**Die Big Data Herausforderungen**

Die 5 V’s fassen die Big Data Herausforderungen zusammen.

1. Volume
2. Variety
3. Velocity
4. Veracity
5. Value

Im konkreten Beispiel (Video: Türöffner für die Connected Car Revolution) versuche ich nun meine Erkenntnisse zu allen Herausforderungen zusammenzufassen.

|  |  |
| --- | --- |
| Herausforderung | Erkenntnis |
| Volume | Das Volumen ist eines der größten Herausforderungen, da es sich um sehr große Datenmengen handelt.   * Über 200 Sensoren generieren 2 PB an Daten / Auto / Jahr * Davon „nur“ 25 GB signifikante Daten / Auto / Stunde, wenn das Auto gefahren wird * Nachdem das Auto nicht rund um die Uhr gefahren wird, kommt man auf 130 TB an signifikante Daten / Auto / Jahr |
| Variety | Die Vielfalt der Daten bereit auch immer eine große Herausforderung, da die Daten strukturiert, semi-strukturiert oder sogar unstrukturiert sein können.   * Streaming (Real-Time) * Batch-weise Übertragung * Multiple Formats |
| Velocity | Sofortige Verarbeitung der generierten Daten nahezu Real-Time.   * Nicht alle generierten Daten werden im Auto gespeichert, sondern nur das Wesentliche * Die Autos sind über Mobilfunknetz mit dem Internet verbunden und verschicken die Daten über das Mobilfunknetz |
| Veracity | Die Richtigkeit der Daten ist auch eine große Herausforderung. Will man blind vertrauen oder kritisch hinterfragen bzw. bereinigen?  Folgende Unsicherheiten gibt es:   * Dateninkonsistenz * Unvollständigkeit der Daten * Unklarheit * Latenz * Täuschung * Schlechtes Modell   Deswegen muss:   * Datenbereinigung durchgeführt werden * Datenqualitätsüberprüfung * Ein gutes Analysemodell muss entwickelt werden * Zusammenführen und Analysieren mehrerer Datenquellen |
| Value | Das Analysemodell muss richtig gewählt sein, damit ein Mehrwert aus den Daten erkannt werden kann.  Dabei können sich unter anderem folgende Vorteile herauskristallisieren, wobei die aufgezählten Punkte sicherlich nur ein Bruchteil davon sind:   * Vorteile für den Besitzer wie z.B. Kostenreduktion, weil kein Parkplatz gesucht werden muss. * Durch Assistenzsysteme wird autonomes Fahren möglich * Unfälle und Probleme auf der Route in Echtzeit erkennen und wenn möglich eine andere Route planen * Vorteile   + Autohäuser   + Reifenhändler, da genauere Forecasts berechnet werden können   + Werkstätte bei Teilebestellung, da die Werkstätte im Vornhinein die Teile bestellen können   + Bessere Analysen für Versicherungen unter Anderem im versuchten Betrug |

**Die 4 Ebenen der Datenverarbeitung**

|  |  |
| --- | --- |
| Data Source  (Data Source Layer) | * Daten von Dingen (Auto, Kamera, Sensoren, IOT…) * Das Wesentliche der Daten erkennen und verarbeiten * Zugriff auf unterschiedliche Datenquellen * Übertragung der Daten über z.B. Mobilfunknetze |
| Data Messaging and Store Layer  (Data Storage Layer) | * Geeignete Art der Speicherung der Daten * Qualitätskriterien * Datenbereinigung * ETL / ELT |
| Analysis Layer  (Processing Layer) | * Einsetzen von geeignetem Werkzeug für die Verarbeitung * Zeitreihenanalyse * Nahezu Real-Time Analysen * Erkennen von Trends * Vorhersagen |
| Consumption Layer  (Data Output Layer) | * Erstellen von Dashboards * Geeignete Art der Visualisierung (Maps, Diagramme, Tabelle…) * Bestimmen der Zielgruppe |