



ParseKit powers critical data operations, making data teams more efficient and data pipelines more reliable. Born out of Enigma's own needs as we continuously onboarded and updated more than 100,000 datasets from thousands of sources to power our business, ParseKit is now a robust language and platform that Fortune 500 companies and leading financial services firms around the globe depend on to acquire, link, and apply data at scale for intelligent operations.



### BOOST EFFICIENCY

Accelerate data pipeline assembly

Work smarter, not harder. ParseKit provides the framework for an efficient, reliable operation, eliminating redundant work and time-consuming maintenance that comes with custom code. Our simplified language—similar to writing steps for parsing in plain English—is backed by a substantial library of optimized, reusable parsing “steps”. Envision access to data on tap, thanks to a system of building blocks supported by your existing workflow. By preventing large-scale technical debt, eliminating key-man dependency, and lowering the technical proficiency necessary to bring in data, ParseKit vastly increases your team's productivity and your bottom line. And less time spent onboarding data means more time for your data scientists to do actual data science.



### SCALE WITH EASE

Leverage a flexible, reliable infrastructure

Both source and destination agnostic, ParseKit is fully extensible with Python and provides built-in support for any format, database, or protocol, from web scrapes to vendor systems. Your own shared library of steps is self-documenting and allows you to natively integrate steps specific to your data environments. And building with re-usable steps means every bug fix and optimization instantly benefits all related pipelines.



### INCREASE RELIABILITY

Automate deployment in a secure environment

Critical business operations demand predictable data delivery. With a streamlined interface to manage scheduling, performance, and execution resources, ParseKit enables data teams to easily grow and maintain operations at scale. Our standardized language, backwards compatibility, and semantic versioning make it simple to quickly spot and resolve errors. And our detailed alerts and metadata mean you know instantly when and where something breaks, down to the precise data point that caused the issue.



### STRENGTHEN COLLABORATION

Create a coordinated, transparent community of data

Improve communication with a shared real-time view of data availability and performance. Team and role-based permissioning give you total control over (and transparency into) what every user can access or view—so you can get data end-users and developers on the same page. ParseKit provides a streamlined framework for data operations, making it possible to turn one team's need for data into an integrated part of your organization's decision making.

- **Assemble parsers quickly and efficiently**

Build web scrapers, transformation pipelines, and more using our domain-specific language and vast library of reusable steps. ParseKit comes with everything you need to deal with common databases, formats, protocols, and systems.

- **Incorporate new functionality**

Extend ParseKit with custom steps – and leverage the 100,000+ open-source packages actively maintained by the Python community. Host new steps on the platform's internal PyPI server to instantly share them.

- **Develop within your existing workflow**

Assemble, debug, and test pipelines on your local machine with familiar tools and conventions like virtualenv and pip. Pull the latest version from your git repository and execute it with the same dependencies in a containerized environment.

- **Manage and automate scheduling**

Run parsers according to specific times or easily set dynamic rules to run recurring jobs, so your data is always fresh.

- **Dynamically provision resources**

Let us do the heavy lifting. ParseKit optimizes parser execution at scale by specifying number of worker nodes, CPU, RAM etc. for each specific project.

- **Respond quickly to issues**

Stay up-to-date on the scheduling, health, and status of every pipeline. Get reports and notifications for runs and errors, including detailed alerts and metadata the moment an issue arises.

- **Easily spot, diagnose, and resolve errors**

Dive into ParseKit's interactive debugging mode to trace ingestion problems to their root cause. Our standardized language makes it simple to identify errors in seconds, regardless of prior context.

- **Look under the hood of your pipeline**

Dive into the performance metrics and logs collected from each execution cycle to understand what went wrong and where improvements can be made.

- **Open communication among developers and data end-users**

Share data status and availability and manage notifications through a centralized interface. Allocate access to view, monitor, and edit parsers with team and role-based permissioning.

**WHAT'S INCLUDED**

<b>Runtime</b>	<ul style="list-style-type: none"> <li>• DSL: Ansible-like descriptive language based on YAML</li> <li>• Interactive debugging with breakpoints, pause on exception and more</li> <li>• Dozens of pre-built steps to support many popular sources and destinations</li> <li>• Powerful Command Line Interface</li> </ul>
<b>Platform</b>	<ul style="list-style-type: none"> <li>• Internal Pypi-like server to host custom and private packaged ParseKit steps</li> <li>• REST API to control and automate every aspect of the platform</li> <li>• Fine-grained scheduling and support for timezones</li> <li>• On-demand execution of any pipeline</li> <li>• Full encryption of data in transit and at rest</li> <li>• LDAP authentication and authorization</li> <li>• Support for Kerberos</li> <li>• Full audit trail of data and users</li> <li>• Instant alerts and notifications for error handling</li> <li>• Detailed parser execution logs</li> </ul>
<b>Support</b>	<ul style="list-style-type: none"> <li>• Robust training portal with product guides, tutorials, and use cases</li> <li>• Detailed documentation library</li> <li>• Custom training sessions and on-demand support as needed</li> </ul>

**DEPLOYMENT****Type**

On-premise or cloud-based deployment on AWS, Microsoft Azure, or Google Cloud

**Provisioning**

Deployment on as few as 3 servers and 1 worker instance and scale to as many nodes as needed

**Package**

Ansible script packages to deploy our components on a set of predefined server instances that you provisioned for us

**Customization**

Settings and parameters adjustable and adaptable to your unique deployment environment

**Updates**

A new version of ParseKit is released every 8 weeks. Extensive testing and strict versioning practices protect you from backward incompatibilities when you upgrade.

**ABOUT ENIGMA**

A new approach to Operational Data Management and Intelligence, Enigma builds an ecosystem of tools to acquire, link and apply data for intelligent operations. Our massive repository of public data informs and trains each of our products. All of our solutions are designed to complement the expertise of our users and change the way they interact with data. Fortune 500 organizations including Merck, American Express and EMD Millipore, rely on Enigma for repeatable, scalable insights.