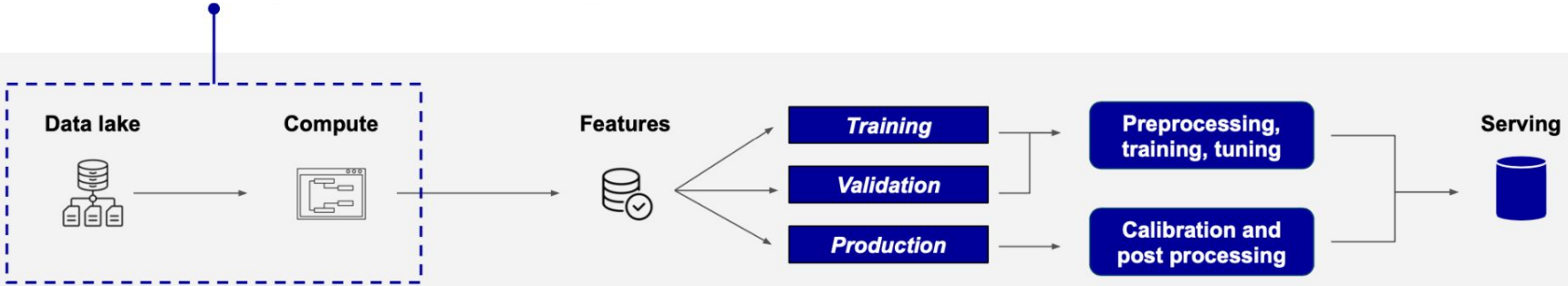
Data processing tools for compute are an integral component of AI/ML workflows



Industry trends

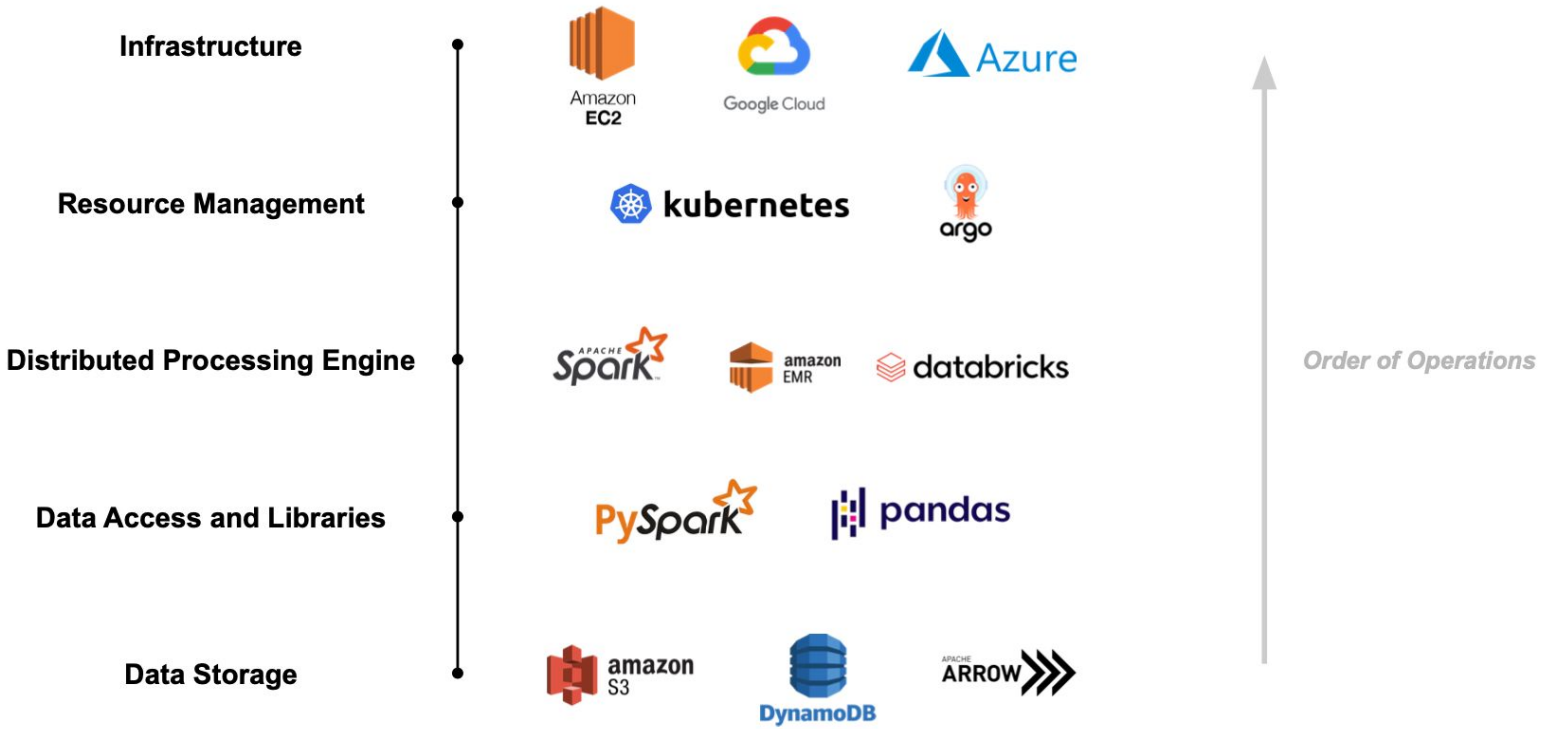
30.4%
Global Machine Learning market CAGR through 2030

19-22%
Increase in global demand for data center capacity

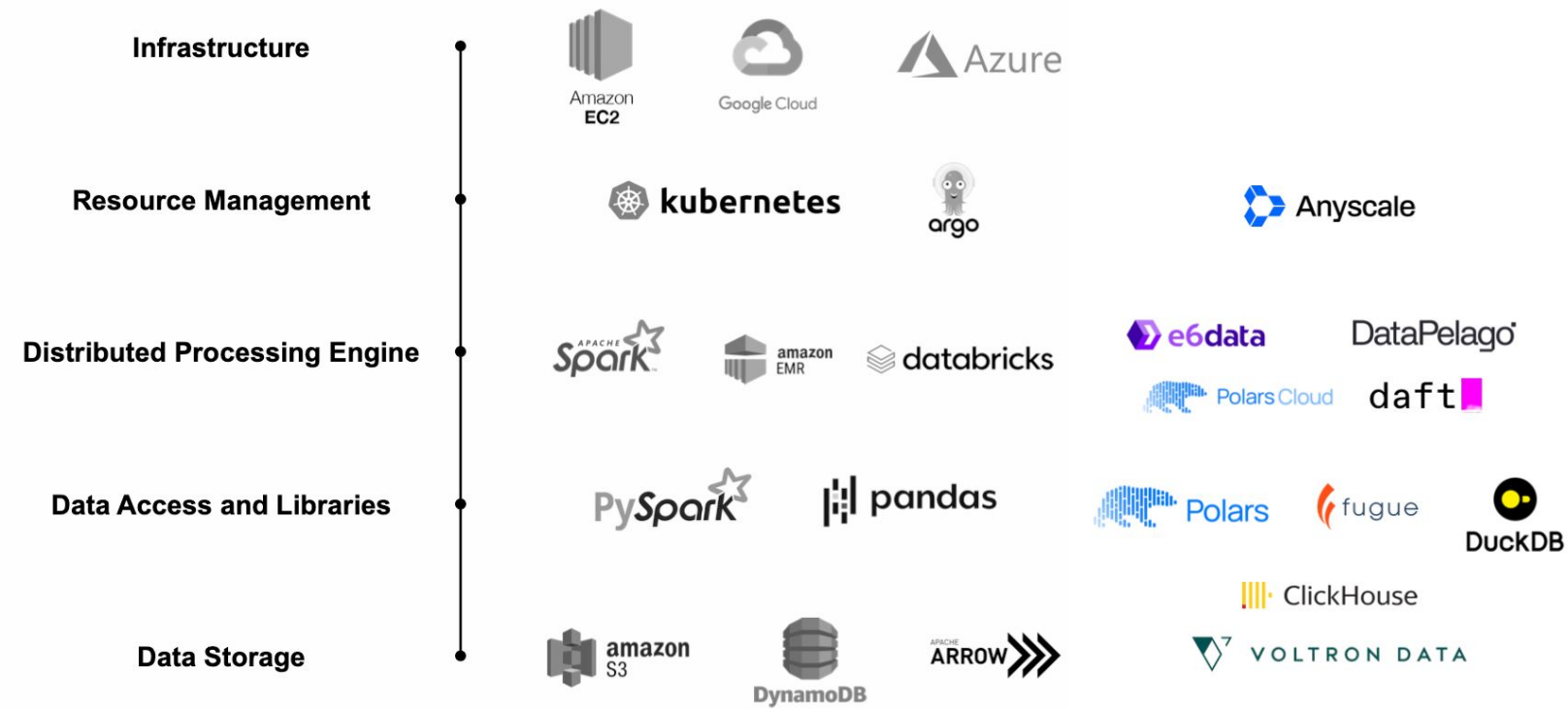
23%
YoY growth of enterprise spending on cloud services

Sources: Grand View Research; McKinsey; Synergy

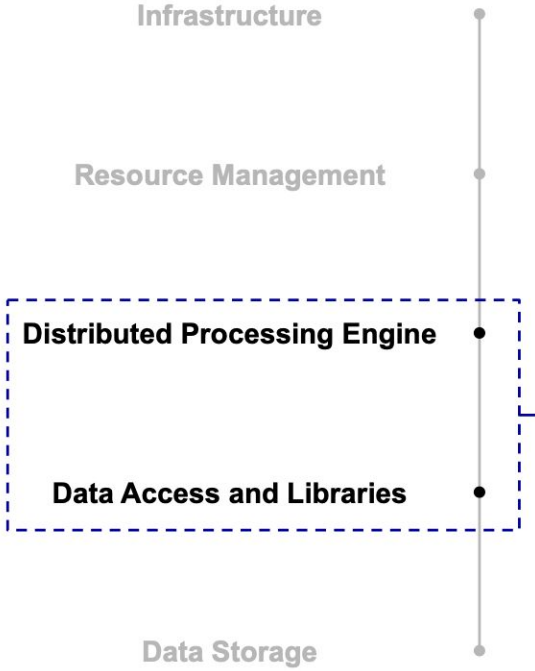
Storage-to-compute stacks are comprised of various hyperscaler and open source offerings



... and emerging startups that offer optimized alternatives to baseline tools



Query engines are the optimization layers for how data is computed



A **Query Engine** is software responsible for interpreting a user's request and executing the corresponding data retrieval, manipulation, and aggregation as efficiently as possible

Not a Query Engine

VOLTRON DATA

fugue

Anyscale

pandas

Query Engine Focus

Polars

e6data

DataPelago

daft

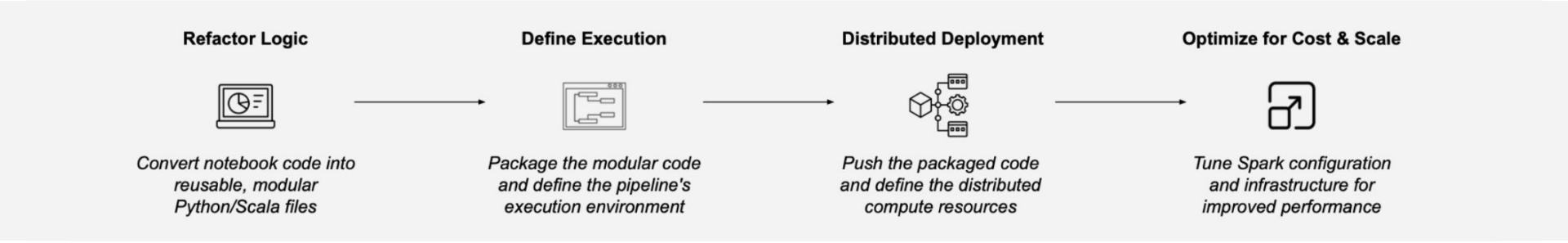
DuckDB

ClickHouse

Market landscape: Data processing startups with a query engine optimization layer

	Polars	e6data	DataPelago	Daft (Eventual)	DuckDB
	Polars Cloud	e6data	DataPelago	daft	DuckDB
Funding *Investors, Stage*	*$25M* *Accel, Series A*	*$13.6M* *Accel, Series A*	*$47M* *Eclipse, Qualcomm, Series A*	*$30M* *Felicis, YC, Citi, Series A*	*OSS* *non-profit foundation*
Employees *Size*	*10*	*95*	*63*	*29*	*24*
Momentum *Customers, Adoption*	*Open Beta on AWS* *OSS 36k+ stars*	*Chargebee and Freshworks*	*McAfee, Hidden Layer and more …*	*OSS 4.7k+ stars* *Amazon, Together.AI and more …*	*OSS 34k+ stars*
Product snapshot	*Python-native distributed data processing engine*	*Kubernetes-native for distributed SQL tasks*	*Accelerated compute platform focused on hardware abstraction*	*Python-native distributed data processing engine*	*Single-Node Analytical Database (OLAP DBMS)*
Compete?	✅	✅	⛔	✅	⛔

Data compute flows are limited by configuration friction for cloud deployments



Option databricks

Pain points

High costs: DBU and Cloud Provider Infrastructure

Learning Curve: Requires mastering Databricks-specific tools (Unity Catalog, Workflows, Delta Lake) beyond Spark

Reduced Visibility: Infrastructure abstraction makes deep performance troubleshooting difficult

Proprietary Optimizations: Superior performance from Databricks Runtime and Photon is platform-locked

Option Polars Cloud

Advantages

Serverless experience
Provisions resources when needed without managing your infrastructure

Unified API
Build data workloads locally or in the cloud without migrating logic

Diagonal scaling
Scales workloads on a single machine or a cluster with the distributed engine

Polars and Polars Cloud

Polars

New era of DataFrames with Polars

Polars is a DataFrame library designed from the ground up for fast and efficient data processing on a single machine. Used in the data ecosystem for fast transformations to replace pandas and PySpark.

- **Intuitive DataFrame API**, enabling full columnar context and reusable blocks to build your queries.
- **Optimizer takes responsibility for query performance**, incorporating decades of database system research to execute as efficient as possible.
- **Different engines for each use case**, process data in-memory, with streaming for larger than RAM datasets, or accelerated using the GPU engine.
- **End to end integrated**, written from scratch to have a tight integration from IO to query engine and full control of performance critical parts.

Growing popularity in the data community

22M+	35k+	250M+
monthly downloads	Github stars	total downloads

Powering production workflows at leading organizations

Optiver · RESEARCH · DB · DECATHLON · JUST EAT Takeaway.com

Integrating with popular data ecosystem partners

Hugging Face · learn · NVIDIA

Polars Cloud

SEAMLESS PROCESSING AT ANY SCALE

Introducing Polars Cloud to run Polars at scale

Deploy and scale data processing workloads with zero infrastructure management. Natively built on top of the popular Polars API that data teams know and love.

- Build data workloads locally or in the cloud. No more migration between tools, write Polars once and run.
- Scale out data workloads on a single machine or a compute cluster with the new distributed engine.
- Spin up the compute resources you need to execute at scale without managing infrastructure.

Product comparisons

	Polars Cloud	e6data	daft
Description	Managed service that deploys the Polars distributed engine for large-scale queries with zero infrastructure management.	Lakehouse Engine with a decentralized, Kubernetes-native architecture designed to optimize large-scale SQL workloads.	An open-source, Python-Native, Distributed Data Engine and DataFrame library built specifically for AI/ML workloads.
Advantages	✓ Maximizes core usage through diagonal scaling (vertical AND horizontal) ✓ Write once, run anywhere (seamless local to prod deployment) ✓ Speed (Rust backend)	✓ 5-10x faster than Databricks/Snowflake ✓ Interoperability with all major lakehouse formats and catalogs (Delta, Iceberg)	✓ Speed (Rust backend) ✓ Native support for multimodal data
Disadvantages	✗ Polars Cloud offering is in Beta ✗ Extended data catalog and job scheduling features are still on the roadmap	✗ Primarily a SQL-centric engine ✗ Does not offer a native, Python-first DataFrame experience	✗ No support for diagonal scaling (horizontal only) ✗ Relies on Ray's object and memory handling framework (dependency)

Spark alternatives is not a new market, but a distributed engine solution that connects seamlessly with a performant upstream dataframe library is rare

Existing tools

pandas · PySpark · Anyscale · kubernetes · Apache Airflow · dask · databricks · amazon EMR

Polars Advantage

Polars · Polars Cloud

Solves API inconsistency and tool fragmentation

Code built and tested locally runs without changes on the distributed cloud cluster.

Solves the configuration complexity

Engine is natively built on the Polars architecture, leading to savings and better performance vs. tuning a Spark job.

Polars team and history

Founding Team

Ritchie Vink

CEO and Co-founder

- Xomnia (data + AI servicing)
- Adidas

Chiel Peters

Co-founder

- CTO @ Xomnia (data + AI servicing)

Key Facts

Founding

- Founded 2020 as an open source, high-performance and multi-threaded DataFrame library to address performance limitations in existing data manipulation tools like Pandas.

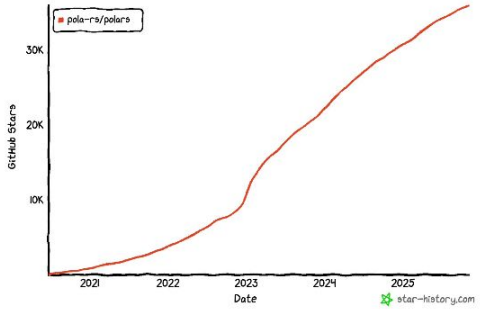
Fundraising

$21M Series A in September 2025

Accel

BainCapital VENTURES

Polars Open Source Star Count (36K+)



Users of Polars Open Source Offering

Capital One appian NETFLIX

and more …