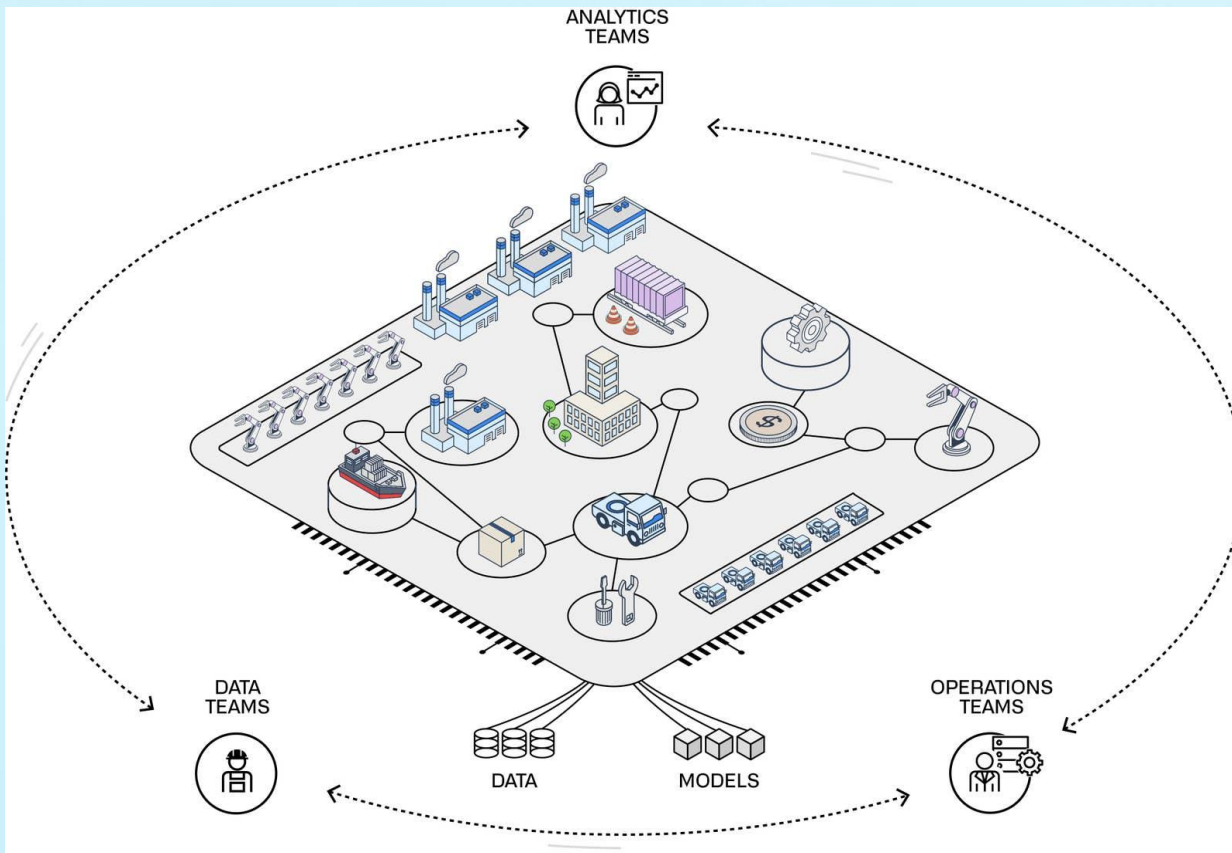


## Система по работе с большими данными

## Описание проекта, реализованного в Swiss Re

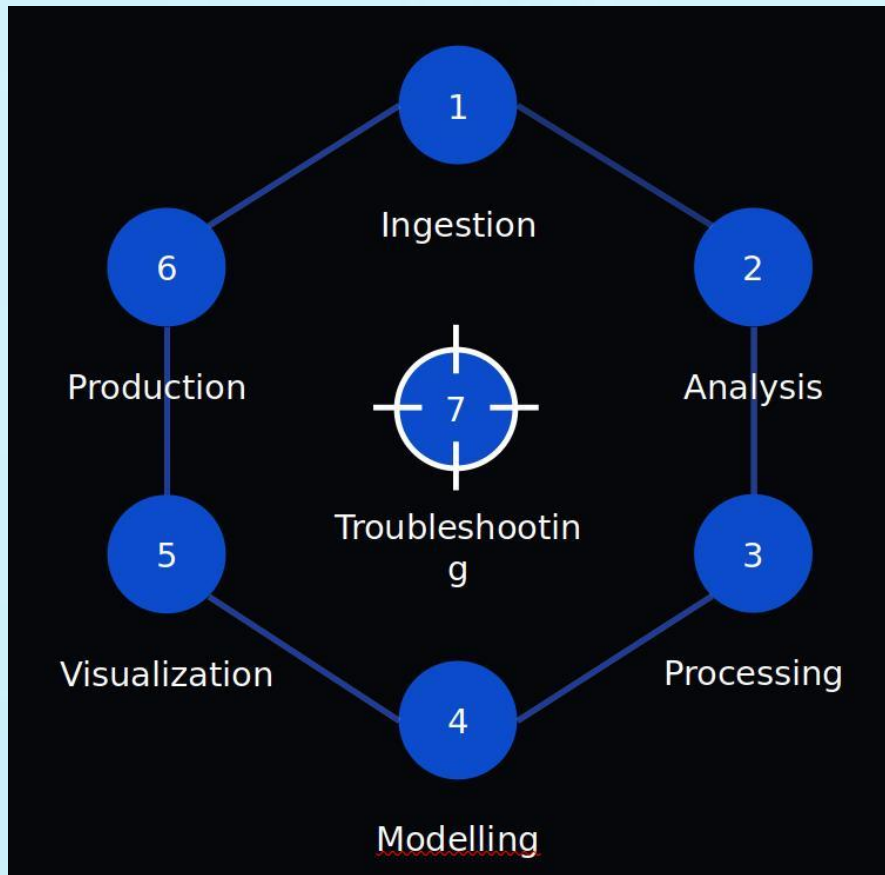
- Проект --- система по работе со страховыми контрактами различных типов.
- Система позволяет производить быстрые вычисления за счёт применения Spark и облачных решений.
- Система позволяет проводить стандартизацию контрактов, приводя их к единому формату.
- Система предоставляет средства просмотра и управления пайплайном работы с данными.
- Система предоставляет удобные средства для визуализации и анализа контрактов.
- Система предоставляет возможность использовать low-code технологии, включая применение машинного обучения к данным.
- Система реализована на особой платформе по работе с большими данными.

## Особенности платформы в Swiss Re



- Платформа подходит для пользователей разных уровней.
- Возможность работы с платформой из браузера.
- Поддержка одновременной работы большого количества пользователей.
- Кластер со Spark для параллельных вычислений.
- Множество мини-приложений для различных задач.
- Ограниченная возможность вывода данных.
- Поддержка загрузки данных в самых разных форматах.
- Поддержка low-code решений на Python, R, SQL, Java, Mesa.

## Стадии работы с данными



- В работе с данными выделяют 6 основных стадий.
- В Swiss Re была возможность получить поддержку в режиме 24 на 7.
- Часто работа с данными идёт по циклу.
- В системе были реализованные решения для каждого этапа работы с данными.

# Загрузка данных

- Поддерживаемые форматы: csv, xls,xlsx, pdf и другие похожие.
- Большое количество файлов в сыром формате.
- Указание методики обработки файлов (релевантные колонки, переименования) в специальных служебных файлах.
- Сырые данные загружаются в платформу. Вся обработка и работа с данными происходят внутри платформы.
- Внутри платформы предоставляются удобные средства работы с сырыми данными.
- Можно просматривать данные, переименовывать столбцы, менять их типы.

Preparing meteorite\_landings\_raw - 2018-05-04 15:36:58

Table Columns

Save as dataset Analyze Save... Share 0

	name String	id Integer	nametype String	recclass String	mass_g Double	fall String	year Timestamp
1	Aachen	1	Valid	L5	21	Fell	1880-01-01T00:00:00...
2	Aarhus	2	Valid	H6	720	Fell	1951-01-01T00:00:00...
3	Abee	6	Valid	EH4	107000	Fell	1952-01-01T00:00:00...
4	Acapulco	10	Valid	Acapulcoite	1914	Fell	1976-01-01T00:00:00...
5	Achiras	370	Valid	L6	780	Fell	1902-01-01T00:00:00...
6	Adhi Kot	379	Valid	EH4	4239	Fell	1919-01-01T00:00:00...
7	Adzhi-Bogdo (stone)	390	Valid	LL3-6	910	Fell	1949-01-01T00:00:00...
8	Agen	392	Valid	H5	30000	Fell	1814-01-01T00:00:00...
9	Aguada	398	Valid	L6	1620	Fell	

Start with:  
meteorite\_landings\_...

name String

Normal 45,711 (99%)

Null 0

Empty 0

Whitespace 0

Needs trim 5 (<1%)

Numeric 0

Non-alpha 0

Uppercase 0

Lowercase 0

Mixed case 45,716 (100%)

LENGTHS 27 by inc. value

VALUES 10,000 by dec. count

Q Filter...

CHANGE DATA (FOR ENTIRE DATASET)

New value  
Change this value.

Find and replace  
Find and replace a value.

Map  
Convert abbreviations or enumerations into human readable values.

Keep or remove  
Keep or remove specific characters.

Extract  
Keep a substring of the value.

Concatenate  
Add column values and free text together.

# Анализ данных

Preparing meteorite\_landings\_raw - 2018-05-04 15:36:58

Table Columns

Save as dataset Analyze Save... Share 0

	name String	id Integer	nametype String	recclass String	mass_g Double	fall String	year Timestamp
1	Aachen	1	Valid	L5	21	Fell	1880-01-01T00:00:00...
2	Aarhus	2	Valid	H6	720	Fell	1951-01-01T00:00:00...
3	Abee	6	Valid	EH4	107000	Fell	1952-01-01T00:00:00...
4	Acapulco	10	Valid	Acapulcoite	1914	Fell	1976-01-01T00:00:00...
5	Achiras	370	Valid	L6	780	Fell	1902-01-01T00:00:00...
6	Adhi Kot	379	Valid	EH4	4239	Fell	1919-01-01T00:00:00...
7	Adzhi-Bogdo (stone)	390	Valid	LL3-6	910	Fell	1949-01-01T00:00:00...
8	Agen	392	Valid	H5	30000	Fell	1814-01-01T00:00:00...
9	Aguada	398	Valid	L6	1620	Fell	

Start with:  
meteorite\_landings\_...

name String

Normal 45,711 (99%)

Null 0

Empty 0

Whitespace 0

Needs trim 5 (<1%)

Numeric 0

Non-alpha 0

Uppercase 0

Lowercase 0

Mixed case 45,716 (100%)

LENGTHS 27 by inc. value

VALUES 10,000 by dec. count

Filter...

CHANGE DATA (FOR ENTIRE DATASET)

New value  
Change this value.

Find and replace  
Find and replace a value.

Map  
Convert abbreviations or enumerations into human readable values.

Keep or remove  
Keep or remove specific characters.

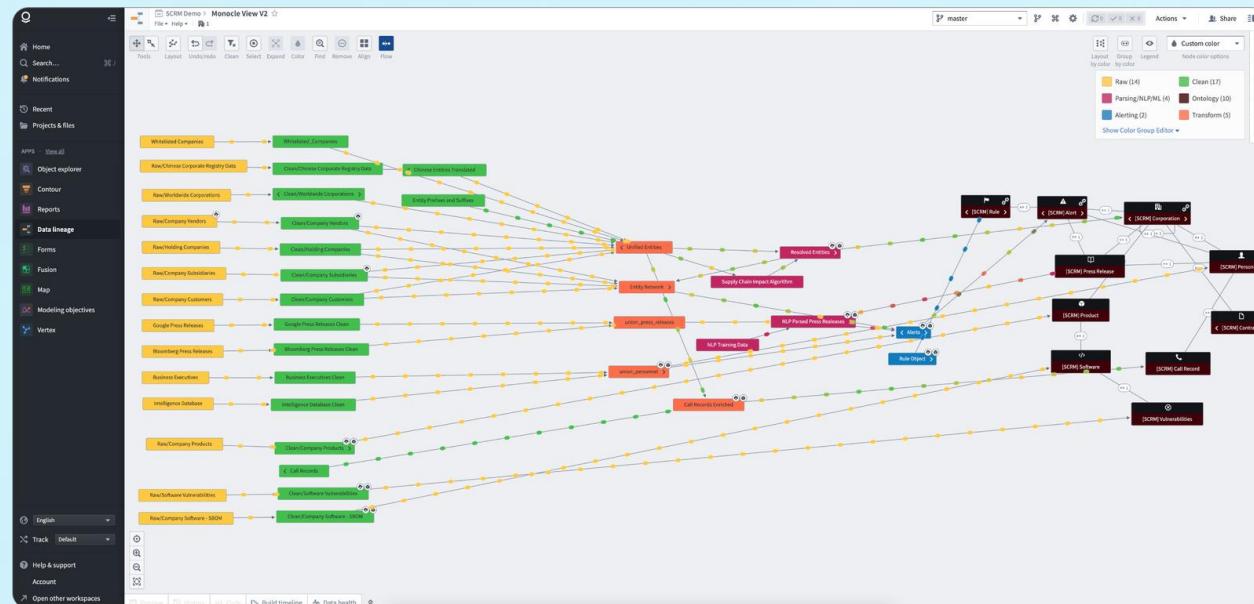
Extract  
Keep a substring of the value.

Concatenate  
Add column values and free text together.

- Предоставляются удобные средства анализа данных.
- Поддержка основных операций обработки данных.
- Поддержка большого количества разнообразных графиков.
- Поддержка различных версий данных.
- Возможность экспорта на любом этапе анализа данных.
- Учёт прав доступа пользователей (row-level security).

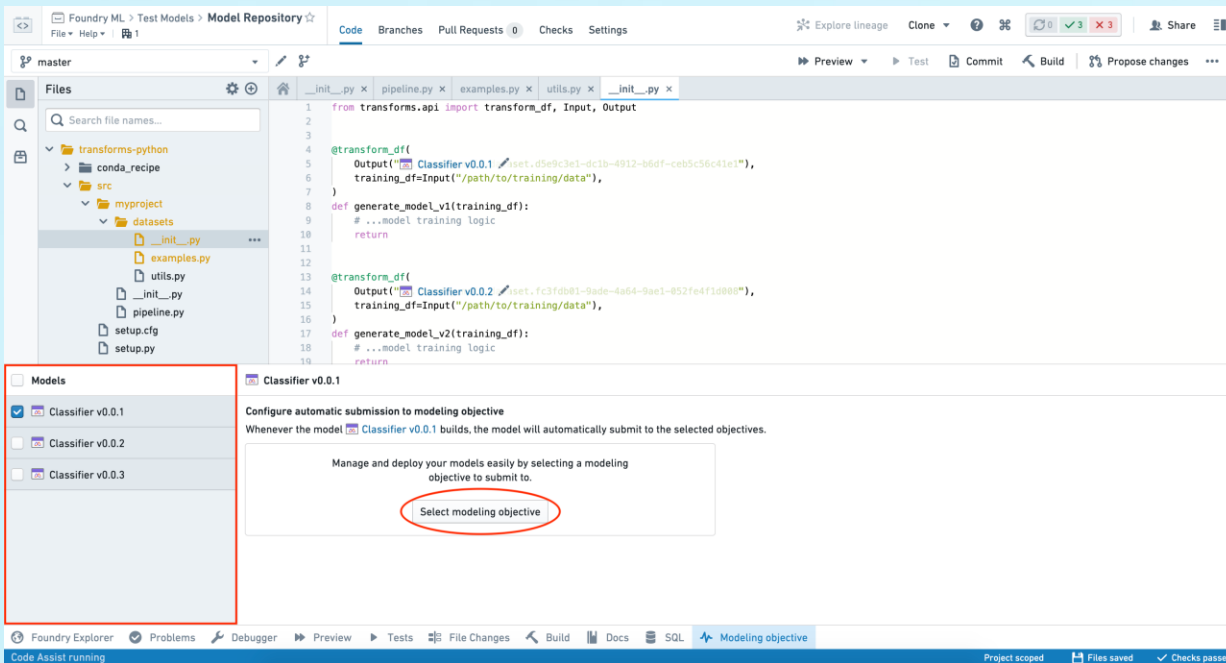


# Обработка данных



- Обработка и объединение данных.
- Удобный запуск пайплайна.
- Low-code решение.
- Поддержка версионирования и сохранения промежуточных шагов.
- Удобный способ тюнинга Spark.

# Моделирование

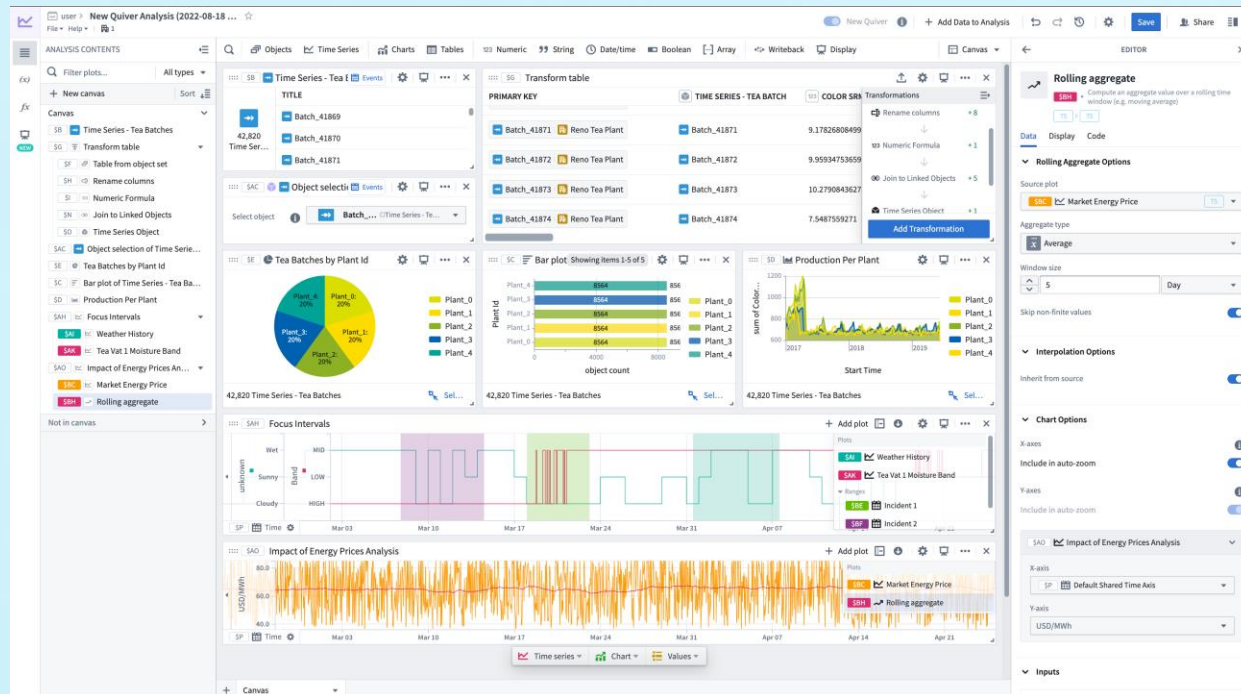


- Поддержка различных библиотек по машинному обучению.
- Удобный способ хранения, сравнения и тестирования моделей.
- Версионирование моделей.
- Модели могут использовать все возможности Spark для ускорения вычислений.
- Low-code решение.

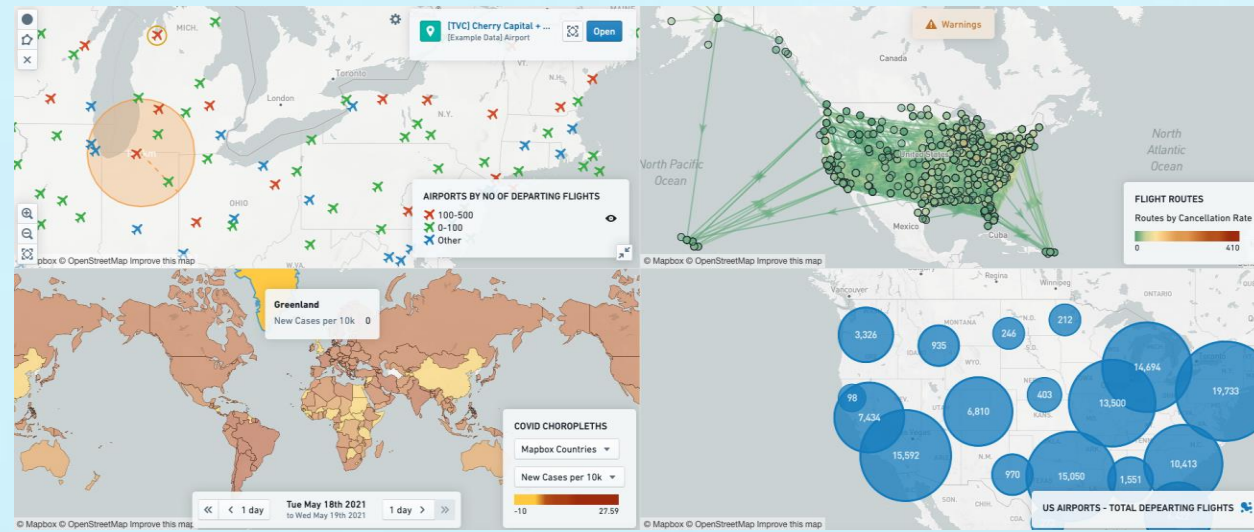


# Визуализация

- Удобный конструктор создания веб-приложений.
- Возможность версионирования.
- Удобная привязка веб-приложения к таблицам.
- Поддержка различных уровней доступа, row-level security.
- Возможность экспорта результатов веб-приложения.
- Возможность использования результатов веб-приложения в пайплайне.



# Production



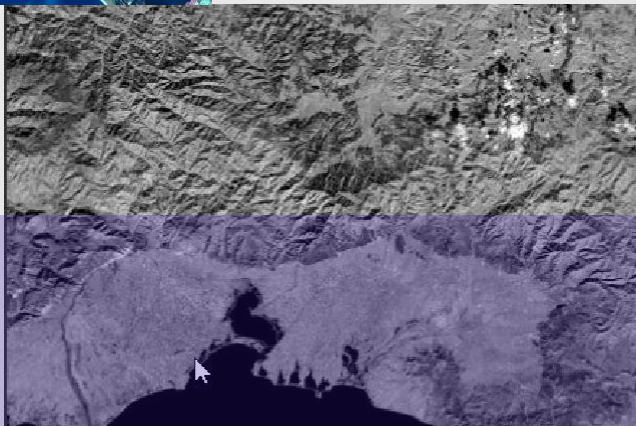
- Предоставление возможности создания low-code веб-приложений с помощью Javascript и SQL.
- Удобный способ привязки к различным датасетам.
- Поддержка различных версий веб-приложения.
- Возможность запуска пайплайнов и обучения моделей из веб-приложения.
- Возможность загрузки данных из веб-приложения.
- Большое количество разнообразных виджетов.

## Предложение

- Сделать платформу по работе с большими данными со Spark.
- С помощью этой платформы сделать проект по обработке и стандартизации контрактов.
- Адаптировать платформу по работе с большими данными к задаче ценообразования.



**Синтезируем науку, бизнес  
и технологии в эффективные ИТ-  
решения**



+7 812 611 06 86  
[info@syncretis.ru](mailto:info@syncretis.ru)

Санкт-Петербург  
Аптекарская набережная, 20 БЦ "Авеню"