The Big Picture for an Analytics Platform

(Note: Does not include the Prescriptive Column yet)

Query and Processing Ingestion and Historical Sources Storage **Predictive** Output **Transformation Dashboards** Connectors **Data Warehouse OLTP Databases** (Fivetran, Stitch, (Looker, Superset, via CDC (Snowflake, BigQuery, Redshift) Matillion) Mode, Tableau) Applications/ERP **Embedded Data Modeling** (Oracle, Salesforce, Analytics (dbt, LookML) Netsuite, ...) (Sisense, Looker, cube.js) **Event Collectors** Workflow **Data Science Platform** (Segment, Snowplow) Manager Augmented (Databricks, Domino, Sagemaker, Dataiku, (Airflow, Dagster, Prefect) DataRobot, Anaconda, ...) **Analytics** (Thoughtspot, Outlier, Anodot, Sisu) Logs **Data Science and ML Libraries** (Pandas, Numpy, R, Dask, Ray, Spark, ... **Spark Platform** Data Lake App Frameworks Scikit-learn, Pytorch, TensorFlow, Spark ML, XGBoost, ...) (Databricks, EMR) 3rd Party APIs (Plotly Dash, Streamlit) (e.g., Stripe) Databricks/ Delta Lake, Iceberg, Ad Hoc Query **Python Libs** Hudi, Hive Acid Engine (Pandas, Boto, **Custom Apps** File and Object (Presto, Dremio/ Dask, Ray, ...) Drill, Impala) Storage Parquet, ORC, Avro **Batch Query** Engine Real-time . . . (Hive) **Analytics** (Imply/Druid, Altinity/ S3, GCS, Clickhouse, Rockset) ABS. HDFS **Event Streaming** (Confluent/Kafka, Pulsar, AWS Kinesis) Stream Processing (Databricks/Spark, Confluent/Kafka, Flink) Metadata **Entitlements** Observability **Quality and Testing** Management and Security (Unravel, Accel Data, (Great Expectations) (Collibra, Alation, Hive, Fiddler) (Privacera, Immuta) Metastore, DataHub, ...)

Architectural Themes

(Note: Does not include the Prescriptive Column yet)

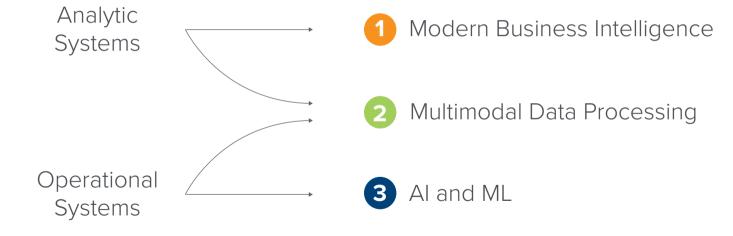
——— Query and Processing ———

Sources	Ingestion and Transformation	Storage	Historical	Predictive	Output
Generate relevant business and operational data	Extract data from operational systems Deliver to storage, aligning schemas between source and destination Transform data to a structure ready for analysis	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	to derive in: Execute queries and da	analysts and data scientists sights (query) ta models against stored sted compute (processing) 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

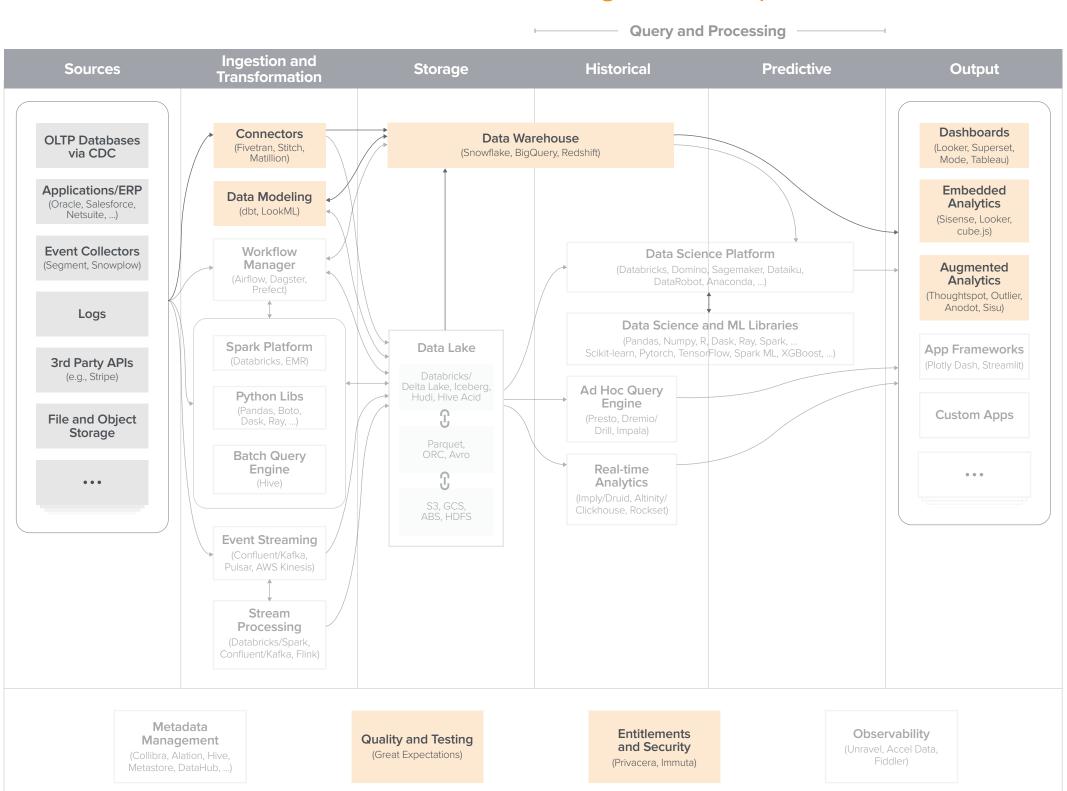
Coordinate the flow of data and the execution of computations across the full lifecycle

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

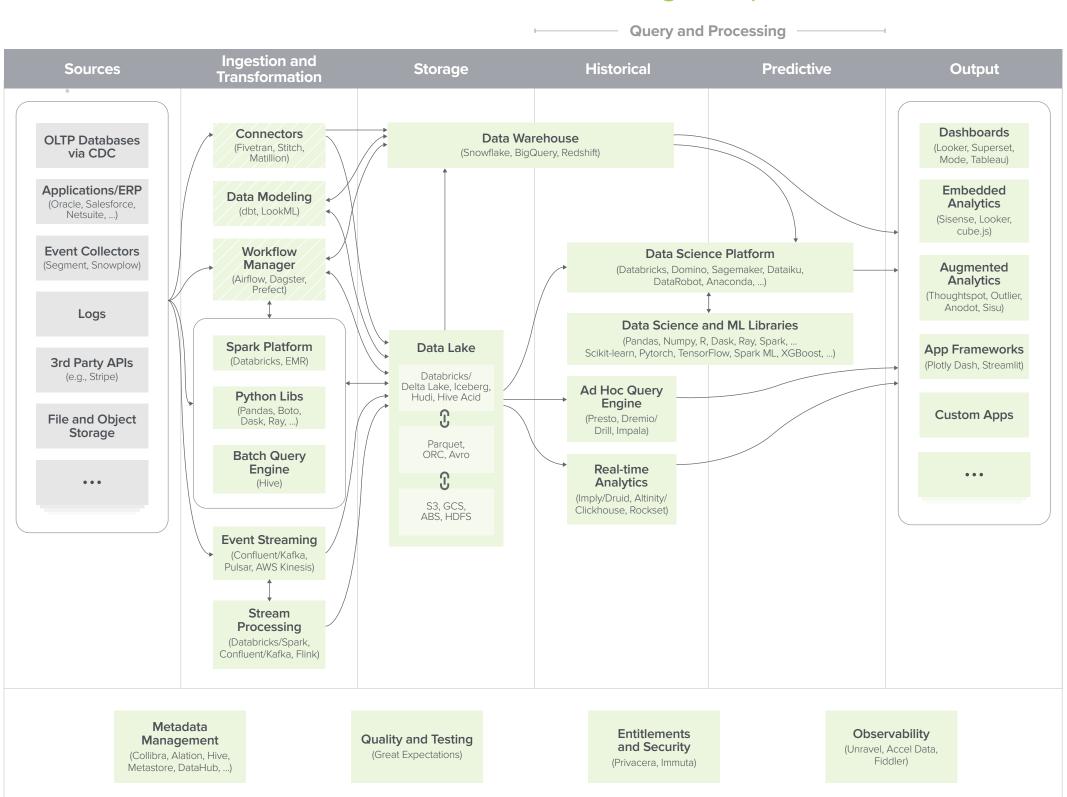
The Common Blueprints



1. Modern Business Intelligence Blueprint



2. Multimodal Data Processing Blueprint



3. Al and ML Blueprint

