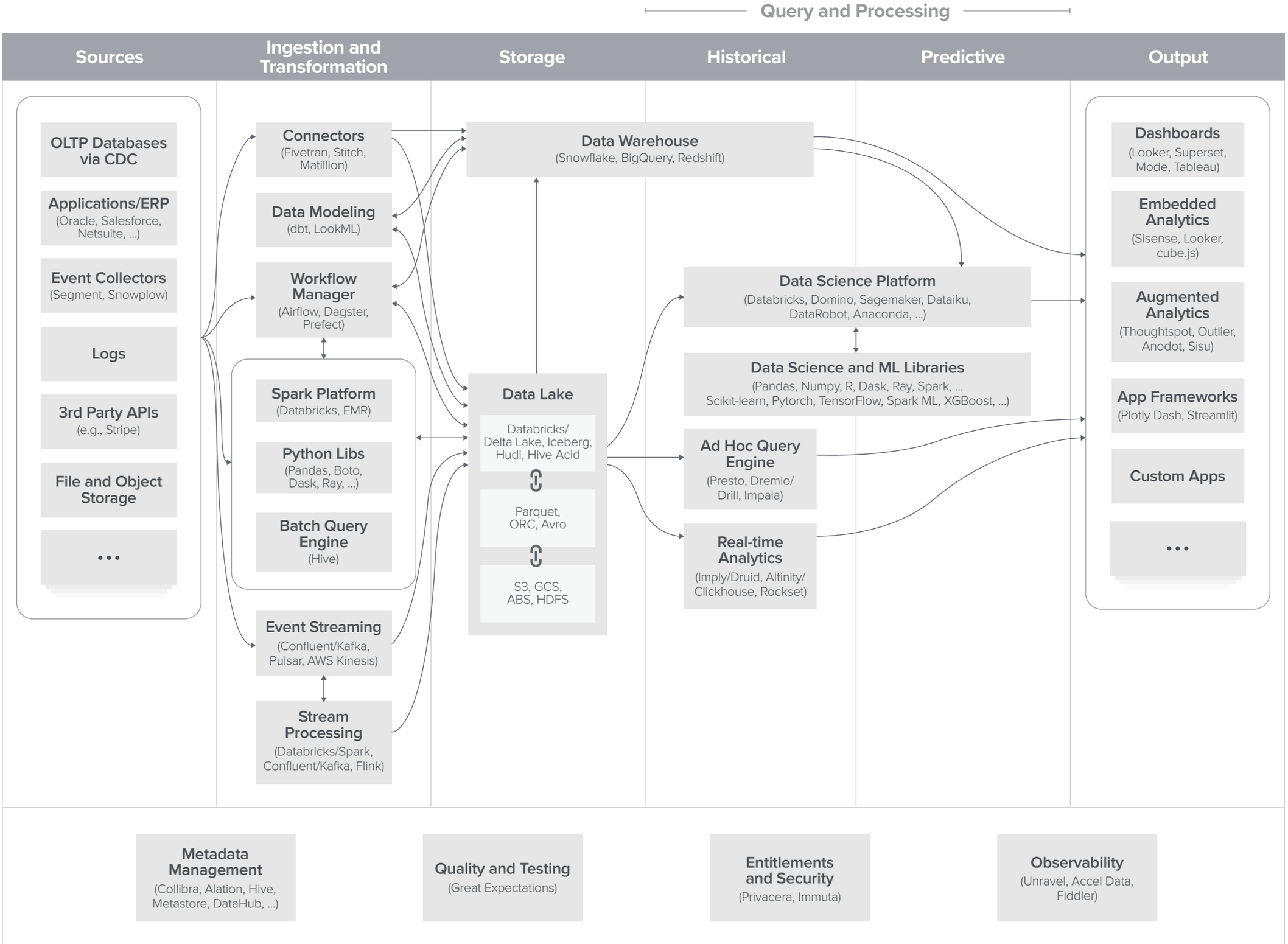
A Unified Data Infrastructure Architecture



Interpreting the Architecture

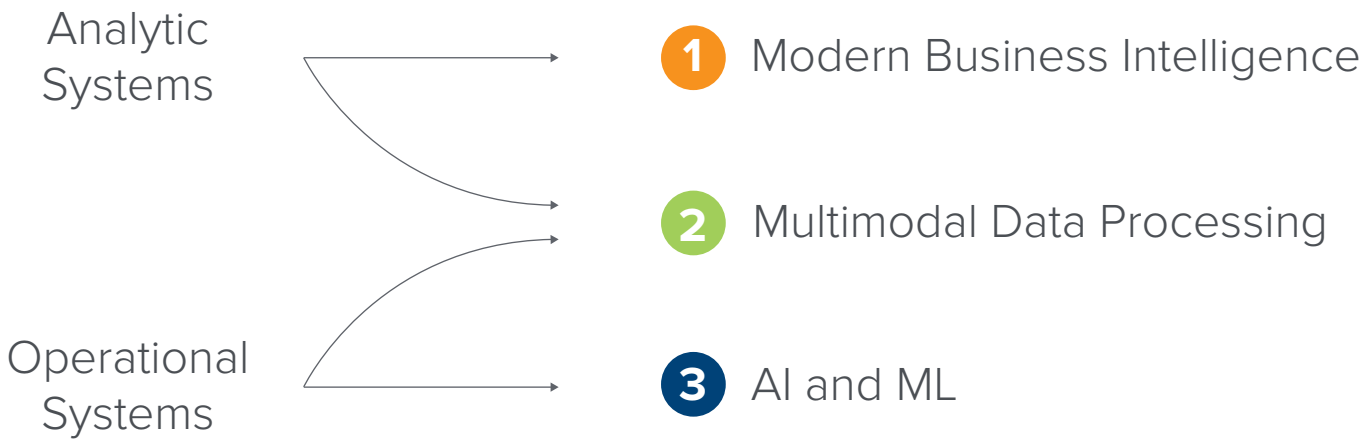
Query and Processing

Sources	Ingestion and Transformation	Storage	Historical	Predictive	Output
Generate relevant business and operational data	Extract data from operational systems (E) Deliver to storage, aligning schemas between source and destination (L) Transform data to a structure ready for analysis (T)	Store data in a format accessible to query & processing systems Optimize for low cost, scalability, and analytic workloads (e.g., column store) In some cases, provide additional data structures or guarantees	Provide an interface for analysts and data scientists to derive insights (query) Execute queries and data models against stored data, often using distributed compute (processing)  Describe what happened in the past (including very recent past)	Predict what will happen in the future Build data-driven/ ML applications	Present results of data analysis to internal and external users Embed data models into operational systems and applications

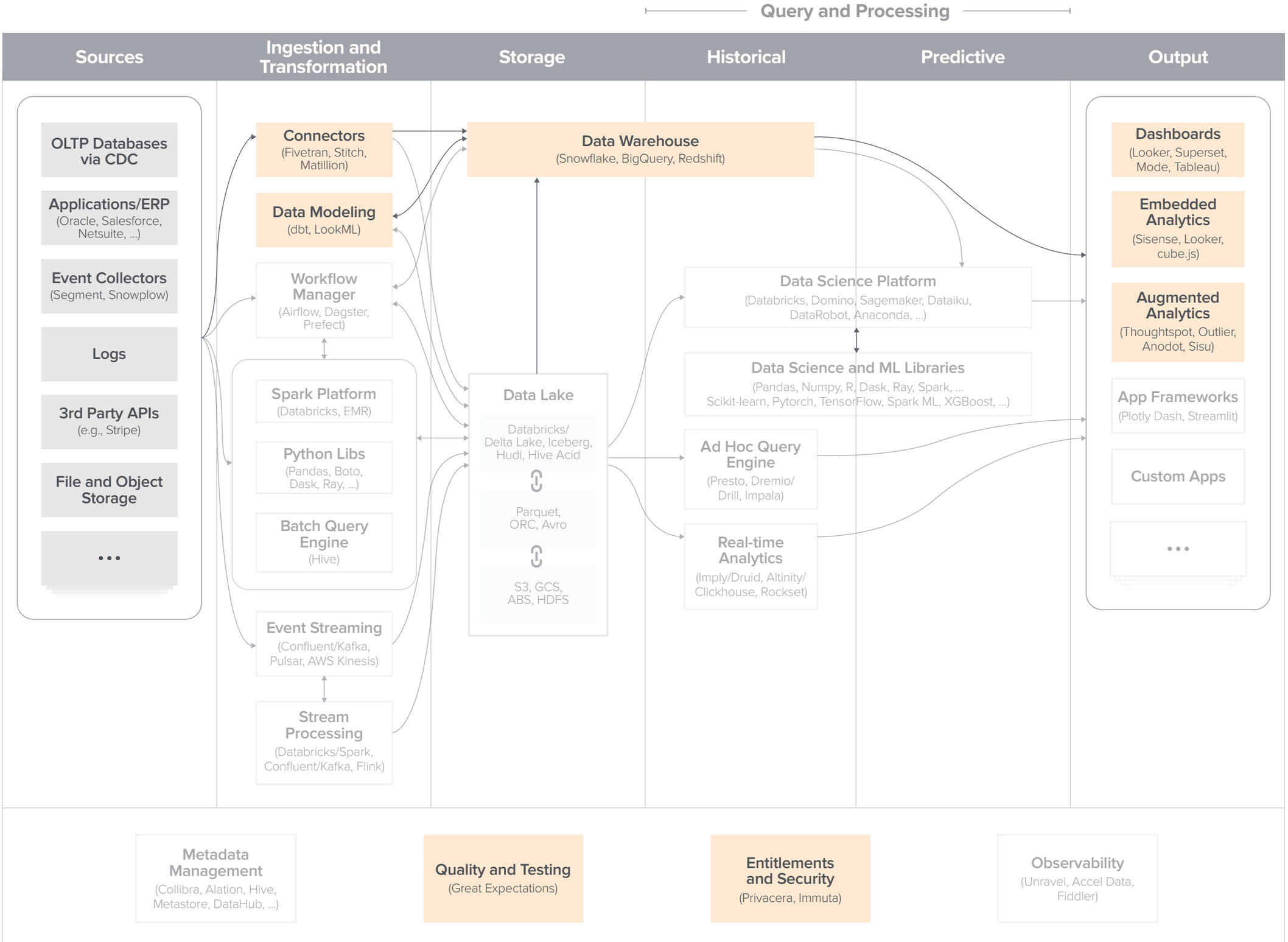
Coordinate the flow of data and the execution of computations across the full lifecycle

Ensure proper data quality, performance, and governance of all systems and datasets

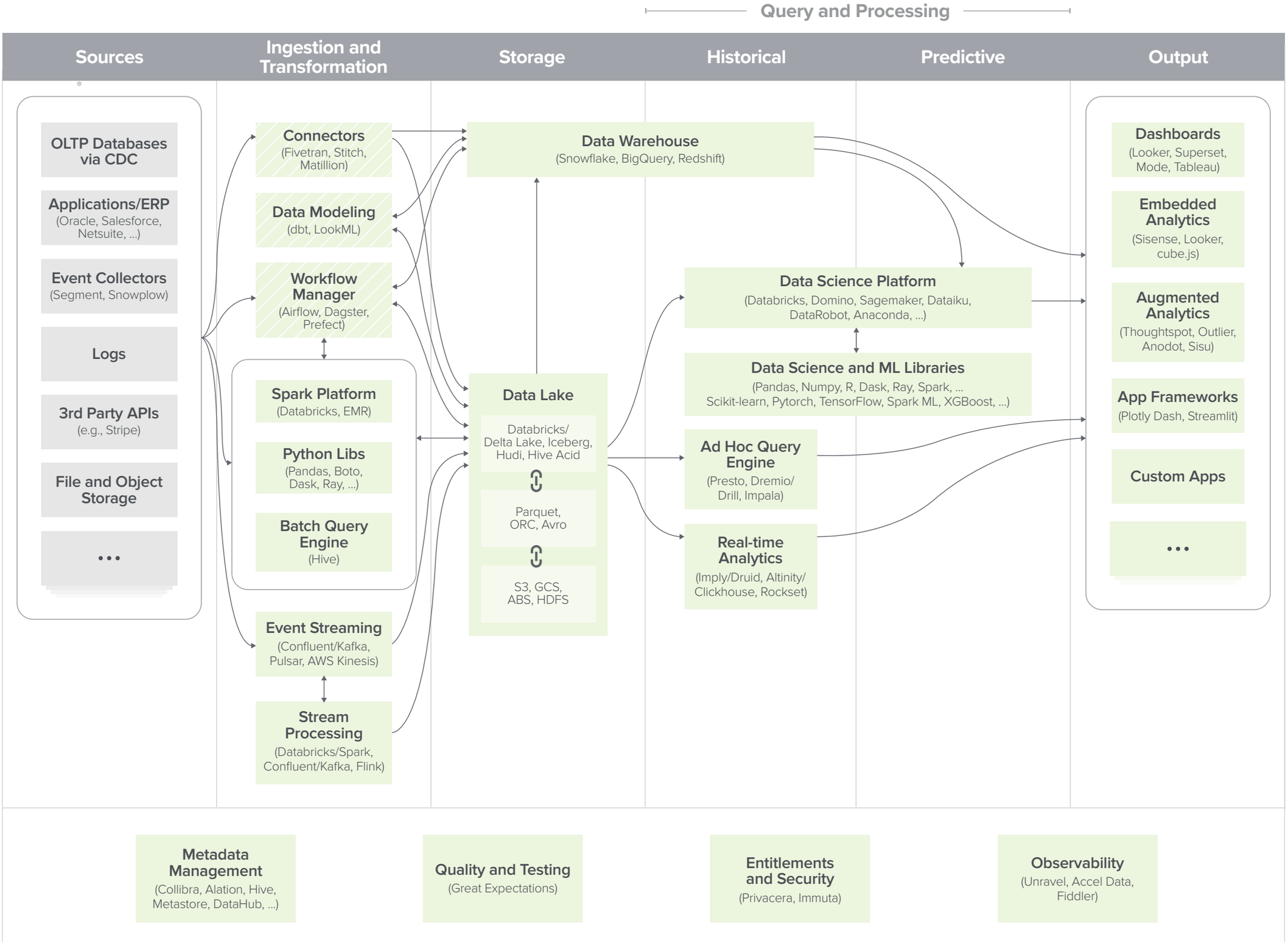
Three Common Blueprints



1. Modern Business Intelligence Blueprint



2. Multimodal Data Processing Blueprint



3. AI and ML Blueprint