Part 1 (BigData & DataScience)

Use Case: Connected Car Revolution

Big Data Challenges

Characteristics	**Challenges**
Volume	Effectively capture and analyze 30.000+ signals and data points from sensors in every car Significant Data Volumes • 25 GB per hour per car • 130 TB per year per car
Variety	Streaming of real-time data Data coming in different formats from multiple IoT applications and resources
Velocity	Drive analytics on 12 million miles of driving data collected every hour
Veracity	Detecting erroneous data in IoT
Value	Consider objectives and economic considerations, such as: • Predictive maintenance • Usage based insurance • Provide recommendations based on traffic patterns, public safety hazards and provide and provide recommendations accordingly

4 Level of Data Handling

1) **Data Source Level:**

How should data be accessed in the use case of Connected Car: Real-time/streamed (Kafka) or in batch mode (Sqoop)

2) **Data Storage Level:**

In what form and to what extent are the data available in the use case of Connected Car: Filesystem (HDPS), Relational (Apache Kudu), NoSQL (HBase)

3) **Processing Level**

Difficult choice of tools, methods and algorithms in the use case of Connected Car: Batch (Spark, MapReduce), SQL (Impala), Search (Solr), SDK (Partners)

4) **Data Output Level**

In the use case of making results visible is done with Cloudera + Arcadia Data.

Part 3 (Git):

https://github.com/sedais/Data-Science-Infrastructure