

**Dheeraj Kumar Vempaty**  
Senior Data Engineer, PayPal  
vempatydheeraj@yahoo.com  
+91-9941735947



---

## Professional Summary

*Senior Data Engineer with **12+** years of experience delivering enterprise-scale data solutions across Finance, HR, and Compliance domains. Specialized in architecting scalable pipelines, cloud-native platforms (GCP), and BI automation using Looker and Tableau. Proven record in driving \$20M+ cost savings, leading LLM-based analytics innovation, and enabling data-driven decision-making across global teams. Known for identifying root causes and developing innovative, scalable solutions. A quick learner who can adopt with new tech stacks*

---

### Technical Skills

**Cloud & Platforms:** Google Cloud (BigQuery, Looker, Notebooks, DataProc)

**ETL Tools:** Informatica, SQL/PLSQL, Python, Dell Boomi

**BI Tools:** Looker, Tableau, SAC Reporting

**Languages:** SQL, Python, Shell Scripting

**Version Control & Orchestration:** GitHub, Autosys, Apache Airflow, Control M, UC4

**Certifications:** Google Cloud Certified Professional Data Engineer, Data Architect

---

### Professional Experience

**PayPal Inc. – Chennai**

**Senior Data Engineer | Jan 2020 – Present**

### Projects –

**Market Insights Tool** – Developed a self-service reporting solution for Global Finance using Looker, enabling analysis of customer transaction data across global markets.

- Engineered and maintained Looker dashboards with **100+ dimensions/measures**, supporting **~80TB** daily data refreshes to deliver near real-time customer transaction insights.

### **LLM-Powered Finance Assistant (Beta) – Internal AI Tool Development**

- Ideation and development of an LLM-based assistant to streamline budgeting and forecasting processes for **Finance Analysts**, reducing reliance on multiple dashboards.
- Trained internally hosted **OpenAI model** by preparing all required KPIs data and sample Queries for expected FAs questions.
- Integrated the **conversational bot** within the **Looker Extension Framework**, enabling analysts to interact with real-time financial data directly from existing dashboards.
- Achieved significant **time(~3hrs per Analyst per week)** and **cost savings** by automating insights retrieval and enhancing data accessibility in a secure, self-hosted environment.

**Learning DataMart** – Built end-to-end data pipelines for employee L&D analytics, leveraging Airflow and REST APIs to unify data across platforms like Udemy, LinkedIn, Cornerstone and O'Reilly.

Highlights –

- Collaborated with external vendors to configure data integrations and build automated daily refresh pipelines using APIs and Apache Airflow.
- Integrated data from Udemy, LinkedIn, O'Reilly, and Cornerstone via vendor APIs, enabling leadership to assess employee engagement and ROI across platforms.
- Derived actionable insights on learning progress, certifications, compliance training, and platform utilization, supporting strategic L&D decisions across business units.

**Real Estate DataMart** – Built Real Estate DataMart from scratch to deliver insights on space utilization and hybrid work models, supporting leadership decision-making in post-pandemic office planning.

- Coordinated cross-functional teams like HR, Finance, Facilities & Procurement teams to define key data attributes for executive-level reporting solutions on real estate and occupancy metrics.
- Identified and implemented optimized reporting on Looker to deliver near real-time occupancy statistics for leadership dashboards.
- Built scalable pipelines processing ~200K of daily records across global PayPal offices, transforming raw data into reporting-ready tables.
- Enhanced the platform based on user feedback and industry best practices, replacing third-party vendor solutions and achieving ~\$20M in direct cost savings within 1<sup>st</sup> year of Go-Live.

**People DataWarehouse & People Analytics** – Single source of truth for all employee related data and reporting.

- Managed data pipelines for employee lifecycle events, including performance, compensation, onboarding, attrition, and HR case metrics, supporting company-wide People Analytics.
- Designed various data and reporting solutions to classify employees as hybrid or virtual based on profile attributes, improving People Analytics reporting.

**Tax DataMart** – Migration of data from legacy systems to GCP and redesigning the ETL flow.

- Took ownership of migrating 15+ TB of structured data from one datacenter to new DataCenter resulting in reduction of renewal costs(~\$8M) on old datacenter.
- Integrated and Automated U.S. tax reporting processes (1099K, CP2100, etc.) diverse data formats from acquired company datalakes (Venmo, Braintree, Honey, Zettle) into a unified Tax DataMart for cross-org compliance reporting.

#### **Key Responsibilities at PayPal –**

- Member of the **Finance Analytical Data Services** team, contributing to enterprise-wide data solutions for Finance and HR domains.
- Integrated data from diverse sources (APIs, flat files, external databases) using tools like Dell Boomi, Python scripting by secure data transfer and encryption methods.
- Built robust data applications for senior leadership to access actionable insights tailored for top-level strategic objectives.
- Experienced in deriving Analytics / Patterns from data.
- Partner with stakeholders (e.g., HR products, Engineering, Leadership) on the Finance & HR Organization to build and improve unique Finance & HR solutions.

- Identify key metrics, conduct rigorous explorative data analysis, create executive-facing dashboards, and build business cases that drive decision making and business prioritization.
- Delivered strategic insights to senior leadership through presentations and **data storytelling**.
- Define and cultivate best practices in analytics and evaluating new tools available in market.
- Managed concurrent projects in a high-speed, results-driven environment with tight deadlines.
- Led efforts to identify inefficiencies and standardize systems for improved stability and scalability.

**Company: Cognizant**

**Client: The Hartford Financial Services Group, Inc., –**

**Duration: Nov 2013 to Jan 2020**

**Role: Data Engineer**

**Database:** Oracle, MySQL

**Tools:** Informatica Power center Toad, Excel, Autosys, Putty, Rally, SVN, GitHub

**Languages:** SQL, PL/SQL, UNIX shell scripting, Python

#### **Key Responsibilities at Cognizant -**

- Engineered full-stack data warehouse solutions from scratch, including ingestion from heterogeneous sources, transformation, cleansing, business logic application, and delivery of reporting views for key business stakeholders.
- Designed and developed Type 1 and Type 2 slowly changing dimensions, fact tables, and star schemas to support historical and point-in-time reporting.
- Performed data cleansing and quality validation using custom checkpoints to ensure accuracy and consistency across ETL workflows.
- Created scalable and reusable data models tailored to performance and evolving business requirements, improving data architecture agility.
- Implemented ad-hoc and permanent solutions for critical batch job failures, minimizing downtime and improving SLA adherence.
- Designed and deployed Azure-based chatbots for high-volume insurance workflows, automating manual processes and saving significant operational costs annually.
- Developed enterprise-level Tableau dashboards to provide real-time visibility into data availability of Enterprise warehouse, enabling faster issue resolution for business users.
- Tuned SQL procedures, optimized ETL workflows, and refactored legacy jobs, cutting application runtimes from hours to minutes for data workloads.

#### **Education**

Sathyabama University, Chennai – B.E (Electronics and Instrumentation Engineering) – **75%**

Sri Vivekananda Jr College, Vijayawada - Intermediate Education – **87%**

City Central High School, Kodad - SSC Education – **90%**