

# Konzept einer Big Data Infrastruktur für zukünftige Leitstellensysteme

*Wednesday, October 11, 2017 4:00 PM (30 minutes)*

Im Rahmen der Projekte Energie Lab 2.0 und Energiesystem 2050 wird eine Forschungsinfrastruktur aufgebaut, die eine ganzheitliche Forschung zu zukünftigen intelligenten Energiesystemlösungen auf Systemebene ermöglicht. Verschiedene Partner aus der Helmholtzgemeinschaft und der Industrie sind dabei involviert. Für die Erforschung der für solche Systeme benötigten smarten Softwaredienstleistungsinfrastrukturen wird eine IT-Plattform auf Basis von „Big Data“-Werkzeugen und modernem Cluster-Computing aufgebaut und durch neuartige Energie bezogene Softwarelösungen ergänzt, welche als Grundlage für die IT-Infrastruktur von zukünftigen Leitstellensystemen bzw. Energiedienstleistungsinfrastrukturen dienen kann.

Diese Infrastruktur besteht aus fünf Ebenen. Unterste Ebene und damit die Basis der Infrastruktur bildet ein Computing Cluster auf Hardwareebene. Auf mehrere Knoten verteilte Linux-Betriebssysteme dienen hierbei als Softwaregrundlage. Auf einer zweiten Ebene werden die Linux-basierten Basisbetriebssysteme auf den einzelnen Rechnerknoten durch betriebssystemnahe Werkzeuge zum Management größerer Cluster von darauf verteilt laufender Softwareanwendungen ergänzt (z.B. Mesos zur Ausführung der Jobs verteilter Anwendungen in Containern auf unterschiedlichen Knoten des Clusters, oder Softwaresysteme zur Containervirtualisierung von Anwendungen, wie z.B. Docker und Kubernetes). Diese Ebene wird z.B. durch Produkte, wie DCOS (Distributed Computing Operating System) der Fa. Mesosphere bereitgestellt oder ist oft auch in „Big Data“-Technologiestacks, wie MapR, Cloudera oder Hortonworks, als betriebssystemnahe Dienstleistungsebene enthalten.

Auf dieser Grundinfrastruktur baut die nächste Ebene auf: ein „Big Data“-Technologiestack, wie er über Hadoop-basierte „Big Data“-Plattformen, wie MapR, Cloudera oder DCOS, bereitgestellt wird. Diese Ebene beinhaltet bereits vorinstallierte „Big Data“-Softwaresysteme, wie z.B. das Hadoop File System (HDFS), Apache Hbase, ZooKeeper oder Yarn als Grundlage.

Diese dritte Ebene wird durch „Big Data“-Softwarewerkzeuge zur Datenspeicherung, Datenanalyse, Kommunikation, Datenaufnahme ergänzt. Für die Datenspeicherung können verschiedene Big Data / NoSQL Datenbanken wie z.B. Neo4j, MongoDB, OpenTSDB, InfluxDB oder Elasticsearch verwendet werden. Für die Datenanalyse können u.a. Apache Spark und Apache Storm genutzt werden.

Die im Rahmen von Forschungsprojekten am IAI entwickelte Serviceebene mit grundlegenden Komponenten für zukünftige großskalige Softwarelösungen für das Energiesystem baut als vierte Ebene auf diesem Softwarestack auf und trägt den Namen Generic Microservice Backend (GMB). Sie beinhaltet ein verteiltes Servicegateway als Single Point of Entry, ein zentrales Sicherheitskonzept (OAuth2), sowie Mechanismen für Load Balancing und Service Discovery und ist als typische Microservicearchitektur aufgebaut.

Wichtige Bestandteile des GMB sind die Generic Data Services (GDS), welche die Funktionalitäten zum Datenmanagement kapseln. Die GDS unterstützen dabei einen polyglotten Ansatz. Unterschiedliche Datenbanken für die spezifischen Aufgaben werden über eine einheitliche Schnittstelle angesprochen. GDS bietet dazu Dienste für Zeitreihen, Stammdaten, Schemata und Digital Assets.

Einen weiteren wichtigen Bestandteil der Serviceebene bilden die Application Services. Dazu gehören Analyse-dienste, Optimierungsdienste und eine (Co-)Simulationsplattform. Sowohl die GDS als auch die Application Services greifen auf die darunter liegenden „Big Data“-Tools und -Datenbanken zu, um hochperformante und hochskalierbare Dienste zu ermöglichen.

Die fünfte und letzte Ebene bildet die Applikationsebene. Im Rahmen der oben genannten Projekte handelt es sich hierbei um diverse Energieapplikationen, über die Funktionalitäten wie Energie Management, Visualisierung von Energiedaten, Vorausplanung des Betriebs unter Nutzung von Datenvorhersage und Optimierung, etc. für den Endnutzer bereitgestellt werden. Die Applikationen greifen über eine einheitliche Serviceschnittstelle im Gateway auf die Cluster-Infrastruktur zu und bieten dem Endbenutzer nutzerfreundliche und hochkonfigurierbare Frontend-Schnittstellen, um auf das Big Data Cluster zuzugreifen ohne die darunterliegende Komplexität kennen zu müssen.

Im Rahmen des Vortrags werden die grundlegenden Konzepte der Gesamtlösung detailliert vorgestellt und die Nutzbarkeit für unterschiedliche Energieprojekte demonstriert.

## **Track**

BDAHM

**Author:** Mr STUCKY, Karl-Uwe (KIT)

**Co-authors:** Dr DÜPMEIER, Clemens (KIT); Dr STUCKY, Karl-Uwe (KIT); Dr SÜSS, Wolfgang (KIT)

**Presenter:** Mr STUCKY, Karl-Uwe (KIT)

**Session Classification:** Platforms