

DATA + ANALYTICS

## Analytics Lab and Starflow

Jinyoung Kim

Head of Development, Analytics Platform

Morningstar, Inc.

# Vision Inspiration

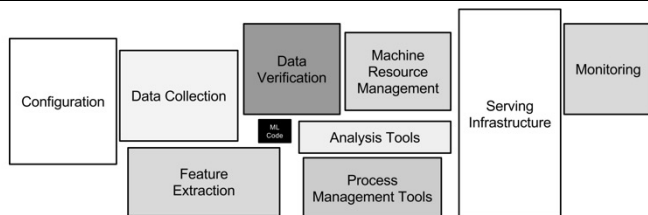
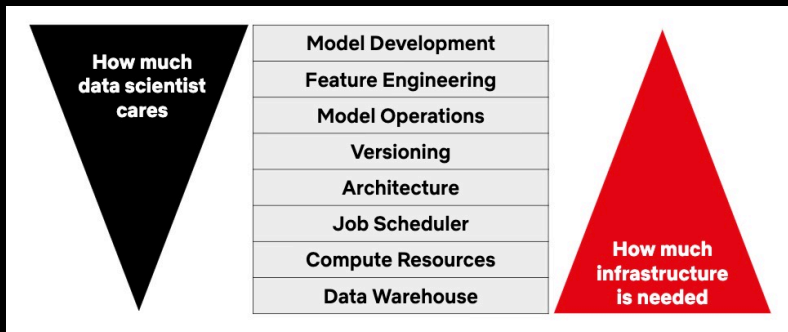


Figure 1: Only a small fraction of real-world ML systems is composed of the ML code, as shown by the small black box in the middle. The required surrounding infrastructure is vast and complex.

<https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf>



<https://docs.metaflow.org/introduction/what-is-metaflow>



# Frameworks

“How do you build your webapp?”

2000



2004



2008



symfony



2015



# Data/ML Frameworks

“How do you build your models?”

2013

SPSS  
STATA

sas  
MATLAB

R  
python



Java



2015

TensorFlow

scikit  
learn

PyTorch

2019

Kubeflow

mlflow

METAFLOW

# Frameworks + Platform as a Service



+



STARFLOW

+

Analytics Platform

# Starflow App Example

## Flask

Flask is a lightweight [WSGI](#) web application framework. It is designed to make getting started quick and easy, with the ability to scale up to complex applications. It began as a simple wrapper around [Werkzeug](#) and [Jinja](#) and has become one of the most popular Python web application frameworks.

Flask offers suggestions, but doesn't enforce any dependencies or project layout. It is up to the developer to choose the tools and libraries they want to use. There are many extensions provided by the community that make adding new functionality easy.

```
from flask import Flask, escape, request

app = Flask(__name__)

@app.route('/')
def hello():
    name = request.args.get("name", "World")
    return f'Hello, {escape(name)}!'
```

```
1 from typing import Dict, Any
2 import starflow
3
4 app = starflow.App()
5
6 @app.run(run_type='add')
7 def add(x: int, y: int) -> Dict[str, Any]:
8     return {"sum": x+y}
9
```

# Starflow API (Use Model)

GET	/status	The status, latency and details of various services on which CAP depend to keep the starflow model app up and running. A 200 is returned even if some dependent services are down.	
GET	/metadata	Metadata information like git commit, branch, version and environment of the starflow model app.	🔒
POST	/files	Generates a signed URL to upload the file.	🔒
GET	/files	Get a paginated list of files.	🔒
GET	/files/{id}	A secure signed url to download the file matching the unique file identifier.	🔒
GET	/validations/{id}	Get validation details for the given validation id.	🔒
GET	/validations	Get a paginated list of validations.	🔒
POST	/validations	Submit data for validation of inputs for different run types.	🔒
GET	/runs/{id}	Get details for the given run id.	🔒
GET	/runs	Get a paginated list of runs.	🔒
POST	/runs	Submit data to queue a run.	🔒
GET	/statistics	App Statistics like average and standard deviation of run-time for all run types of the starflow model app. Statistics are calculated from executions in the last 6 months.	🔒
POST	/schedules	Schedule a Starflow run.	🔒
GET	/schedules	Get a paginated list of schedules.	🔒
GET	/schedules/{id}	Get details for the given schedule id.	🔒
DELETE	/schedules/{id}	Archive a schedule for a Starflow run. Archived schedules do not trigger runs. They can still be accessed with GET /schedules/{id} but cannot be updated by PATCH /schedules/{id}.	🔒
PATCH	/schedules/{id}	Update a schedule for a Starflow run.	🔒

# Starflow App Management API (Deploy Model)

POST	<b>/apps</b> Create a starflow application in the Core Analytics Platform.	🔒
GET	<b>/apps/{appName}</b> Get app information	🔒
PATCH	<b>/apps/{appName}</b> Update partial starflow application info	🔒
POST	<b>/apps/{appName}/environments</b> Create an environment to run the starflow application.	🔒
PATCH	<b>/apps/{appName}/environments/{envName}</b> Update the starflow application runtime environment	🔒
GET	<b>/apps/{appName}/environments/{envName}</b> Get app environment level information	🔒
POST	<b>/gitevent</b> Send a git event notification of the starflow application repository.	
GET	<b>/apps/environments</b> Get app name and environment name for the matching S3 bucket path	🔒
POST	<b>/apps/{appName}/environments/{envName}/execute</b> Trigger the app build in the desired environment.	
POST	<b>/apps/{appName}/environments/{envName}/promote</b> Promote a non-prod app environment to production.	



# Analytics Lab

The screenshot displays the Analytics Lab (beta) interface. The top menu bar includes File, Edit, View, Run, Kernel, Git, Tabs, Settings, and Help. The left sidebar features a 'Data Explorer' tab with 'Select Dataset' (Equity) and 'Select Database' (All databases) options. Below this is a 'Tables' section with a search bar and a list of aggregation metrics such as 'Aggregation Daily Dividend Metrics', 'Aggregation Efficiency Ratios', 'Aggregation Enterprise Value Calculations', 'Aggregation Financial Health Ratios', 'Aggregation Financial Statements Simple Summation', 'Aggregation Financial Statements Weighted Average', 'Aggregation Leverage', 'Aggregation Margins', 'Aggregation Market Capital', 'Aggregation Monthly Dividend Metrics', 'Aggregation Price Multiples', 'Aggregation Price Yields', 'Aggregation Profitability', 'Aggregation Residual Risk And Return Sensitivity', 'Aggregation Trailing Return', 'Average Rates', and 'Average Rates'.

The main area is titled 'Analytics Lab' and features the Morningstar logo. It includes a 'Launcher' tab and a 'Basic Notebook' section with the text: 'Starting out fresh? Use this blank notebook with the Morningstar data integration established for you.' Below this are sections for 'Direct Notebook' and 'Presentation Studio Notebook'. The 'Morningstar Notebooks' section displays a grid of notebooks: 'Firm Diversity', 'Investment Product Launches', 'Manager History', 'Security Ownership Analysis', and 'Time Series Regression Tools'. The 'Utilities' section displays a grid of utility icons: 'Show Contextual Help', 'Terminal', 'Python 3 (ipykernel)', 'Octave', 'R', 'Data Quality', 'Text File', 'Markdown File', and 'VS Code IDE [x]'. A 'Feedback' button is located on the right side of the interface.

# Easy Access to Data and Compute

The screenshot shows the Analytics Lab (beta) interface. On the left, there's a sidebar with 'Data Explorer' and 'Direct User Objects'. The 'Data Explorer' shows a tree view of databases and tables. The 'Direct User Objects' section shows a list of tables, including 'Benchmark Regulation Profile Indicators', 'Benchmark Regulation Profiles', 'Board Member Diversity Demographic Profile Aggregate Numbers', 'Board Member Diversity Demographic Profile Granular Numbers', 'Carbon Custom', 'Carbon Emissions', 'Carbon Emissions Profiles', 'Carbon Fossil Fuel Profiles', 'Carbon Rating', 'Carbon Research', 'Carbon Risk Rating Benchmarks Company Level', and 'Carbon Risk Rating Benchmarks Company Level Details'.

The main area displays a Jupyter Notebook titled 'Untitled20.ipynb'. The notebook has two cells. The first cell contains the code:

```
[3]: import analyticslab as al
```

The second cell contains a SQL query and the resulting DataFrame:

```
[4]: query_string = """
SELECT
    "entity_id", "profile_id", "profile_type_id", "date_valid_from",
    "date_valid_to", "date_created", "date_updated"
FROM
    "company_esg_prd"."benchmark_regulation_profiles" LIMIT 10
"""

df = al.query(query_string)
df
```

The resulting DataFrame is shown below:

	entity_id	profile_id	profile_type_id	date_valid_from	date_valid_to	date_created	date_updated
0	1007896757	265466	2	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
1	1007896757	265467	3	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
2	1007896757	265468	4	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
3	1007896757	265469	5	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
4	1007896757	265470	6	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
5	1007896757	265471	7	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
6	1007896757	265472	8	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
7	1007896757	265473	9	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
8	1007896757	265474	10	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640
9	1007896757	265475	11	2021-10-04	2021-10-04	2021-10-04	2021-10-05 13:20:17.640

The screenshot shows the documentation page for `analyticslab.direct`. The page title is `analyticslab.direct`. The page content includes a list of functions and their descriptions:

- `calculate_performance_report(report_id: str) -> pandas.DataFrame`  
Calculate performance report based on report id.  
**Parameters:** `report_id` - The performance report id that want to calculate.  
**Returns:** `DataFrame` object.  
**Raises:** `DnaLabException` - An error occurred if param `report_id` is wrong.
- `get_all_lists()`  
Getting All Direct Lists.  
**Returns:** `DataFrame` object.
- `get_all_performance_reports() -> pandas.DataFrame`  
Get all performance reports.  
**Returns:** `DataFrame` object.
- `get_all_search_criterias()`  
Get all Direct Search Criteria.  
**Returns:** `DataFrame` object.
- `get_all_universes() -> pandas.DataFrame`  
Get all universes.  
**Returns:** `DataFrame` object.
- `get_all_user_datasets() -> pandas.DataFrame`  
Get all user created datasets.  
**Returns:** `DataFrame` object.
- `get_asset_flow_data(investment_ids: list = None, list_id: str = None, search_criteria_id: str = None, market_id: str = None, datapoint_settings: pandas.DataFrame = None) -> pandas.DataFrame`  
Get asset flow data.  
**Parameters:**
  - `investment_ids` - The investment id list.
  - `list_id` - The investment list id.
  - `search_criteria_id` - The search criteria id.
  - `market_id` - The market id.

# Speed of Delivery of Insights

**Forbes**[Subscribe](#)[Sign In](#)

Sep 28, 2021, 11:43am EDT | 25,905 views

## Speed Matters: How Morningstar Accelerates Innovation With Data

**Amazon Web Services** **BRANDVOICE** | Paid Program  
Leadership

When the New York Stock Exchange (NYSE) announced plans to delist several large Chinese telecom firms in January 2021, investors grew concerned. The move was a response to an executive order targeting companies suspected of helping the Chinese military, and it required investors to divest of the affected stocks by November.

### Reinvent your business with data

Leverage data insights to drive better, faster outcomes

[Learn how it works](#)

# Empowering Investors

WIRED INSIDER

## Pioneering a new way to empower investors

*The financial services company Morningstar has always been a leader in investment research and insight. But the way it was storing and using its data was starting to hold it back.*

Search Quotes and Site

MORNINGSTAR

Contact Us

Software & Data Services ▾ Money Management ▾ Solutions for Business All Products & Services

### Morningstar Notebooks in Analytics Lab

Automate your innovation with advanced analytics at the click of a button

Discover Notebooks with your free Morningstar Direct™ trial today!

Get Morningstar Notebook Email Updates

To succeed in the financial services industry, you need to transform vast amounts of data into actionable insights every day, which can be challenging and time-consuming. Our newest Direct workspace centralizes data, analytics, and

