



LATVIJAS
UNIVERSITATE
ANNO 1919

Datorzinātņu doktorantūras zinātniskais seminārs

Atrašanās vietas intelīģences metodes datu noliktavu mobilai lietotnei



1.k.doktorante: **Daiga Plase**

Darba vadītāja: prof., Dr. sc. comp.
Laila Niedrīte

Saturs

- 1) Atrašanās vietas intelīģence (BI konteksts)
- 2) Esošie risinājumi (piemēri)
- 3) Problēma
- 4) Darba mērķis
- 5) Metodes un problēmas risinājumi
- 6) Secinājumi
- 7) Turpmākie darbi

Biznesa intelīģence

teoriju, metožu, procesu,
arhitektūru un tehnoloģiju kopums, kas
pārveido izejas datus
jēgpilnā un biznesa mērķu sasniegšanai
noderīgā informācijā

Ko?
Kad? Cik?
Kas? Kur?

* Guru Vashist, „Bussiness Intelligence and Tools”, 2013

BI risinājumi

* Gartner 2015:
Magic Quadrant
for Business
Intelligence and
Analytics
Platforms



1. Datu integrācijas slānis

- Metadatu pārvaldība
- BI administrācija
- Sadarbība
- ETL

2. Datu analīzes slānis

- Pārskati
- OLAP
- Interaktīva vizualizācija
- **Ģeotelpiskā un atrašanās vietas intelīģence**
- KPI
- Datizrace

3. Informācijas piegādes slānis

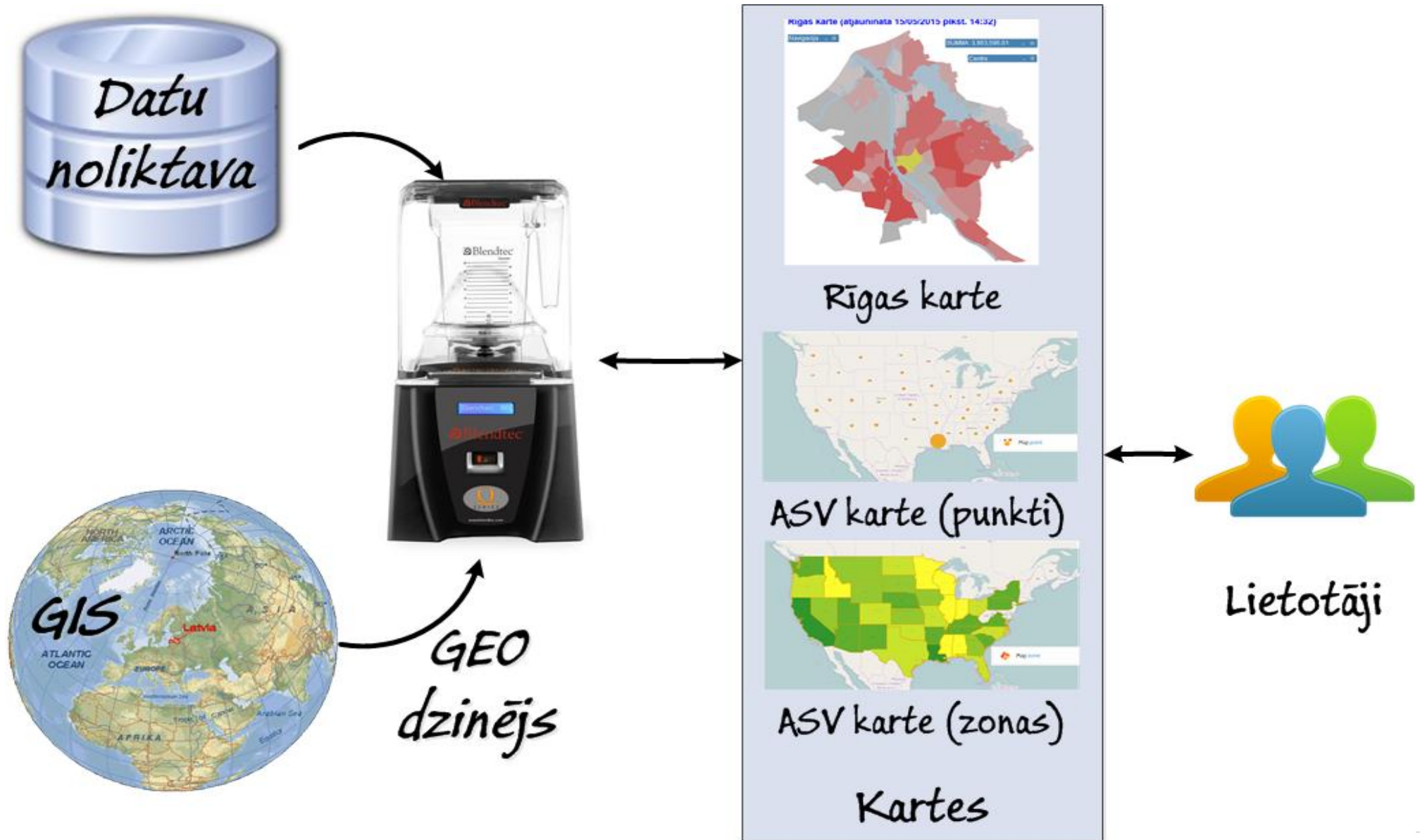
Datu vizualizācija:

- Vadības paneļi (dashboards)
- Ekspromtus vaicājumi
- Ms Office Integrācija
- Mobilā BI

Atrašanās vietas intelligence

rīku un metožu kopums, kas
sasaista biznesa datus ar ģeogrāfisko informāciju,
lai identificētu modeļus un sakarības lēmumu pieņemšanai,
ko citādi (bez telpiska attēlojuma) būtu sarežģīti atklāt

Biznesa vajadzība



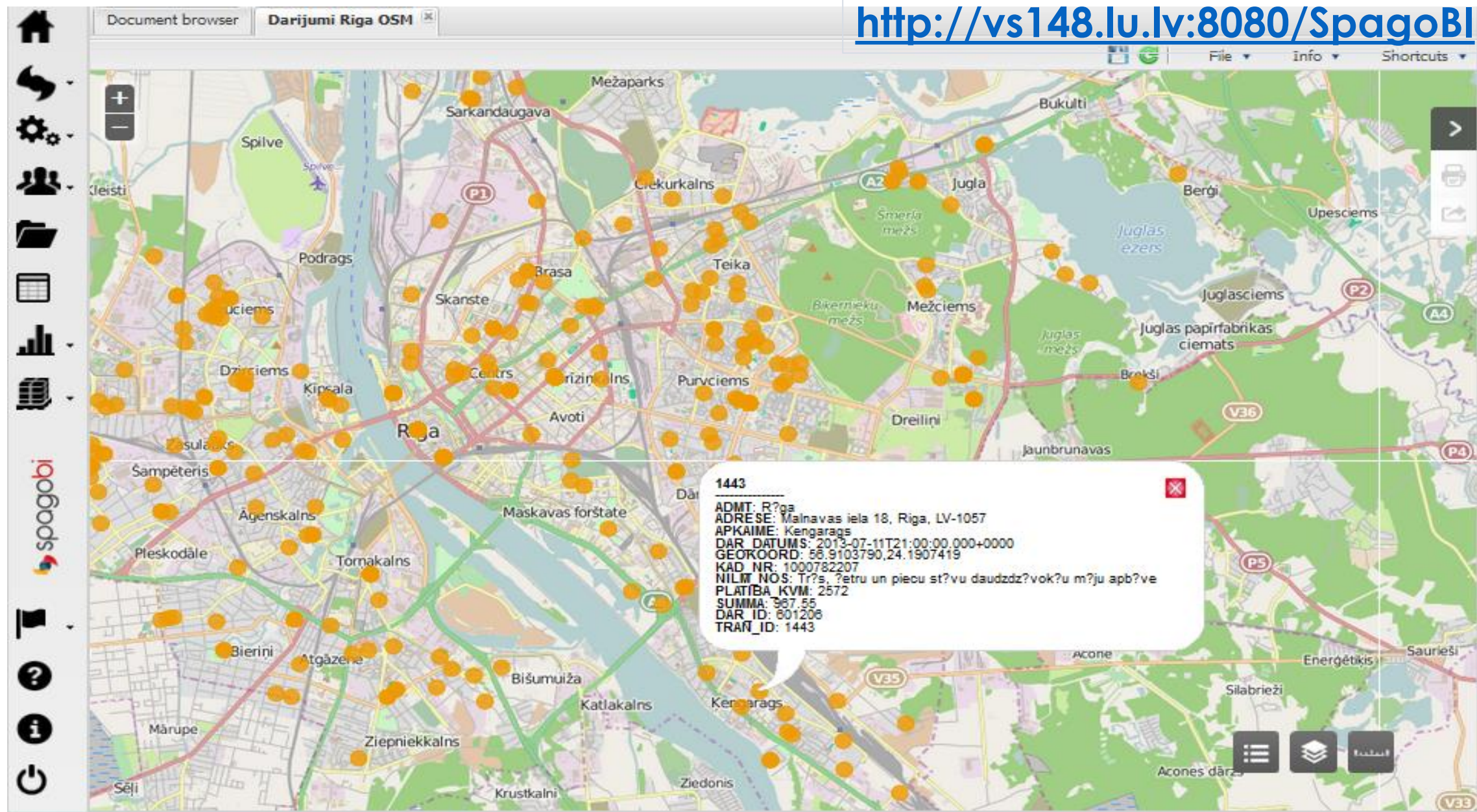
Location Intelligence

Piemēri ...

(no prakses)

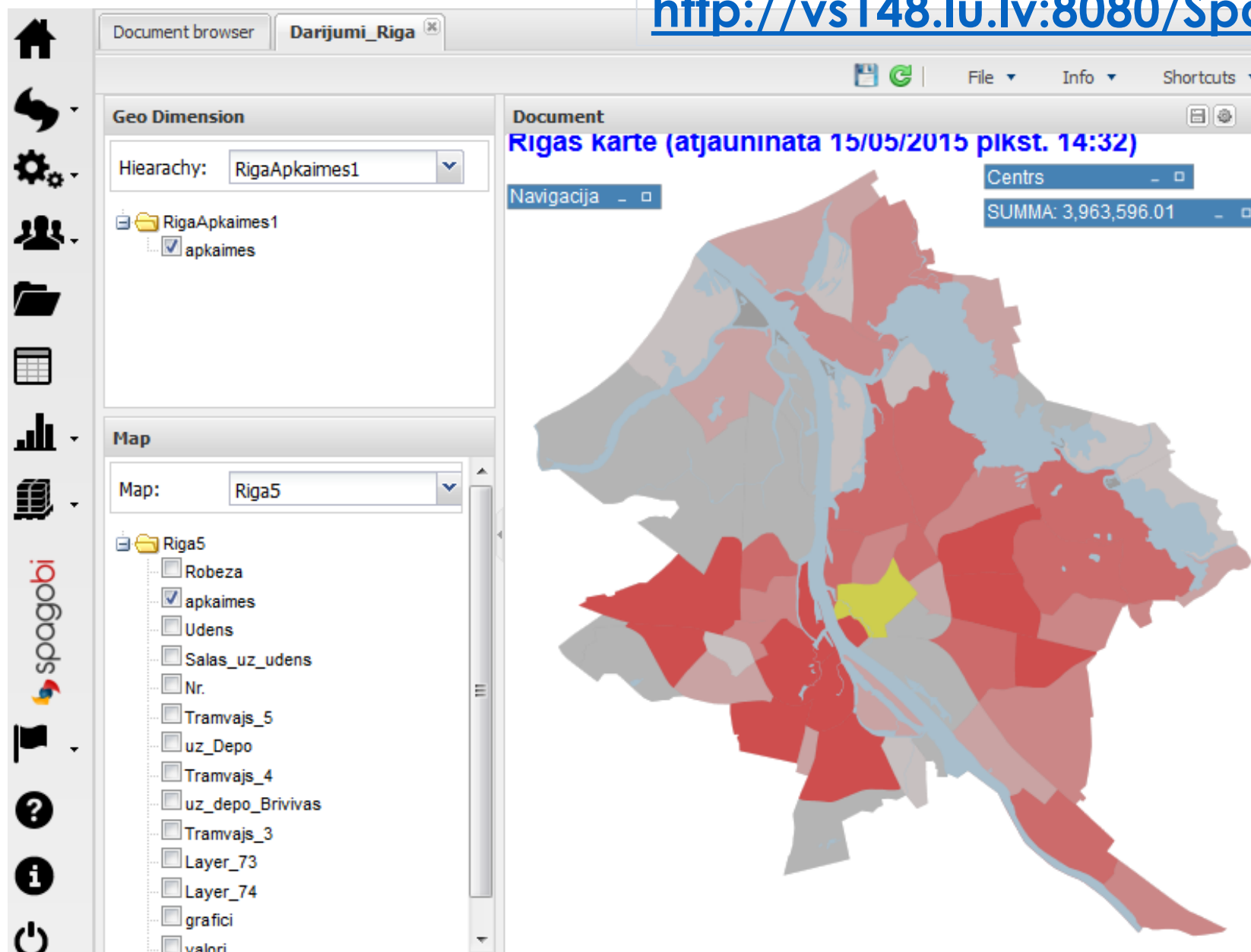
Nī darījumu karte

<http://vs148.lv.lv:8080/SpagoBI>



Tematiskā karte

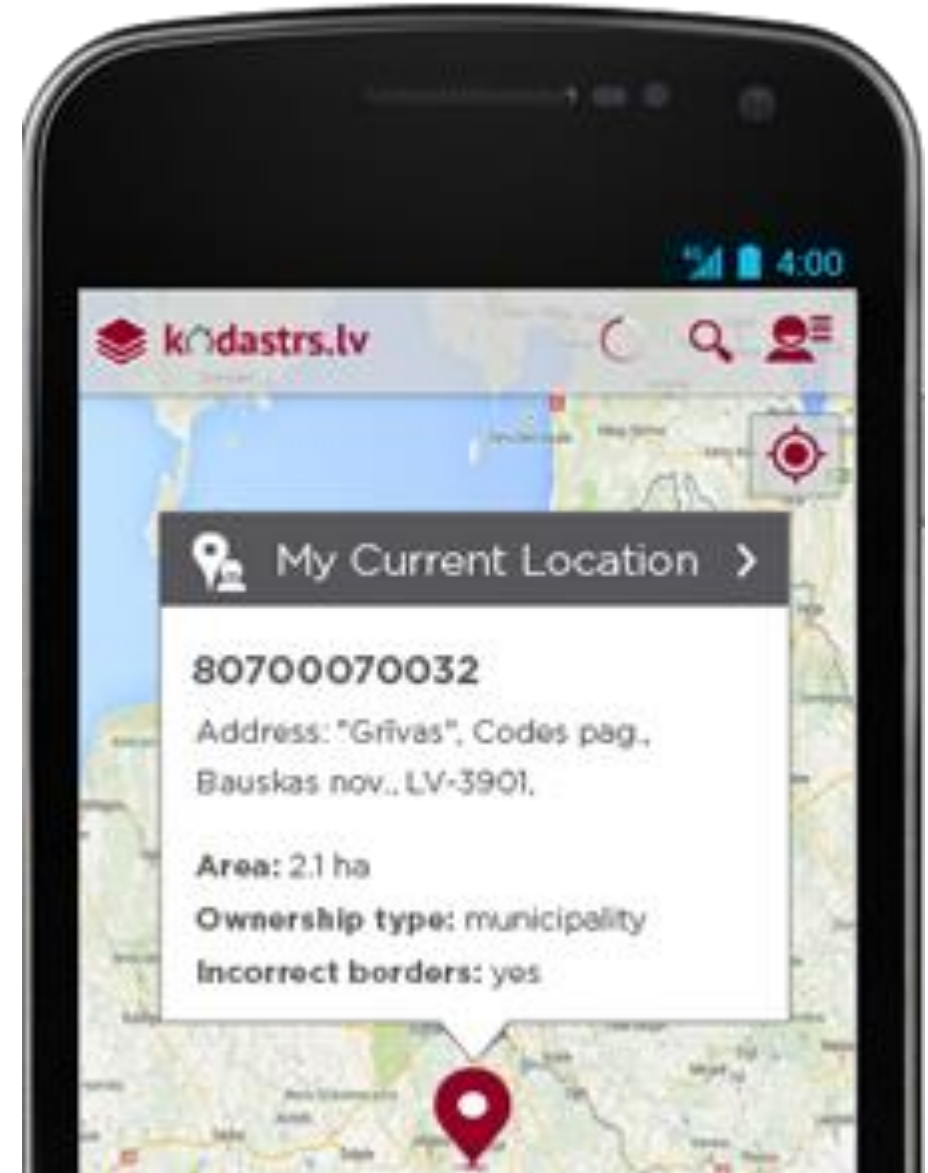
<http://vs148.lv.lv:8080/SpagoBI>



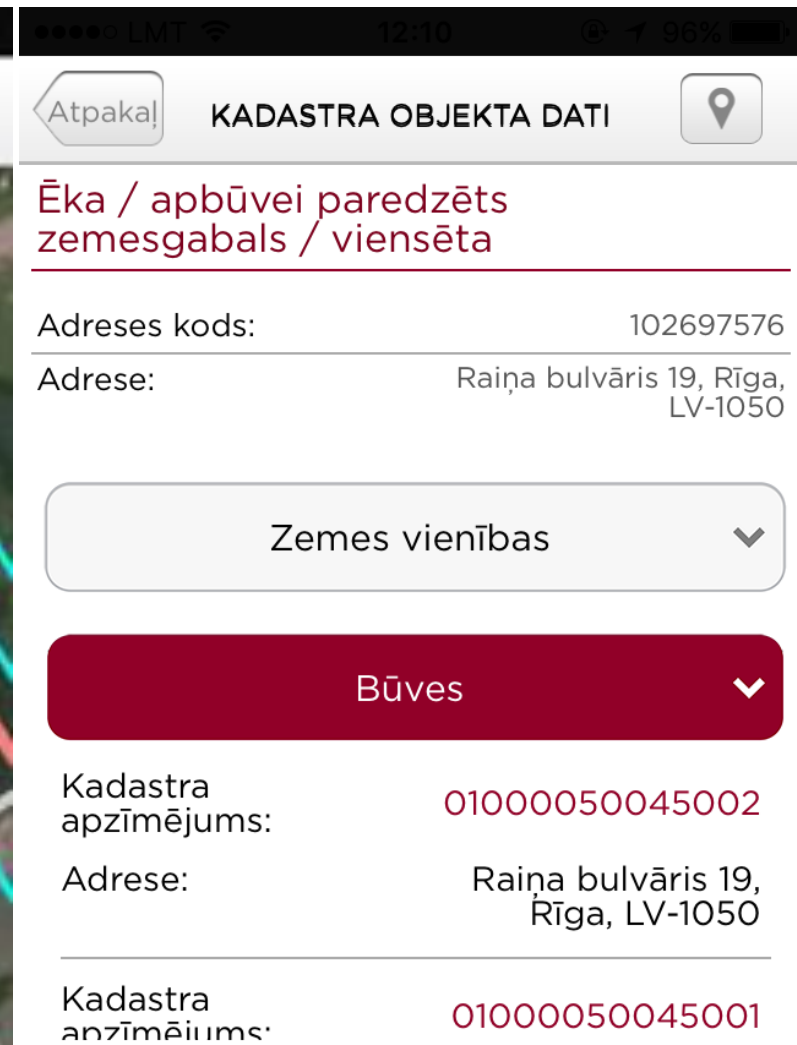
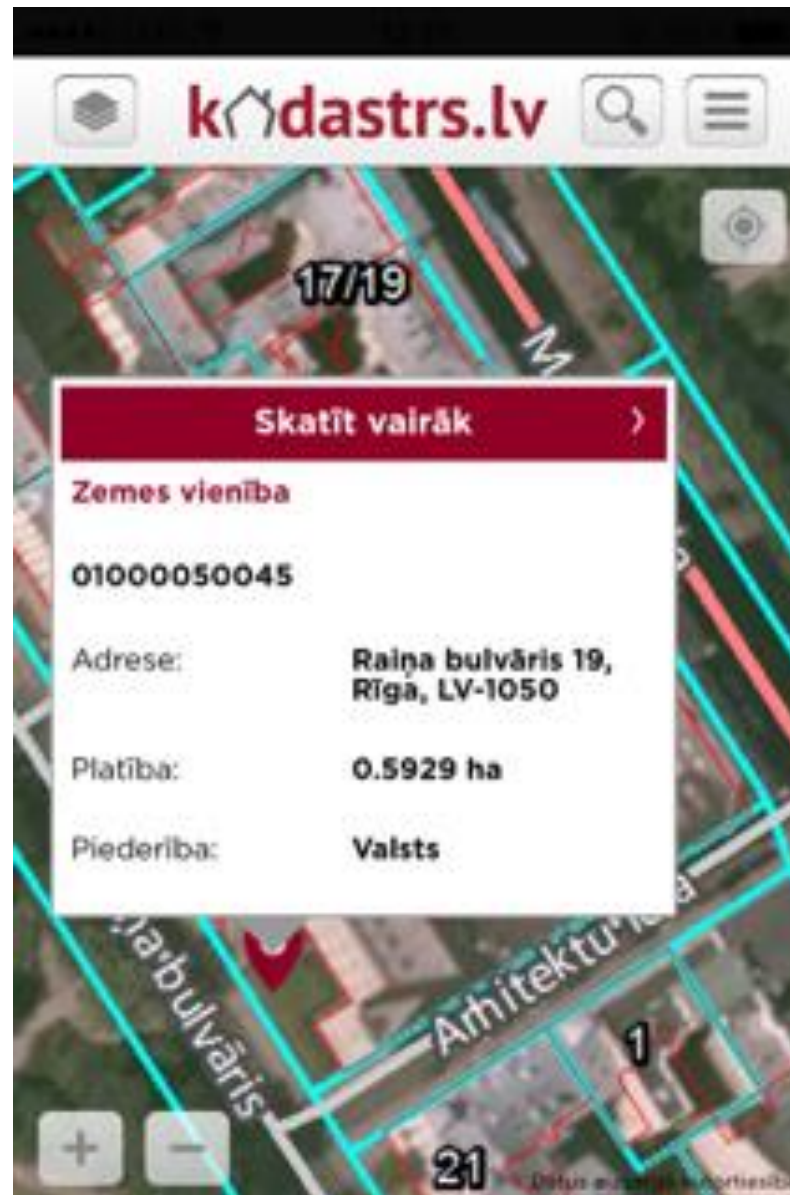
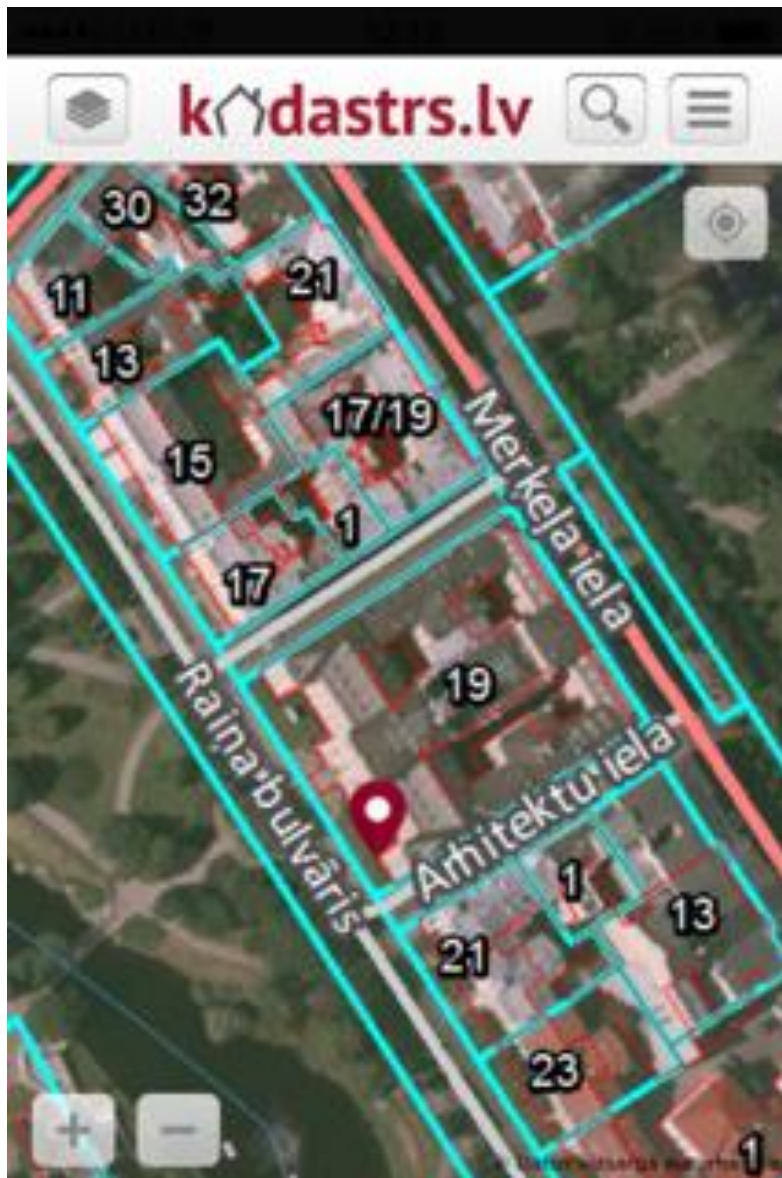
Mobilā lietotne kadastrs.lv



kadastrs.lv

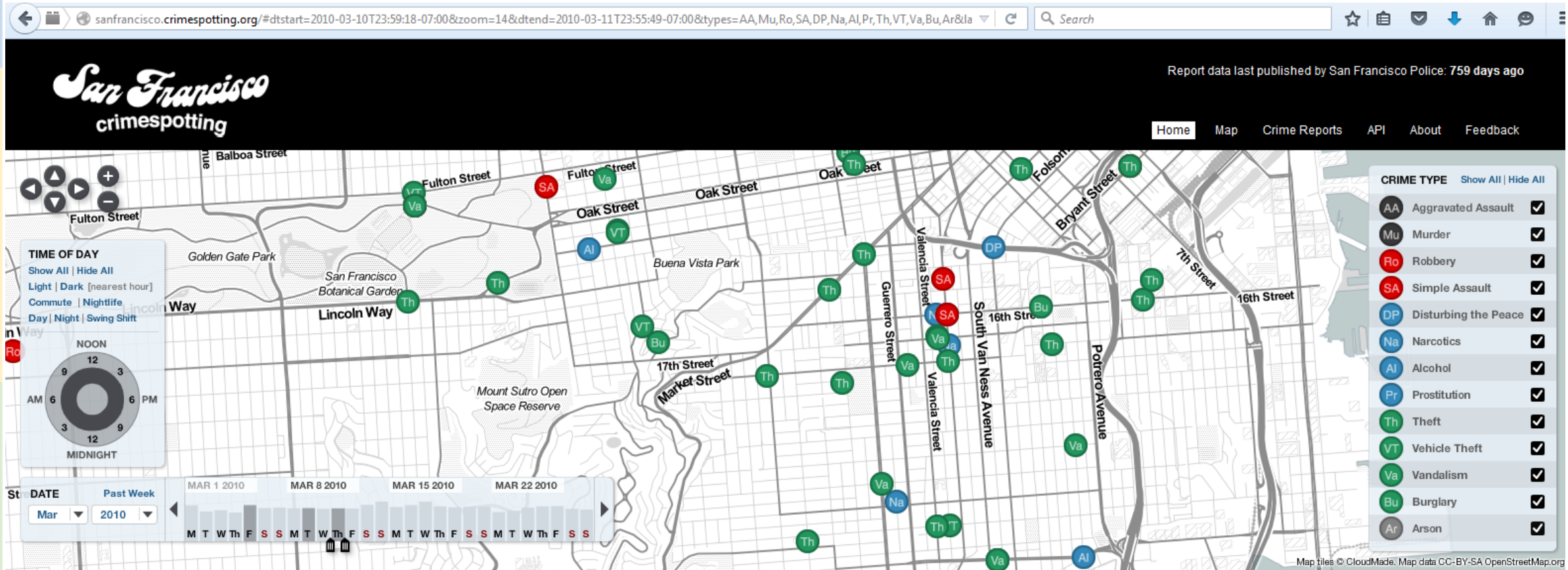


Raiņa bulvāris 19



Example – Crime data by location

<http://sanfrancisco.crimespotting.org>



Problēma(s)

- 1) Datu noliktavās **trūkst ģeogrāfiski referencēti dati**
- 2) **Ātrdarbība** (īpaši mobilajās ierīcēs)
- 3) Liels **datu apjoms** (kur un kā šos datus uzkrāt?)

Darba mērķis

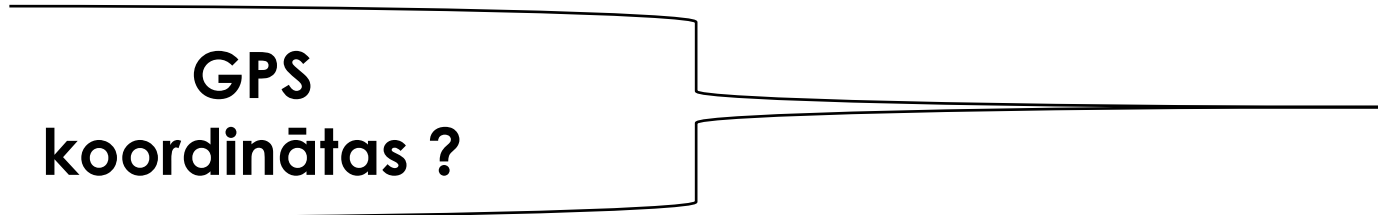


izstrādāt metodes
atrašānās vietas inteliģences pielietojumam

- 1)** datu noliktavām, kas nesatur ģeogrāfiski
referencētus datus,
- 2)** mobilajā lietotnē

Atrašanās vieta

- 1) **adrese**
- 2) **ip adrese**
- 3) pārdošanas teritorija (piem., Latvija, Somija, Zviedrija)
- 4) maršruts (piem., RIX→FRA)
- 5) detalizēts piegādes maršruts (piem., Rīga→Bauska→Jelgava)
- 6) administratīvās robežas (piem., Rīgas apkaimes)



Metodes 1

datu noliktavām, kas nesatur ģeogrāfiski referencētus datus:

- 1) Geocoding - ģeokodēšana
- 2) IP geolocation
- 3) ēku centrālo punktu vai zemes vienību ģeometrijas koordinātu izmantošana (kadastra dati)
- 4) pilsētu, valstu, pasta nodaļu centrālie punkti
- 5) sociālo tīklu iespējas (Twitter, Flickr, Instagram ģeomarkērie dati)
- 6) viedtālrunu iespējas (sensoru dati)
- 7) HTML5 Geolocation API

un citas

Koordinātu piesaistes risinājums

<http://vs148.lv/Cadastre/>

Numurs	01001070195
Platība	1,200.00
Kadastrālā vērtība	10,602.00
Persona	
Name	20127810560
Facebook	raivo.junkers
Adrese	Bērslapu iela 27, Rīga, LV-1058
Koordinātes	Meklēt koordinātes

1

Izmantojot Google Geocoding API sameklēt koordinātes un ievadīt tās laukā «Koordinātes»

Numurs	01001070195
Platība	1,200.00
Kadastrālā vērtība	10,602.00
Persona	
Name	20127810560
Facebook	raivo.junkers
Adrese	Bērslapu iela 27, Rīga, LV-1058
Koordinātes	56.9045184,24.0894203 Meklēt koordinātes

2

Saglabā

Atcelt

3

Saglabāt ierakstu

Ieraksts pēc saglabāšanas:

Numurs:	01001070195
Platība kvm:	1,200.00
Kadastrālā vērtība EUR:	10,602.00
Īpašnieks:	20127810560
Facebook:	Open Facebook
Adrese:	Bērslapu iela 27, Rīga, LV-1058
Koordinātes:	56.9045184,24.0894203 Google Maps

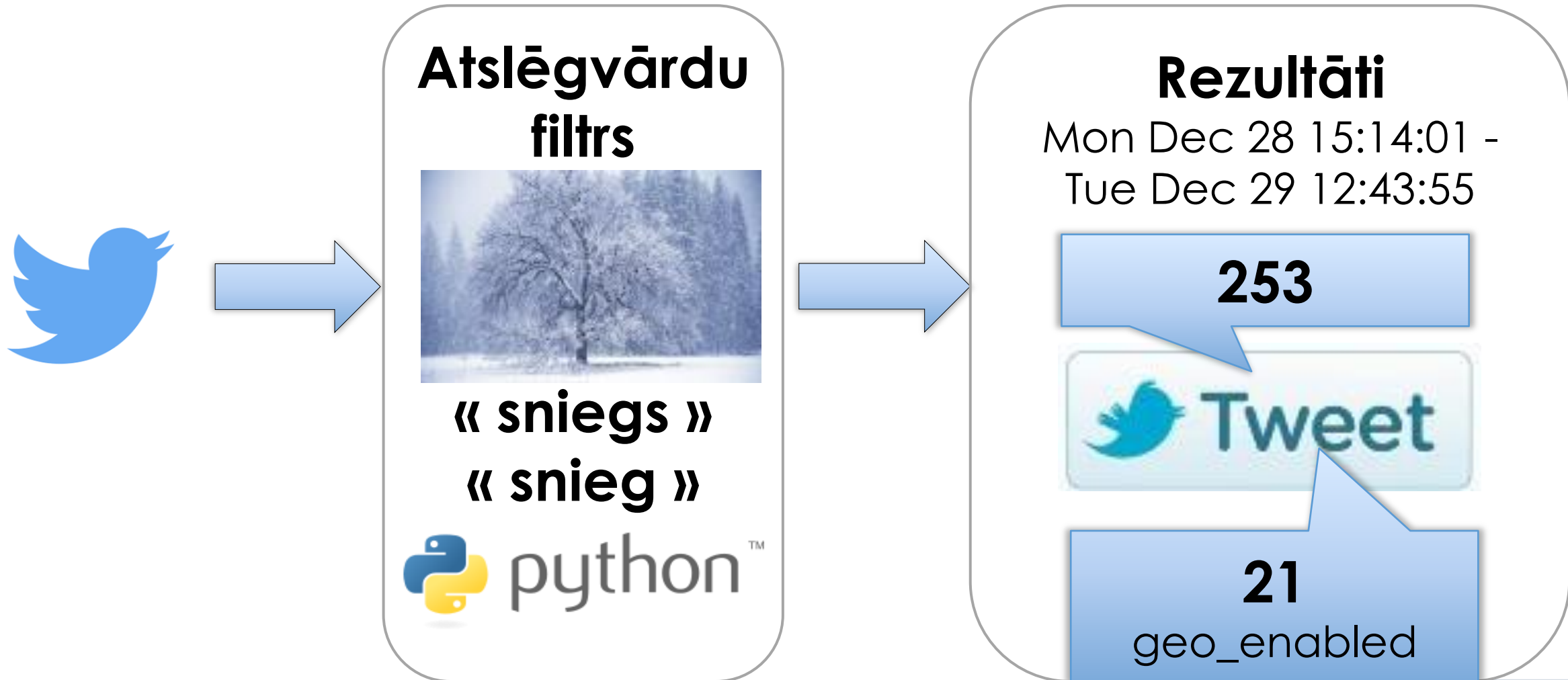
```
protected void ButtonGetCoordinates_Click(object sender, System.EventArgs e)
{
    //izveido HTTP pieprasījumu
    WebRequest request = WebRequest
    .Create("http://maps.googleapis.com/maps/api/geocode/xml?sensor=false&address="
    + HttpUtility.UrlEncode(this.TextBoxAddressName.Text));
    //sagatavo no Google saņemtās atbildes objektu
    using (WebResponse response = request.GetResponse())
    {
        //sagatavo plūsmu XML atbilžu lasīšanai
        using (Stream stream = response.GetResponseStream())
        {
            //ielādē XML dokumentu no plūsmas
            XDocument document = XDocument.Load(new StreamReader(stream));

            //nolasa koordinātes longitude, latitude
            XElement longitudeElement = document.Descendants("lng").FirstOrDefault();
            XElement latitudeElement = document.Descendants("lat").FirstOrDefault();

            //ja koordinātes satur vērtību, tad ieraksta to teksta laukā
            if (longitudeElement != null && latitudeElement != null)
            {
                this.TextBoxCoordinates.Text = latitudeElement.Value + ", " + longitudeElement.Value;
            }
        }
    }
}
```


Geomarkētie dati sociālajos tīklos

Twitter geotagging feature



Twitter –JSON



```
{ "created_at": "Mon Dec 28 16:28:39 +0000 2015",
```

```
"id": 681512093976293376,
```

```
"text": "Sniegs , saule , putenis ,Saule ... Tagad  
es zinu kapēc ir MĀTE daba nevis TĒVS daba.
```

```
Sievietes nekad nezina, ko īsti vēlas.",
```

```
"user": {
```

```
  "id": 1063264742,
```

```
  "screen_name": «XXX»,
```

```
  "followers_count": 737,
```

```
  "friends_count": 1927,
```

```
  "favourites_count": 332,
```

```
  "statuses_count": 1842,
```

```
  "created_at": "Sat Jan 05 15:34:49 +0000 2013",
```

```
  "geo_enabled": true,
```

```
  "lang": "lv",
```

```
},
```

```
"place": {
```

```
  "id": "43fae2766ad1e571",
```

```
  "place_type": "city",
```

```
  "full_name": "Sigulda, Latvija",
```

```
  "country_code": "LV",
```

```
  "country": "Latvija",
```

```
  "bounding_box": {
```

```
    "type": "Polygon",
```

```
    "coordinates": [[[24.783954, 57.001370],
```

```
                      [24.783954, 57.218559],
```

```
                      [25.157951, 57.218559],
```

```
                      [25.157951, 57.001370]]]]
```

```
  }, },
```

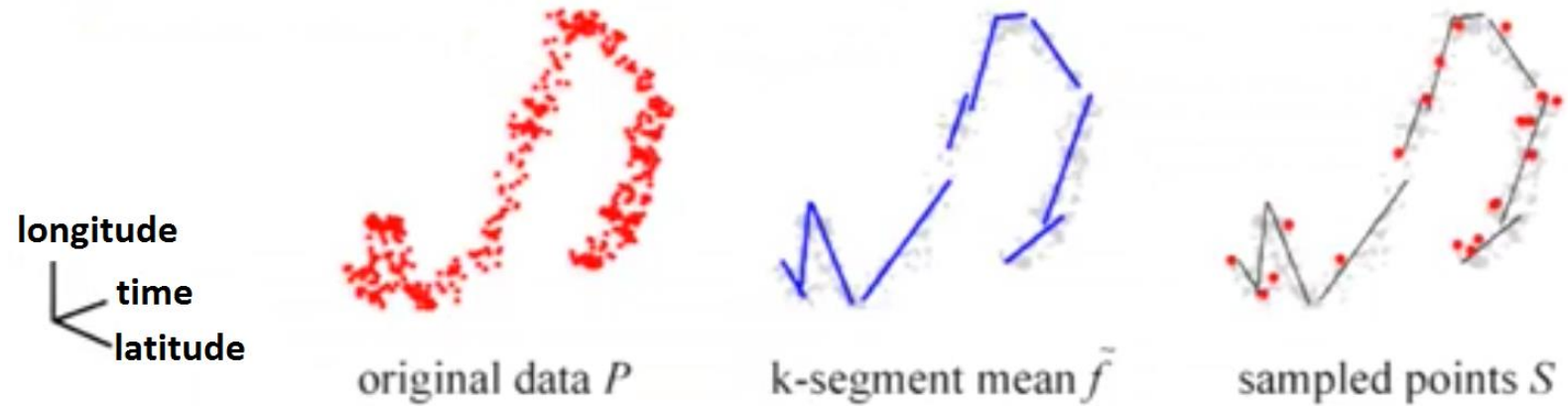
```
"timestamp_ms": "1451320119743 "
```

```
}
```

Metodes 2

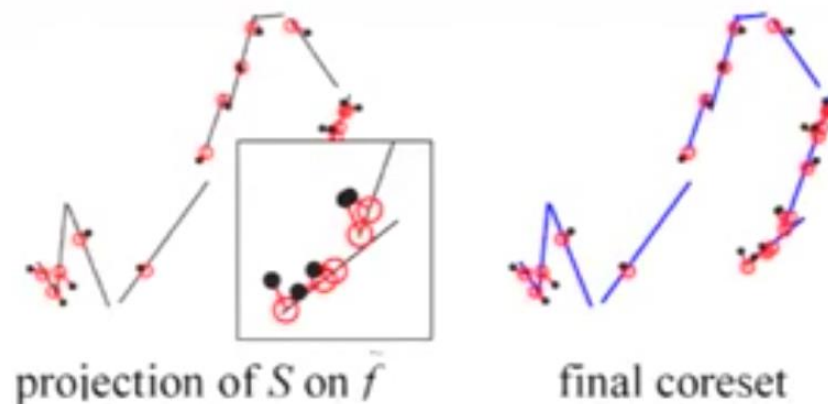
ātrdarbība un LI pārskatu pieejamība mobilajās ierīcēs

1) coresets



P: 1733 points $\rightarrow$ 42 points

2) GPU



Metodes 3

Liels datu apjoms (izaicinājums mobilajām ierīcēm)

- 1) datu uzkrāšana viedtālrunī ir dārga;
- 2) dārga ir arī datu pārraide;
- 3) neapstrādātus datus ir grūti interpretēt;
- 4) dinamiska reāllaika datu straumēšana.

Metodes un risinājumi:

- 1) kompresija – coresets
- 2) uzglabāšana hadoop
- 3) kompakti, ātri, binārie datu formāti (avro, parquet)

GPS datu pakete



1 GPS pakete = 100 baiti

latitude, longitude, time

```
{  
  "time": "2016-01-13T14:33:37+00:00",  
  "longitude": 24.783954,  
  "latitude": 57.001370  
}
```

~0,4 Mb / stundā vai ~10 Mb / dienā

~0,01 Gb / dienā / 1 ierīce

Liels GPS datu apjoms



~1,2 mljd.viedtālrunu pārdoti 2014.gadā

<http://www.gartner.com/newsroom/id/2996817>

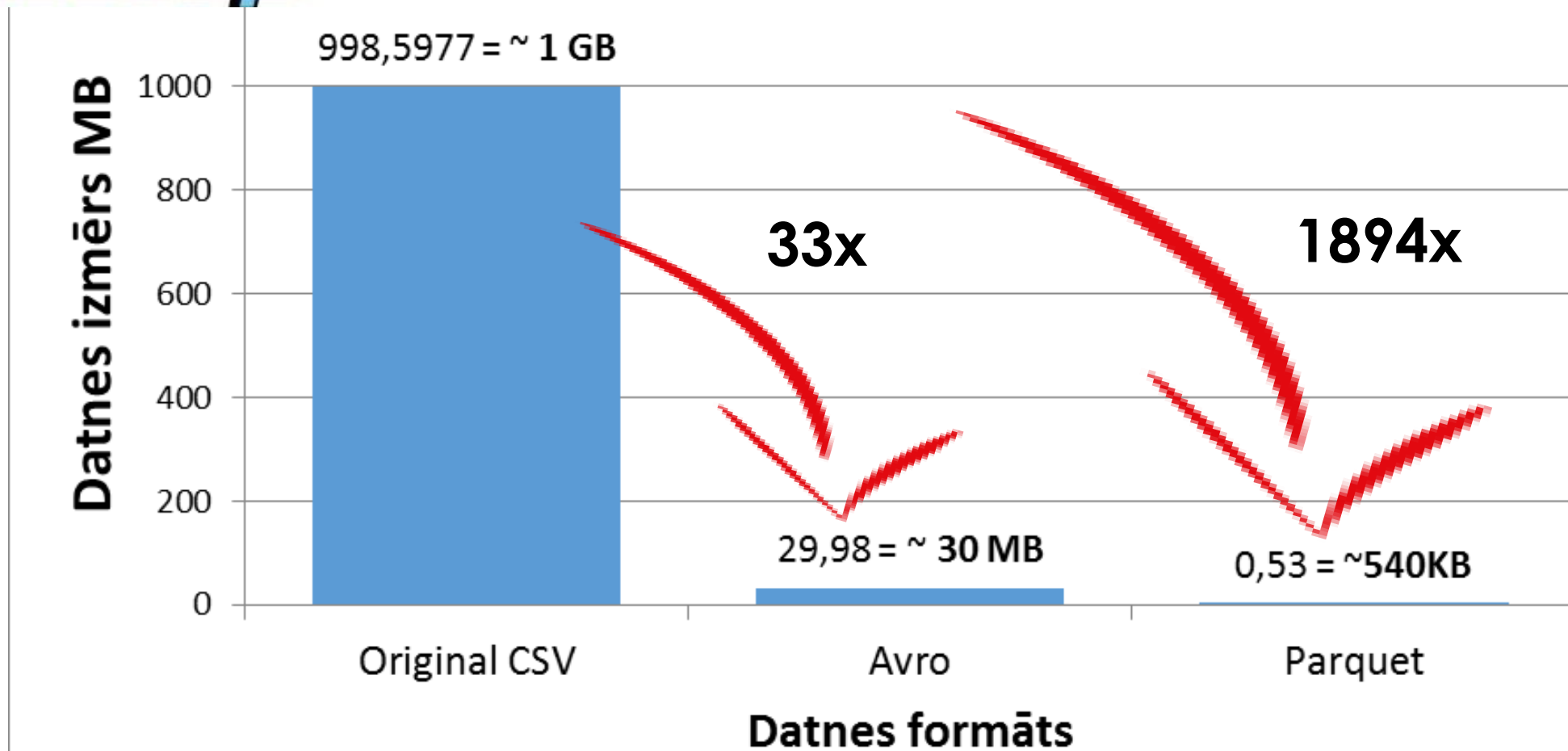
~1 mljd.viedtālruni → ~ 1PB/ dienā

Jautājums: kur un kā šos datus glabāt?

Avro / Parquet apjoms



Datnes izmērs MB~

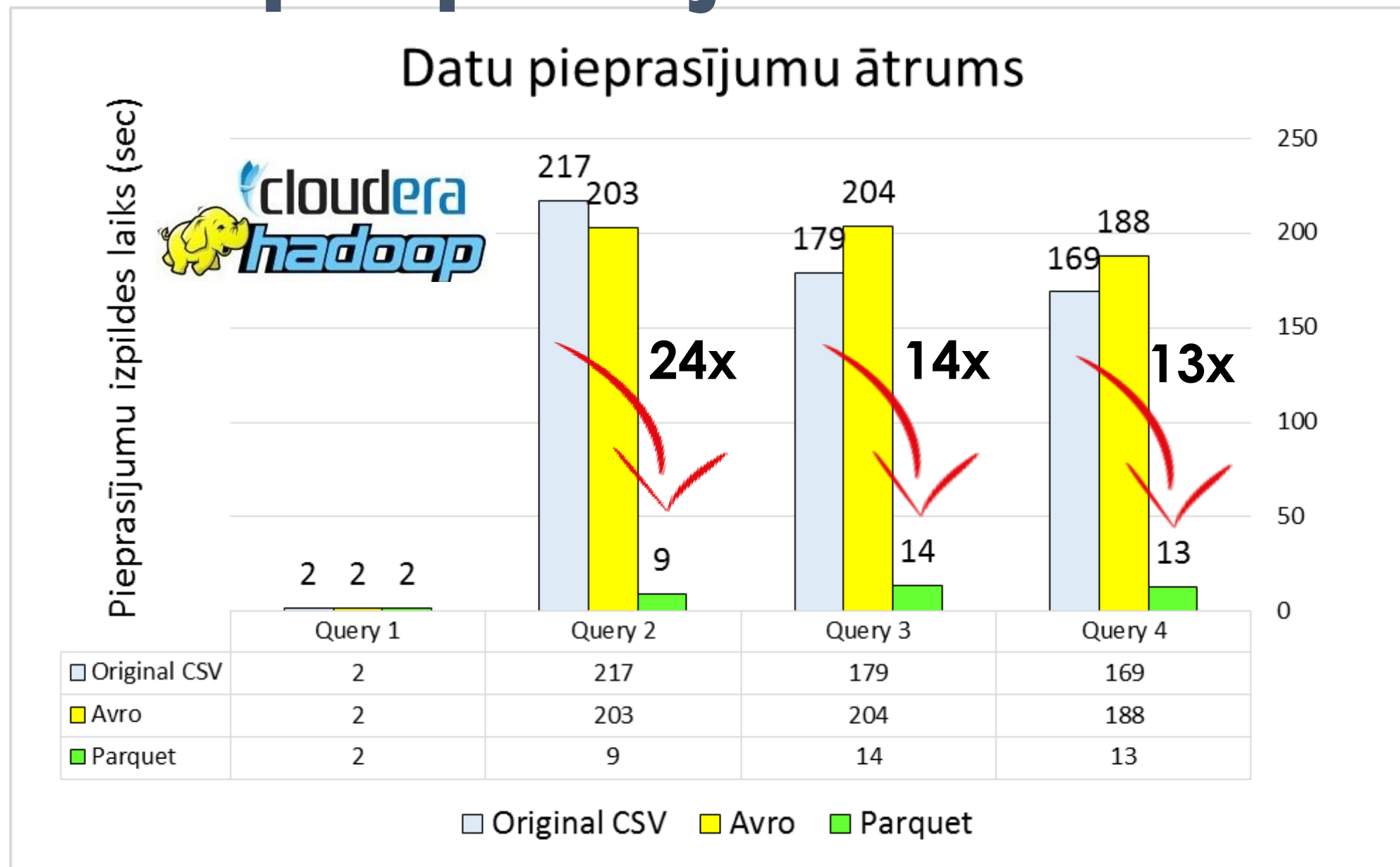


Datu struktūra un vaicājumi

- 1) select * from dataset
- 2) select **count**(*) from dataset
- 3) select route, **SUM**(price_value) as pv from dataset **group** by route
- 4) select origin, **sum**(price_value) as pv from dataset where price_value < 100 **group** by origin

- route **STRING**,
- origin **STRING**,
- destination **STRING**,
- direct_flight **STRING**,
- type **STRING**,
- departure_flight **STRING**,
- arrival_time **STRING**,
- airlines **STRING**,
- duration **STRING**,
- price_type **STRING**,
- **price_value DOUBLE**,
- discount **STRING**,
- departure_date **STRING**,
- days_long_trip **INT**,
- date_when_gathered **STRING**,
- time_when_gathered **STRING**,
- flight_id **STRING**,
- seats_left **STRING**

Datu pieprasījumu ātrums



Secinājumi

- 1) ne visi adreses ieraksti DN ir pareizi strukturēti → ģeokodēšanas metode nav precīza, izmantojot bezmaksas serviss;
- 2) IP geolocation metode darbojas ar zemu precizitāti;
- 3) nav attīstītas citas metodes atrašanās vietas noteikšanai (ģeoreferencēšanai), piem., pēc zemes vienības vai būves kadastra apzīmējuma;
- 4) sociālo tīklu un sensoru datu laikmets dod jaunas iespējas datu analīzei un atrašanās vietas inteliģences risinājumu attīstībai;
- 5) mobilā BI pieprasa viedtālruniem un planšetdatoriem pielāgotus atrašanās vietas inteliģences risinājumus ar jaunām iespējām.

Tālākie darbi

- 1) Jāpabeidz izstrādāt risinājums lielāka apjoma GPS datu ievākšanai;
- 2) Jāievāc dati eksperimentu veikšanai (coresets, GPU);
- 3) Jāsagatavo datu pieprasījumu šabloni kompakto failu formātā (Avro un Parquet) padziļinātai izpētei;
- 4) Jāsagatavo pirmais zinātniskais raksts un jāiesniedz DB&IS 2016 konferencei.

Accelerating data queries on Hadoop framework by using compact data formats

Jautājumi