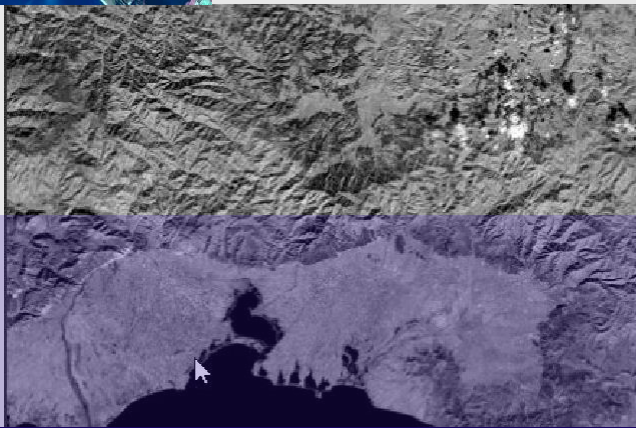




## DASY для РНПК



## Проблематика

- В Swiss Re есть большое количество данных в самых разных форматах. Задача приведения этих данных к единому формату, выделения из одного источника данных информации, связанной с другим источником, очень сложна.
- Swiss Re --- крупная перестраховочная компания, у которой большое количество клиентов --- страховых компаний. В период обновления страховых договоров Swiss Re нужна любая информация, связанная с портфелями компаний, чтобы помочь в принятии решения, продлевать ли договора, менять ли условия договоров.
- Страховые компании, клиенты Swiss Re, обладают большей информацией о своих клиентах, чем Swiss Re. Однако Swiss Re может использовать различные данные многих страховых компаний для того, чтобы устранить эту асимметрию в информации.

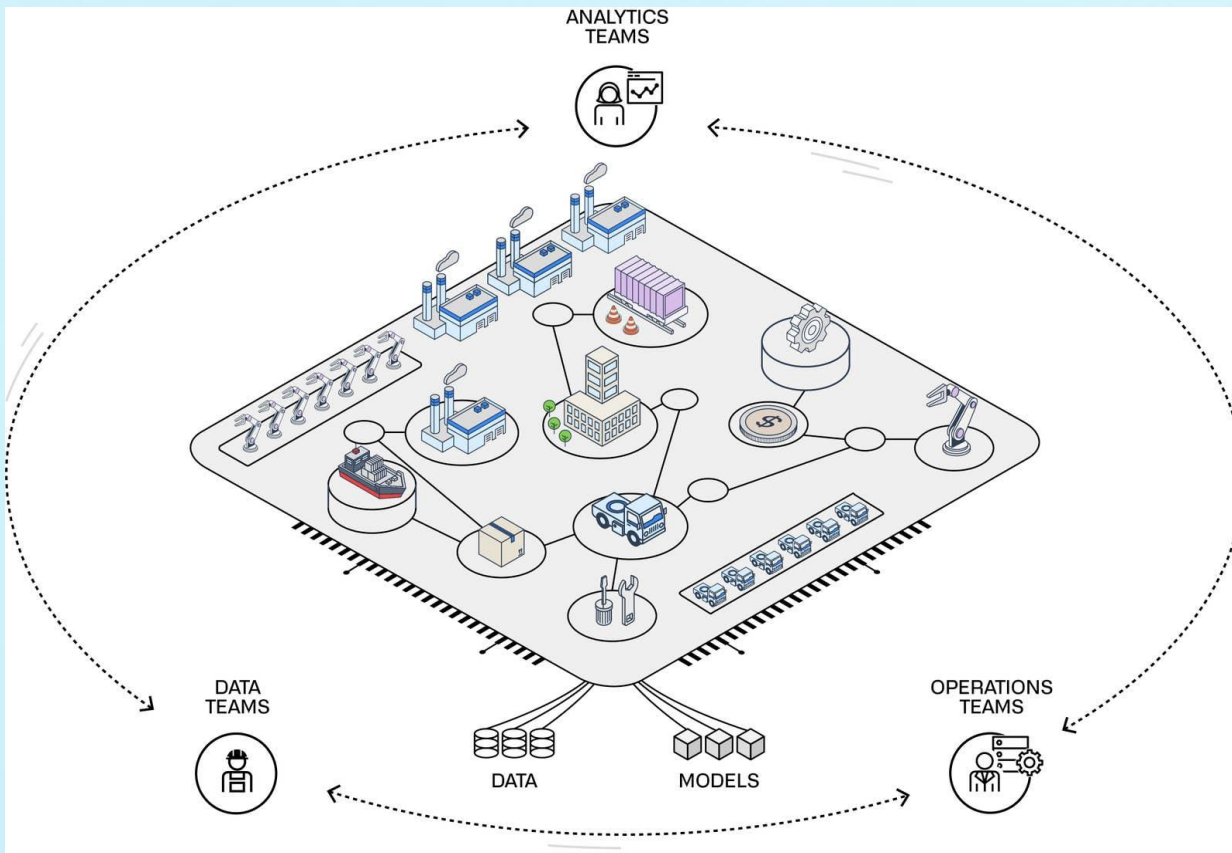
## Проект DASY в Swiss Re

- Проект DASY или Data Assymetry был призван устранить ассиметрию в информации между Swiss Re и её компаниями-клиентами, а также попытаться решить проблемы с предыдущего слайда.
- Проект работал с различными страховыми данными (бордеро, договора, контракты) в различных форматах (csv, xls, xlsx).
- Данные приводились к единому виду, объединялись и обратывались, насыщались другими источниками, анализировались.
- Данные к проекту поступали из разных источников. Часто напрямую от андеррайтеров. Были случаи, когда перед продлением договора проект сигнализировал о подозрительных компаниях, тем самым помогая в принятии решений андеррайтерам.

## Этапы развития DASY

- В проекте DASY можно чётко выделить 3 этапа развития.
- **Запуск на локальных машинах.** На этом этапе делалась библиотека по обработке и стандартизации данных, всё обрабатывалось локально.
- **Перенос на платформу StarGate.** Потом данных стало слишком много, возникло желание использовать источники данных, уже загруженные на платформу **StarGate**. Обработка и стандартизация данных проводились внутри платформы **StarGate**.
- **Развитие аналитического блока (Cyber-Dasy).** После того как процесс включения новых данных (бордеро) был хорошо отлажен (добавление нового бордеро и получение всех данных по нему занимало 1 день) основные усилия тратились на усиление аналитического блока, для того чтобы лучше помочь андеррайтерам.

## Особенности платформы StarGate



- 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.

# Загрузка и анализ данных

Загрузка и предварительный анализ данных проводились с помощью специальных приложений Contour и Data preparation.

- Поддерживаемые форматы: csv, xls, xlsx, pdf и другие похожие.
- Можно просматривать данные, переименовывать столбцы, менять их типы.
- Поддержка основных операций обработки данных.
- Поддержка большого количества графиков.
- Возможность экспорта на любом этапе анализа данных.
- Учёт прав доступа пользователей (row-level security).

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

	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	

This dataset has 10 columns and 45,716 rows.

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

Yamato 792753 1

Diep River 1

Tambakwatu 1

Wuzhi 1

Acfer 216 1

Acfer 290 1

Allan Hills 83017 1

Allan Hills 84050 1

Allan Hills 84065 1

Allan Hills 84134 1

Allan Hills 84179 1

Allan Hills 84198 1

Allan Hills 84210 1

Allan Hills 85009 1

Allan Hills A77094 1

Allan Hills A77139 1

Allan Hills A78009 1

CHANGE DATA (FOR ENTIRE DATASET)

New value

Find and replace

Map

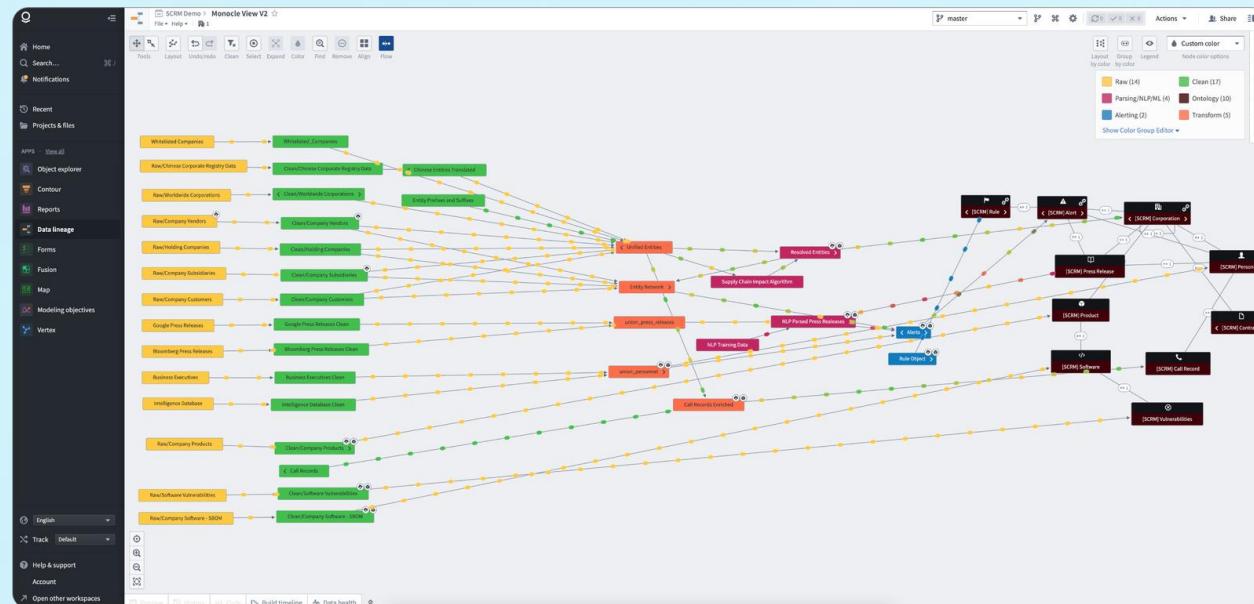
Keep or remove

Extract

Concatenate



# Обработка данных



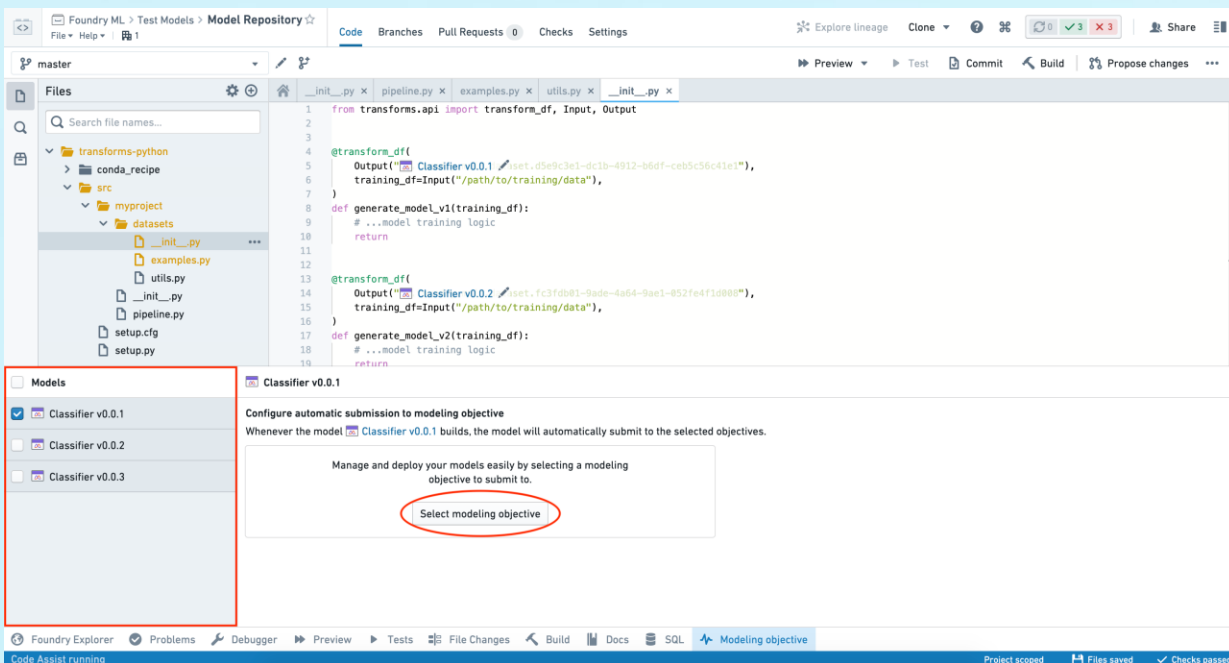
Таблиц было достаточно много. Поэтому требовалось удобное средство просмотра связей между ними lineage graph.

- В этом приложении можно было запускать пайплайн.
- Можно было писать небольшое количество кода.
- Была поддержка версионирования и сохранения промежуточных шагов.

# Моделирование

Для объединения с другими источниками данных применялись data science модели. Кроме того, они применялись для прогнозирования.

- Всё это делалось в специальных приложениях (code repositories, code workbooks).
- Была поддержка различных библиотек по машинному обучению.
- Был удобный способ хранения, сравнения и тестирования моделей.
- Модели могут использовать все возможности Spark для ускорения вычислений.
- Low-code решение.

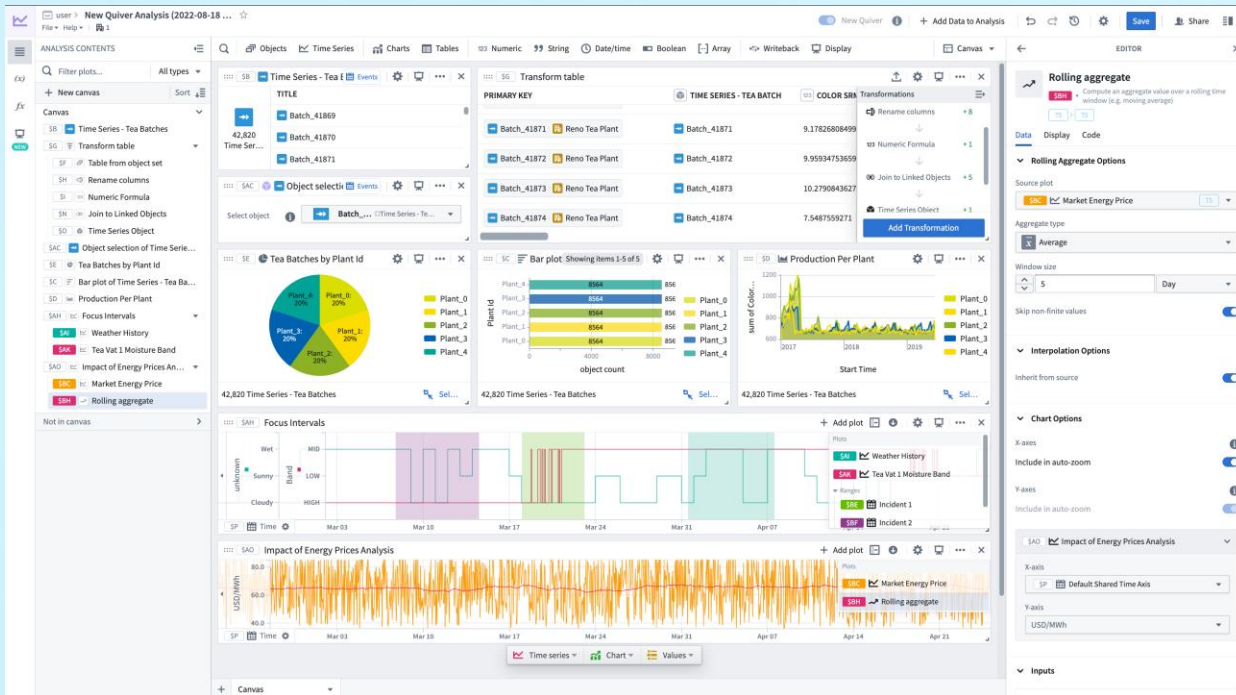




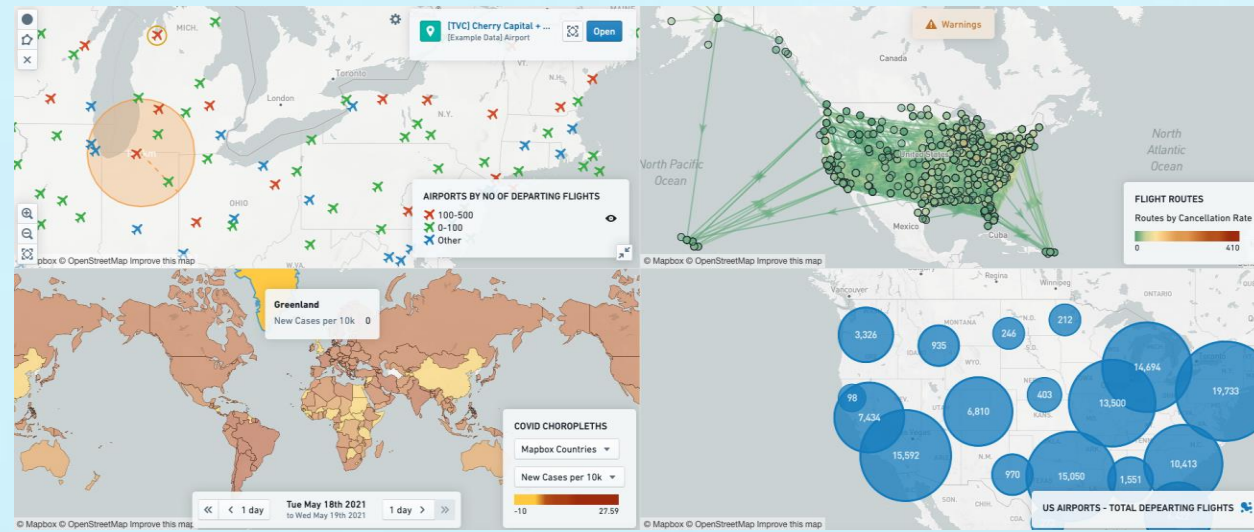
# Визуализация

Визуализация проводилась с помощью приложения под названием Quiver. Для его использования не была нужна особая подготовка.

- Это удобный конструктор создания веб-приложений.
- Реализована удобная привязка веб-приложения к таблицам.
- Есть поддержка различных уровней доступа, row-level security.
- Есть возможность экспорта результатов веб-приложения.
- Есть возможность использования результатов веб-приложения в пайплайне.



# Production



При выходе в production в команде проекта появились профессиональные FE-разработчики, поэтому перешли с Quiver на Slate.

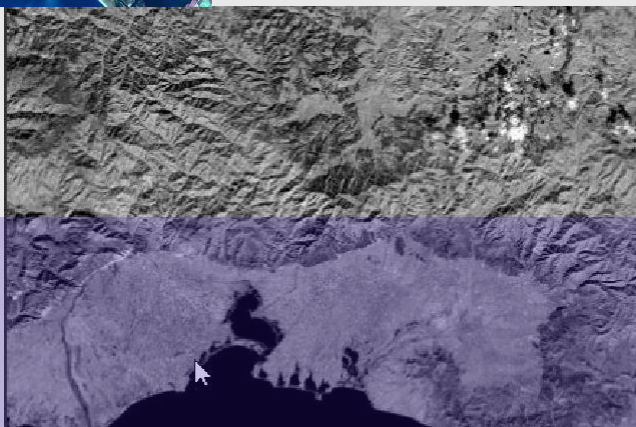
- Slate отличает предоставление возможности создания low-code веб-приложений с помощью Javascript и SQL.
- Есть удобный способ привязки к различным датасетам.
- Есть поддержка различных версий веб-приложения.
- Есть возможность запуска пайплайнов и обучения моделей из веб-приложения.
- Есть возможность загрузки данных из веб-приложения.
- Есть большое количество разнообразных виджетов.

## Предложение

- У РНПК есть большое количество данных в разных форматах (бордеро, договора, статистические данные).
- Можно реализовать проект DASY для РНПК, ориентируясь на эти форматы данных, с прицелом на задачу ценообразования.
- Можно двигаться по этапам развития проекта DASY, притом после 1-го этапа развития понадобится имплементация частей платформы StarGate со Spark.



**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"