

# State of Data Engineering 2024

[Explore the table](#) ↓

Presented by  lakeFS



INGEST	DATA LAKE	DATA VERSION CONTROL	COMPUTE ENGINES	PIPELINES	PRACTITIONER APPS	GOVERNANCE
<p><b>Ingest Tech</b></p> <ul style="list-style-type: none"><li>Google Cloud Pub/Sub</li><li>APACHE STORM</li><li>amazon KINESIS</li><li>beam</li><li>kafka</li><li>StreamNative</li><li>SPARK</li><li>nifi</li><li>Flink</li><li>Informatica</li><li>Redpanda</li><li>BENEATH</li><li>memphis.dev</li><li>Azure Data Factory</li><li>WarpStream</li><li>PULSAR</li><li>Airbyte</li><li>CONFLUENT</li><li>meltano</li><li>KASKADA</li><li>upsolver</li></ul> <p><b>Ingest SaaS</b></p> <ul style="list-style-type: none"><li>Airbyte</li><li>decodable</li><li>POLYTOMIC</li><li>Fivetran</li><li>twilio segment</li><li>Dataddo</li><li>rudderstack</li><li>Stitch</li><li>datacoral</li><li>SNOWFLOW</li><li>TROCCO</li><li>Keboola</li><li>Rivory</li><li>MATILLION</li></ul> <p><b>Reverse ETL</b></p> <ul style="list-style-type: none"><li>Census</li><li>hightouch</li><li>HeadsUp</li><li>POLYTOMIC</li><li>SEEKWELL</li><li>Grouparoo</li><li>omnata</li><li>rudderstack</li><li>Octolis</li><li>TROCCO</li></ul>	<p><b>Object Storage</b></p> <ul style="list-style-type: none"><li>Microsoft Azure Blob Storage</li><li>Google Cloud Storage</li><li>amazon S3</li><li>ORACLE CLOUD</li><li>IBM Cloud</li><li>Alibaba Cloud</li><li>hadoop</li><li>MINIO</li><li>zadara</li><li>CloudFiles</li><li>DigitalOcean</li><li>CLOUDFLARE</li><li>ceph</li><li>wasabi</li><li>PURESTORAGE</li><li>StorageGRID</li><li>VAST</li><li>STORJ</li></ul>	<p><b>lakeFS</b></p> <ul style="list-style-type: none"><li>xethub</li><li>Nessie</li><li>DVC</li><li>OXEN.AI</li><li>Underhive</li><li>Git LFS</li><li>Pachyderm</li></ul> <p><b>METADATA</b></p> <p><b>Metastores</b></p> <ul style="list-style-type: none"><li>databricks Unity Catalog</li><li>Azure Purview</li><li>Tabular</li><li>Cloud Dataproc</li><li>AWS Glue</li><li>CLOUDERA</li><li>OpenHouse</li><li>Gravitino</li></ul> <p><b>Open Table Formats</b></p> <ul style="list-style-type: none"><li>ONEHOUSE</li><li>Apache hudi</li><li>ICEBERG</li><li>DELTA LAKE</li><li>orc</li></ul>	<p><b>Distributed Compute</b></p> <ul style="list-style-type: none"><li>databricks</li><li>CLOUDERA</li><li>Cloud Dataproc</li><li>SPARK</li><li>RAY</li><li>hadoop</li><li>ASCEND.IO</li><li>Google DataFlow</li><li>dask</li><li>Azure HDInsight</li><li>amazon EMR</li><li>trino</li><li>akka</li><li>bodo.ai</li><li>Azure Databricks</li><li>Azure Synapse Analytics</li></ul> <p><b>Analytics Engines</b></p> <ul style="list-style-type: none"><li>DATA FUSION</li><li>Google Big Query</li><li>amazon REDSHIFT</li><li>snowflake</li><li>databricks Data Lakehouse</li><li>Apache Impala</li><li>StarRocks</li><li>FIREBOLT</li><li>HAZELCAST</li><li>ClickHouse</li><li>amazon ATHENA</li><li>dremio</li><li>SingleStore</li><li>Materialize</li><li>QuestDB</li><li>Redis</li><li>KX</li><li>pinot</li><li>Polars</li><li>trino</li><li>[ROCKSET]</li><li>imply</li><li>Pentaho</li><li>star+tree</li><li>Qubole</li><li>druid</li><li>Starburst</li><li>Yellowbrick</li><li>dbt</li><li>VOLTRON DATA</li><li>HYDRA</li><li>DORIS</li><li>presto</li></ul>	<p><b>Orchestration</b></p> <ul style="list-style-type: none"><li>Apache Airflow</li><li>PREFECT</li><li>ASTRONJMER</li><li>MLtwisj</li><li>Shipyard</li><li>seqera</li><li>dagster</li><li>DAGWORKS</li><li>kestra</li><li>Flyte</li><li>union</li><li>sematic</li><li>MAGE</li></ul> <p><b>Data Quality / Observability</b></p> <ul style="list-style-type: none"><li>KENSU</li><li>WHYLABS</li><li>great expectations</li><li>re</li><li>metaplane</li><li>SODA</li><li>lightup</li><li>Datafold</li><li>Bigeye</li><li>acceldata</li><li>Anomalo</li><li>!oloClean</li><li>union Pandera</li><li>MONTE CARLO</li><li>APACHE GriffIn</li><li>unravel</li><li>Databand</li><li>elementary</li><li>awslabs/deequ</li><li>foundational</li><li>timeseer.AI</li><li>Datorios</li></ul>	<p><b>MLOps End-to-End</b></p> <ul style="list-style-type: none"><li>OctoAI</li><li>ABACUS.AI</li><li>METAFLOW</li><li>baseten</li><li>Verta</li><li>AIBLE</li><li>DataRobot</li><li>Vertex AI</li><li>cnvrg.io</li><li>Hugging Face</li><li>mlflow</li><li>SELDON</li><li>Wallaroo.AI</li><li>Domino</li><li>kedro</li><li>Weights &amp; Biases</li><li>graft</li><li>Kubeflow</li><li>FEAST</li><li>CLEAR ML</li><li>H2O.ai</li><li>dotData</li><li>SaturnCloud</li><li>neptune.ai</li><li>comet</li><li>ZenML</li><li>Modular</li><li>Azure Machine Learning</li><li>BentoML</li><li>DagsHub</li><li>data iku</li><li>Qwak</li><li>Valohai</li><li>databricks Data Intelligence Platform</li><li>Uber Michelangelo</li><li>DVCX</li><li>TACTON</li><li>HOPSWORKS</li><li>deci</li><li>Amazon SageMaker</li></ul> <p><b>Data Centric AI/ML</b></p> <ul style="list-style-type: none"><li>DVCX</li><li>scale</li><li>GRAVITI</li><li>Snorkel</li><li>ALECTIO</li><li>dataloop</li><li>stack</li><li>Pachyderm</li><li>VOXEL51</li><li>dstack</li><li>DVID</li><li>DataLdd</li><li>Galileo</li><li>tasq.ai</li><li>hudi</li><li>Labelbox</li><li>V7</li><li>Toloka</li><li>databricks Mosaic AI</li></ul> <p><b>ML/AI Observability &amp; Monitoring</b></p> <ul style="list-style-type: none"><li>deepchecks</li><li>mona</li><li>Superwise</li><li>Arthur</li><li>fiddler</li><li>FirstEigen</li><li>arize</li><li>truera</li><li>census</li><li>WHYLABS</li><li>EVIDENTLY AI</li><li>Galileo</li><li>GANTRY</li><li>badook</li><li>AIMon</li><li>hydrosphere.io</li><li>Giskard</li><li>HOWSO</li><li>aporia</li><li>nannyML</li><li>ROBUST INTELLIGENCE</li><li>LatticeFlow</li><li>Patronus AI</li><li>distributional</li><li>BRAINTRUST</li><li>HoneyHive</li></ul> <p><b>Vector DBs</b></p> <ul style="list-style-type: none"><li>FeatureBase</li><li>Pinecone</li><li>MongoDB</li><li>milvus</li><li>Chroma</li><li>activeloop</li><li>drant</li><li>elastic</li><li>Vespa</li><li>Google ScaNN</li><li>pgvector</li><li>KX</li><li>FAISS</li><li>ClickHouse</li><li>Vald</li><li>Weaviate</li><li>Apache CASSANDRA</li><li>OpenSearch</li></ul> <p><b>Notebooks</b></p> <ul style="list-style-type: none"><li>jupyter</li><li>Apache Zeppelin</li><li>HEX</li><li>count</li><li>Deepnote</li><li>COCALC</li><li>kaggle</li><li>CloudBank</li><li>databricks Notebooks</li><li>nteract</li><li>Google colab</li><li>DL Database</li></ul> <p><b>Analytics Workflows</b></p> <ul style="list-style-type: none"><li>Querybook</li><li>databricks Delta Live Tables</li><li>Google Cloud Dataform</li></ul>	<p><b>Data Catalog / Governance</b></p> <ul style="list-style-type: none"><li>Meta Cat</li><li>ckan</li><li>Open Metadata</li><li>iguazio</li><li>Nemo</li><li>Amundsen</li><li>DataHub</li><li>atlan</li><li>Alation</li><li>OKERA</li><li>BigID</li><li>boomi</li><li>data.world</li><li>Collibra</li><li>zeenea</li><li>satori</li><li>Platform</li><li>raito</li><li>Apache Atlas</li><li>MARQUEZ</li><li>Google Cloud Dataproc Metastore</li><li>IMMUTA</li><li>magda</li><li>IBM Knowledge Catalog</li><li>Acryl Data</li><li>Gravitino</li><li>SELECT STAR</li></ul>