Yandex Weather

How to combine physical models, machine learning and production performance

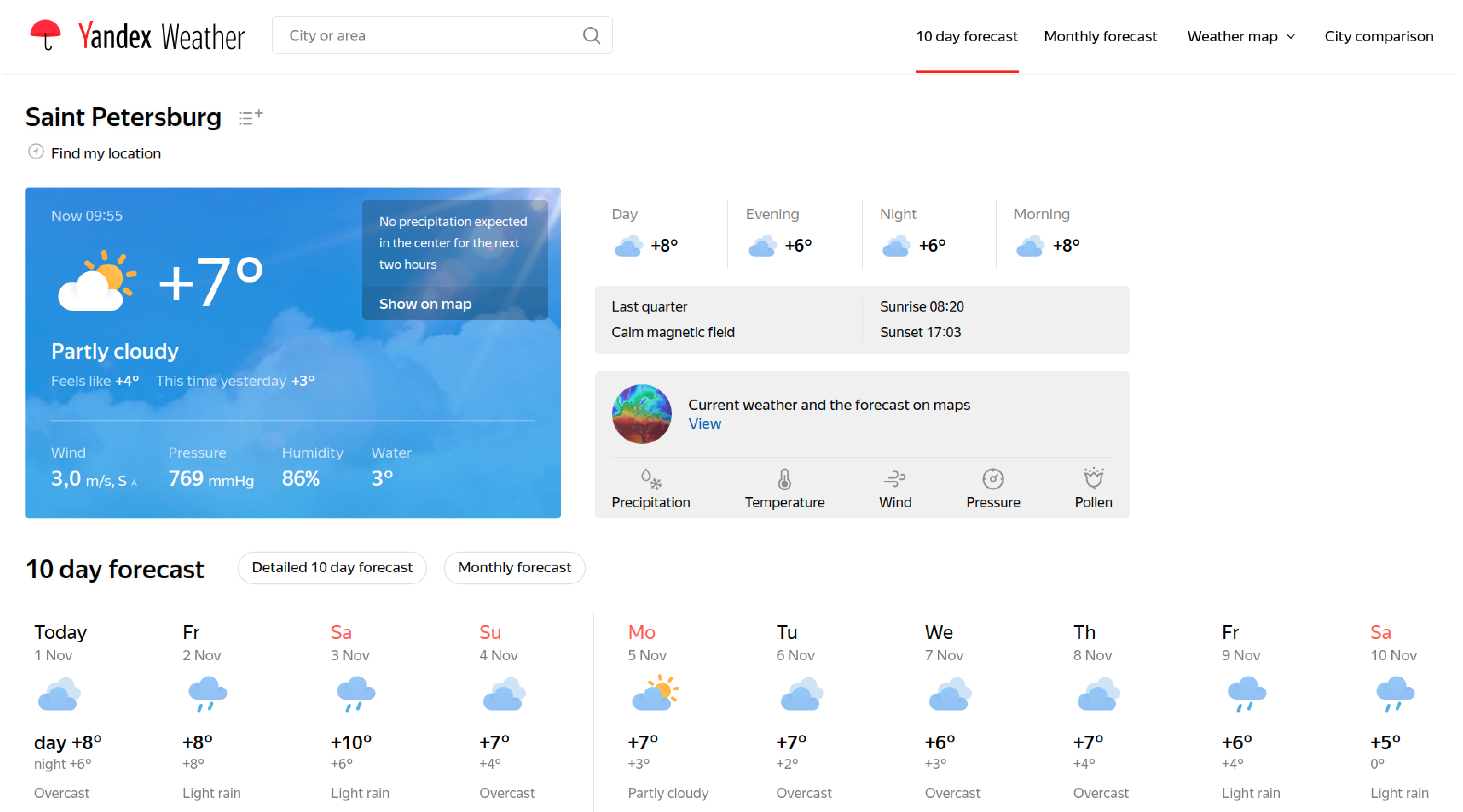
Lena Volzhina
Developer @ Yandex.Weather

Overview

› What is Yandex.Weather
› Forecasting with machine learning
› Architecture of the service
› Fast calculation of forecasts

Overview

- **What is Yandex.Weather**
- Forecasting with machine learning
- Architecture of the service
- Fast calculation of forecasts



Yandex Weather

City or area

10 day forecast | Monthly forecast | Weather map | City comparison

Saint Petersburg

Find my location

Now 09:55

+7°

Partly cloudy

Feels like +4° This time yesterday +3°

No precipitation expected in the center for the next two hours

Show on map

Wind	Pressure	Humidity	Water
3,0 m/s, S	769 mmHg	86%	3°

Day	Evening	Night	Morning
+8°	+6°	+6°	+8°

Last quarter
Calm magnetic field

Sunrise 08:20
Sunset 17:03

Current weather and the forecast on maps
View

Precipitation | Temperature | Wind | Pressure | Pollen

10 day forecast

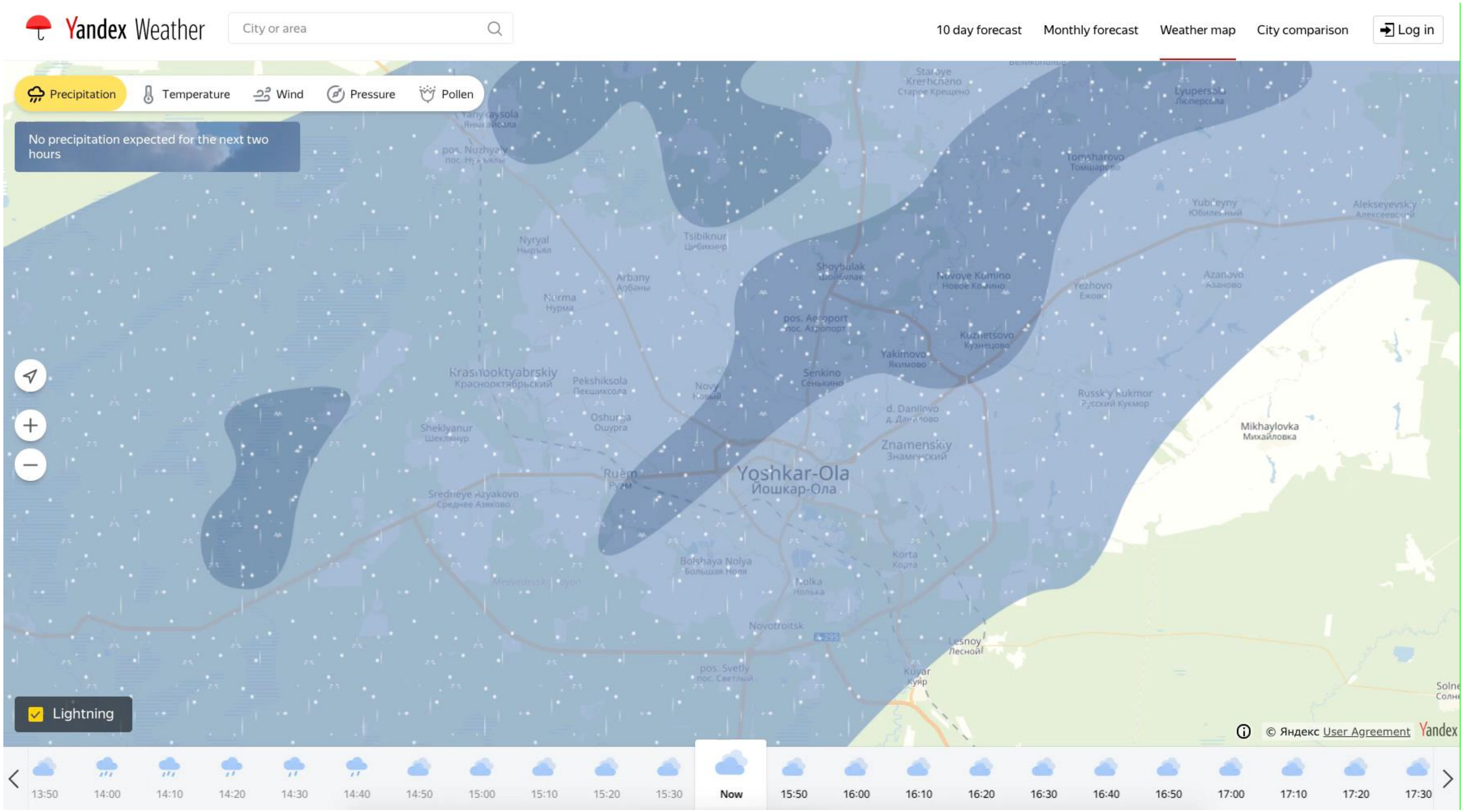
Detailed 10 day forecast | Monthly forecast

Today	Fr	Sa	Su	Mo	Tu	We	Th	Fr	Sa
1 Nov	2 Nov	3 Nov	4 Nov	5 Nov	6 Nov	7 Nov	8 Nov	9 Nov	10 Nov
day +8°	+8°	+10°	+7°	+7°	+7°	+6°	+7°	+6°	+5°
night +6°	+8°	+6°	+4°	+3°	+2°	+3°	+4°	+4°	0°
Overcast	Light rain	Light rain	Overcast	Partly cloudy	Overcast	Overcast	Overcast	Light rain	Light rain

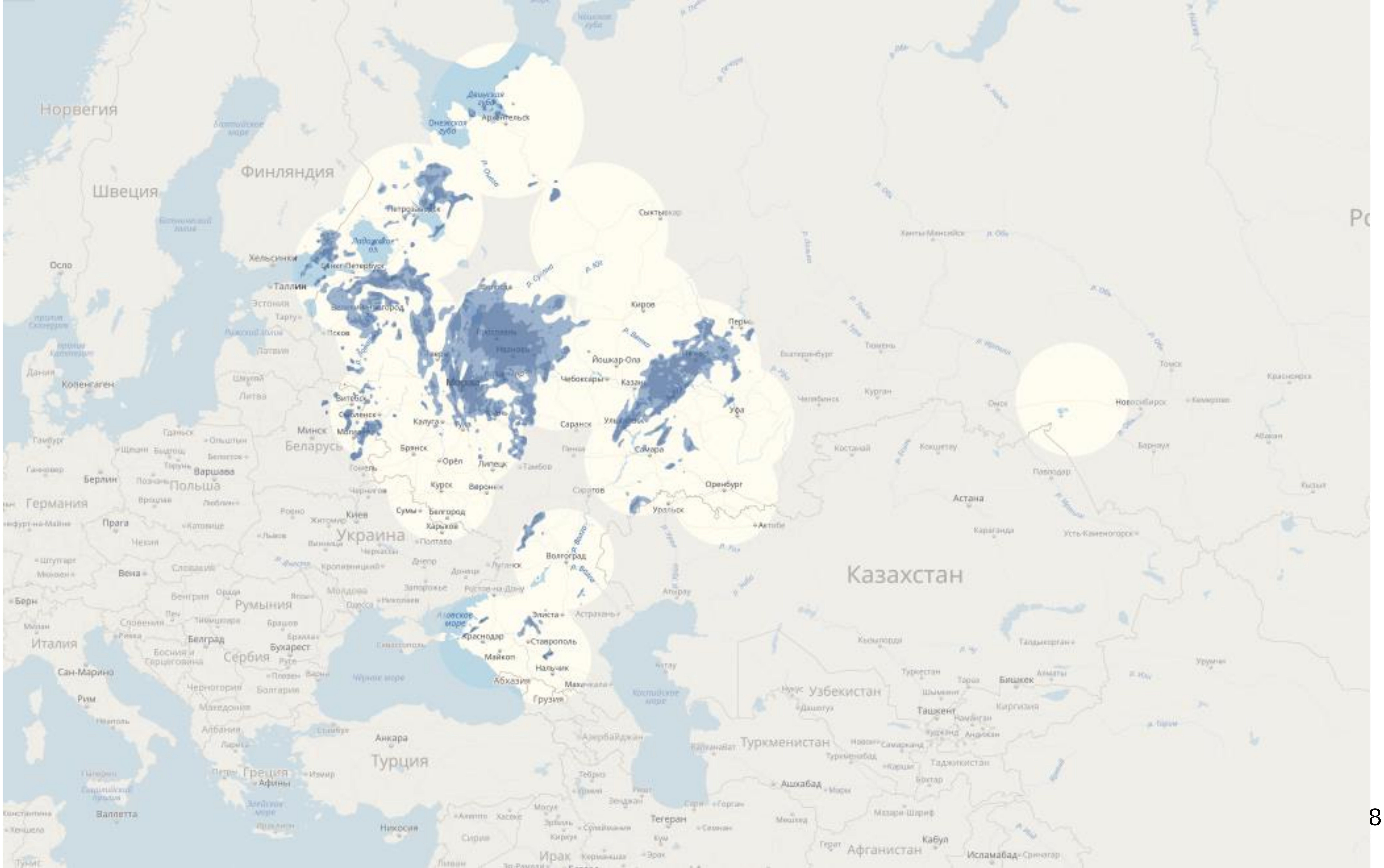
https://yandex.com/weather/saint-petersburg/

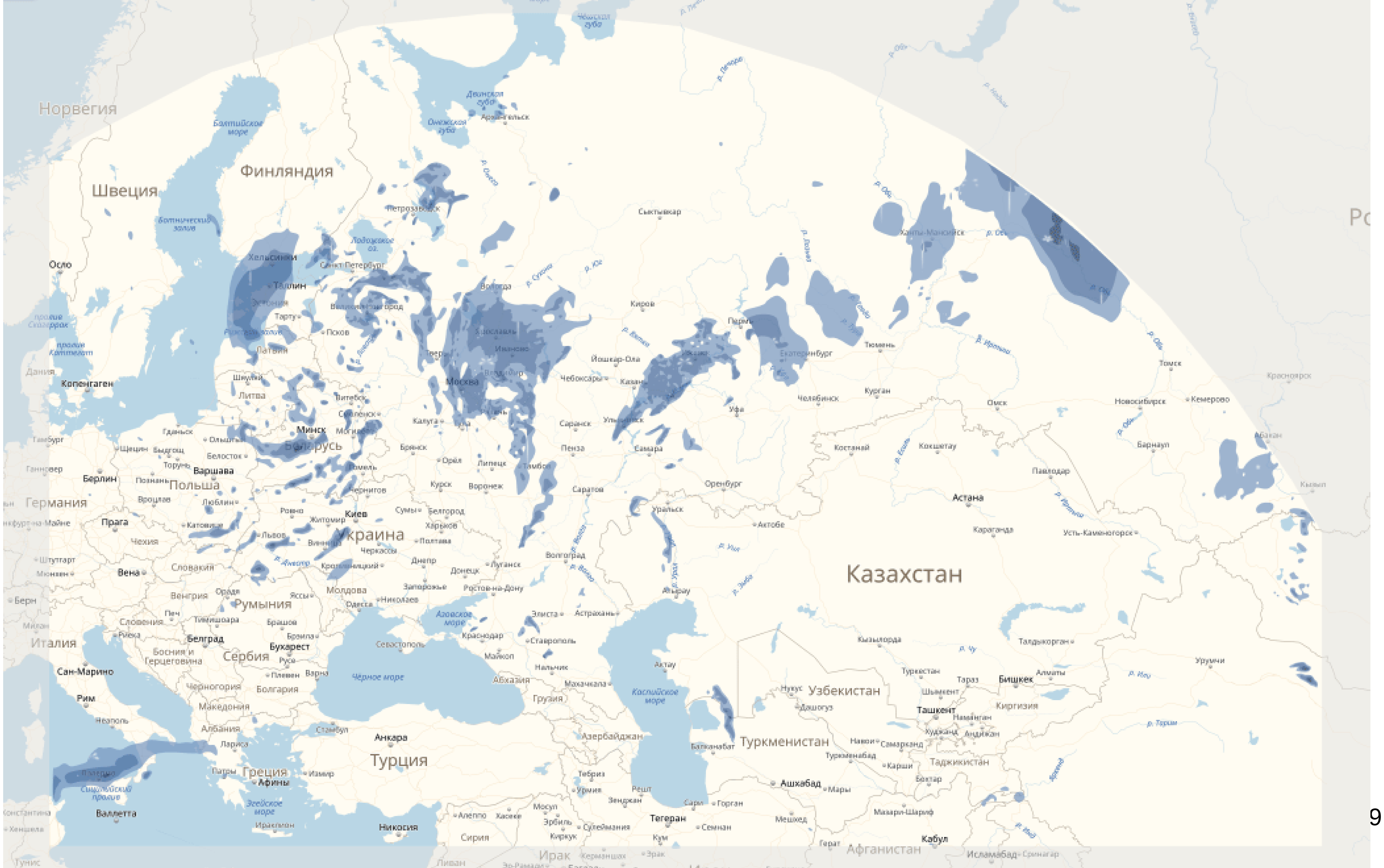
Yandex Weather
City or area
10 day forecast
Monthly forecast
Weather map
City comparison
Log in
Precipitation
Temperature
Wind
Pressure
Pollen
Markers
Animated wind
-40 -30 -20 -10 0 10 20 30 40 °C
Wed, 31 06:00 09:00 12:00 15:00 18:00 21:00
Thu, 1 00:00 03:00 06:00 09:00 12:00 15:00 18:00 21:00
Fri, 2 00:00 03:00 06:00 09:00 12:00 15:00 18:00 21:00
Sat, 3 00:00 03:00
© Яндекс User Agreement
Yandex

Yandex Weather
City or area
10 day forecast
Monthly forecast
Weather map
City comparison
Log in
Precipitation
Temperature
Wind
Pressure
Pollen
Animated wind
525 613 678 703 734 737 755 780 mmHg
Russia
Kazakhstan
Mongolia
China
Finland
Sweden
Norway
Ukraine
Belarus
Poland
Germany
France
Spain
Italy
Romania
Turkey
Iran
Moscow
Wed, 31 06:00 09:00 12:00 15:00 18:00 21:00
Thu, 1 00:00 03:00 06:00 09:00 12:00 15:00 18:00 21:00
Fri, 2 00:00 03:00 06:00 09:00 12:00 15:00 18:00 21:00
Sat, 3 00:00 03:00
© Яндекс User Agreement



No precipitation expected for the next two hours


Норвегия
Финляндия
Швеция
Архангельск
Петрозаводск
Санкт-Петербург
Хельсинки
Таллин
Осло
Эстония
Великий Новгород
Псков
Латвия
Литва
Копенгаген
Дания
Сыктывкар
Киров
Пермь
Ярославль
Иваново
Йошкар-Ола
Чебоксары
Казань
Москва
Тверь
Витебск
Смоленск
Минск
Калуга
Тула
Беларусь
Брянск
Орёл
Липецк
Тамбов
Саранск
Ульяновск
Самара
Уфа
Пенза
Саратов
Оренбург
Уральск
Актобе
Курск
Воронеж
Белгород
Харьков
Сумы
Киев
Украина
Полтава
Волгоград
Элиста
Астрахань
Ставрополь
Краснодар
Майкоп
Нальчик
Махачкала
Абхазия
Грузия
Польша
Варшава
Берлин
Германия
Прага
Вена
Румыния
Бухарест
Сербия
Белград
Италия
Рим
Турция
Анкара
Греция
Афины
Казахстан
Астана
Новосибирск
Екатеринбург
Челябинск
Курган
Тюмень
Омск
Узбекистан
Ташкент
Туркменистан
Ашхабад
Бишкек
Киргизия
Таджикистан
Тегеран
Афганистан
Кабул
Исламабад



What is Yandex.Weather

- Over 10 million users per week
- More than 25k RPS to API

- Tbs of numeric data daily
- Dozens of ML models calculated

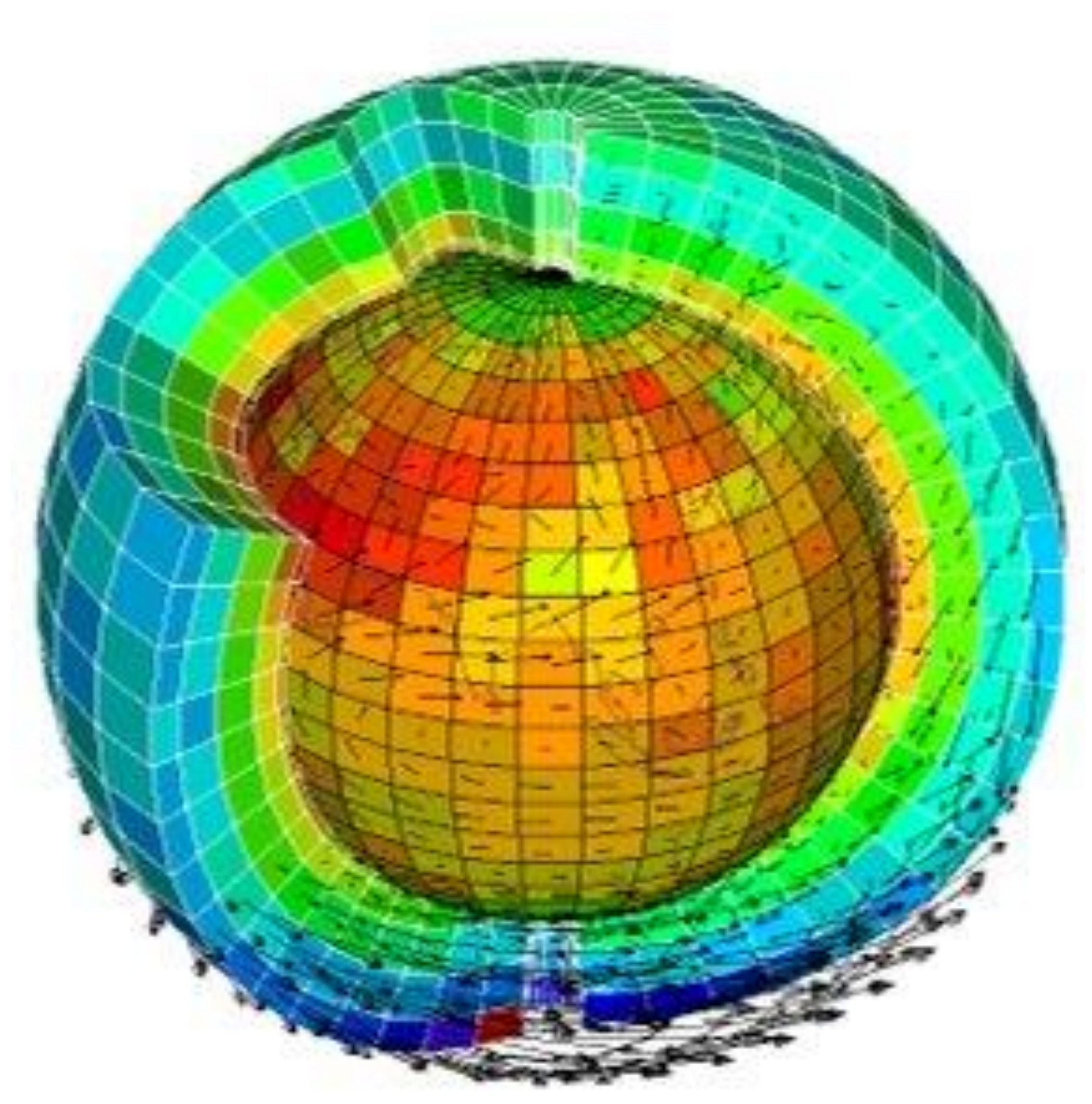
Overview

› What is Yandex.Weather
› **Forecasting with machine learning**
› Architecture of the service
› Fast calculation of forecasts

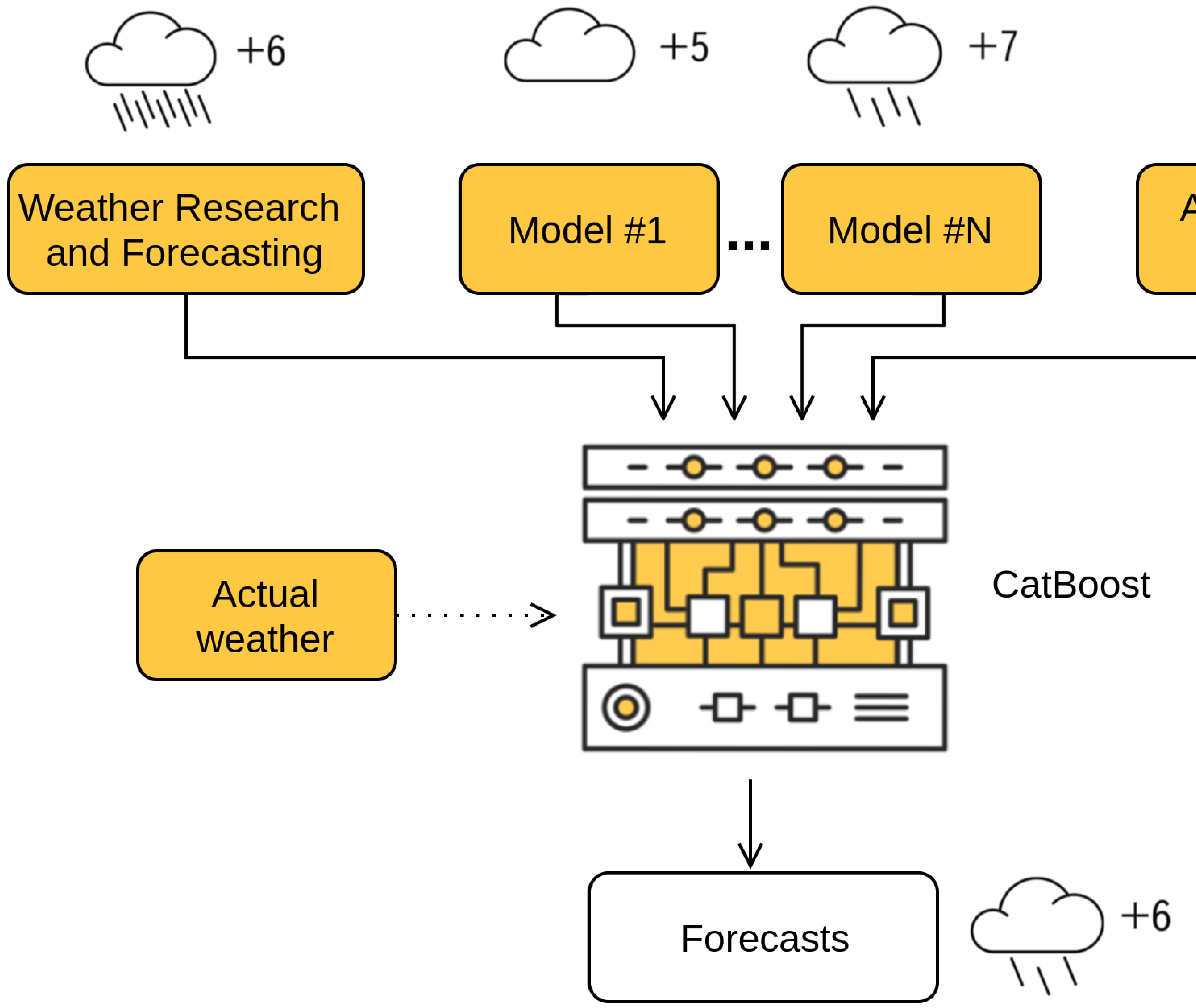
Forecasting with machine learning

Regular forecasts

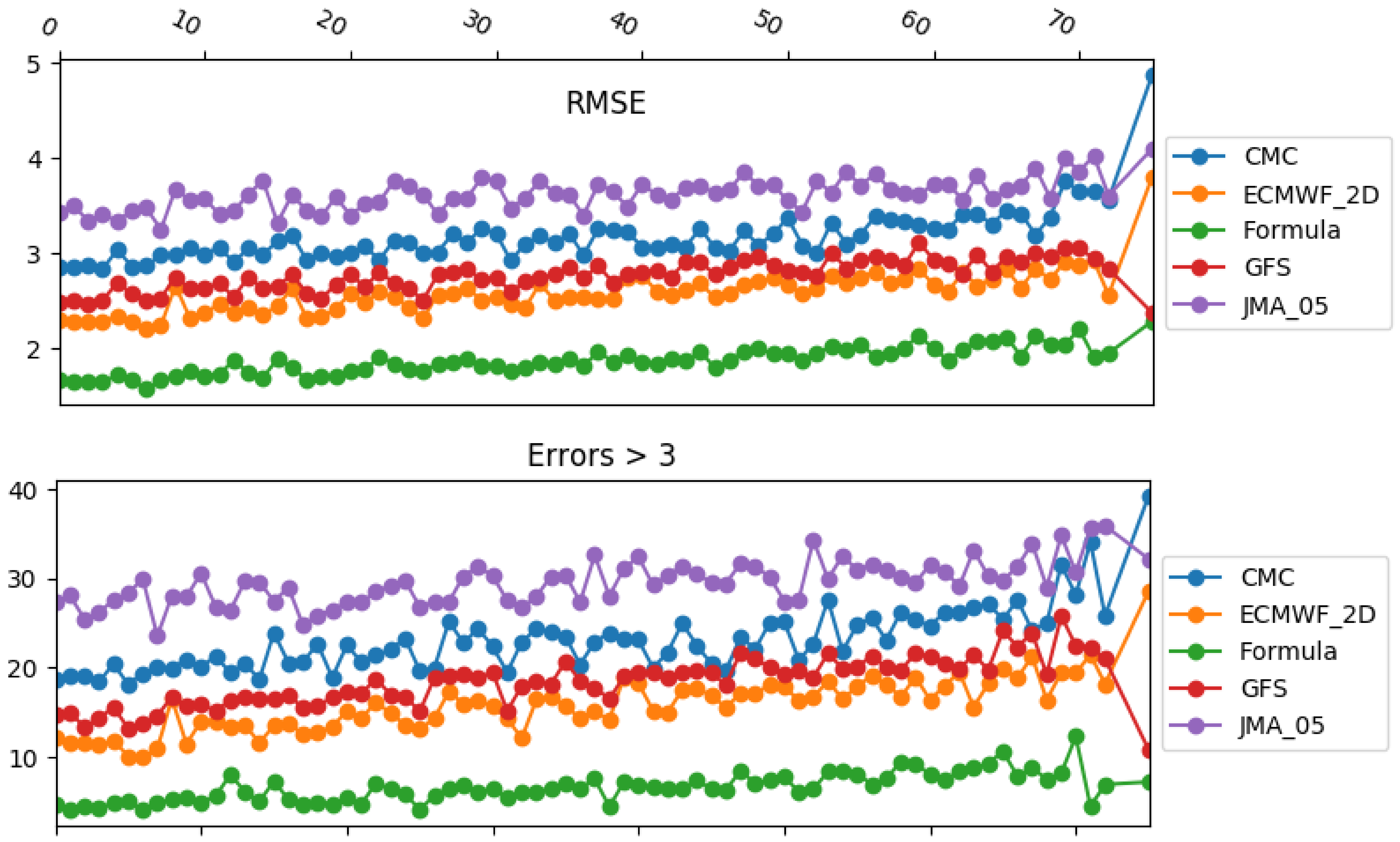
- Numerical weather prediction
- Global / regional models
- Gradient boosting over decision trees: CatBoost



Forecasting with machine learning



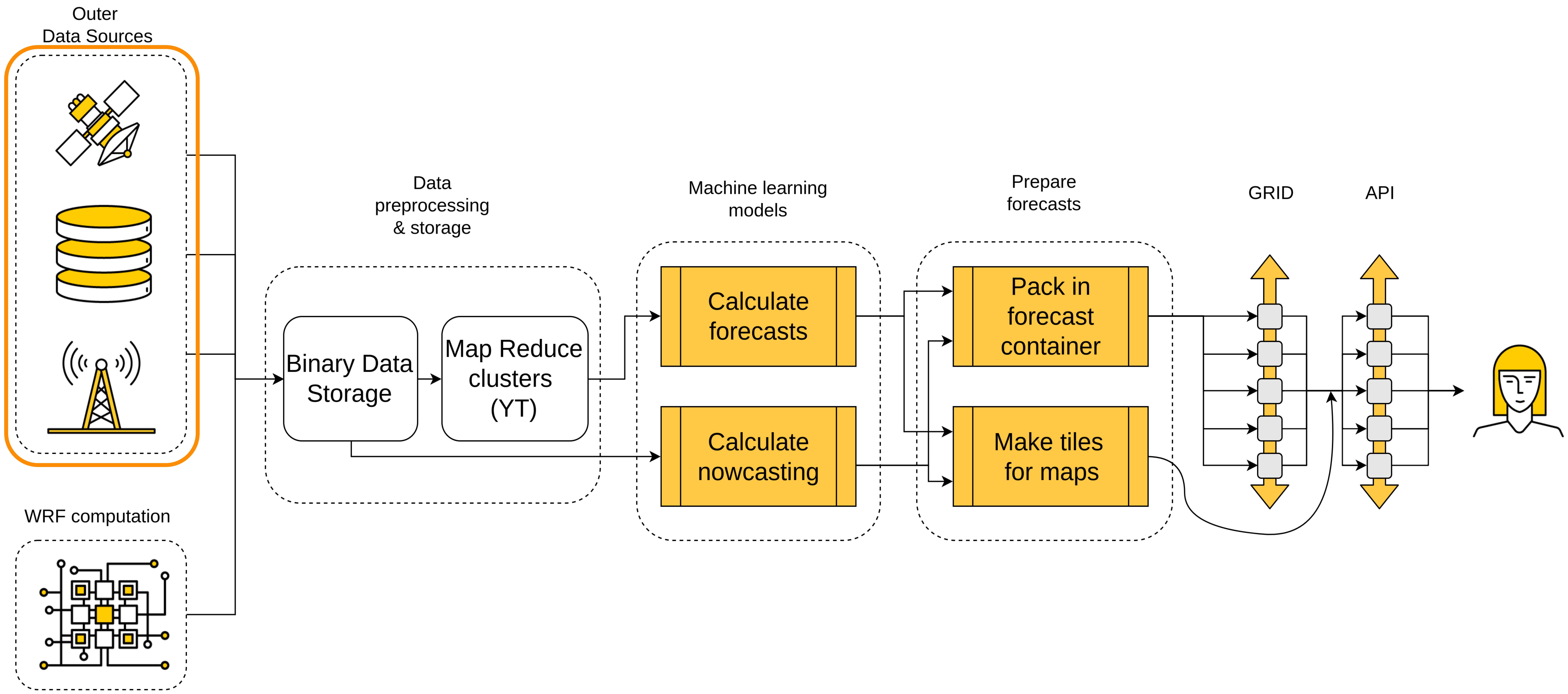
Forecasting with machine learning



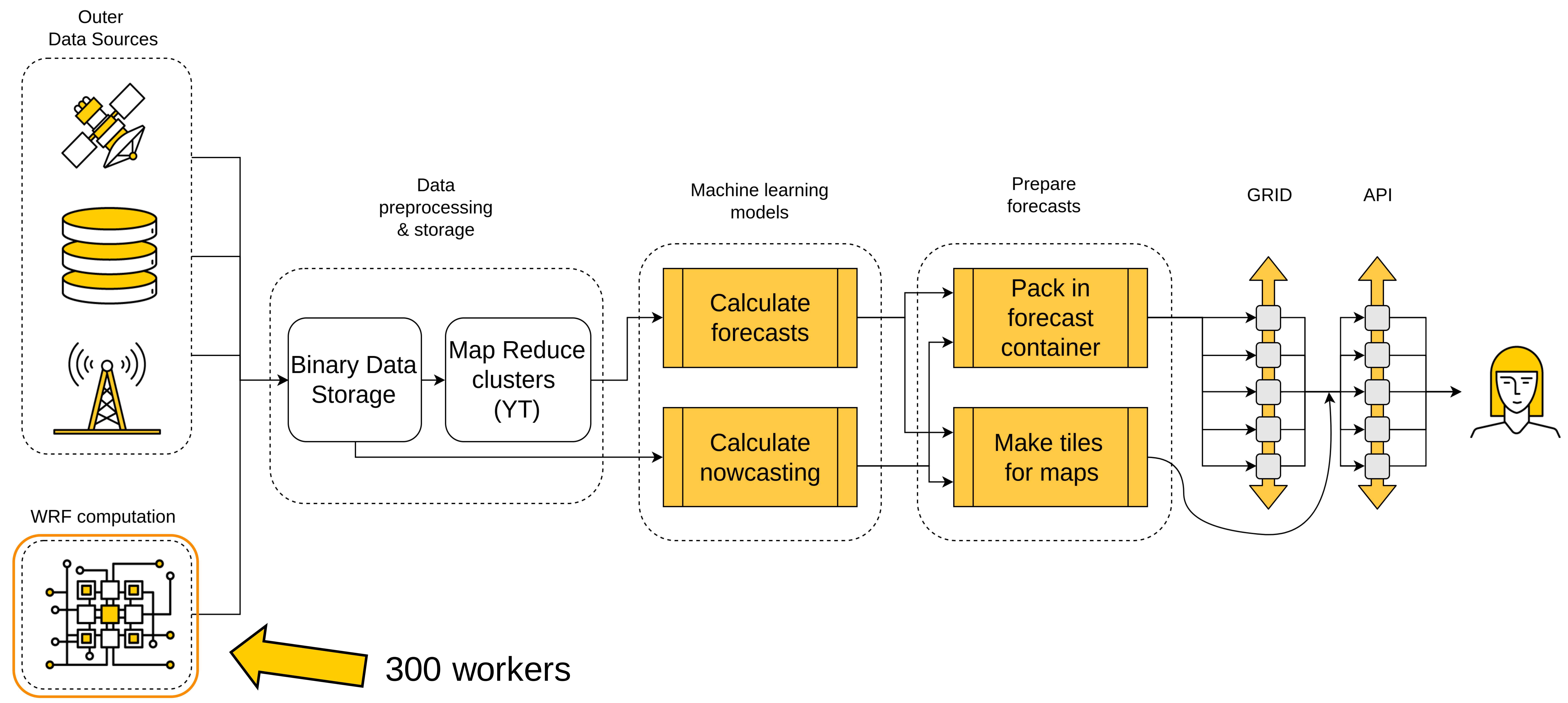
Overview

- What is Yandex.Weather
- Forecasting with machine learning
- **Architecture of the service**
- Fast calculation of forecasts

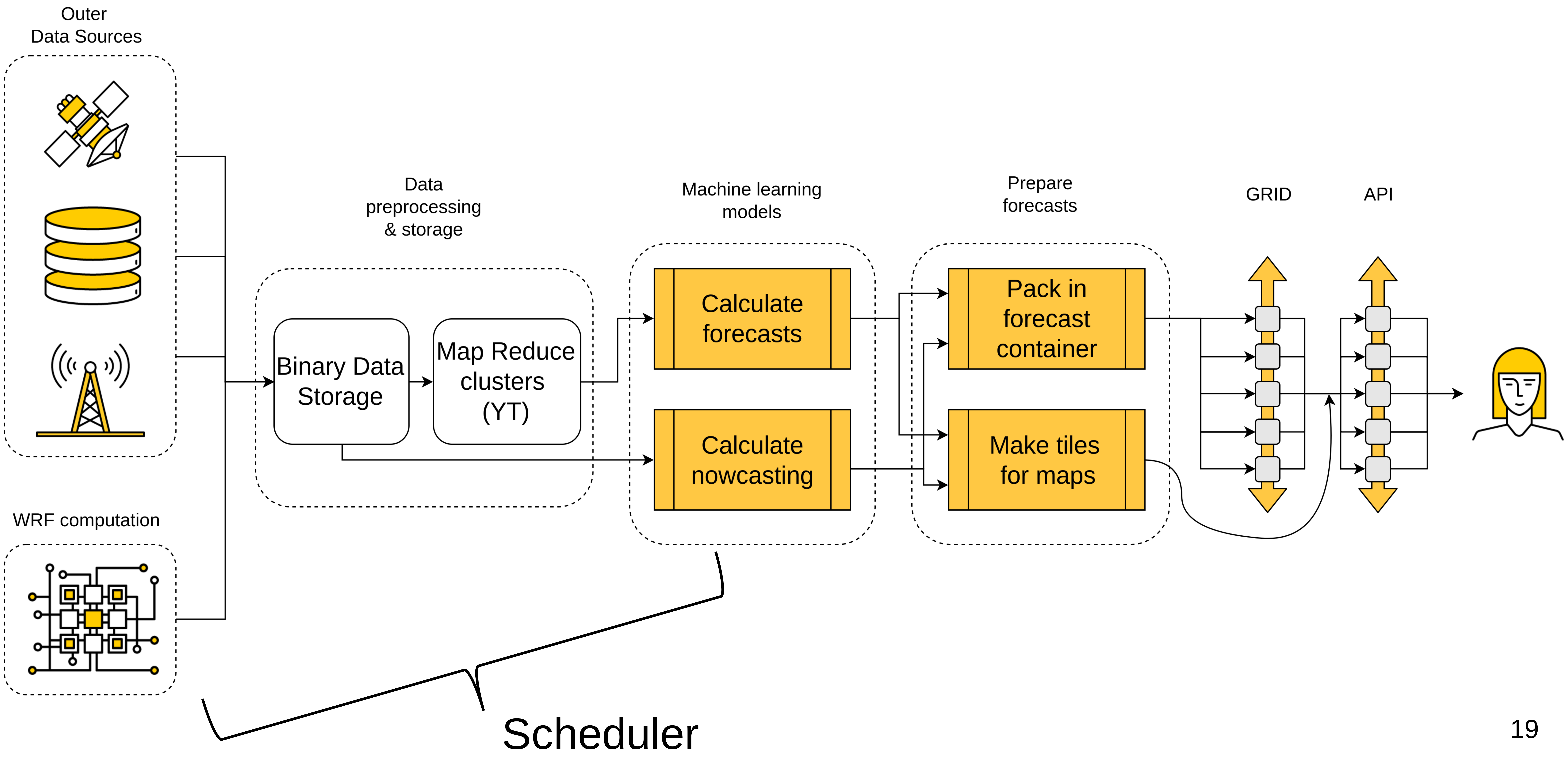
Architecture of the service



Architecture of the service



Architecture of the service

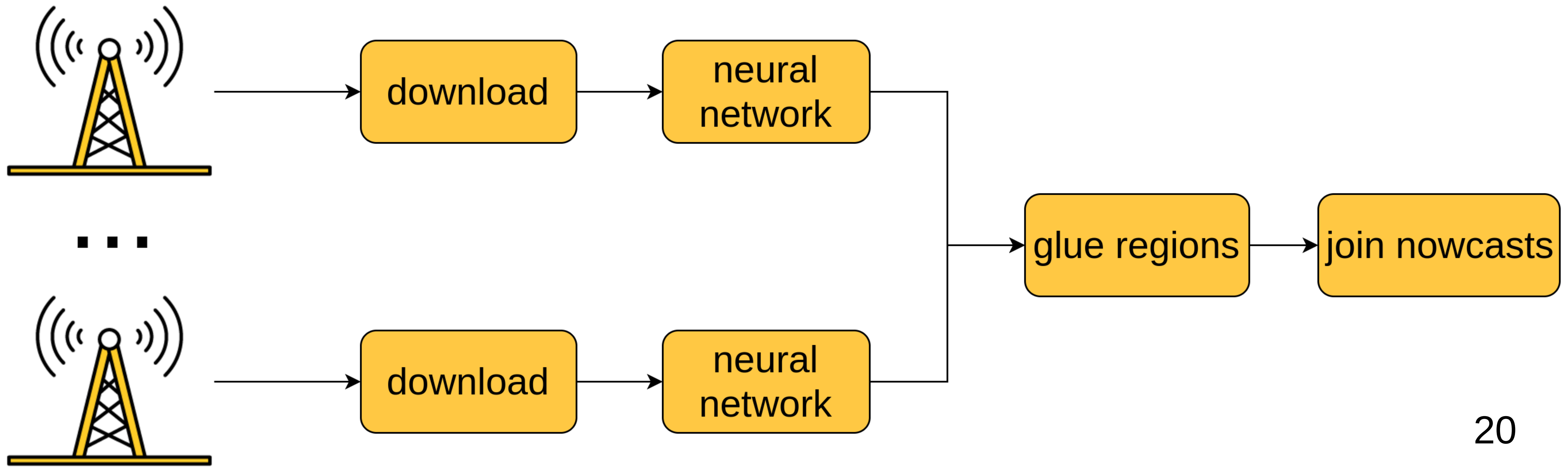


Scheduler

- Python, Celery, PostgreSQL
- YT for distributed locks

Tasks:

- import various data (GRIB, NetCDF, TIFF, BUFR, HDF5)
- export data to MR clusters
- nowcasting pipeline
- ...



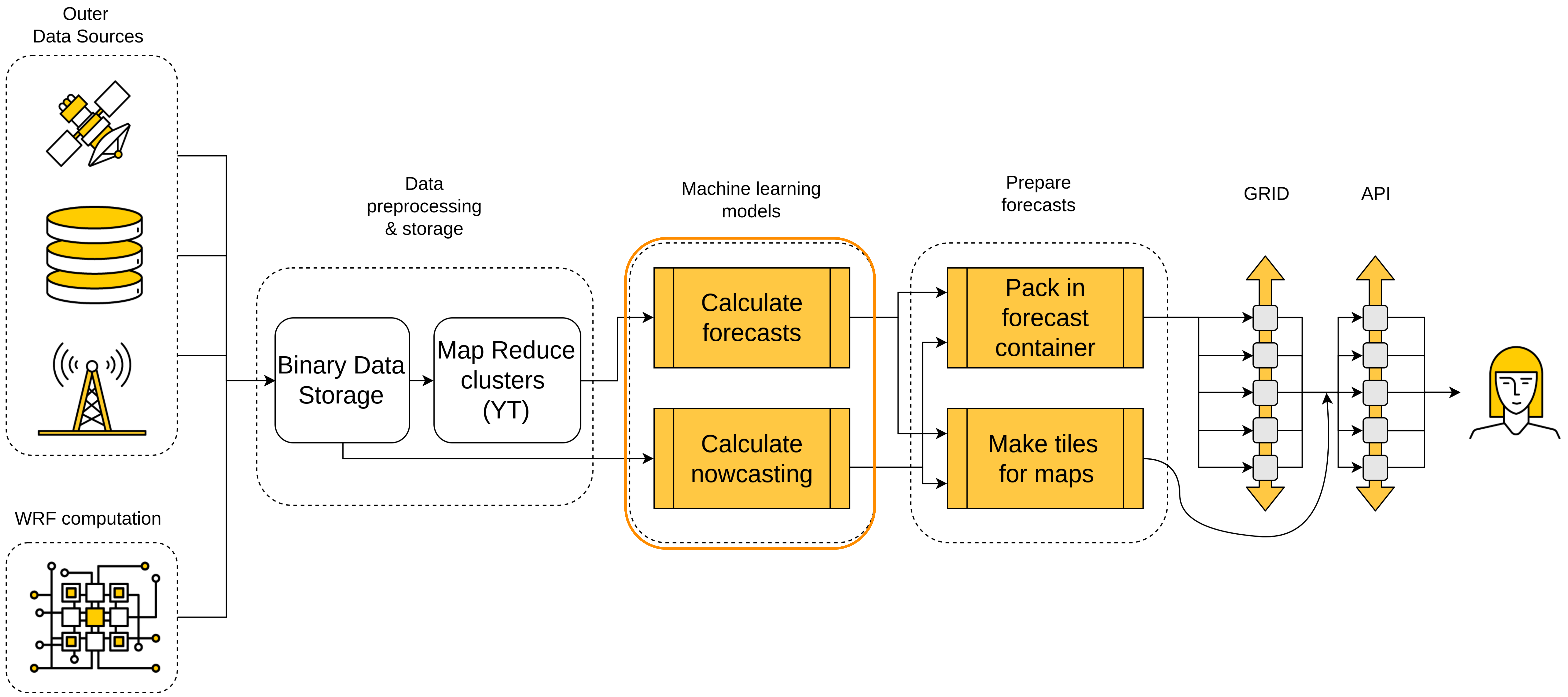
Scheduler

```python
from celery import Task

class BaseTask(Task):
    """Logging, progress calculation"""
    ...

class LockedTask(BaseTask):
    """Distributed locks to keep tasks unique"""
    ...

class MeteoFlowExtractorTask(LockedTask):
    """Implements some data processing"""
    ...
```

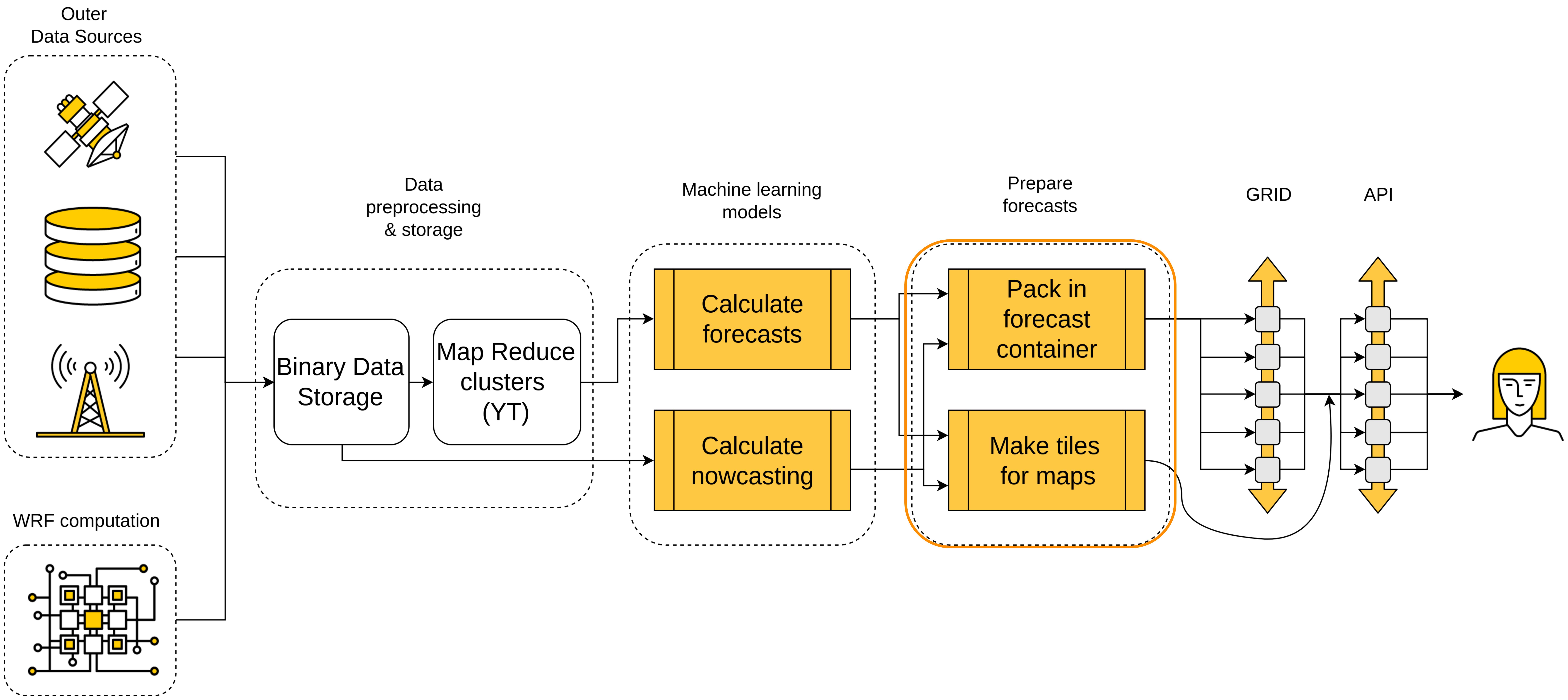
Scheduler

```
from celery.canvas import chain, chord, group, shared_task

@shared_task(base=MeteoFlowTask)
def nowcasting_master_task(…):
    ...
    workflow = chord(
        (get_radar_task.s(radar, **config) for radar in radars),
        chain(
            optical_flow_task.s(gen_time, **config).set(expires=deadline),
            write_to_grid_task.s(**config).set(expires=deadline),
            group(after_tasks),
        )
    )
    self.replace(workflow)
```

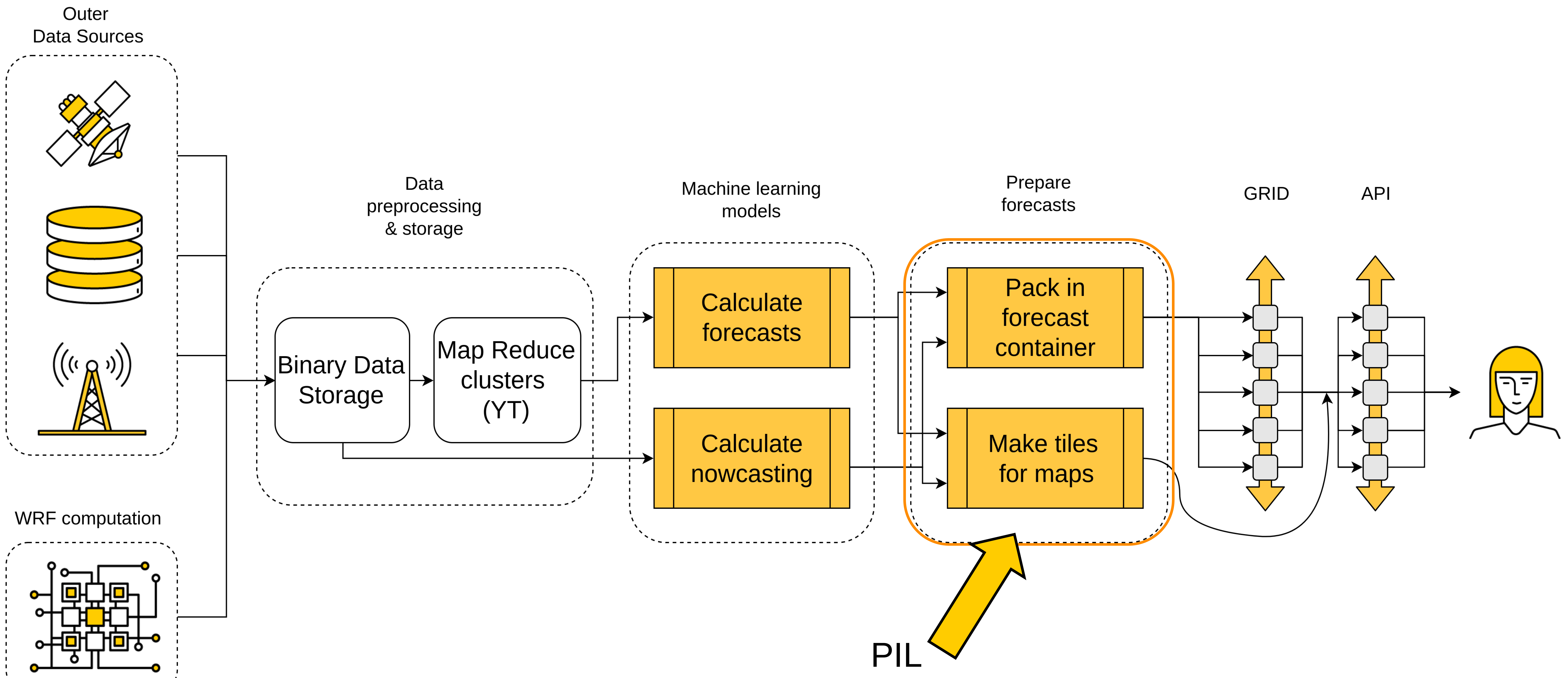
Architecture of the service

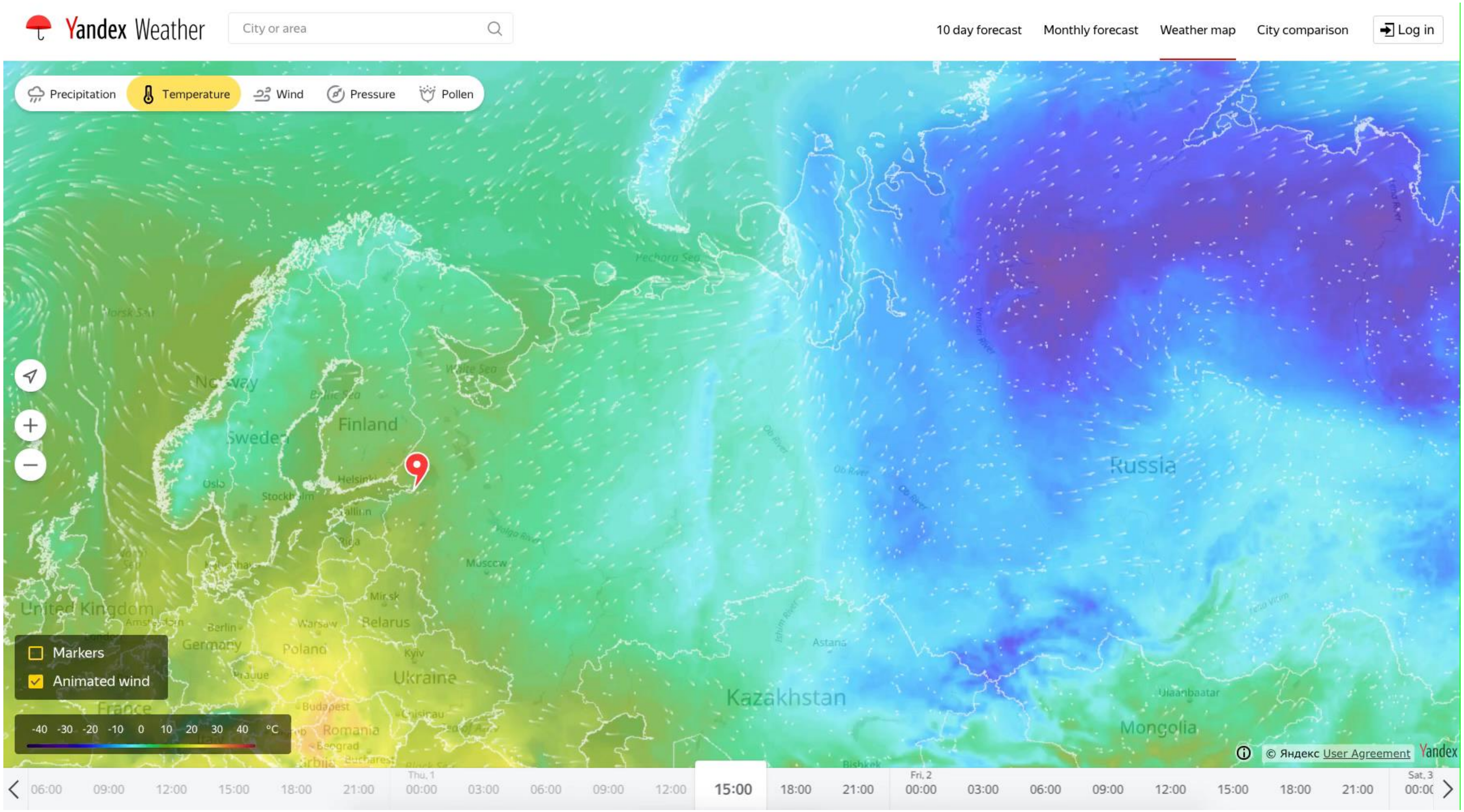


Architecture of the service



Architecture of the service

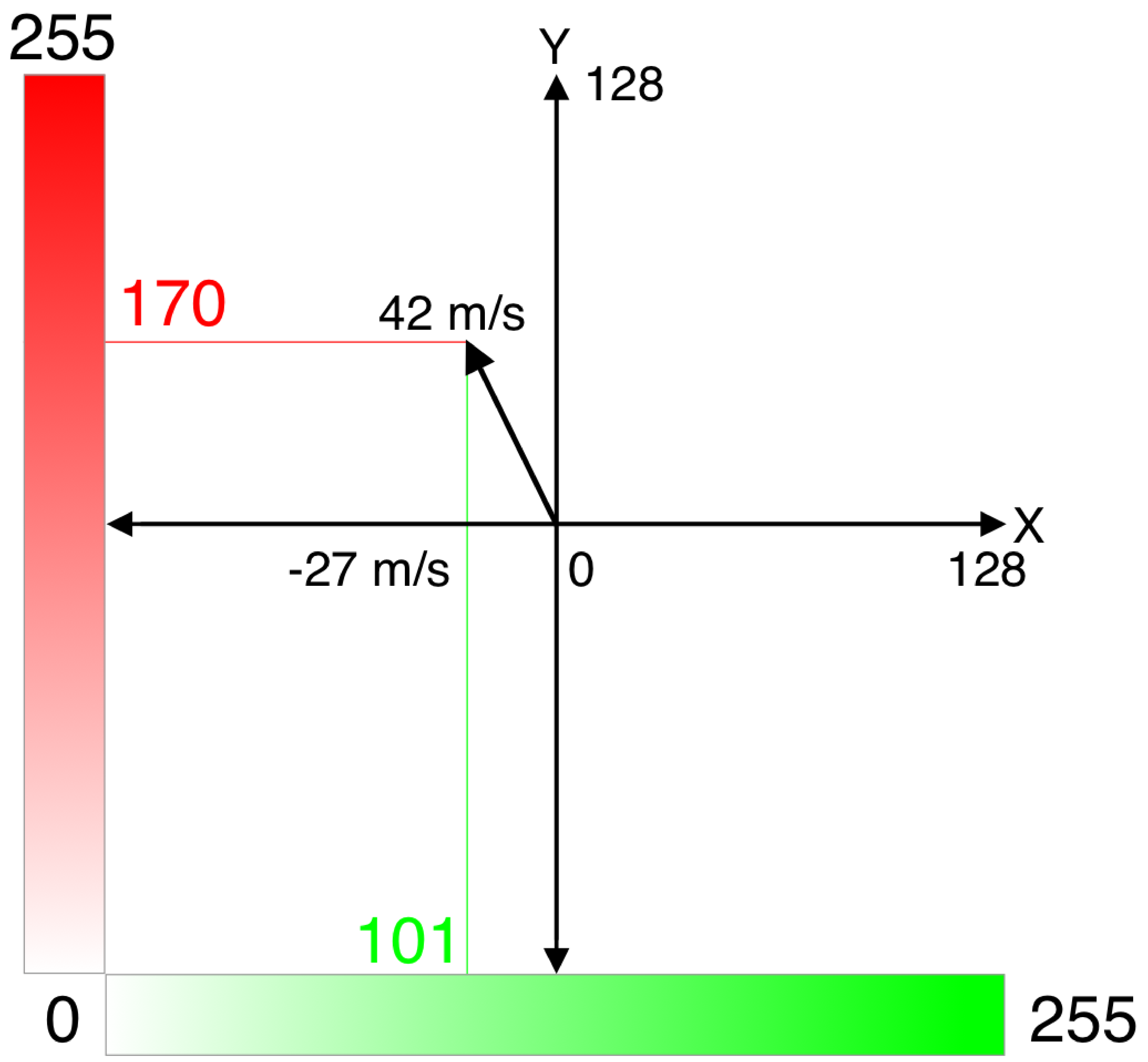


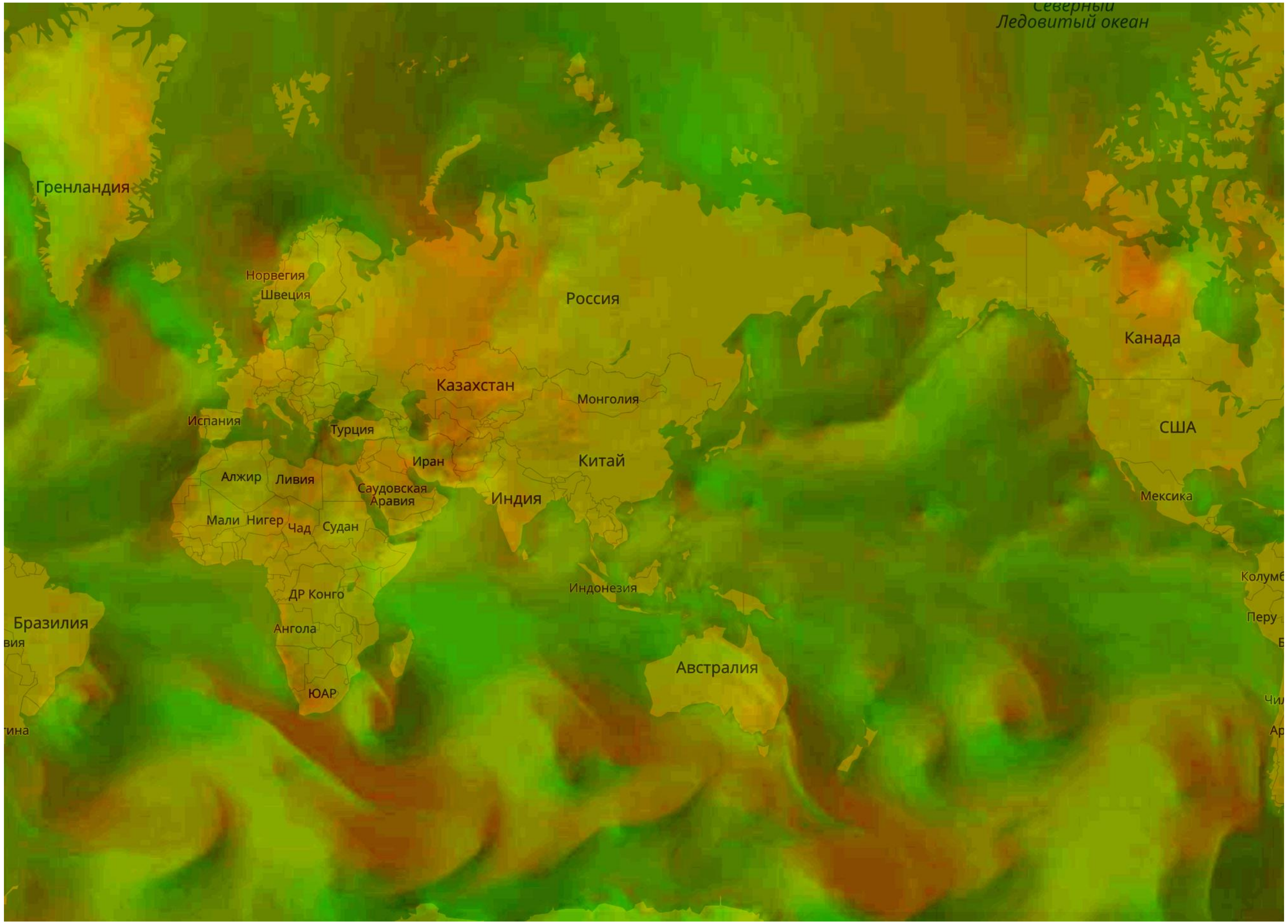


Temperature [✓]

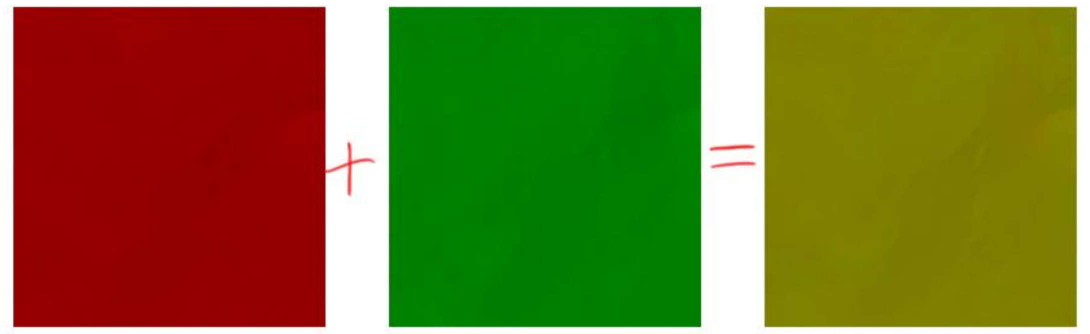
☐ Markers

☑ Animated wind

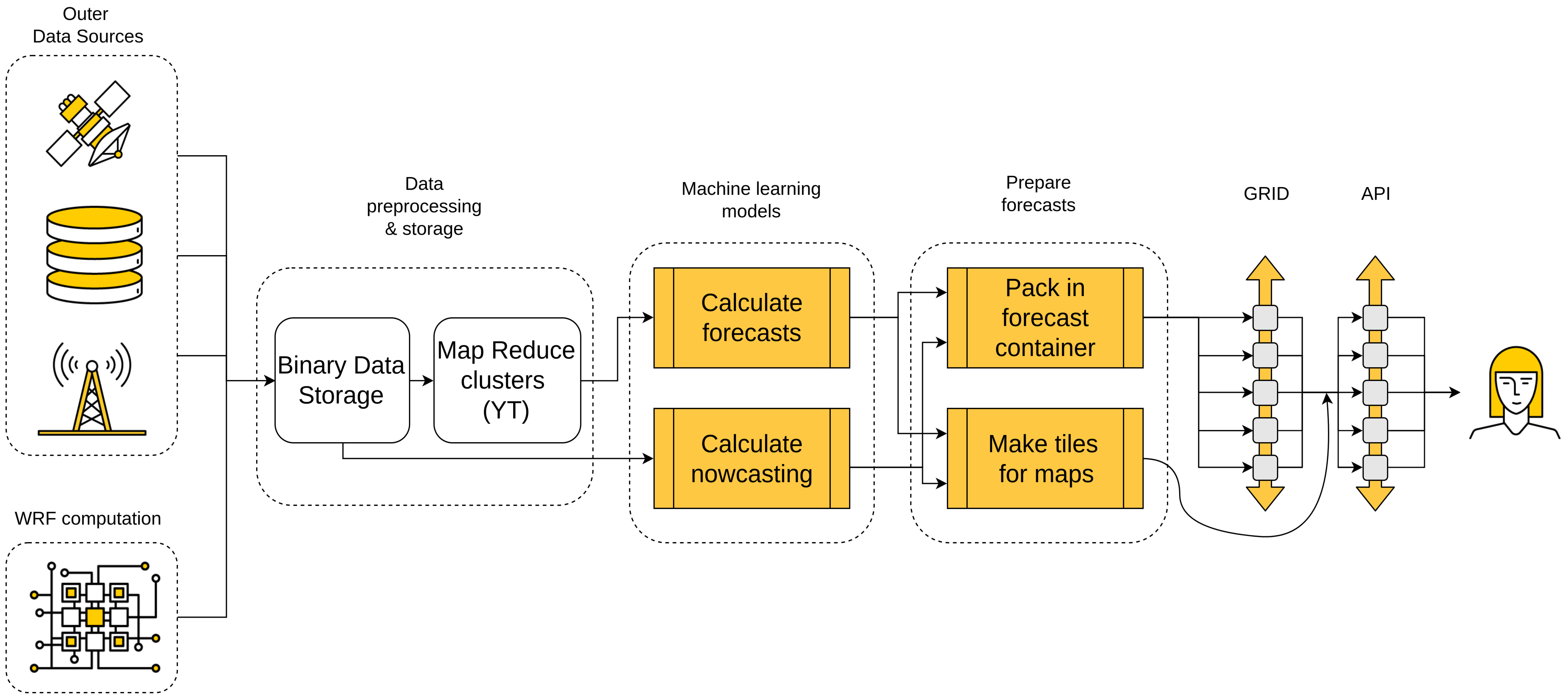

255
Y
128
170
42 m/s
X
-27 m/s
0
128
101
0
255



-68	-48	-28	0	28	52	80м/с	
60	80	100	128	156	180	200	R
60	80	100	128	156	180	200	G



Architecture of the service



Overview

- What is Yandex.Weather
- Forecasting with machine learning
- Architecture of the service
- **Fast calculation of forecasts**

Fast calculation of forecasts

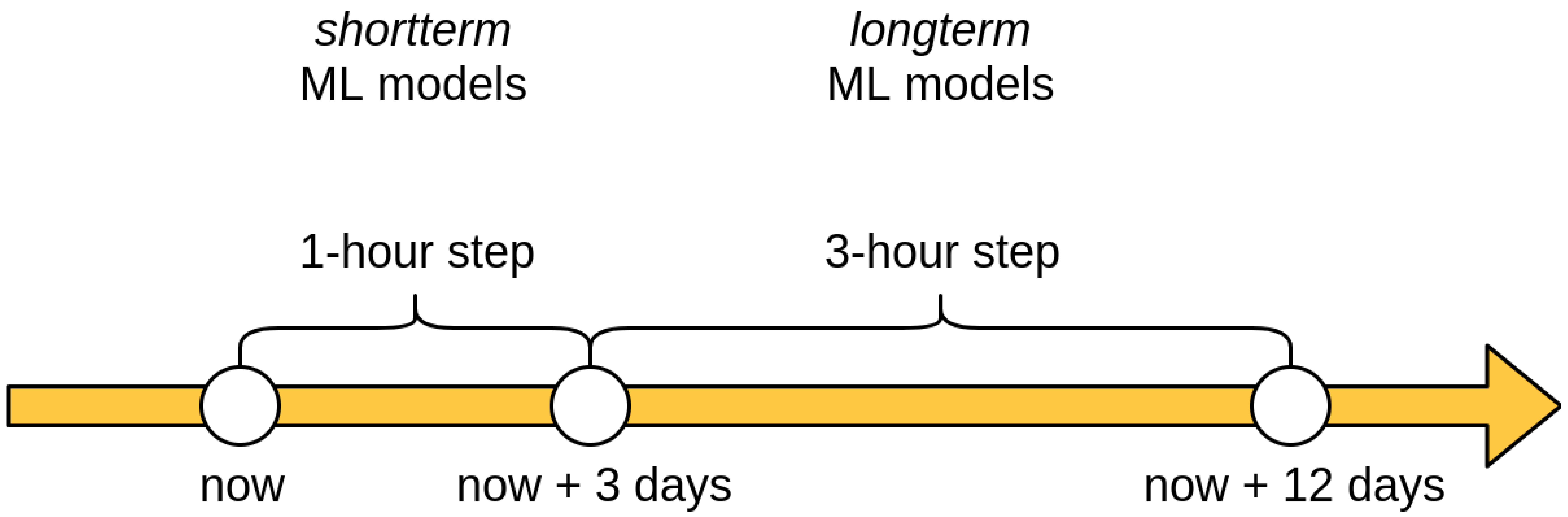
Machine learning models

- 10-15 *default* models
- Dozens of *active* models

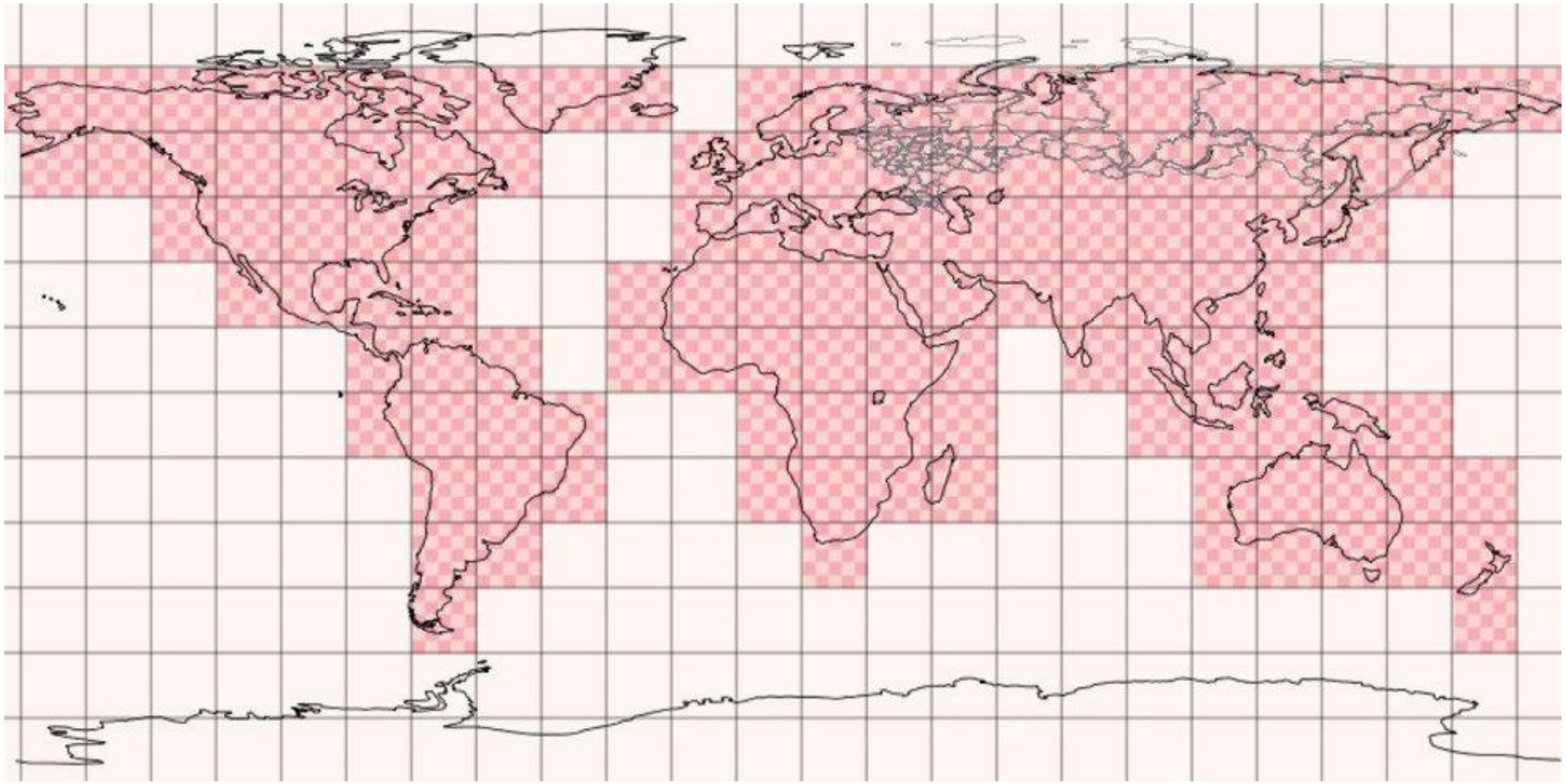
Input data

- Non-uniform through time
- 5 models, 10 days, each table ~1Gb (compressed npz)

Fast calculation of forecasts. Horizons



Fast calculation of forecasts. Geography



Fast calculation of forecasts. Features

```
features_dict = {
    'wrf_available': ('is_not_nan', 'wrf/wrf_temperature'),

    'wrf_rain': (
        'safe_div',
        ('sub', 'wrf_next/wrf_rain', 'wrf/wrf_rain'),
        'horizon_const/wrf_period',
        'matrix_const/-9999.'
    ),
}
```

Fast calculation of forecasts. Features

```
class Executor():
     def __init__(self, expressions, operations, sources):
         self.expressions = expressions
         self.operations = operations
         self.sources = sources

    def __getitem__(self, key):
        return self._execute(self.expressions[key])
```

Fast calculation of forecasts. Features

```
class Executor():
    ...
    def _execute(self, expr):
        if type(expr) is DataSourceDescriptor:
            # feature from data
            if expr.provider == 'self':
                return self[expr.varname]
            return self.sources[expr.provider][expr.varname]
        else:
            # need to calculate operation
            operation, *args = expr
            if not callable(operation):
                operation = self.operations[operation]
            return operation(*map(self._execute, args))
```

Fast calculation of forecasts

Principles:

- Prioritization
- MapReduce parallelization
- Lazy calculations + cache

Resources:

- 3K CPU + 1K CPU for forecasts
- 300 workers for WRF

Fast calculation of forecasts

Apply for 50 horizons

- Time: median about 15-20 minutes
- 3216 jobs, total time: over 1 week
- 200+ tables, 7Tb of data in intermediate steps

Yandex.Weather API

```
import json
import requests

result = requests.get(
    "https://api.weather.yandex.ru/v1/forecast?lat=60.0&lon=30.0",
    headers={"X-Yandex-API-Key": API_KEY},
)

data = json.loads(result.content)
data.keys()
```

Result: ['info', 'now_dt', 'now', 'fact', 'forecasts']

Links (RU)

- Nowcasting: https://habr.com/company/yandex/blog/317626/
- Nowcasting with radars: https://habr.com/company/yandex/blog/425517/
- Maps: https://habr.com/company/yandex/blog/343518/

- GRID architecture: https://youtu.be/4As-5fhDvsU?t=7218
- MapReduce clusters https://habr.com/company/yandex/blog/311104/

- Yandex.Weather API: https://tech.yandex.com/weather/doc/dg/concepts/about-docpage/

Thank you

Lena Volzhina

Developer @ Yandex.Weather

✉ Lenavolzhina@yandex-team.ru

@lenavolzhina