

# Big data and research Opportunities: Project OPENS / Tekes

Jerker Björkqvist  
Åbo Akademi University

# Challenge

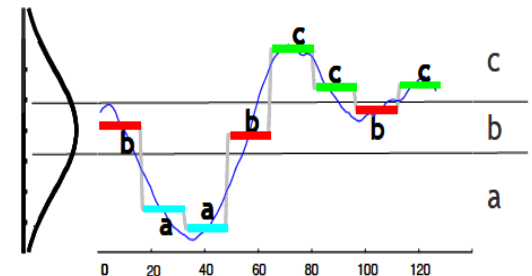
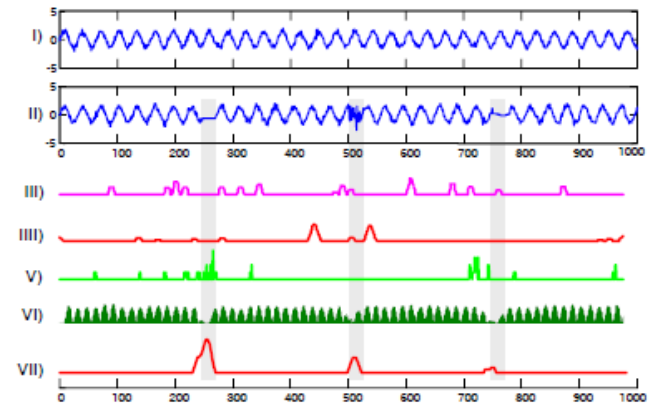
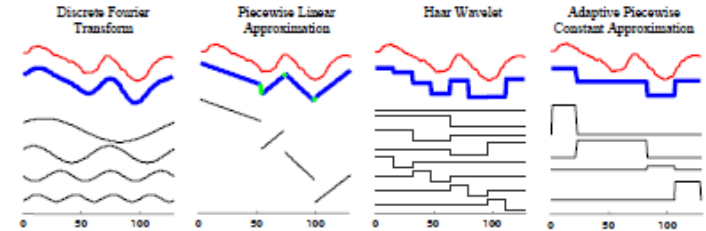
- Big Data
  - Industry has been collecting data
  - Industry wants to collect data – but what data?
- How to monetize on data?
- For companies in service business
  - **Somebody WILL** be using Big Data to provide
    - better service
    - increase their own margins
  - That **somebody** should not be somebody else...



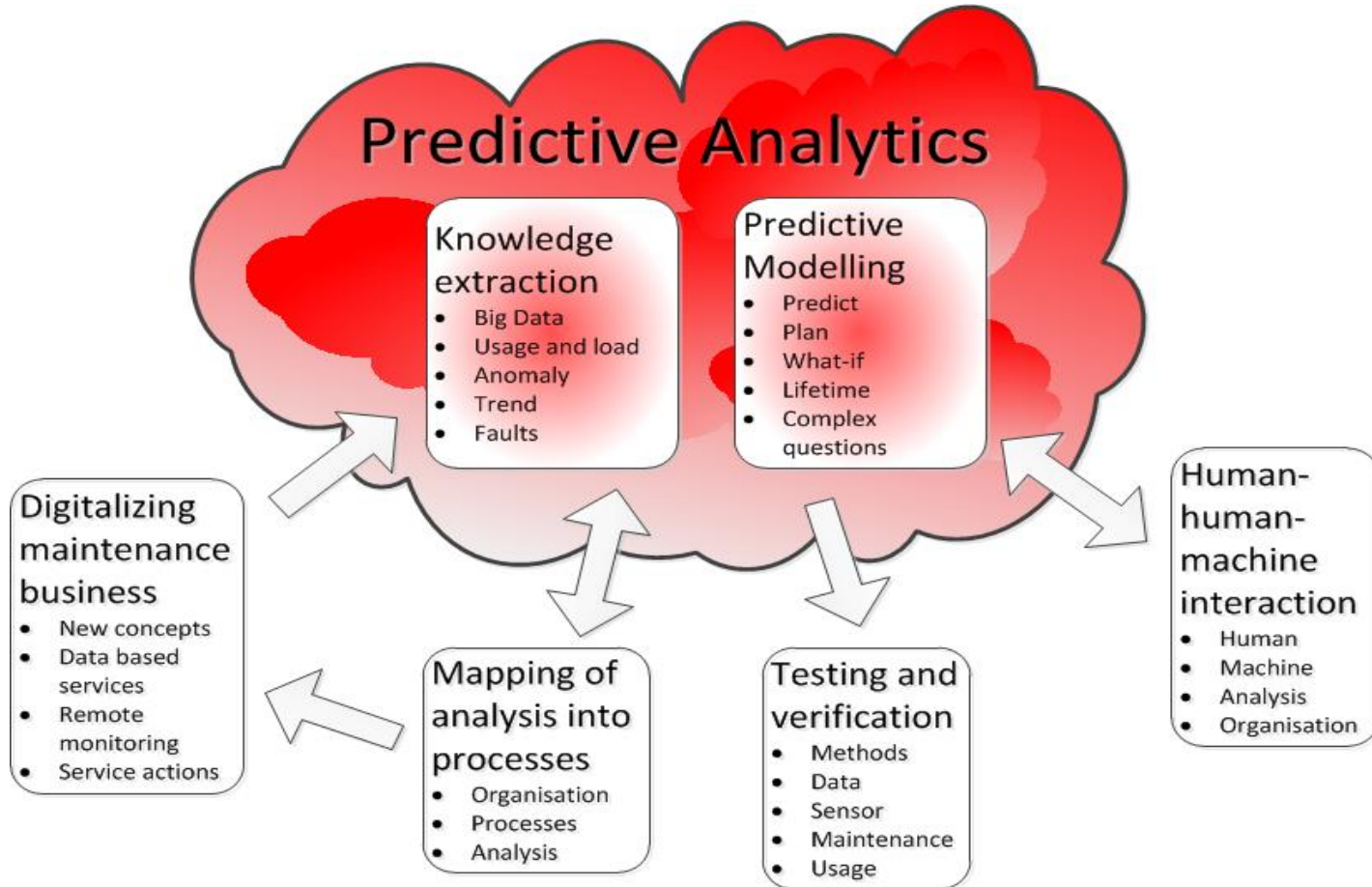
# Transferring data into knowledge

- Automated data processing
  - Time series data mining
    - Anomaly detection
    - Symbolic Approximation
  - Human-human machine interaction
    - How to get expert knowledge back into automated system?

**RESEARCH**



# Usage: predictive analytics for maintenance business



# Project OPENS

- Novel Predictive Analytics Technologies for Future Maintenance Business
  - Demonstrator driven research project
    - To reach new business possibilities using data driven predictive analytics



TAMPERE  
UNIVERSITY OF  
TECHNOLOGY



# Finding popular routes quickly from mobile sports tracking application data

Jan Westerholm  
Jani Sainio  
Information Technologies, Åbo Akademi University  
Project Supra, partially funded by Tekes.



# Background

- Mobile sports data are accumulated daily from walking, running, cycling, paddling, mountaineering, etc.
- Users wish to display their sports activities online for themselves, among friends, or even publicly.
- What does it take to create an online server application to display **user criteria selected heatmaps of popular sports routes while preserving user privacy?**

# Objective

- We received 1017551 CSV files from Sports-Tracker Oy, now part of Amer Oyj. The data was marked public by the user.
- One file = one track, total size 114 GB.
- Objective: From this data generate a heatmap representing the popularity of routes.
- The user zooms to the desired geographical area, selects sports type, time of year and time of day.



# Implementation

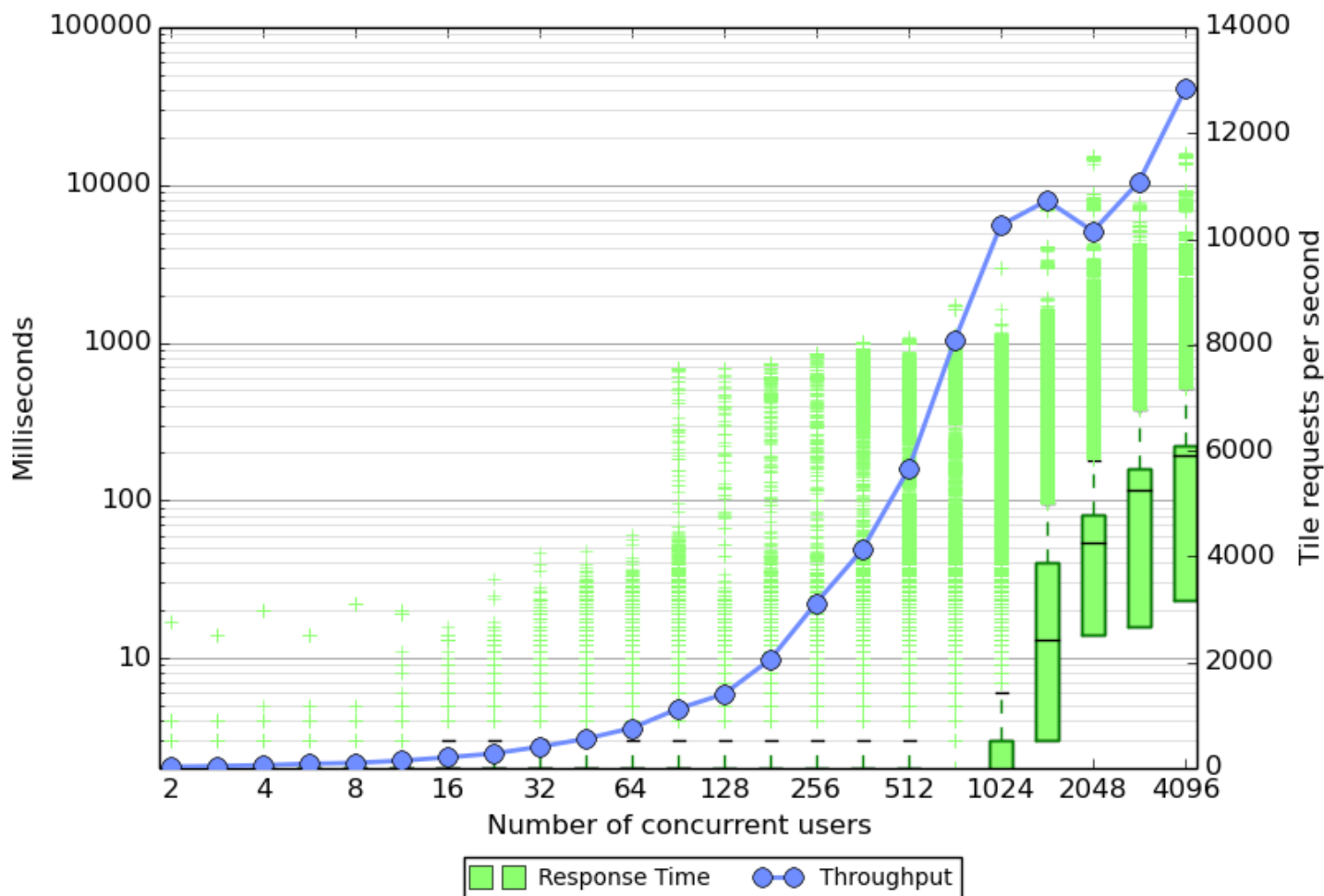
- We used tiled maps from Google Maps, OpenStreetMap and OpenCycleMap, with 14 zoom levels.
- Preprocess data as much as possible: divide tracks into tiles, rasterize tracks on all zoom levels.
- Find all tracks matching search parameters, discard the results if less than five tracks match the criteria.

# Heatmaps





# Stress tests



# Conclusion

- Very large data sets can be processed and displayed in millisecond with appropriate pre-processing and fast retrieval of in-memory data.
- We also "replicated" the system in a cloud environment for increased loads.
- We expect to be able to enlarge the server application to multinode systems to handle increased amounts of data.