

Prompt Engineering for Data Engineering : Unlocking Natural Language Access to Cloud Data Platforms

Jimish Kadakia - Solutions Architect, Snowflake

Conf42.com Prompt Engineering 2025 - November 6, 2025





The Evolution of Data Access

Traditional Approach

Data engineers and analysts have historically needed deep technical expertise to interact with cloud data platforms. SQL queries, script writing, and pipeline management required specialized knowledge that created bottlenecks and limited data accessibility across organizations.

The LLM Revolution

Large language models are fundamentally changing this paradigm. Natural language interfaces now enable users to interact with complex data ecosystems without extensive technical training, democratizing data access across the enterprise.



Beyond Chatbots: LLMs in Data Engineering

While most discussions of prompt engineering focus on conversational AI and content generation, the potential for data engineering applications is transformative. LLMs can bridge the gap between business users and technical data infrastructure.



Query Generation

Natural language converts to optimized SQL



Pipeline Orchestration

Trigger workflows through conversation



Real-Time Analytics

Instant insights without technical barriers

The Core Components of Data-Focused Prompt Engineering

01

Context Optimization

Provide the LLM with schema information, business logic, and data relationships to ensure accurate query generation and pipeline management.

03

Validation Layers

Implement safeguards to verify generated queries and pipeline commands before execution, protecting data integrity.

02

Prompt Design Patterns

Develop reusable templates that guide users toward successful interactions while maintaining consistency across the data platform.

04

Feedback Loops

Continuously refine prompts based on user interactions and model performance to improve accuracy over time.

Democratizing Data Access

The most powerful outcome of applying prompt engineering to data platforms is democratization. Business users who previously depended on data teams for every query can now independently explore datasets, generate reports, and derive insights.



Business Users

Ask questions in plain language



LLM Interface

Translates to technical operations



Cloud Data Platform

Executes and returns results



Real-World Applications

Ad-Hoc Querying

Users describe the data they need in natural language, and the system generates optimized queries against cloud data warehouses.

Example: "Show me customer retention rates by region for Q4"

Pipeline Management

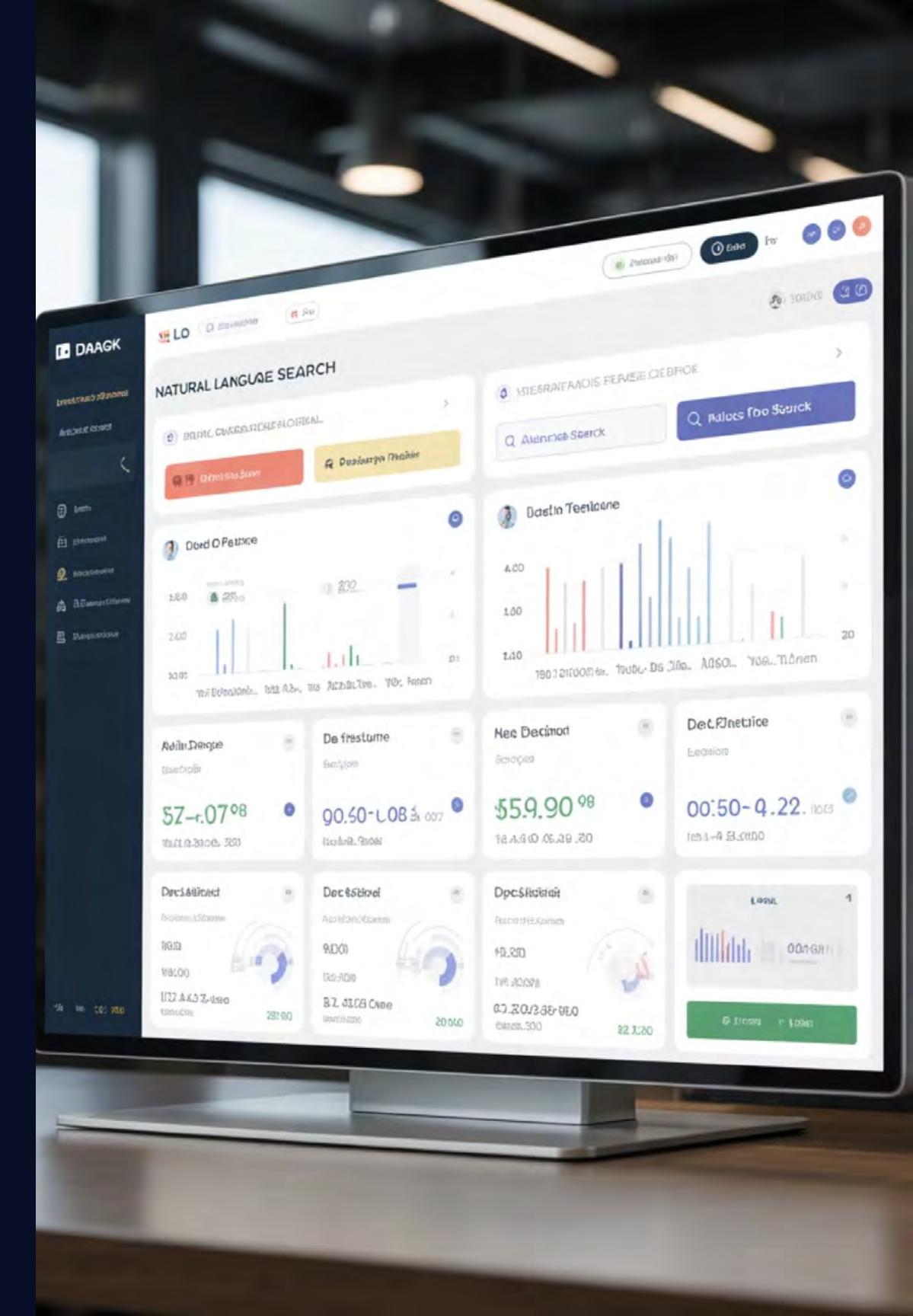
Trigger ETL processes, schedule jobs, and monitor data flows through conversational commands.

Example: "Run the nightly customer aggregation pipeline"

Data Exploration

Discover relationships, identify patterns, and generate hypotheses without writing complex analytical code.

Example: "What factors correlate with high-value purchases?"



Critical Challenges to Address

Accuracy and Hallucinations

LLMs can generate plausible but incorrect queries or misinterpret business logic. Rigorous validation and testing frameworks are essential to prevent erroneous data operations.

Data Governance and Security

Natural language interfaces must respect existing access controls, data privacy regulations, and compliance requirements. Permission models need careful integration.

Model Fine-Tuning

Generic LLMs require customization for enterprise data environments. Fine-tuning on domain-specific schemas and business terminology improves performance significantly.

Performance at Scale

Complex queries and large datasets require optimization strategies to maintain responsive user experiences without overwhelming compute resources.

Governance: The Foundation of Trust

Successfully deploying LLM-powered data interfaces requires robust governance frameworks. Without proper controls, the accessibility benefits can introduce significant risks to data security and integrity.

Access Control

Row-level security and column masking enforced automatically

Audit Trails

Complete logging of all LLM-generated queries and operations

Approval Workflows

Critical operations require human review before execution



Implementation Strategy



Start with Limited Scope

Begin with **read-only queries** on non-sensitive datasets to build confidence and refine prompts.



Collaborate Across Teams

Data engineers, business analysts, and security teams must work together to define requirements and guardrails.



Test Extensively

Validate model outputs against known queries and edge cases before expanding access.



Scale Incrementally

Gradually expand capabilities based on user feedback and performance metrics.



The Future of Data Interaction

As LLMs continue to evolve, the boundary between business users and data platforms will blur further. Prompt engineering represents a fundamental shift in how organizations think about data accessibility and technical expertise.

Faster Decision-Making

Eliminate the wait time for data team support. Business users get answers when they need them, accelerating decision cycles.

Smarter Data Practices

More people interacting with data means more questions asked, more insights discovered, and more data-driven culture.

Key Takeaways



Prompt engineering extends far beyond chatbots

It is a powerful tool for democratizing access to complex cloud data ecosystems



Start small, test rigorously, and scale incrementally

Build trust and confidence before expanding capabilities



Success requires careful attention to accuracy, governance, and fine-tuning

These challenges are solvable with proper frameworks and collaboration



The future of data interaction is conversational

Organizations that embrace this shift will empower teams and accelerate innovation

Thank you!

v

Jimish Kadakia

Conf42.com Prompt Engineering 2025 – November 6, 2025