Department of Health Care Access and Information
Duty Statement

Employee Name Vacant	**Organization** Office of Technology Services Infrastructure and Data Operations Branch Data Operations Section
Position Number 441-175-1414-012	**Telework Option** Hybrid
Classification Information Technology Specialist II	**Working Title** Lead Data Engineer, Data Operations Section
Supervision Exercised None	**Location** Sacramento
Conflict of Interest: Yes ☒ No ☐	**Fingerprint/Live Scan:** Yes ☐ No ☒
Revision Date **September 2026**	**Effective Date**

Mission and Vision

HCAI is a leader in collecting data and disseminating information about California's healthcare landscape, promoting an equitably distributed health workforce, and publishing valuable information. The Department does this through five program areas - Affordability, Workforce, Data, Facilities, and Financing.

HCAI's mission is to expand access to quality, equitable, affordable health care for all Californians by supporting high-value delivery systems, resilient health facilities and workforces, and actionable health information and strategies.

HCAI's vision is a healthier California where all receive equitable, affordable, and quality health care.

General Description

Under the general direction of the Data Operations Section Manager, the Information Technology Specialist II (ITS II) serves as the Lead Data Engineer within the Office of Technology Services (OTS), Infrastructure Data Operations Branch (IDOB), Data Operations Section (DOS). This role provides senior-level technical leadership across HCAI's enterprise data platforms - including the Microsoft SQL Server-based Enterprise Data Warehouse (EDW), the Databricks Lakehouse (Apache Spark / Delta Lake) environment, and the Informatica PowerCenter and Intelligent Data Management Cloud (IDMC) ingestion platforms - to enable reliable, timely, and governed data movement across HCAI's systems.

The Lead Data Engineer leads end-to-end ingestion, validation, and governance, ensuring stable data transfer to support operational processes, analytics, reporting, regulatory obligations, and cross-program integration needs. As the application-tier owner of ETL/ELT platforms, the incumbent leads design, modernization, configuration integrity, performance optimization, release

and patch planning, permission-model management, workflow execution, and content lifecycle activities that keep data operations secure, resilient, and aligned with governance standards.

The incumbent serves as the engineering-tier owner of the Databricks environment - leading pipeline development, transformation logic, and lakehouse design - and coordinates closely with the Databricks platform administration/management function for backend and platform-tier activities, including workspace configuration, cluster and compute provisioning, platform security, and version upgrades. Similarly, the incumbent collaborates with the Infrastructure and Operations (I&O) teams and program teams to ensure underlying infrastructure, security controls, and integration requirements fully support application-tier capabilities, and advances enterprise modernization by guiding the evolution from traditional ETL toward a modern lakehouse/ELT paradigm.

Essential Job Functions

35% Enterprise Data Pipeline Engineering and Modernization

The Lead Data Engineer serves as the application-level owner for HCAI's ingestion and data-processing platforms across both the Microsoft SQL Server EDW and the Databricks Lakehouse. On the MS SQL EDW, the incumbent leads T-SQL development, stored procedures, SSIS packages (where applicable), indexing and partitioning strategies, and query/performance tuning. On Databricks, the incumbent leads Spark-based ingestion, Delta Lake and medallion (bronze-silver-gold) architecture design, notebook and job-cluster development, and both batch and streaming pipelines, coordinating with the Databricks platform and analytics administration/management function for backend platform provisioning, workspace configuration, and platform security. Using Informatica PowerCenter and IDMC as enterprise ingestion tooling, the incumbent leads the architecture, design, development, testing, deployment, and optimization of ETL/ELT pipelines that move, transform, and validate high-value data across HCAI systems. The Lead Data Engineer engineers robust error-handling frameworks, automated verification steps, idempotent recovery processes, and restart-safe sequences; modernizes ingestion for reliability, performance, transparency, and maintainability; and defines modular, reusable transformation patterns across staging, integration, EDW, and lakehouse environments. The incumbent collaborates with platform administrators to optimize compute sizing, storage formats, partitioning, caching, and workflow orchestration.

30% Data Validation, Quality, Governance and Controls

The Lead Data Engineer establishes and enforces validation and quality standards across both the SQL EDW and the Databricks lakehouse, including transformation logic, schema checkpoints, lineage capture, metadata requirements, and retention compliance. The incumbent leads data validation and reconciliation between the relational EDW and the Databricks lakehouse to ensure consistency across platforms. The incumbent governs the data catalog

and lineage spanning both environments (e.g., Unity Catalog for Databricks alongside EDW metadata), manages configuration integrity, workflow orchestration, service publishing, and upgrade planning, and administers pipeline-level permission models and workflow-execution rights to ensure secure, least-privilege access, coordinating with the Databricks platform administration/management function on backend platform-level security and access controls. The role includes overseeing content lifecycle for staging, integration, and lakehouse layers, ensuring continuous operational uptime, and working with the Infrastructure and Operations (I&O) and Information Security Teams to ensure application-tier operations adhere to departmental security standards and infrastructure requirements.

20% Collaboration, Integration, and Cross-Team Support

The Lead Data Engineer partners with program teams, data owners, platform administrators, enterprise architects, upstream system specialists, and security operations staff to align ingestion workflows with business, analytical, regulatory, and technical requirements across both the SQL EDW and the lakehouse. The incumbent ensures changes in source structures, schemas, interfaces, or update cadences are accurately translated into pipeline requirements, and coordinates ingestion sequencing, update windows, and transformation requirements across integrated relational and lakehouse systems. The role includes facilitating troubleshooting of ingestion anomalies, schema evolution, interface issues, and transformation mismatches, and ensuring platform configuration, permissions, and workflow execution remain properly coordinated with the Databricks platform administration/management function and with the infrastructure and security controls provided by the Infrastructure and Operations (I&O) Teams.

10% Leadership, Best Practices, and Continuous Improvement

The Lead Data Engineer provides senior-level leadership in establishing and advancing best practices for enterprise data engineering across relational and lakehouse platforms, and leads modernization from traditional ETL toward a lakehouse/ELT paradigm. The incumbent defines standardized approaches to pipeline design, workflow orchestration, metadata integration, validation logic, and transformation patterns for MS SQL, Databricks, and Informatica. The incumbent mentors engineering staff on both SQL and Spark/Databricks patterns - offering technical guidance, reviewing solution designs, validating transformation logic, and ensuring adherence to approved architectural standards - and promotes structured testing, peer review, CI/CD-aligned deployment, and automation. The role evaluates emerging methods and platform features, develops reusable components and templates, and drives continuous improvement through documentation, training, and collaborative engagement.

Marginal Job Functions

5% Performs related duties, special assignments, and program support activities consistent with the classification and level of responsibility of the position.

Physical Demands

Must possess and maintain sufficient strength, agility, endurance, and sensory ability to perform the duties contained in this duty statement with or without reasonable accommodation.

Working Conditions

Requires use of computing devices and phones, frequent face-to-face contact with management, staff, consultants, and the public, verbal, written, and digital (e-mail) communication, extensive review, analysis, and preparation of electronic and written documents, assessment of practical demonstrations, mobility to various areas of the Department, occasional travel and overnight stays to training/conferences or the Los Angeles field office may also be required, and work hours may deviate from core business hours based on the service requirements of the Department.

Employee Statement

I have reviewed and discussed the duties and responsibilities of this position with my supervisor and have received a copy of this duty statement.

Employee Name	Employee Signature	Date Signed

Supervisor Statement

I have discussed the duties of this position with them and have provided a copy of this duty statement to them.

Supervisor Name	Supervisor Signature	Date Signed