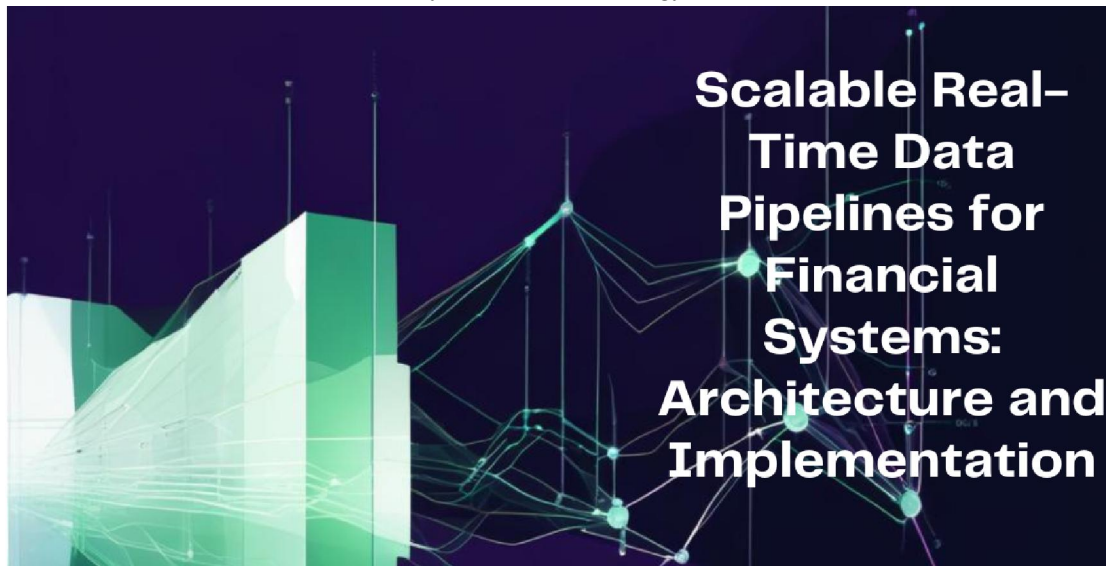


Scalable Real-Time Data Pipelines for Financial Systems: Architecture and Implementation

Anandan Dhanaraj

New Jersey Institute of Technology, USA



Abstract: *This article presents a comprehensive architectural framework for building high-performance data pipelines that meet the demanding requirements of modern financial systems. The article explores the complete pipeline lifecycle, from initial data ingestion through enrichment and storage to analytical consumption, with particular emphasis on achieving real-time processing capabilities while handling the volatile workloads characteristic of financial markets. The architecture employs canonical schemas to normalize diverse data sources, sophisticated enrichment processes to enhance raw transactions with contextual information, and optimized storage strategies utilizing sharded database implementations and distributed caching. The article incorporates automatic scaling mechanisms at both application and database tiers to maintain performance during market peaks without overprovisioning during quieter periods. Performance benchmarks demonstrate the architecture's capacity to process hundreds of thousands of transactions per second with sub-100ms latency, while our production implementation case study validates substantial improvements in processing capacity, system stability, and business outcomes. As financial data volumes continue to grow exponentially, this architectural pattern provides a scalable foundation that can evolve to incorporate emerging technologies and deliver increasingly sophisticated analytical capabilities.*

Keywords: Real-time Data Pipeline, Financial Transaction Processing, Canonical Schema Architecture, Scalable Data Enrichment, Cloud-native Database Scaling

