Working with geodata in Go

Andrew Minkin

Who am I

- TeamLead at maddevs.io
- 10 years of experience
- https://github.com/meshbird/meshbird

Bishkek

- 1kk citizens
- 100++ taxi services

Namba Taxi

- 8k orders
- 600+ online drivers
- 500k clients

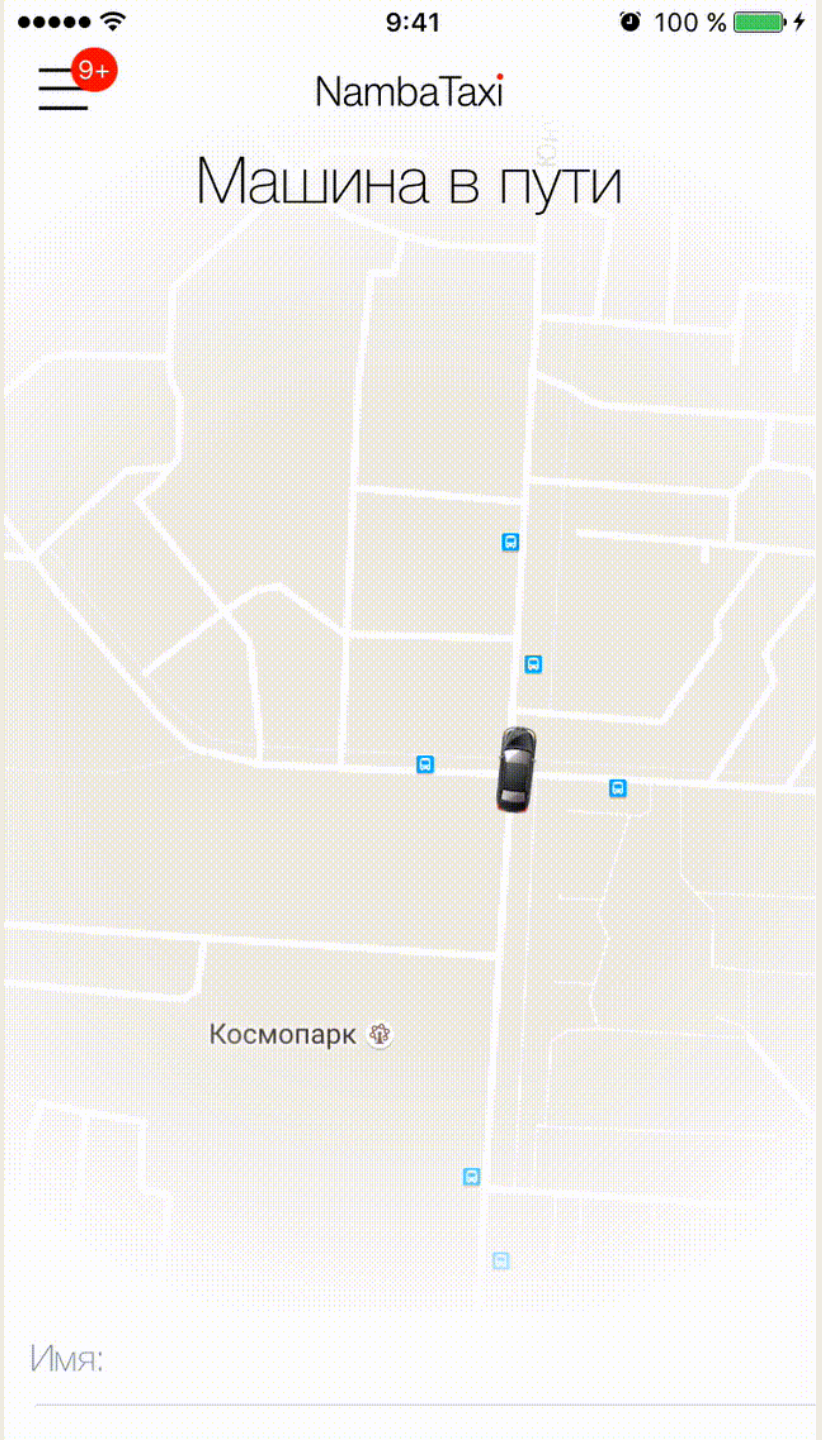
What’s a taxi

- Clients
  - Voice
  - Text
  - Mobile App
  - ....
- Drivers
- Operators
- Managers

AVG response time

- Drivers 20 ms
- Operators 2.5 ms

Prehistory



MAD DEVS
MADDEVS.IO

First challenge

- Decrease update interval on driver’s device
  - Interval 15 secs

MAD DEVS
MADDEVS.IO

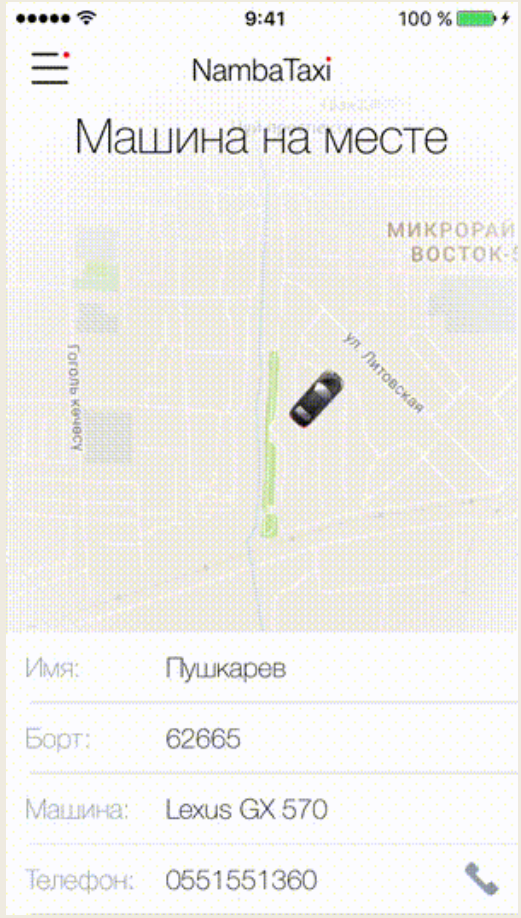
First try

- Make request -> save coordinates
- Make request -> animate the car

First problems

- We can’t animate car properly
- Car moves through fields, forests and quarters

First problems



MAD DEVS
MADDEVS.IO

Solution

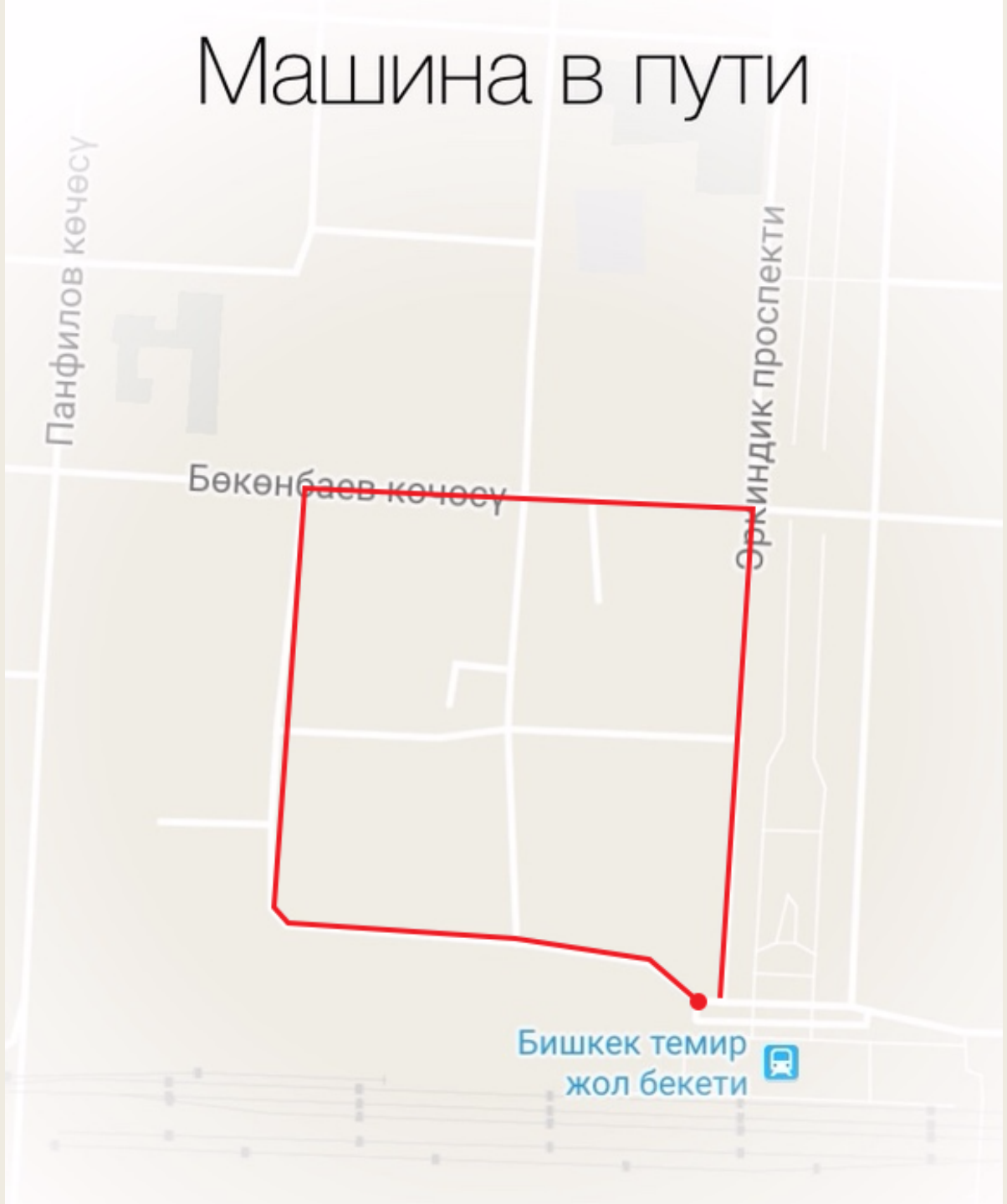
- OSRM

Try again

- Timeout – 15 secs
- Make request -> save coordinates
- Send coordinates
- Building route via OSRM
- Get route -> animate marker

MAD DEVS
MADDEVS.IO

Машина в пути



MAD DEVS
MADDEVS.IO

Solution

- Check for 20 meters

Application Released!

Needs improvement

- Trip cost calculations on driver side
- We need more tracks and 1 track at 15 seconds is few
- GPS problems on driver

MAD DEVS

MADDEVS.IO

GPS problems

- Bad device/Bad module
- GPS going to die during the time
- Pits and “Jams”

Tasks

- Start to collect more tracks from drivers
- Show animated cars on the main screen
- Store intermediate trip cost on the server side
- Save mobile data
- Collect each track per one second

MAD DEVS
MADDEVS.IO

What’s the track?

- Latitude
- Longitude
- Session
- OrderID
- TripCost

MAD DEVS

MADDEVS.IO

Trafic economy

- 1 track = 100 bytes

Tell me all the options, please

- HTTP
- WebSockets
- TCP
- UDP

Tell me all the options, please

- HTTP
- WebSockets
- TCP
- **UDP**

MAD DEVS

MADDEVS.IO

Why we choosed UDP?

- We send only datagrams
- We don’t need guarantees
- Minimalism
- Save lots of data
- We have only 20 bytes overhead
- Not blocked in our country

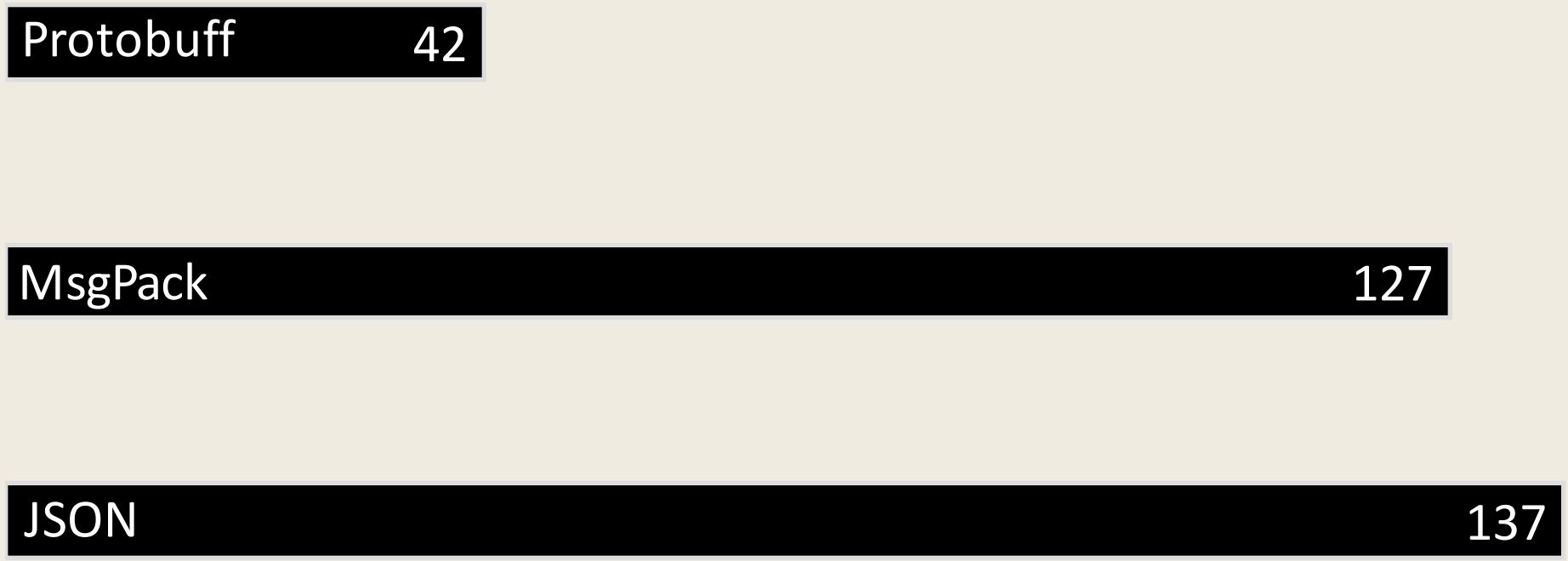
MAD DEVS

MADDEVS.IO

What about data serialization?

- JSON ☹
- MsgPack ☹
- Protocol Buffers ☺

Data size



MAD DEVS
MADDEVS.IO

Total

- 42 bytes of payload
- + 20 bytes of IP headers
- = 62 bytes per track
- = PROFIT!

MAD DEVS
MADDEVS.IO

Data storage

What data to store?

- Driver’s session
- Cab number
- Order ID
- Trip cost
- Last location
- N last locations

Which storages do we use?

- Percona
- Redis
- Elasticsearch

MAD DEVS

MADDEVS.IO

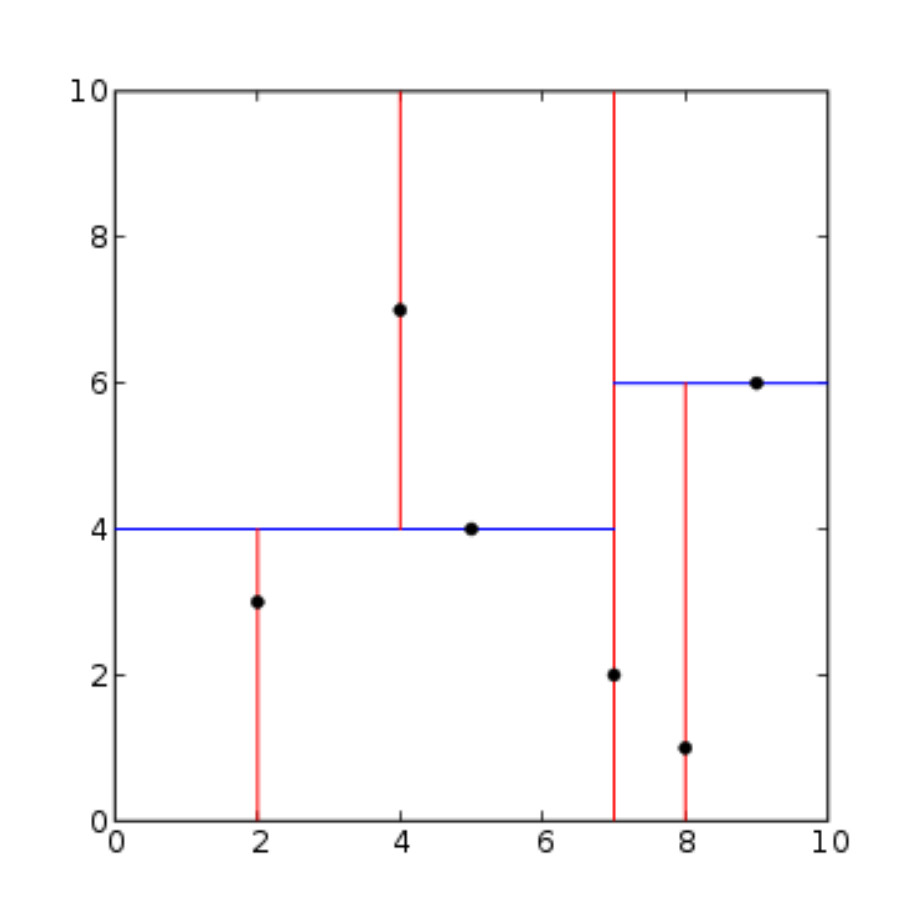
We need geoindex

- KD-tree
- R-tree

What requirements for geoindex

- Search of N nearest points
- Balanced tree

KD-Tree

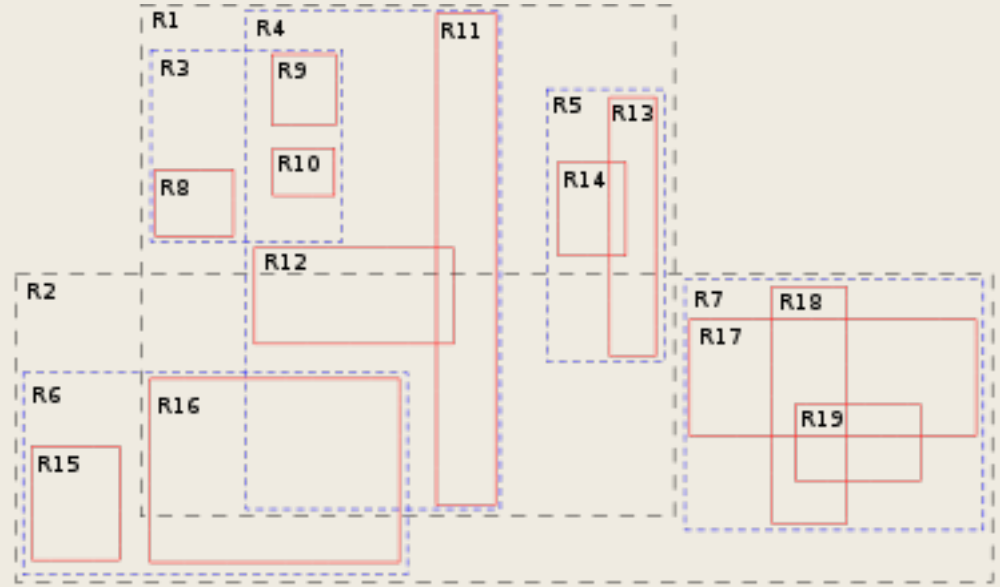


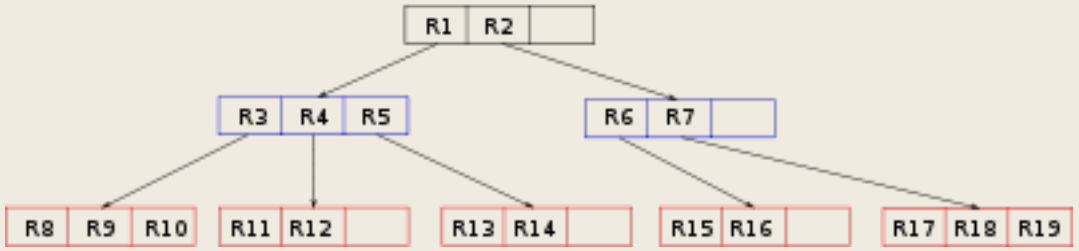
MAD DEVS
MADDEVS.IO

KD-tree cons

- Unbalanced tree
- Can search only one nearest point

R-tree





MAD DEVS
MADDEVS.IO

Where can I get it?

- https://github.com/dhconnelly/rtreego

What’s wrong with drivers?

- Bad internet connectivity
- Turned off the phone
- Low battery
- Removed application from RAM
- And lots of other reasons

MAD DEVS

MADDEVS.IO

- We need expire mechanism
- We need LRU data structure for storing coordinates

Storage architecture

- In-memory
- R-tree
- Map with drivers(key is the session)
- Map with drivers(key is cab number)

What algorithm on backend

- Get data by UDP
- Try to get driver from storage
- If doesn’t exist – get driver from redis
- Check and validate data
- Set driver to storage
- If doesn’t exist – initialize LRU
- Update r-tree

MAD DEVS
MADDEVS.IO

Go

- Strong typed and compiled
- Small size of docker containers
- Few resource usage
- Less own zoo ☺

HTTP API

- Return nearest drivers
- Remove driver from storage(by cab number or session)
- Get information about trip
- Get information about driver

How to maintain

- Logging into stdout/stderr
- Metrics to Graphite
- Checks to sensu
- Usefull /status
- Bots

MAD DEVS

MADDEVS.IO

Usefull /status

- Uptime since
- HTTP Statuses counters
- Total requests

MAD DEVS

MADDEVS.IO

Bots

- Emulate of driver’s workflow
- Emulate of client’s workflow
- Running near Marokko or Congo

MAD DEVS

MADDEVS.IO

How it looks now?

- We get client location from sensors
- We get nearest drivers with routes
- Animate each car
- Update interval 15 seconds

Main slide of first story

- UDP+Protobuf for data savings
- In-memory storage
- R-tree for nearest drivers
- LRU cache for storing last locations
- OSRM for map matching and building routes

MAD DEVS
MADDEVS.IO

Challenges in geocoding

What challenges do we have?

- Lack of data
- Crossroads navigation
- We can’t trust GPS

MAD DEVS
MADDEVS.IO

Which map providers available?

- Yandex
- Google
- 2GIS

What we have to solve issue

- OpenStreetMap data
- Own database with addresses with coordinates

Search format example

- Chui ave 139
- Chui / Manasa
- Red centre
- AUCA
- 5-15
- 7 marksa 181

MAD DEVS
MADDEVS.IO

We need own geocoder

- Available to search in different formats
- Support for auto translated values
- Support of synonyms

Elasticsearch index

- Addresses(alias)
  - OSM_timestamp
  - Drivers_data

MAD DEVS
MADDEVS.IO

./ariadna update

- Download file
- Create index in Elasticserach
- Populate with data from osm
- Search intersections and populate data
- Change aliases
- Removes old index

MAD DEVS

MADDEVS.IO

Features

- Geocode
  - By the name
  - By crossroad
  - By the name of institution
  - Synonyms support
- Reverse geocode

MAD DEVS

MADDEVS.IO

Where to get it?

- https://github.com/maddevsio/ariadna

Questions?

- https://github.com/maddevsio
- https://github.com/maddevsio/ariadna

- @gen1us2k (twitter,github,facebook,skype,telegram)

MAD DEVS
MADDEVS.IO