Project Aurora — Technical Specification

1. Executive Summary

Project Aurora is an internal data platform designed to unify three legacy systems: Hermes (customer records), Titan (transaction logs), and Orion (analytics dashboards). The project was approved by the engineering steering committee on January 15, 2026, with a total budget of $2.4 million and a target completion date of September 30, 2026.

The primary objective is to reduce data latency from the current average of 47 minutes to under 5 minutes for all downstream consumers. Secondary objectives include eliminating redundant ETL pipelines (currently 23 active pipelines across teams), standardizing the data schema to a single canonical model, and reducing infrastructure costs by 35%.

The project is led by principal engineer Maya Chen, with technical oversight from VP of Engineering David Park. The core team consists of 8 engineers, 2 data scientists, and 1 product manager. Three external contractors from Nexus Consulting will assist during the migration phase scheduled for Q3 2026.

2. System Architecture

Aurora uses an event-driven architecture built on Apache Kafka as the central message broker. All three legacy systems publish change events to dedicated Kafka topics. A stream processing layer built with Apache Flink consumes these events, applies transformations, and writes to a unified PostgreSQL database (version 16.2) serving as the system of record.

The PostgreSQL instance runs on AWS RDS with a db.r6g.2xlarge instance type, providing 8 vCPUs and 64 GB of memory. Read replicas are deployed in us-east-1 and eu-west-1 regions. Connection pooling is handled by PgBouncer with a maximum of 200 connections per replica.

2.1 Data Flow

The data flow follows a strict sequence: source system emits event, Kafka ingests with guaranteed at-least-once delivery, Flink processes with exactly-once semantics using checkpointing at 30-second intervals, and PostgreSQL receives the transformed record. Average end-to-end latency measured during load testing was 3.2 minutes under normal load and 4.8 minutes at peak (2x normal throughput).

2.2 API Layer

A RESTful API layer built with FastAPI (Python 3.12) serves as the primary access point for downstream applications. The API supports pagination, filtering, and field selection. Authentication uses OAuth 2.0 with JWT tokens issued by the internal identity provider. Rate limiting is set to 1000 requests per minute per client.

3. Migration Strategy

The migration from legacy systems to Aurora follows a three-phase approach. Phase 1 (February-March 2026) focuses on Hermes customer records. There are 12.4 million active customer records and 89 million archived records. The migration tool performs row-level checksums using SHA-256 to verify data integrity after transfer.

Phase 2 (April-June 2026) migrates Titan transaction logs. This is the largest dataset at 340 million records totaling 2.1 terabytes. The migration will run during off-peak hours (Saturday 2 AM to Sunday 6 AM EST) over 8 consecutive weekends. A rollback procedure has been documented that can restore the previous state within 4 hours.

Phase 3 (July-August 2026) decommissions Orion dashboards and replaces them with Aurora's built-in analytics module. All 47 existing Orion dashboards have been cataloged. Of these, 31 will be recreated in Aurora, 9 are deprecated with no replacement needed, and 7 require redesign based on feedback from the analytics team.

Each phase includes a two-week parallel-run period where both old and new systems operate simultaneously. During parallel run, automated comparison jobs validate that query results match within a 0.01% tolerance threshold.

4. Security and Compliance

Aurora must comply with SOC 2 Type II and GDPR requirements. All data at rest is encrypted using AES-256 with keys managed by AWS KMS. Data in transit uses TLS 1.3 exclusively. Database columns containing personally identifiable information (PII) are additionally encrypted at the application level using envelope encryption.

Access control follows the principle of least privilege. Four roles are defined: viewer (read-only access to non-PII fields), analyst (read access including PII with audit logging), engineer (read-write access to non-production environments), and admin (full access with mandatory two-person approval for production changes).

GDPR data subject requests (access, deletion, portability) are handled by a dedicated microservice called Guardian. Guardian processes deletion requests within 72 hours and generates a compliance certificate for each completed request. In 2025, the predecessor systems processed 4,200 deletion requests. Aurora's Guardian service is designed to handle up to 50,000 requests per year.

Penetration testing is scheduled for August 2026, conducted by CyberShield Partners. A bug bounty program will launch post-deployment with rewards ranging from $500 to $15,000 depending on severity classification.

5. Budget and Timeline

The total approved budget is $2.4 million, broken down as follows: infrastructure and cloud services account for $840,000 (35%), personnel costs including the core team and contractors total $1,200,000 (50%), software licensing fees are $192,000 (8%), and a contingency reserve of $168,000 (7%) is held for unforeseen expenses.

As of February 2026, $380,000 has been spent (16% of budget). The project is currently on schedule and $22,000 under the planned spend for this period. The primary cost risk identified is potential overtime during the Phase 2 transaction log migration, estimated at $45,000-$80,000 if weekend migrations encounter failures requiring extended windows.

5.1 Key Milestones

The six key milestones are: Hermes migration complete by March 31, Titan migration complete by June 30, Orion dashboard replacement complete by August 15, penetration testing passed by August 31, production launch on September 15, and legacy system decommission by September 30, 2026. Each milestone has a designated owner and a documented escalation path if the deadline is at risk.

5.2 Risk Register

Three risks are classified as high priority. First, Kafka cluster instability during peak migration could cause data loss; mitigation is pre-provisioned excess capacity at 3x expected load. Second, key personnel departure (specifically Maya Chen or David Park) would delay the project by an estimated 6-8 weeks; mitigation is documented knowledge transfer sessions recorded bi-weekly. Third, GDPR audit triggered before Guardian service is production-ready; mitigation is an interim manual process staffed by the compliance team.