityUtility

Responsibility: Provides security services across the system.

Key Functions:

- Implements encryption and decryption operations
- Provides security validation functions
- Manages security logging and auditing
- Implements data masking for sensitive information
- Provides security-related helper functions

MonitoringUtility

Responsibility: Provides monitoring and observability services.

Key Functions:

- Implements metric collection and reporting
- Manages distributed tracing
- Provides health check mechanisms
- Implements logging standardization
- Offers performance monitoring tools