



A DATA ORCHESTRATION FRAMEWORK

By Mina Sonbol

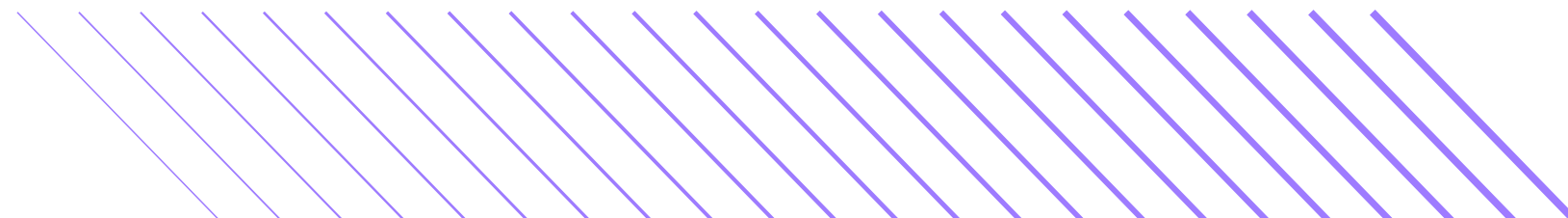




TABLE OF CONTENTS

- A bit about myself
- Mage history & background
- Data orchestration
- Mage design principles
- Mage abstractions
- Demo



A BIT ABOUT MYSELF

-  Now
 - Data Analyst at Afiniti
-  Previously
 - Data Analyst @ Persado
 - IT operations and support (SRE-like roles) @ France Telecom
 - B.Eng in computer engineer from AAST, MSc in DA/DS from UCF



HISTORY & BACKGROUND



- Mage.ai was created in 2020 by Tommy Dang and Xiaoyou Wang, formerly of Airbnb
- Initially, Mage was designed as a no-code AI tool that automated some of the work involved in training an AI model, such as data cleaning, feature engineering, hyperparameter tuning ([link](#))
- In June 2022, Mage pivoted from an AI tool to an open-source data pipeline tool ([link](#))



DATA ORCHESTRATION



- The process of managing, designing, and automating data related activities across various stages (collection, transformation, loading)
- A good data orchestrator enables ([link](#)):
 - ⚙️ Workflow management & automation
 - 🚨 Error handling & debugging
 - 🌱 Recovery
 - 🔍 Observability, monitoring, alerting
 - 🗄️ Resource optimization
 - ✅ Compliance / Auditing

- Available tools:



Airflow



Mage



Prefect



Fivetran



Dagster



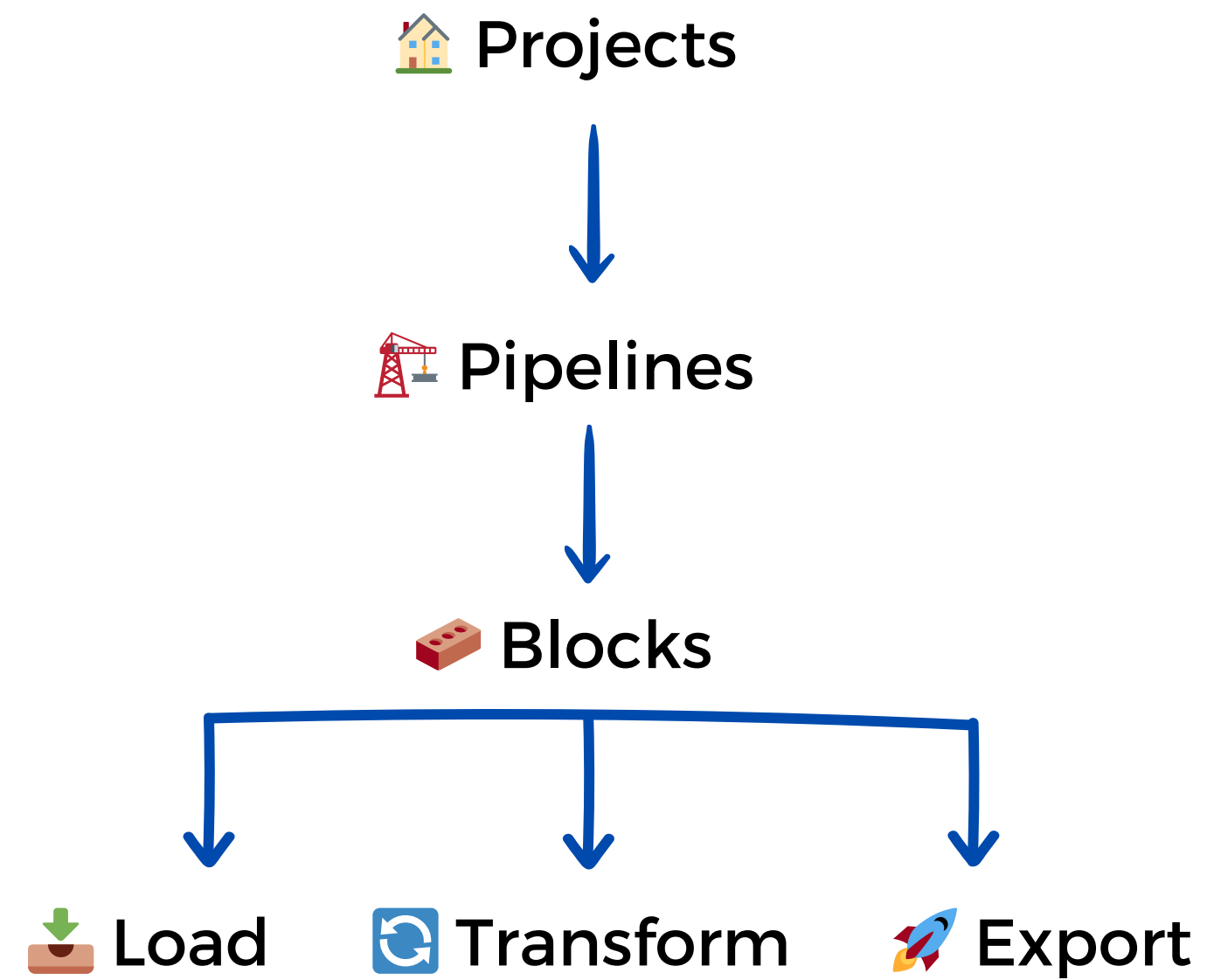
MAGE DESIGN PRINCIPLES



- 🤝 User-friendly experience
- 🔄 Quick feedback loop through a notebook-like UI
- 🧪 Engineering best practices are built-in
- 🎯 Data at the center
- 🚀 Scalable
- + Extensible

Source

ABSTRACTIONS



Source

THANK YOU