Big data analytics : “technik”

Ian Marsh, SICS-RISE
December 6-7th
2017

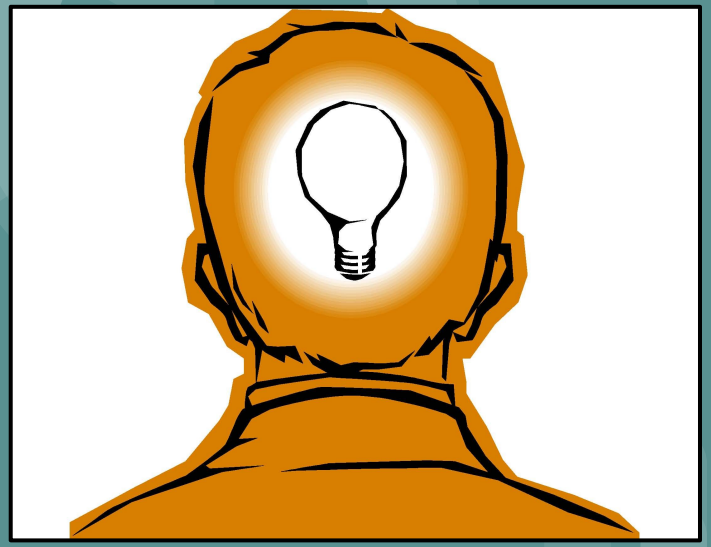
Your project : Volvo Polestar

1 Introduction (Ian)

- Bachelors in Physics (Manchester)
- Masters in Computer Science (Manchester)
- Ph.D in electrical engineering (KTH)
  - Post doc in traffic flow theory (Porto)
- 20 years in Sweden
- Telecommunications background
- Recently working in big data and analytics
- Colloboration with Ahmad -> system design expert important

Slightly new take on ‘teknik’ within BADA

1. *Experience is important*
   - Talking is cheap -> “get our hands dirty”
   - “Kött på benen”
2. *FFI conference 2 weeks ago (Stockholms massen)*
   - AI and ML appeared 20+ times
3. *Help our colleagues*
   - With analysis, and computation
   - ML/AI is multi-disciplinary
   - Example is Ahmad (systems) and Ian
4. *Always need auxiliary datasets*
   - For traffic
     - Weather, friction, events (black friday)



Next 3 hours, … 5 presentations, 5 nationalities

1. *Introduction*
   - Bjorn Bjurling, SICS-RISE
2. *Large data processing: Stored & Streaming*
   - M. Neumann, ÅF
3. *HopsWorks, a scalable platform for analysis*
   - Guatier Barthou, HopsWorks developer
4. *Traffic flow analysis using Hops+Data,*
   - Ian Marsh, Researcher, SICS-RISE
5. *Short Term Traffic Prediction,*
   - Zainab Abbas, Ph.D student KTH
6. *Traffic safety using ElasticSearch,*
   - Ahmad Al-Shishtawy, Researcher, SICS-RISE
7. *Wrap up, summary & achievements* , Bjorn Bjurling

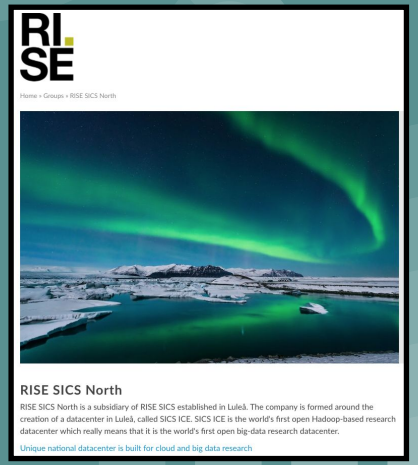
2 Large data processing (Martin)

- Alot of experience in many different fields
- We have heard about Big data in the morning ("lift it up to the board level")
- We will give an overview for the people not involved in the field daily
- Requires some competence **and** experience **and** computer science
- Streamed data adds some security

Break

3 HopsWorks (Gautier)

- Import platform for BADA
- *Secure* platform for sharing & processing
- Very fast
- Developed at SICS-RISE
- Local help available (to BADA)
- Data sets being shared & worked on
- Commercial arm of HopsWorks
  - Logicalclocks AB

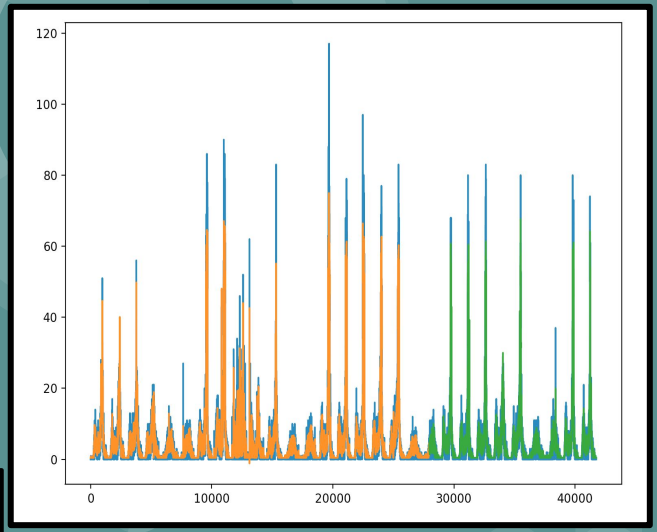


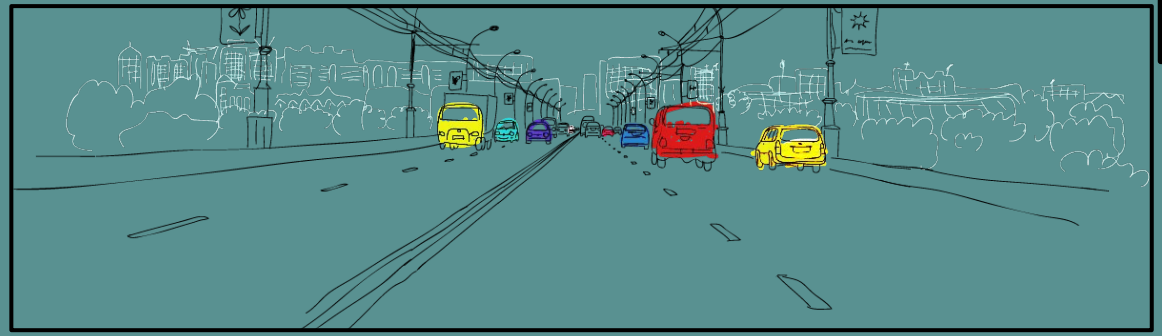
4 Traffic flow example (Ian)

- Instead of hands on we discussed
  - We will have “eyes on”
- Two examples
  - http://bada.sics.se/MFD/line.html
  - https://www.youtube.com/watch?v=Suugn-p5C1M
- Ian will go through a notebook example
- Point is to show it is easy to get started
  - Use an example from flow data
- We can be available after
- Or to present at your site(s)

5 Deep learning (Zainab)

- Heard about deep learning in the morning
- In the morning about traffic sensors
- Presentation on learning for traffic prediction
- Technique, applications and future
- ongoing @ SICS-RISE and KTH

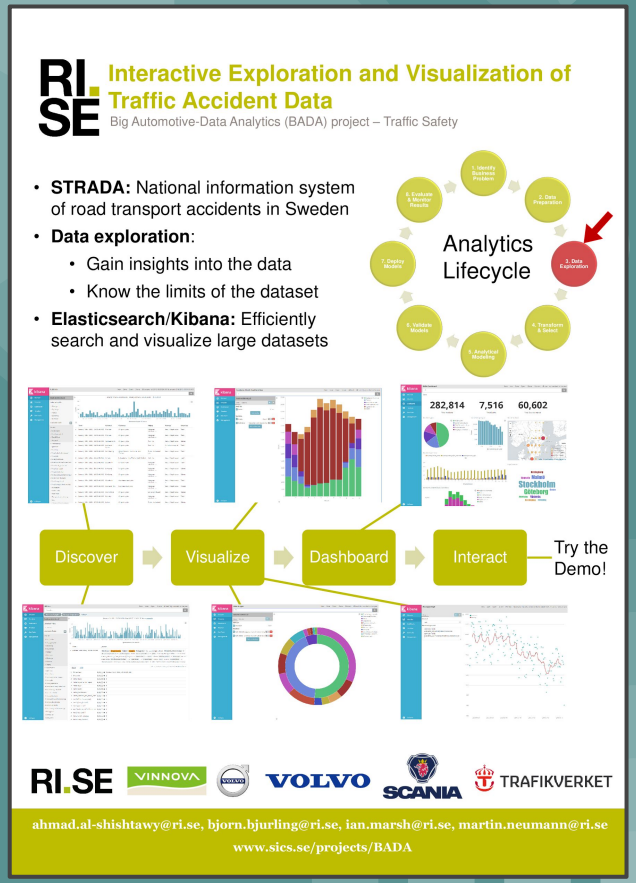




6 Traffic Safety (Ahmad)

- Traffic safety with standard tools
- Very good for data exploration
- Very important phase with large data sets
  - And visualisation
- Again we are available after …

elastic + logstash + kibana



Summary and take home(s)

1. *Start early*
   - Everything takes time
2. *Technical to technical contacts critical*
3. *Be good friends with all in project*
   - *Always need something!*
4. *Continual develoment important*
   - *Tools : Git, Website, Slack*
   - *Meet and discuss often*
   - *Results based, not "what shall we do"*
5. *Read up on latest developments, e.g.*

Deep Learning for Short-Term Traffic Flow Prediction

First Draft: December 2015
This Draft: February 2017

Real-time applications are going places

Data streaming is the paradigm behind applications that can process data and act upon insights on the fly. Apache Flink, one of the key platforms in this area, is releasing new versions and getting traction. But there may be other players in this game soon, and not the ones you think.

By George Anadiotis for Big on Data | November 29, 2017 -- 19:21 GMT (19:21 GMT) | Topic: Big Data Analytics