Master's Thesis: Descriptive study of the ELK stack applicability for data analytics use cases in the mobility industry

October 6th, 2015

Oemer Uludag, Prof. Dr. Matthes, Thomas Reschenhofer, M.Sc.

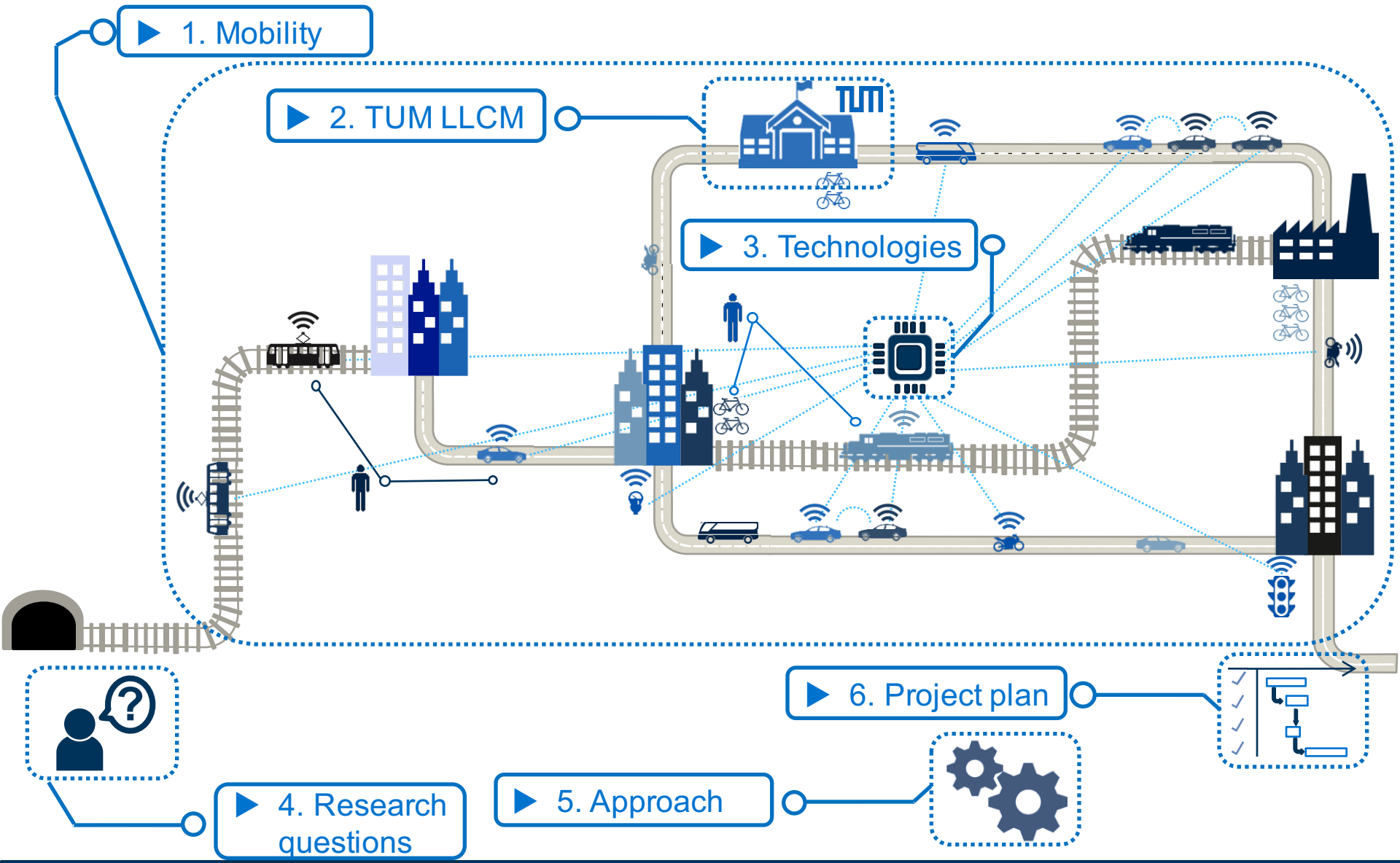
Software Engineering for Business Information Systems (sebis)
Department of Informatics
Technische Universität München, Germany

wwwmatthes.in.tum.de

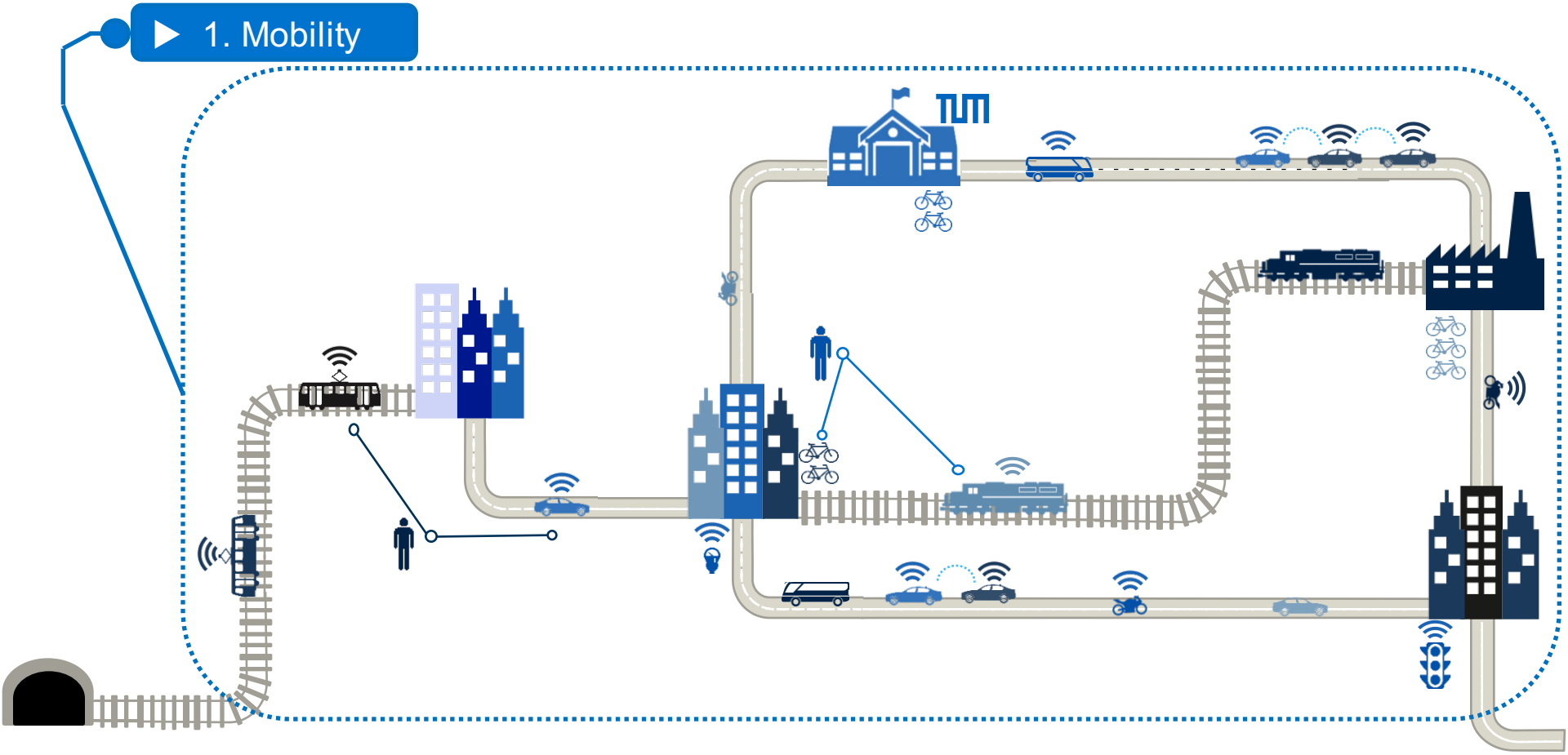
Agenda

1. **Mobility**
   - Society in transition and urbanization
   - Effect of the global urbanization on Munich
2. **TUM Living Lab Connected Mobility (TUM LLCM)**
   - Research project TUM LLCM overview
3. **Technologies**
   - Big Data and search-based discovery tools
   - The Elasticsearch, Logstash and Kibana (ELK) stack
4. **Research questions**
   - Overview of relevant research questions of the master's thesis
5. **Approach**
   - Approach for answering the research questions
6. **Project plan**
   - Structured project plan for realizing the approach

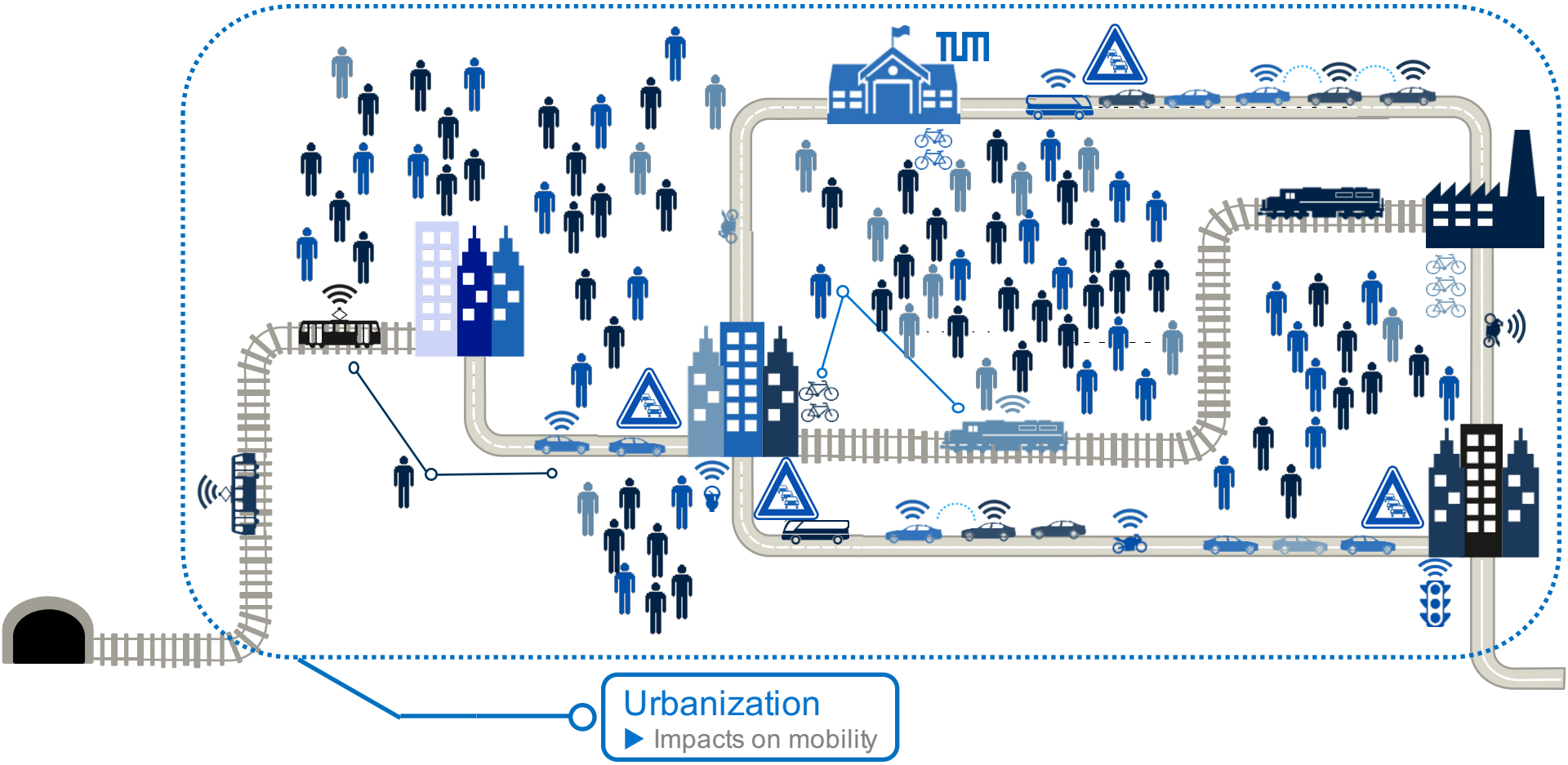
Building up a framework for the master's thesis



1. Mobility



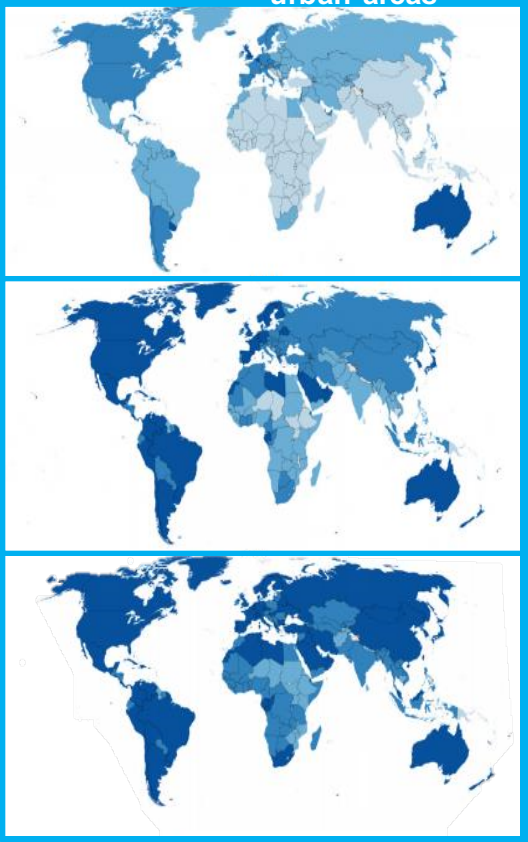
Society in transition

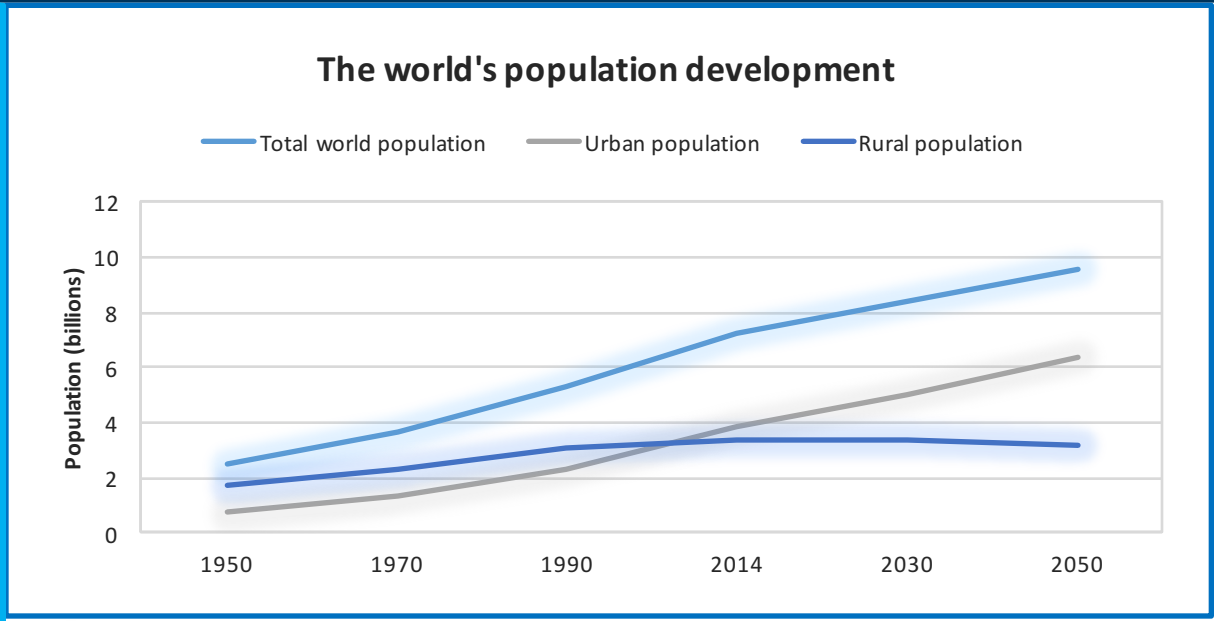


Urbanization
- Impacts on mobility

Profound changes in the coming decades for the world's population

Percentage of the population residing in urban areas





Implications

In 2007, for the first time in history, global urban population exceeded the global rural population.

66% of the world's population is expected to be urban by 2050.

54% of the world's population was urban in 2014.

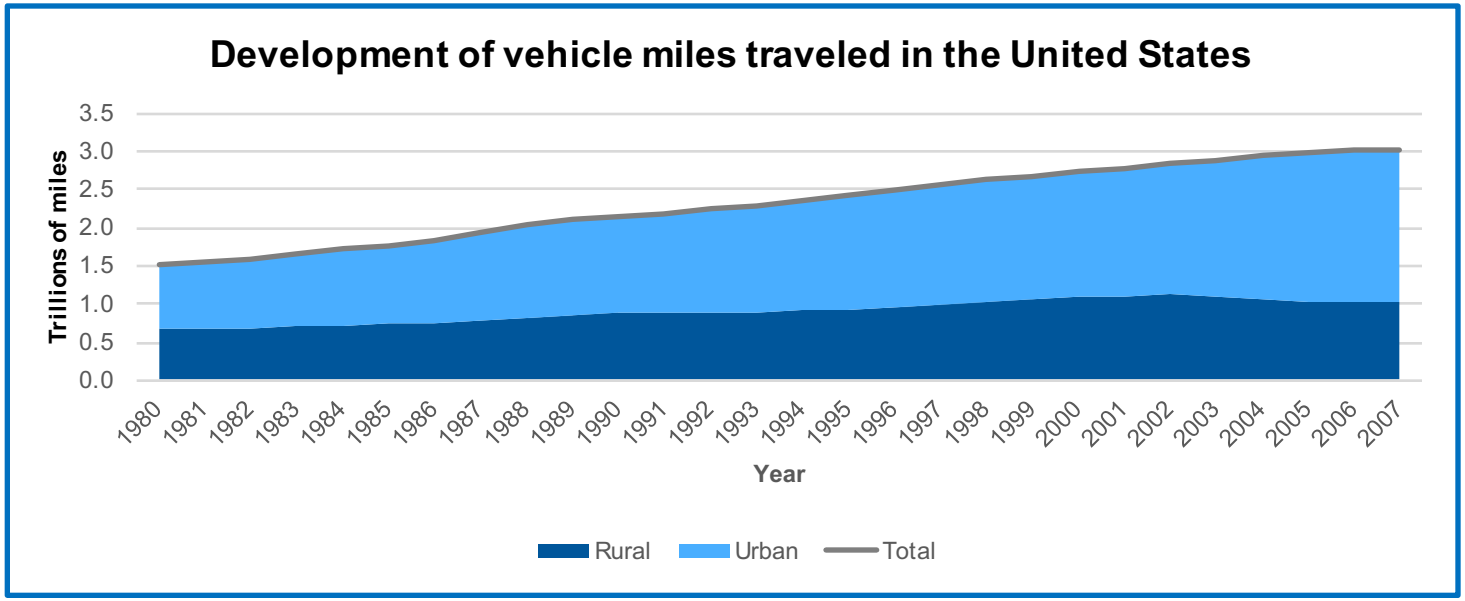
Sources: based on [1]

Rising urbanization impedes mobility

The increasing population in urban areas involves serious challenges for mobility in urban areas:

1. Increasing vehicle miles traveled

- Given that the population increasingly resides in urban areas, it is clear that vehicle miles traveled will increasingly accrue in urban areas.



Sources: based on [2, 3]

Rising urbanization impedes mobility

The increasing population in urban areas involves serious challenges for mobility in urban areas:

2. Increasing congestion in urban areas

- Rising population in urban areas leads to increasing congestion in urban areas.

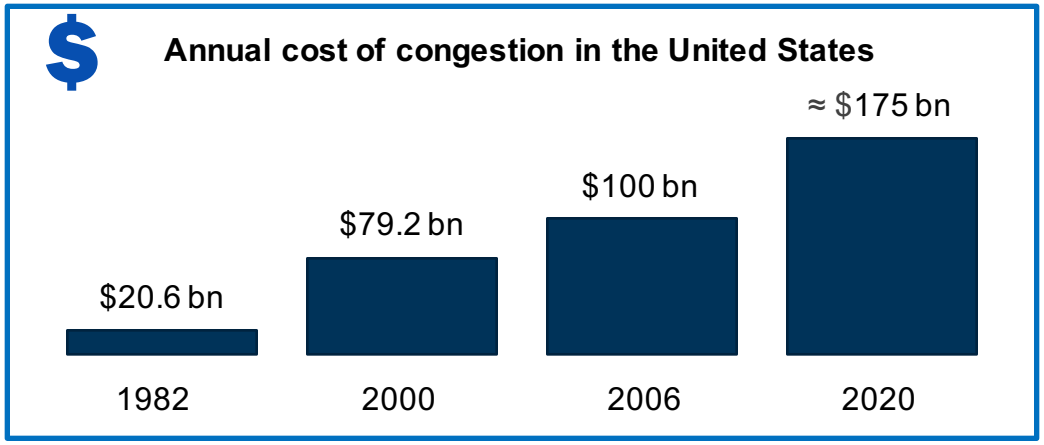
Increasing delay per U.S. commuter in traffic (hours):

- 1982: 14.4
- 2000: 34.8
- 2006: 39.1
- 2020: ≈ 41.0

How does this situation look in Munich?

Increasing fuel waste (billions of gallons):

- 1982: 0.36
- 2000: 1.63
- 2006: 2.18
- 2020: ≈ 3.2

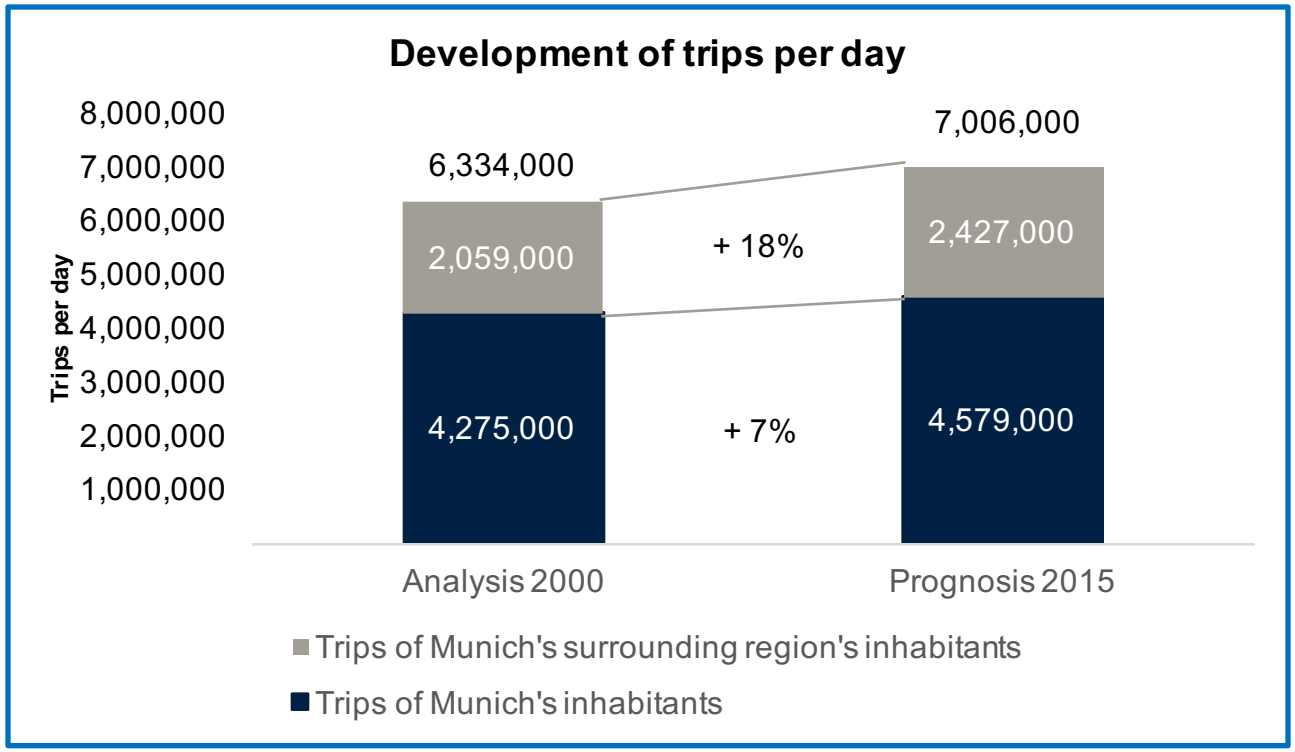


Sources: based on [4]

Munich is also affected by the global urbanization trend

Some facts about Munich's transition

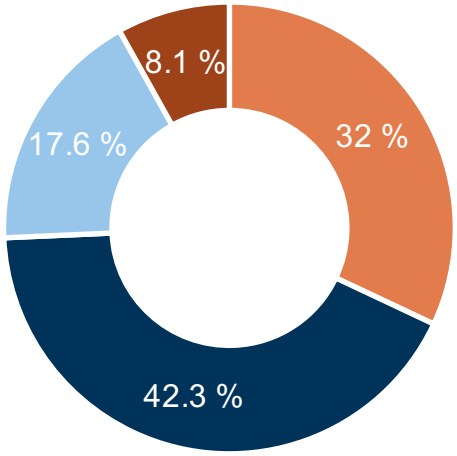
- A population growth of 15.4% is expected between 2013 and 2030
- Based on historical and predicted data of Munich's population and employment development, a great increase of its traffic is expected



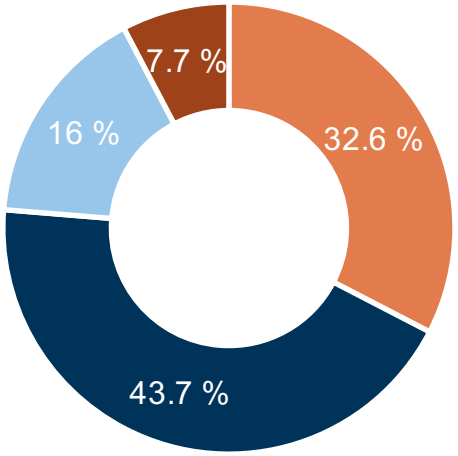
Sources: based on [5, 6]

Munich is also affected by the global urbanization trend

Overview of preferred means of transport 2000



Overview of preferred means of transport 2015

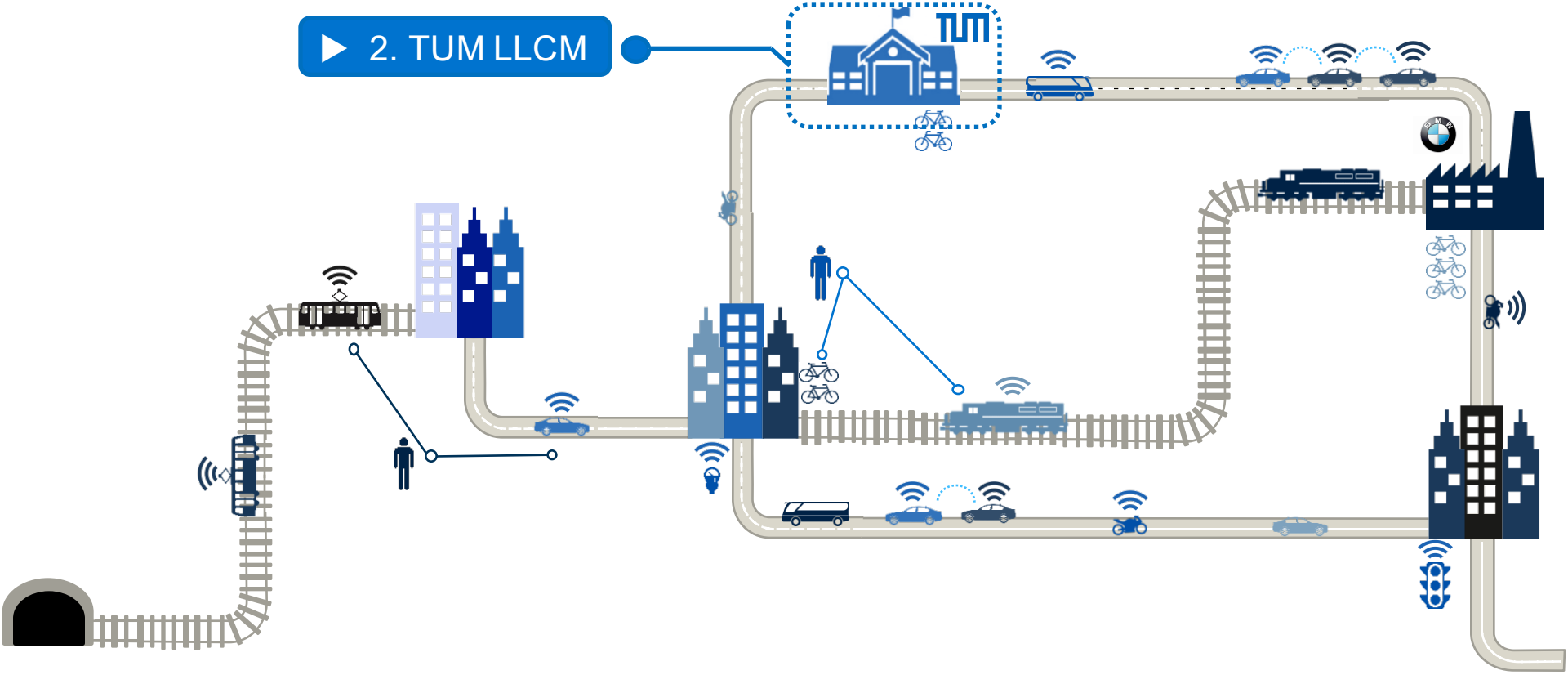


Intelligent use of data as an opportunity to improve the mobility situation in Munich

- TUM Living Lab Connected Mobility (TUM LLCM)

Sources: based on [6]

2. TUM LLCM



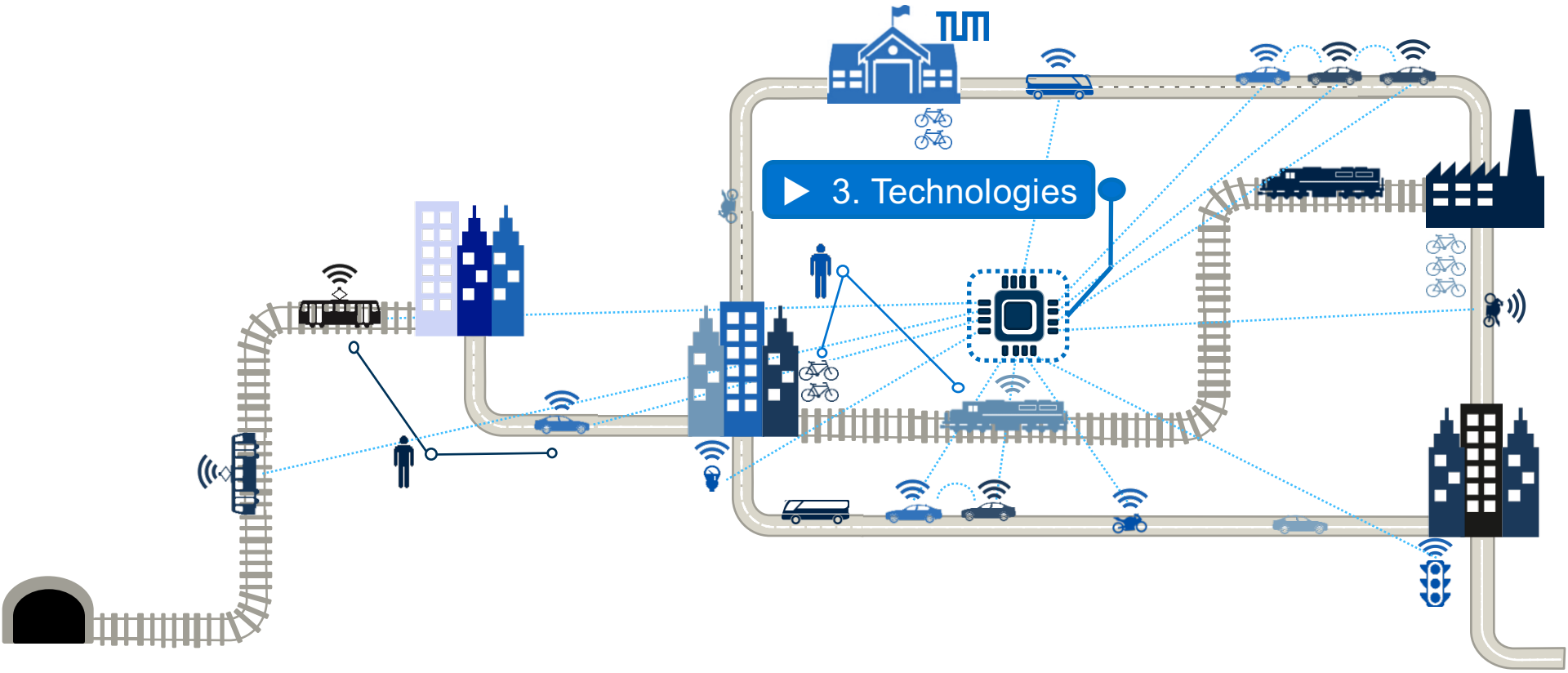
The TUM Living Lab Connected Mobility (TUM LLCM)

- aims to make a significant contribution to an open service platform digital mobility in Bavaria by researching, developing, and evaluating innovative platform services and applications for digital mobility platforms.
- targets to promote the design and prototypical implementation of an open usable digital mobility platform.
- is carried out in close collaboration with leading industrial suppliers of digital mobility services.
- is used as an innovation platform for simplified and accelerated exchange with the development of digital mobility services between university, industry and end-users.

The TUM LLCM aims to merge different mobility-related technologies into a platform that collects and processes data. Once merged, these data are then provided as a basis for an efficient, safe and comfortable mobility.

Sources: based on [7]

3. Technologies



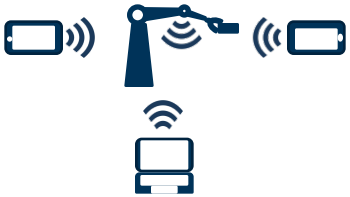
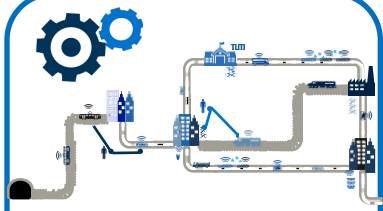
Technological transformation in the area of mobility

- Ubiquitous computing is also in the area of mobility, promoting to new technologies and leading to a rapid and disruptive technological transformation in this area.
- Various kinds of vehicular sensors generated by the Internet of Things and a new generation of strongly networked and integrated systems contribute continuously to the expansion of huge mounds of data.
- The ability to process and analyze this data and to extract insight and knowledge that enable intelligent services is a critical capability.

Sources: based on [8, 9, 10]

Technological transformation in the area of mobility

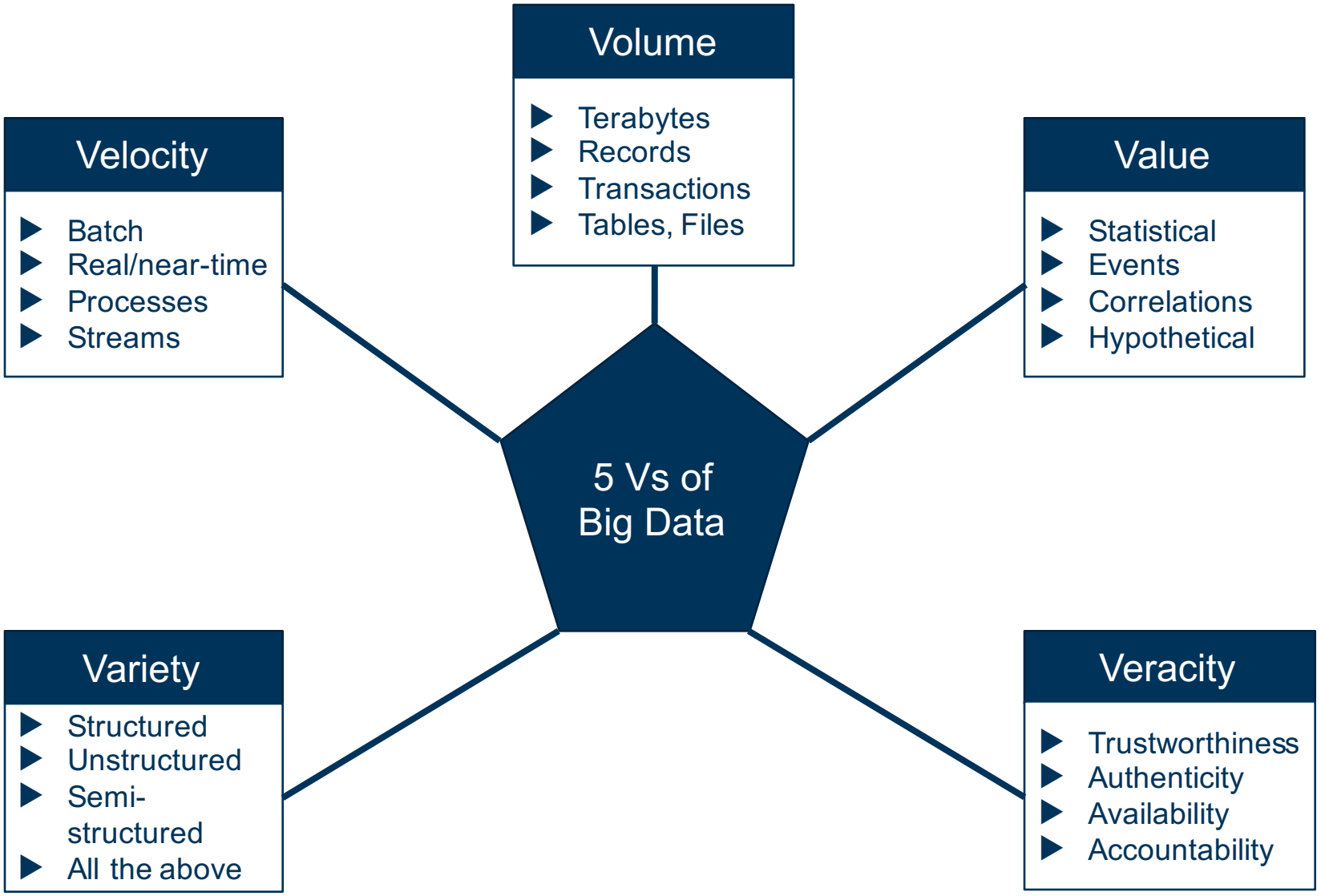
Example of these kinds of applications in the mobility industry comprise:

Connected vehicle	Autonomous driving	Smart manufacturing	Industrial internet	Mobility services
				

The 5 Vs of volume, velocity, variety, veracity, and value are often used to describe the requirements of Big Data applications and the characteristics of Big Data.

Sources: based on [10]

Big Data and search-based data discovery tools



Sources: based on [11, 12]

Search-based data discovery tools

- raise huge expectations and promise high benefits for organizations among Big Data and analytics technologies.
- facilitate users to develop and refine views and analyses of multi-structured data using search term and to find relationships across structured, unstructured, and semi-structured data.
- feature a performance layer to lessen the need for aggregates and pre-calculations.
- are vended by i.e. Attivio, IBM, Oracle, Splunk, and ThoughtSpot

The combination of the three open source projects Elasticsearch, Logstash and Kibana (ELK), also known as the ELK stack is an outstanding alternative to commercial search-based data discovery tools.

Sources: based on [13]

The Elasticsearch, Logstash and Kibana (ELK) stack

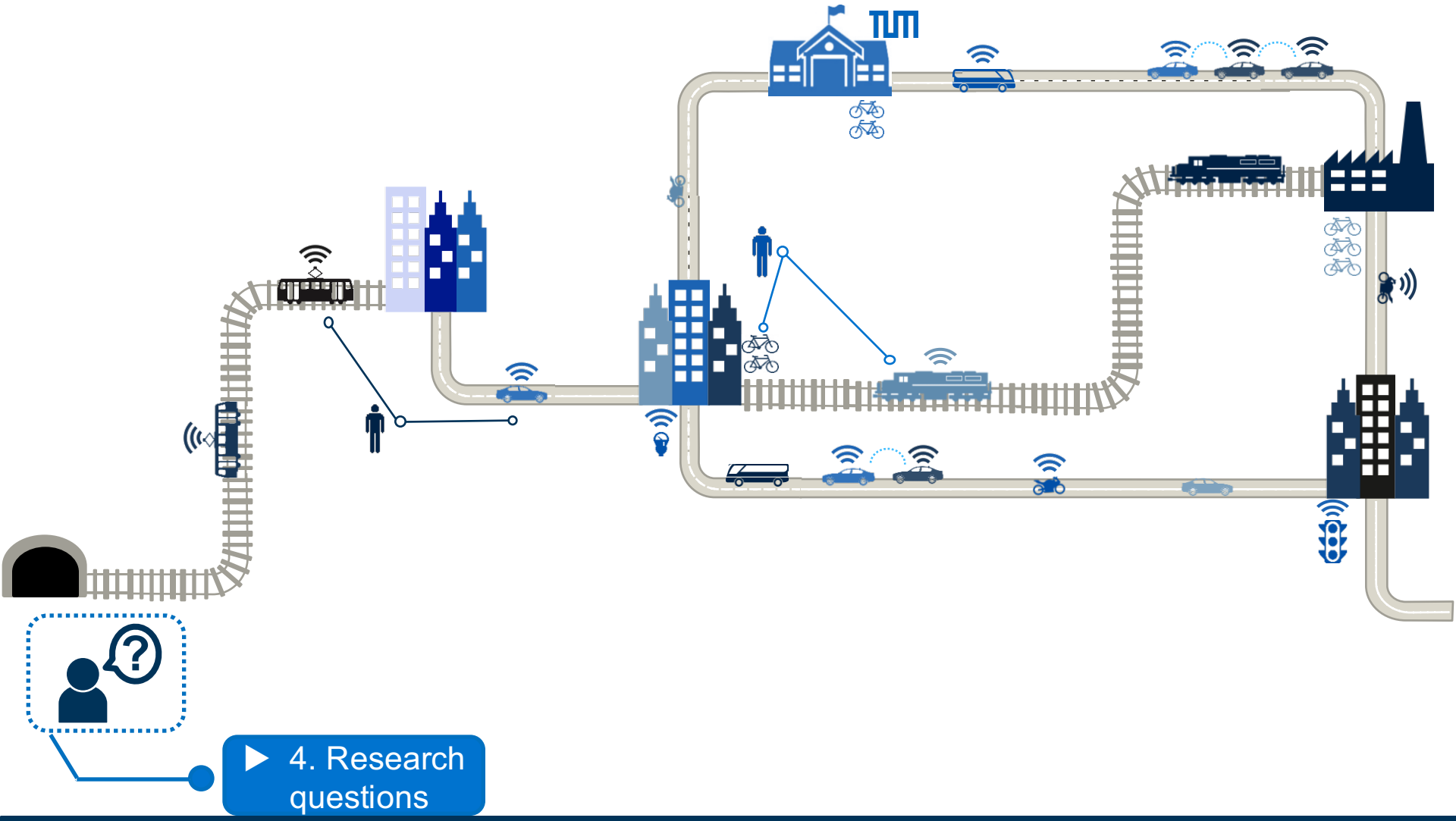
The ELK stack as an outstanding search-based data discovery tool

- The ELK stack is and end-to-end stack that glean actionable insights in real time from almost any type of structured and unstructured data source.
- **Elasticsearch**: performs deep search and data analytics.
- **Logstash**: is responsible for centralized logging, log enrichment and parsing log files.
- **Kibana**: is used to visualize data from Elasticsearch.

Due to the fact that the ELK stack is used by many organizations for a variety of business critical functions, an evaluation of its applicability in the mobility industry seems auspicious and indispensable.

Sources: based on [14, 15]

4. Research question



4. Research questions

Overview of relevant research questions

Research question 1:

What are capabilities and key features of the ELK stack for data analytics?

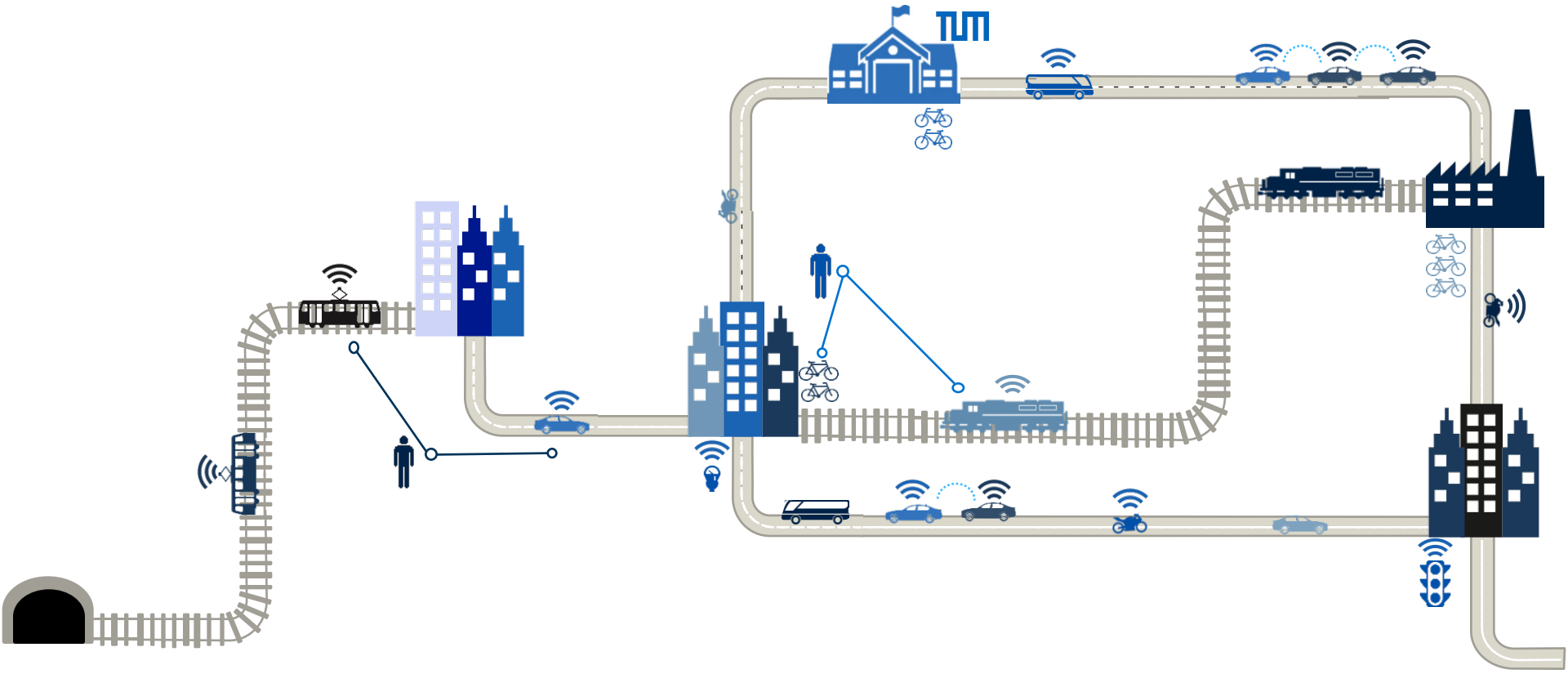
Research question 2:

What are data analytics use cases in the mobility industry?

Research question 3:

For which type of data analytics uses cases is the ELK stack applicable?

5. Approach

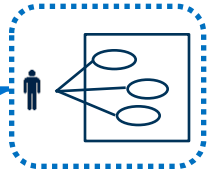


▶ 5. Approach

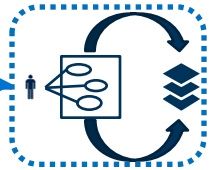
Approach for answering the research questions

1. Implementation of the ELK stack
   - Deliverable: Configured ELK stack on the cluster

2. Analysis of the ELK stack
   - Deliverable: Derivation of capabilities and features of the ELK stack



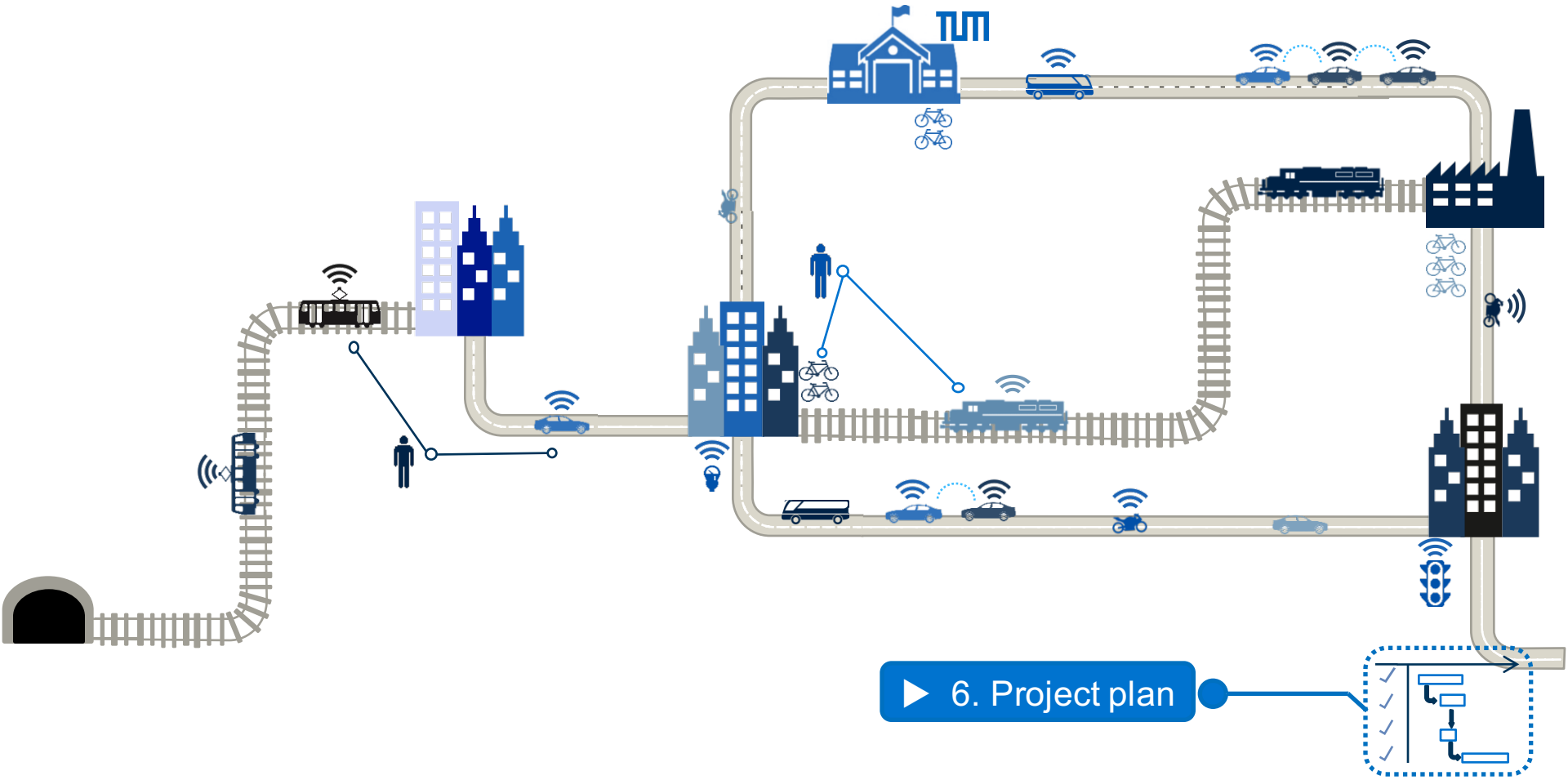
3. Ascertainment of data analytics use cases in the mobility industry
   - Deliverable: Collection of data analytics use cases



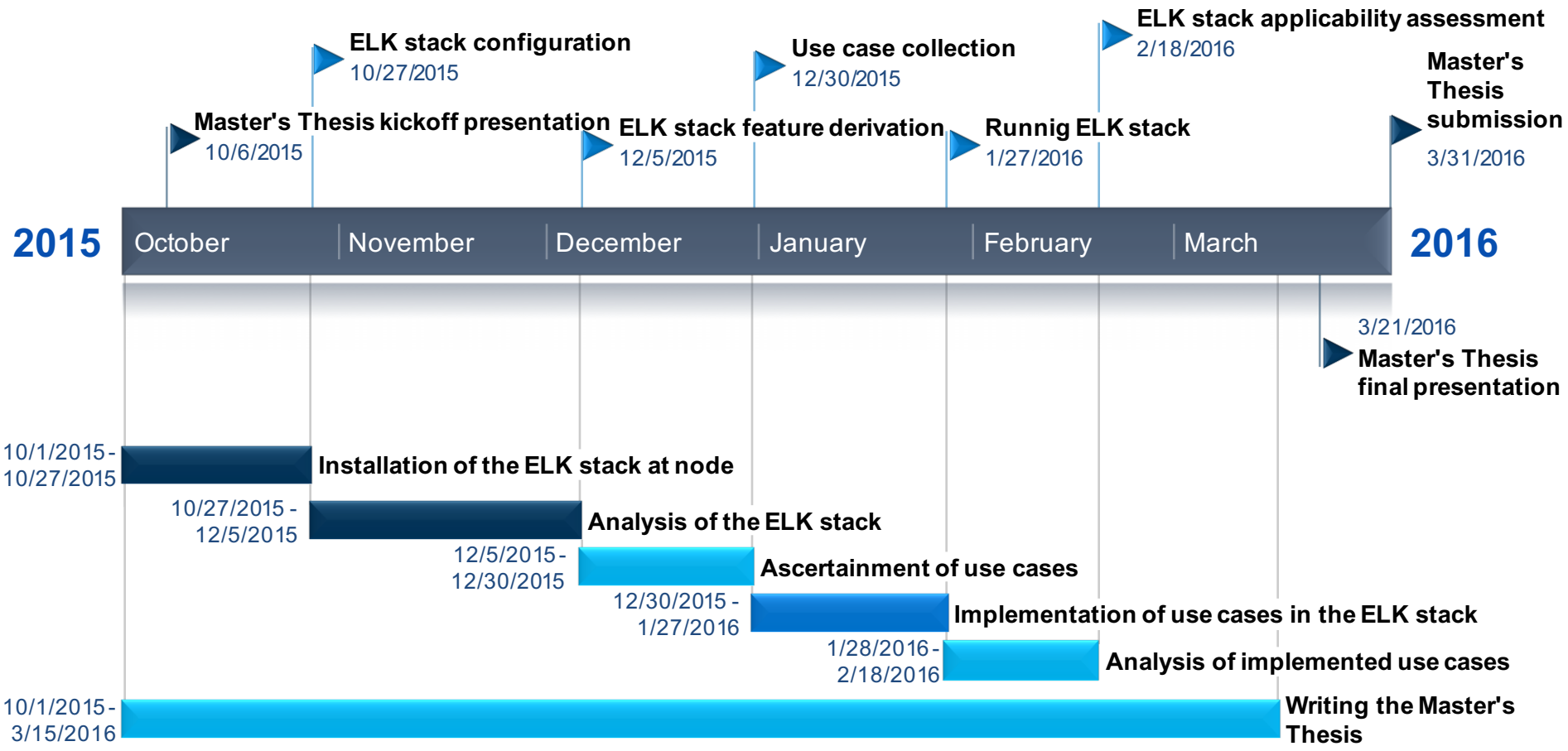
4. Implementation of data analytics uses cases in the ELK stack
   - Deliverable: Running ELK stack with real data

5. Analysis of implemented data analytics use cases
   - Deliverable: Derivation of ELK stack applicability for data analytics use cases for the mobility industry

6. Project plan



Structured project plan for realizing the approach



Thank you very much for your attention!

Do you have any questions?

References

[1] United Nations, Department of Economic and Social Affairs, Population Division. 2015. *World Urbanization Prospects: The 2014 Revision*. New York, USA.

[2] William J. Mitchell, Christopher E. Borroni-Bird, and Lawrence D. Burns. 2010. Reinventing the Automobile: Personal urban mobility for the 21st century. MIT Press, Cambridge, MA, USA.

[3] US Department of Transportation, Federal Highway Administration, Office of Highway Policy Information. 2014. Annual Vehicle – Miles of Travel, 1980 – 2007. Washington, DC, USA. http://www.fhwa.dot.gov/policyinformation/statistics/vm02_summary.cfm. Accessed: October 6, 2015.

[4] David Schrank, Tim Lomax, and Bill Eisele. 2011. The 2011 Urban Mobility Report. Texas Transportation Institute, Texas A&M University, TX, USA.

[5] Landeshauptstadt München. 2015. Demografiebericht München – Teil 2, Kleinräumige Bevölkerungsprognose 2013 bis 2030 für die Stadtbezirke. Munich, Germany.

[6] Landeshauptstadt München. 2006. Verkehrsentwicklungsplan. Munich, Germany.

[7] TUM Living Lab Connected Mobility. 2014. Technische Universität München, Munich, Germany. http://tum-llcm.de. Accessed: October 6, 2015.

[8] Michael Friedewald and Oliver Raabe. 2010. Ubiquitous computing: An overview of technology impacts. *Telematics Inform*, 28, 2 (2011), 55 – 65.

[9] Günther Sagl, Martin Loidl, and Bernd Resch 2012. Visuelle Analyse von Mobilfunkdaten zur Charakterisierung Urbaner Mobilität. In *Geoinformationssysteme*, Wichmann Verlag, Berlin, Germany, 72 – 79.

[10] Andre Luckow, Ken Kennedy, Fabian Manhardt, Emil Djerekarov, Bennie Vorster and Amy Apon. 2015 In *Proceedings of IEEE Conference on Big Data*. IEEE. Santa Clara, CA, USA.

[11] Philip Russom. 2011. Big Data Analytics. *TDWI Best Practices Report*, *Fourth Quarter*. The Data Warehouse Institute, Renton, WA, USA.

[12] Yuri Demchenko, Paola Grosso, Cees de Laat, and Peter Membrey. 2013. Addressing Big Data Issues in Scientific Data Infrastructure. In *Proceedings of the 2013 International Conference on Collaboration Technologies and Systems*. IEEE. San Diego, CA, USA, 48–55.

[13] Bart De Muynck. 2015. How to Derive Value From Big Data in Transportation. Gartner Research.

[14] An Introduction to the ELK stack. Elastic. http://www.elastic.co/webinars/introduction-elk-stack. Accessed: October 6, 2015.

[15] The ELK Stack in a DevOps Environment. Elastic. http://www.elastic.co/webinars/elk-stack-devops-environment. Accessed: October 6, 2015.