Flow Overview

Data Discovery	Researchers often begin by identifying datasets that align with their objectives.	
	Data Catalog	Categorization: Datasets are systematically organized by domain, type (e.g., raw, processed), and frequency.
		Metadata: Detailed descriptions, data sources, and update schedules accompany each dataset, enabling informed decision-making.
		Search and Filtering: Advanced search features with keyword tagging and filtering options (e.g., by timeframe or region) allow researchers to pinpoint relevant datasets quickly.
	example: Example: A researcher exploring volatility factors can search for "historical stock prices" or filter datasets tagged with "equities," "options," or "intraday volatility" on our platform.	
Data Access and Retrieval	Once datasets are identified, the platform simplifies access through the Data Lake and Data Catalog	
	Data Lake	Centralized Repository: The Data Lake stores both raw and processed datasets in various formats (e.g., CSV, Parquet), ensuring that all data is accessible from a single location.
		Secure Access Controls: Role-based permissions ensure that only authorized users can access sensitive data, maintaining compliance and security protocols.
		APIs and Data Connectors: Programmatic access via APIs enables integration with tools like Python, R, or SQL interfaces, allowing for automated data retrieval within custom scripts or applications.
	Data Catalog	Integration with Data Catalog: Direct links from the catalog entries allow researchers to access or request datasets without navigating away, streamlining the retrieval process.
	Example: A researcher running an event study can retrieve clean historical earnings data with a single query, specifying a date range for pre- and post-event windows.	
Data Preparation and Transformation	The Data Workbench is an interactive environment supporting data manipulation tasks	
	Data Workbench	Cleaning and Validation: Automated checks identify and correct inconsistencies, outliers, or missing values.
		Aggregation and Filtering: Researchers can group data by time intervals (e.g., monthly returns) or specific attributes (e.g., sector-level data).
		Custom Workflows: Custom scripts or predefined templates streamline repetitive tasks, ensuring consistency.
	Example: For factor modeling, a researcher aggregates daily returns into monthly returns, filters out illiquid stocks, and applies a z-score normalization—all within the workbench.	
Analysis Integration	The platform provides Python APIs to enable seamless integration with researchers' analysis environments, ensuring smooth transitions from data preparation to analysis	
	Data Lake	Seamless Access via Python API: Researchers can easily access raw or processed data from the Data Lake using Python APIs. These APIs also support automated workflows, including fetching external data and performing interactive operations within the Data Lake.
	Data Workbench	Integration with Analytical Tools: The exported data integrates seamlessly with popular libraries and frameworks, such as pandas for data manipulation, scikit-learn for machine learning, and statsmodels for statistical analysis. This compatibility ensures a smooth handoff from the workbench to analysis tasks.
		Streamlined Data Transfer: Data from the workbench is prepared for export through structured APIs. Researchers can export data in commonly used formats or as structured objects compatible with analytical tools. The platform enables a direct pipeline to Python environments, ensuring data integrity and eliminating manual handling errors.
Conclusion	Our data platform is meticulously designed to support quant researchers at every step:	
	Efficient Discovery: The Data Catalog makes finding relevant datasets intuitive and fast.	
	Easy Access: The integration of the Data Lake and Catalog ensures seamless retrieval of data.	
	Powerful Preparation Tools: The Data Workbench provides robust capabilities for data cleaning and transformation.	
	Smooth Integration: Easy movement of data into analysis environments accelerates the research cycle.	