

Deploying the Python army in your digital water battlefield

CASE STUDY

By Greg Johnston

Digitization has arrived as a battle cry in the daily operations of water utilities everywhere. Thousands of sensors are pumping out millions of time-series data points, vendor solutions compete to provide reams of analytic reports, and vast amounts of public data beckon that can inform wastewater flows, moisture impacts and other vital parameters.

There's a ready, willing and able army of data-driven engineers armed with Python, Julia and R programming skills, keen to using predictive analytics and Machine Learning (ML) automation to tackle the challenges wrought by climate change-induced storm events, urban population concentration and ageing infrastructure.

What top-gun coders do at Christmas break

It's worth noting Python's popularity in particular. With a name inspired by the surreal TV sketches and movies of the British comedy troupe Monty Python, Python is a high-level open source programming language conceived by the Dutch programmer Guido Van Rossum in the late 80s - during his Christmas break at university!

While he has since pursued a brilliant technical career working in the coding armies at Google, DropBox and most recently, Microsoft, he shouldered lead developer responsibilities for Python as "benevolent dictator for life". This title was bestowed upon him after he passed his torch in 2019 to a five-member Steering Council that now leads the Python project.

Unstoppable Python rated world's most popular programming language

The August, 2022 edition of InfoWorld reported that "unstoppable Python" topped the chart as the world's most popular programming language, registering a 15.42% market share - an all-time high for the language. It's the only programming language besides Java and C to hold that No. 1 position.

Python courses are ubiquitous at Canadian college and university engineering programs not to mention those at MIT, Stanford and Carnegie Mellon. And the Coursera "IT Automation Python Professional Certificate" has a current enrollment of over 450,000 students who pay an average of US\$300 to complete that course in six months.

Extraordinary open source ML libraries

Python power and popularity has been built with extraordinary open source Python libraries for data science that are used by programmers every day in solving problems. For the uninitiated, Python libraries are collections of modules that contain useful codes and functions, eliminating the need to write them from scratch. There are tens of thousands of Python libraries that help ML developers, data science and data visualization engineers, and more.

Python is the preferred language for machine learning because its syntax and commands are closely related to English, making it efficient and easy to learn. Compared with C++, R, Ruby, and Java, Python is one the simplest languages, enabling accessibility, versatility, and portability. It can operate on nearly any operating system or platform.

Data modelling algorithms in Scikit-learn

As an example of vast Python resources, Scikit-learn is an actively used library for ML. It includes easy integration with other ML programming libraries like NumPy and Pandas. Scikit-learn comes with the support of various algorithms such as classification, regression, clustering, dimensionality reduction, model selection, and preprocessing.

Built around the idea of being easy to use but still be flexible, Scikit-learn is focussed on data modelling. It is considered sufficient enough to be used as an end-to-end ML toolkit, from the research phase to the deployment.

Pandas library for data manipulation and analysis

Pandas is another notable Python library. It is primarily used for data manipulation and analysis. Pandas make working with time series and structured multidimensional data effortless for machine-learning programmers. Some of the great features of Pandas when it comes to handling data are dataset reshaping and pivoting, merging and joining of datasets, handling of missing data and data alignment, hierarchical axis indexing, fancy indexing, and data filtration options.

As founder of the company that brought the flowworks flow monitoring, analysis and reporting software platform to life over 70 Smart City water utilities including Los Angeles County, Toronto, Seattle, and Dallas, I led a four-million-dollar Digital Supercluster R&D consortium project that brought our technical team deep into Python's potential for ML and other advanced data calculation applications.

Fresh Water Data Commons Project Python roots

The Fresh Water Data Commons project, now deployed in Andersen Creek in the Columbia Basin near the city of Nelson, BC, uses an ML-enhanced version of the flowworks platform to transform data using Python scripts to build real-time analytics, predictive analytics and ML models.

The objective of the Canadian Federal Government’s Digital Supercluster initiative is to identify global industry leaders and where Canadians benefit from the prosperity and growth that comes from creating new products and services that are meaningful across the country and around the world. We did precisely that with a product launched at WEFTEC 2022 called face pro. Some of our Beta customers – among them the world’s leading engineering services outfits – refer to it as “the Python tamer”.

The Python tamer – face pro

A well-used toolset within flowworks, face stands for “flowworks advanced calculation engine”. Using face, municipal engineers create edit and combine data channels with mathematical functions to face empowers users with data manipulation tools, allowing them to create and define new datasets from incoming raw channels using advanced math, statistics and logic equations. Users can create, edit, delete and combine data channels with powerful mathematical functions for sophisticated real-time analysis.

With face, users can do things like convert measurement units, define lookup tables for weirs, flumes, and pipe cross-sectional areas, create rolling averages and sums, build time-weighted averages of irregularly spaced data, and move data forward or backward in time to compare it with previously collected datasets.

Cut-and-paste Python scripts

Infinittii face pro, as the name suggests, takes advanced calculations and related automation to another level.

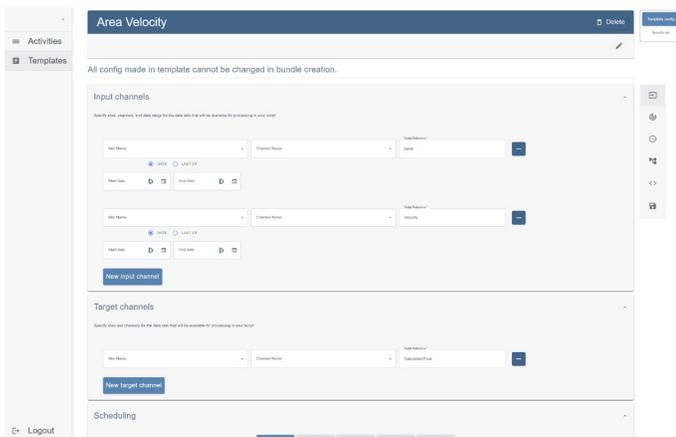


Figure 1: infinittii face pro input data channel and target output data channel for Area Velocity

Users can literally cut and paste Python (or Julia or R) scripts into flowworks production systems and then seconds later ML algorithms defined by those scripts are running system wide. After a recent demo of infinittii face pro, a senior software developer working at one of our flowworks engineering services partners declared that “I can do my entire job inside this software!”

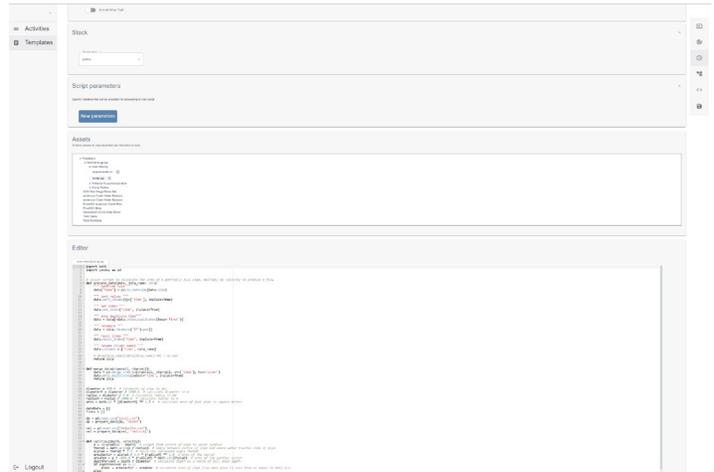


Figure 2: infinittii face pro Python script automation interface

With infinittii face pro, flowworks users can easily deploy existing algorithms and calculations from the vast Python libraries technologists use every day. It’s as simple as cutting and pasting Python code into an intuitive interface, and then minutes or even seconds later seeing that code run in production systems.

Predictive maintenance, failure prediction and more

Use cases for this Python-taming calculation engine include forecasting, anomaly detection, predictive maintenance and failure prediction. The types of advanced calculations easily performed with infinittii face pro include Soil and Water Integrated Model (SWIM) calculations that track and predict climate and land use change impacts at a regional scale, and Evapotranspiration (ET) calculations used to estimate soil-moisture storage based on precipitation deficit and the maximum water-holding capacity of the soil.

Breaks down versus puts up barriers to other data

Working alongside the flowworks platform, infinittii face pro provides a utility-wide data monitoring and reporting environment with state-of-the-art streaming data calculations that take advantage of the latest ML algorithms available in the vast Python libraries or created in-house to solve very specific problems. Significantly, flowworks “plays well with others,” and breaks down versus puts up barriers to working with data provided by other vendors.

You can plot several channels in parallel on a time series. A scatter graph requires at least 2 and at most 5 channels to be selected.

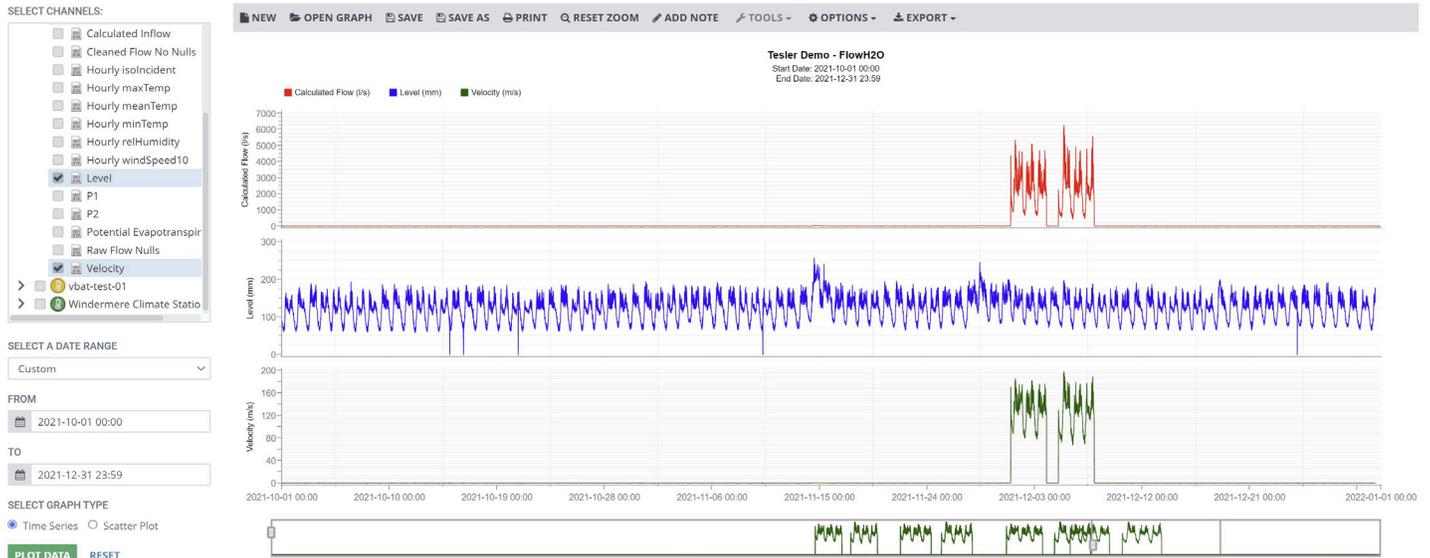


Figure 3: infinitii face pro Area Velocity results graph calculated from new data channel

Flowworks accepts all types of data from any source. Real-time, historic, wireless, satellite, SCADA, public data sets including USGS, NOAA and weather forecasts – it doesn't matter where the data originates or from which sensor. You can connect it. Customers use flowworks daily to monitor water flow, rainfall, and other vital parameters in thousands of collection system metering points.

Eric Corey, the Smart Wastewater Network Project Manager at Core & Main, has stated that with flowworks, "We now have the unique ability to treat data like data, not like countries."

flowworks as data Switzerland

He continued, saying "Given the complexity of the Smart Wastewater Network and the number of different software solutions being utilized, what flowworks provides us is a data Switzerland. This neutrality gives a required level of comfort to our other technology and manufacturing partners that allows us to truly innovate with their technologies versus being locked in or locked out from utilizing strategically important data."

Infinitii face pro users working with the flowworks platform can build analytic models and transform raw sensor data in order to generate new data or output events in near real-time. They can simultaneously filter, aggregate, enrich and analyze high volumes of time-series data from multiple sources, then present it using flowworks graphing and Geographic Information System (GIS) mapping tools like K2 Geospatial's JMap. Flowworks alarming and notification tools are readily employed, and infinitii face pro can also leverage Microsoft Power BI Dashboards.

Greg Johnston is President of Infinitii ai, a leader in predictive analytics for industrial and Smart City infrastructure applications that rely on time-series data. For more information visit www.infinitii.ai.