Little Big Data
Analyze your own data with the Elastic Stack

inovex Meetup Köln, 18.09.2017
Wolfgang Schoch
Data Management & Analytics
inovex GmbH
wolfgang.schoch@inovex.de

Fun:
Marathon
Triathlon
Lactate
Garmin

Work (can be fun, too):
Search
Elastic Stack
Crawling
Geo Data
You create a lot of data every day
Some of them on purpose
So many devices
So many possibilities

Vendor Apps
- SUUNTO
- POLAR
- GARMIN
- [Other icons]

Third-Party Apps
- STRAVA
- SPORTTRACKS
- TRAINING PEAKS
- RUNALYZE
- [Excel icon]
Not enough?
Reclaim your data!
Just 10 clicks to create a cool biking tour

Thanks to billions of tracks uploaded by millions of users.
Motivation

• Explore your data
• Discover new patterns
• Combine data from different sources
• Customize. Everything.
• Most important: Why not?
Example: Garmin Forerunner 920xt

- GPS Multisport Watch
- Integrates in Garmin Connect online portal and mobile app
- Writes natively .fit
- Integrates in Garmin Connect
- Other formats (tcx, gpx) available via Garmin Connect
Mission

- Pull whole history of activities from Garmin Connect
- Normalize and transform it as needed
- Load it into Elasticsearch
- Visualize the hard work in Kibana
- Gain new insights
Why Elasticsearch?

- The Swiss Army knife for search and analytics
- My everyday technology
- Easy to use
Export activities as CSV

<table>
<thead>
<tr>
<th>Activity Name</th>
<th>Activity Type</th>
<th>Course</th>
<th>Start</th>
<th>Max HR</th>
<th>Time</th>
<th>Distance</th>
<th>Elevation Gain</th>
<th>Avg Speed AVG Pace</th>
<th>Max Speed</th>
<th>Max Speed Blk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Katerina Radlffen</td>
<td>Cycling</td>
<td>--</td>
<td>Wed, 31 May 2017 7:31</td>
<td>--</td>
<td>12:35</td>
<td>4.22</td>
<td>2</td>
<td>23.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Balmeranveldtinen</td>
<td>Lap Swimming</td>
<td>--</td>
<td>Tue, 30 May 2017 6:24</td>
<td>--</td>
<td>5:00:30</td>
<td>2,000 m</td>
<td>--</td>
<td>2:10 min/100 m</td>
<td>1:57 r</td>
<td></td>
</tr>
<tr>
<td>Katerina Radlffen</td>
<td>Cycling</td>
<td>--</td>
<td>Fri, 26 May 2017 8:21</td>
<td>--</td>
<td>14:07</td>
<td>4.29</td>
<td>8</td>
<td>17.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Katerina Radlffen</td>
<td>Cycling</td>
<td>--</td>
<td>Wed, 23 May 2017 6:36</td>
<td>--</td>
<td>15:03</td>
<td>5.12</td>
<td>19</td>
<td>24.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Results 1 - 20 of 2372

Export to CSV
<table>
<thead>
<tr>
<th>Aktivitätsname</th>
<th>Datum</th>
<th>Uhrzeit</th>
<th>Zeit</th>
<th>Distanz (m)</th>
<th>Geschwindigkeit (km/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Karlsruhe Radfahren</td>
<td>Mi, 31. Mai 2017</td>
<td>7:21</td>
<td>12:38</td>
<td>4.20</td>
<td>2.00</td>
</tr>
<tr>
<td>Karlsruhe Radfahren</td>
<td>Di, 30. Mai 2017</td>
<td>19:09</td>
<td>16:17</td>
<td>1:02:49</td>
<td>13.01</td>
</tr>
<tr>
<td>Bahnenschwimmen, Bahnenschwimmen</td>
<td>Di, 30. Mai 2017</td>
<td>6:24</td>
<td>53:30</td>
<td>2.000 m</td>
<td>2:12 min/100 m</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Mo, 29. Mai 2017</td>
<td>17:25</td>
<td>14:07</td>
<td>4.21</td>
<td>8.17</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Mo, 29. Mai 2017</td>
<td>8:36</td>
<td>15:03</td>
<td>5.12</td>
<td>18.20</td>
</tr>
<tr>
<td>Krafttraining, Krafttraining</td>
<td>Mo, 29. Mai 2017</td>
<td>7:04</td>
<td>1:05:57</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>So, 28. Mai 2017</td>
<td>15:12</td>
<td>164.19</td>
<td>4.08</td>
<td>6.40</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>So, 28. Mai 2017</td>
<td>17:57</td>
<td>161.49</td>
<td>4.52</td>
<td>65.100</td>
</tr>
<tr>
<td>Rheinstetten Radfahren, Radfahren</td>
<td>Sa, 27. Mai 2017</td>
<td>17:01</td>
<td>0.21</td>
<td>0.3</td>
<td>0.9</td>
</tr>
<tr>
<td>Rheinstetten Freiwasserschwimmen, Freiwasserschwimmen</td>
<td>Sa, 27. Mai 2017</td>
<td>20:04</td>
<td>43:24</td>
<td>2.16</td>
<td>0.20:08</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Sa, 27. Mai 2017</td>
<td>19:25</td>
<td>18:49</td>
<td>6.94</td>
<td>19.22</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Sa, 27. Mai 2017</td>
<td>11:19</td>
<td>142.442</td>
<td>18.122.24</td>
<td>922</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Fr, 26. Mai 2017</td>
<td>14:12</td>
<td>164.405</td>
<td>19.92</td>
<td>4.29</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Fr, 26. Mai 2017</td>
<td>7:37</td>
<td>13:59</td>
<td>4.28</td>
<td>8.18</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Do, 25. Mai 2017</td>
<td>15:55</td>
<td>168.1.295</td>
<td>15.18</td>
<td>37.5</td>
</tr>
<tr>
<td>Rheinstetten Radfahren, Radfahren</td>
<td>Do, 25. Mai 2017</td>
<td>14:19</td>
<td>0.20</td>
<td>0.21</td>
<td>1.633</td>
</tr>
<tr>
<td>Rheinstetten Freiwasserschwimmen, Freiwasserschwimmen</td>
<td>Do, 25. Mai 2017</td>
<td>14:19</td>
<td>8.410</td>
<td>4.140</td>
<td>4.01</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Do, 25. Mai 2017</td>
<td>12:39</td>
<td>18:09</td>
<td>7.07</td>
<td>19.23</td>
</tr>
<tr>
<td>Karlsruhe Radfahren, Radfahren</td>
<td>Mi, 24. Mai 2017</td>
<td>18:26</td>
<td>181.1.12:23</td>
<td>13.23</td>
<td>45.55</td>
</tr>
</tbody>
</table>
Export Single Activity

- Original is binary .fit
- GPX is XML, contains only route
- TCX is XML, contains everything
TCX export works, but…

- there is no bulk export (I created my own with some Python magic)
- can be a huge amount of data (contains GPS data)
- The X in TCX stands for XML, and nobody likes XML (except machines)
I ❤️ XML
Why not use the REST API?

• Did, was every expert should do: googled for help and advice

• Found a couple of semi-official APIs from Garmin

• Endpoint for bulk export of activities
Normalize & Transform

- JSON is a good starting point
- Different units for attributes like speed or cadence per activity type
- Select relevant attributes
Load into Elasticsearch

- Straight forward via _bulk API
- Dynamic mapping is a good starting point
- Define minimal template, e.g. to cover GeoPoints
Add another data source

- Discover relations between different data
- Gain insights
- This is where the fun begins
Example: Weight Data from Withings

- Export CSV from vendor portal
- Import with Logstash into Elasticsearch
- Play with Timelion and Visual Builder
Learnings

• Everyone produces a lot of interesting data every single day
• It is always interesting to have a close look to that
• Start with a small amount of data, avoid confusion
• Try to find answers on questions like „Is there a relation between X and Y?“
Next Steps

- Use Spark / Zeppelin to analyze and preprocess data
- Try to build heat maps on my preferred GPS tracks
Questions?