MAKING A SINGLE DRUG TODAY

$2.6 BILLION

15 YEARS
Deliver to patients highly effective small-molecule drugs, unlikely to be discovered in any other way.
Enable efficient and accurate exploration of vast chemical space containing $10^{12}$ unique small-molecules, to support discovery of novel drug candidates.
Transition from a sequational, “greedy algorithm” approach, optimizing individual drug properties at a time, to a multi-parameter optimization, co-optimizing for multiple steps in parallel.
DEEPCURE’S PLATFORM ANALYZES A SPACE OF ONE TRILLION COMPOUNDS AND ACCOUNTING FOR MULTIPLE DRUG PROPERTIES WHICH IS 10,000 TIMES LARGER THAN THE BIGGEST VIRTUAL SEARCH EVER DONE

Such a campaign would have taken 2 weeks of world compute and a Billion USD with any existing tool.
Project 1: Web application for scheduling, monitoring and analyzing cloud-based hyper-parameter optimization runs

**Motivation:** At DeepCure, we train and evaluate hundreds of machine learning models for different projects every day. We have developed algorithms that use cloud-based, distributed computing to perform efficient hyper-parameter optimization, i.e. selecting the best parameters for each model. Currently the algorithms are run by command line scripts, and the results are analyzed by manual inspection of a database.

**Goal:** Have a web-based platform that allows the machine learning scientists at DeepCure to schedule, monitor and analyze the results of a hyper-parameter optimization run. We expect the platform to minimize the overhead in problem solving and increase the speed of our hypothesis testing cycle.

**Programming Languages:** Python, Javascript, Node.JS

**Frameworks:** Flask, Vue.JS, Amazon Web Services (optional: Plotly.JS, D3.JS)
Project 2: RESTful API for executing complex, AI-driven queries on hybrid columnar- and row-format data stores

**Motivation:** At DeepCure, we are constantly generating data, computationally and experimentally, and storing them in our databases. The data types can range from simple integers to complex array structures. The stored data can be used to train machine learning models, and/or generate insights for decision making in the drug discovery process. Currently, the various databases sit behind disparate APIs, leading to the segregation of data analysis and preparation routines into multiple separate queries and ETL (extract-transform-load) workflows.

**Goal:** Have a web application with a RESTful API that can receive, parse and execute complex queries on the DeepCure data warehouse, and return the results. The queries can range from simple SQL to arbitrary UDFs (user-defined functions).

**Programming Languages:** Python, SQL (optional: Javascript, Node.JS)

**Frameworks:** Flask, Apache {Spark, Hadoop, Impala}, Pandas