

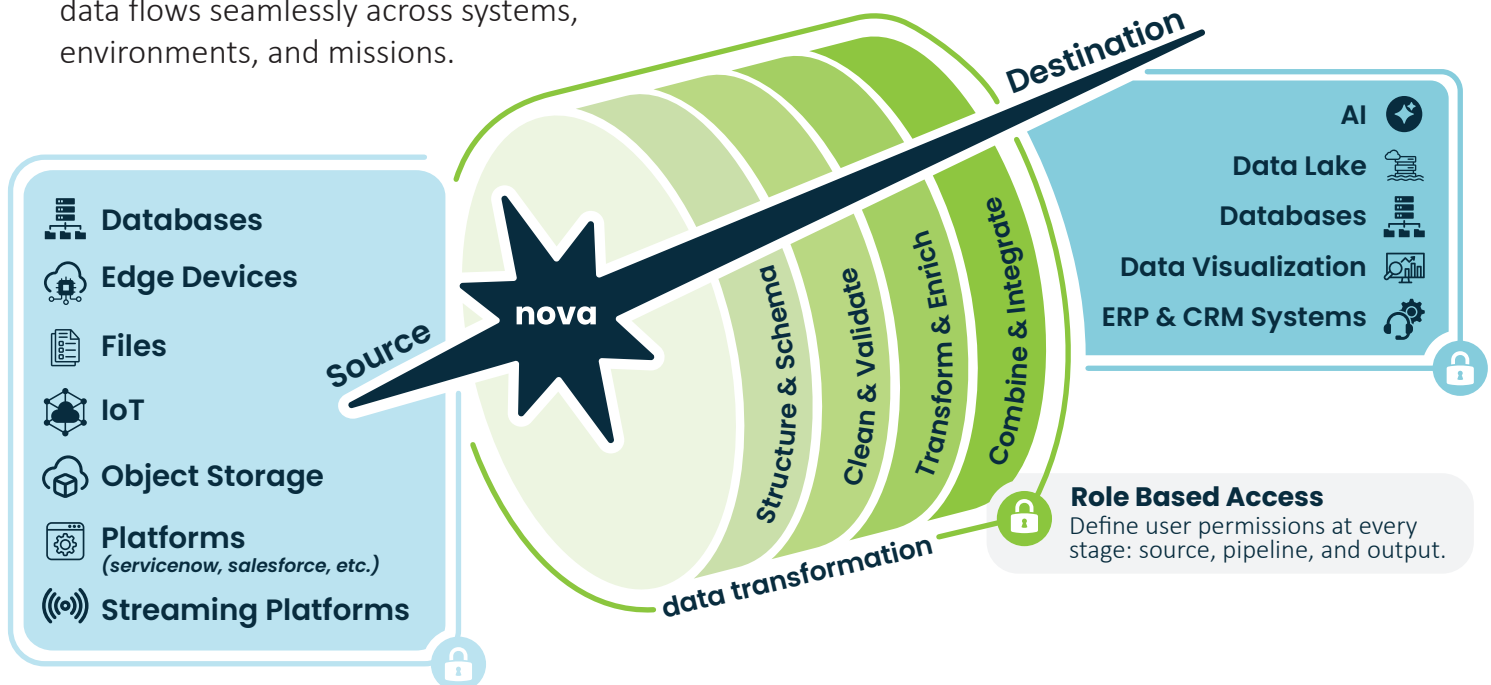


nova

nova powers clean, reliable data — from any source, any type, at scale — for mission-critical insights.

Fragmented, siloed data systems slow decision-making, complicate governance, and increase security risks. Legacy platforms lack the scalability and compliance needed to manage growing data volumes across distributed environments.

nova delivers that solution. nova is your enterprise data pipeline platform, to simplify how you transform, filter, validate, and route data across any source, any type, and at scale for any mission. With nova, organizations gain centralized creation, management, and monitoring, empowering them to scale data operations without limits and ensure data flows seamlessly across systems, environments, and missions.



Features

Process Large Data Loads

Handle massive datasets with speed and reliability.

Any Data Source. Any Data Format.

Integrate data from any format and environment —on-prem, cloud, hybrid.

Unlimited Transformations

Perform countless transformations within a single pipeline.

Reuse Source Buckets

Share the same source bucket across pipelines without duplication.

Flexible Execution

Run your real-time pipeline on demand, on schedule, or triggered by events.

Centralized Oversight

Monitor and manage all pipelines from a single hub.

Data Enrichment

Improve data quality by using advanced customization within the pipeline.

Unlimited Data Retention

Maintain historical data for error resolution and compliance.

Pipeline Metrics & Health

Monitor throughput, back pressure, and lag.

Role-Based Access Control

Apply granular permissions at every pipeline stage.

Zero Trust Security

Combine strict identity controls with full encryption in transit & at rest.

Real-time Pipeline Testing

Test changes to transformation rules before launch.

Transformation Catalog

STRUCTURE & SCHEMA

ChangeSchema	Map property keys, rename, modify data types, or drop keys.	Rename cust_id to customer_id and change type from string to integer.
SelectFields	Keep only selected data property keys.	Keep only customer_id, name, email columns.
DropFields	Remove specified data property keys.	Drop temp_column from dataset.
RenameField	Rename a single data property key.	Rename fname to first_name.
SplitFields	Split keys into two DynamicFrames (<i>selected vs remaining</i>).	Split personal_info and transaction_info into separate frames.
SelectFromCollection	Choose one DynamicFrame from a collection.	Select the first DynamicFrame from a list of frames.
Flatten	Extract fields inside structs into top-level fields.	Flatten address.city into city.
UUID	Add a Universally Unique Identifier column.	Add uuid column for each row.
Identifier	Add a numeric identifier column.	Add row_id starting from 1.
To timestamp	Convert a column to timestamp type.	Convert order_date from string to timestamp.
Format timestamp	Convert timestamp column to formatted string.	Format order_date as YYYY-MM-DD.

CLEAN & VALIDATE

FillMissingValues	Impute missing values with suggested values.	Replace missing age with average age.
Drop Null Fields	Remove columns where all values are null.	Drop middle_name column if all values are null.
Drop Duplicates	Remove duplicate rows based on keys or entire rows.	Remove duplicate rows based on customer_id.
Filter	Split dataset based on a filter condition.	Filter rows where status = 'active'.
Remove Null Rows	Remove rows where all columns are null or empty.	Drop rows where all fields are null.
Record Matching	Apply ML-based record matching for classification.	Match duplicate customer records using ML.

TRANSFORM & ENRICH

SQL	Use SparkSQL queries to transform data.	SELECT name, SUM(amount) FROM orders GROUP BY name.
Add Timestamp	Add processing timestamp for auditing.	Add processed_at column with current timestamp.
Aggregate	Perform calculations (sum, avg, min, max) and create new fields.	Calculate total_sales per region.
Array To Columns	Extract array elements into separate columns.	Convert [red, blue] into color1, color2.
Autobalance Processing	Redistribute data for better parallel processing.	Balance skewed data across workers.
Concatenate Columns	Combine column values into a new string column.	Combine first_name and last_name into full_name.
Conditional Router	Route rows based on multiple conditions.	Send rows with amount > 1000 to one frame, others to another.
Custom Transform	Apply custom code for transformations.	Write Python code to normalize phone numbers.
Derived Column	Create new column using formula or SQL expression.	Add discounted_price = price * 0.9.
Explode Array or Map Into Rows	Flatten nested arrays/maps into individual rows.	Explode items array into separate rows.
Extract JSON Path	Extract columns from JSON string using path.	Extract \$.city from JSON column.
Extract String Fragments	Create new columns from regex matches.	Extract digits from order_code using regex.
Lookup	Add columns from a catalog table based on matching keys.	Add region_name from lookup table using region_id.
Parse JSON column	Convert JSON string column to struct or array.	Parse {city:NY,zip:10001} into columns.
Pivot Rows to Columns	Rotate unique values into new columns.	Pivot month values into columns for sales data.
Split String	Break a string into tokens using regex.	Split tags column by comma into an array.
Unpivot Columns To Rows	Convert columns into rows for normalization.	Convert Jan, Feb, Mar columns into rows under month.

COMBINE & INTEGRATE

Join	Combine two datasets using join conditions	Join customers and orders on customer_id.
Union	Merge rows from multiple sources with the same schema.	Union sales_2024 and sales_2025 datasets.
Spigot	Write samples of data to Amazon S3 for integration/output.	Export 10% sample of dataset to S3 bucket.