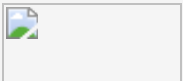# Pydantic v2

Technical hour
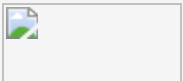
Sunniva Indrehus

Pydantic v2

Technical hour

Sunniva Indrehus

```
$ whoami
```

NGI scientist's aim
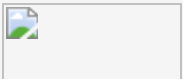
Understand a simplified version
of the real world

Engineering example

Define a valid **data model**
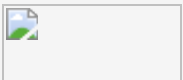
Perform operation X

Iterpret results

Pydantic

💥 v2.0 released 30.06.2023 💥

What's the fuzz?

v2 promise 5-50x speed up compared to Pydantic v1

⭐ GitHub stars 15.4k (22.08.23)
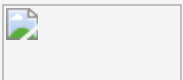
📦 pr week ≈ 22M (week 34 2023)

Pydantic

Why?

🐍 Python is dynamically typed, and PEP484 Type Hints is not enforced during run time

What?

> Pydantic is the most widely used data validation library for Python.
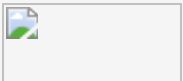
From the official docs

Pydantic v2

A new core

Validations outside of Python

Recursive function calls with Rust
and small overhead

Figure credit: *Prashanth Rao*

Pydantic v2

(Some) new features

- Functionality for discriminated unions

- `pydantic.functional_validators` let you do validation without `base_model` (*e.g* a function)
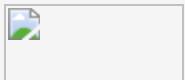
- Possibility to use `TypeAdapter` instead of `BaseModel`

- You can define `field_serializer` you can do *custom serialization*

Examples and inspiration

- Migration guide

- v1 to v2 video example

Let's get our hands dirty

Data

Wine reviews from Kaggle

```
{'country': 'France',
 'description': 'Ripe in color and aromas, this chunky wine delivers heavy '
                'baked-berry and raisin aromas in front of a jammy, extracted '
                'palate. Raisin and cooked berry flavors finish plump, with '
                'earthy notes.',
 'id': 45100,
 'points': 85,
 'price': 10.0,
 'province': 'Maule Valley',
 'taster_name': 'Michael Schachner',
 'taster_twitter_handle': '@wineschach',
 'title': 'Balduzzi 2012 Reserva Merlot (Maule Valley)',
 'variety': 'Merlot',
 'vineyard': 'The Vineyard',
 'winery': 'Balduzzi'}
```
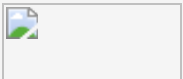
Data model

Pydantic's job

Ensure the `id` field always exists (this will be the primary key), and is an integer

Ensure the `points` field is an integer

Ensure the `price` field is a float

Ensure the `country` field always has a non-null value – if it's set as null or the country key doesn't exist in the raw data, it must be set to Unknown. This is because the use case we defined will involve querying on country downstream

Remove fields like designation, `province`, `region_1` and `region_2` if they have the value null in the raw data – these fields will not be queried on and we do not want to unnecessarily store null values downstream

Demo

Speed-up for *free* 🔓

v1

```
root_validator
```
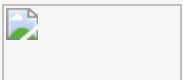
```
wine = Wine(**sample_data)
```

```
pprint(wine.dict(exclude_none=True,
by_alias=True))
```

v2

```
@model_validator/@field_validator
```

```
wine = Wine(**sample_data)
```

```
pprint(wine.model_dump(exclude_none=True,
by_alias=True))
```
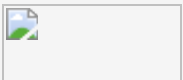
Speed-up with *investements* 💰

v2

```
class
Wine(BaseModel)
```

```
wine =
Wine(**sample_data)
```

v2-optimized

```
class Wine(TypedDict)
```

```
wines =
WinesTypeAdapter.validate_python([sample_data])
```
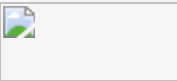
Demo some benchmarking

Example run times

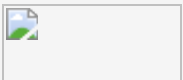129971 records · 5 = 649855 validation operations

	v1	v2	v2-optimized
All cases	45.380	6.944	3.582
Average pr run	9.076	1.389	0.716
Times speed up (v1)	1	6.534	12.678

Sometimes tricky with linting

`mypy` **integration**

[Explanation](#) of why `TypedDict` work with a `field_validator` , but violates PEP589

Credit

Pydantic v2 Plan (accessed 22.08.2023)

Terrence Dorsey and Samuel Colvin Pydantic v2 Pre Release (accessed 22.08.2023)

Prashanth Rao, The Data Quarry, Obtain a 5x speedup for free by upgrading to Pydantic v2 (accessed 22.08.2023)

Yaakov Bressler in Medium (accessed 22.08.2023)

Samuel Colvin at PyCon US 2023 (accessed 22.08.2023)

Configuring mypy with Pydantic (accessed 22.08.2023)