



MAN Anfrage

## **Ausschreibung MAN Big Data Rahmenvertrag Data Science, Data Engineering etc.**

## 0. Inhaltsverzeichnis

0.	Inhaltsverzeichnis .....	2
1.	Einleitung.....	3
2.	Notwendigkeit der Beauftragung .....	4
3.	Standortwahl.....	4
4.	Ausschreibungsgegenstand .....	5
4.1.	Ablauf der analytischen Vorhaben .....	5
4.2.	Inhalt der analytischen Vorhaben .....	5
4.3.	Abgrenzung.....	7
4.4.	Menge/Umfang .....	7
4.5.	Laufzeit .....	8
4.6.	Mitwirkungspflichten Auftraggeber.....	8
4.7.	Material .....	8
4.8.	Umwelt/ Energie/ Verbrauchswerte .....	8
4.9.	Verfügbarkeit und Abruf .....	8
4.10.	Tätigkeitsprofile .....	9
4.11.	Beschreibung Profile .....	9
4.12.	Allgemeine Anforderungen .....	16
5.	Dokumentation .....	16
6.	Logistik .....	17
7.	Garantie / Gewährleistung .....	17
8.	Umwelt .....	17
9.	Technische Vorgaben.....	18
9.1.	Big Data Plattform-Umgebung .....	18
9.1.	Entwicklungsumgebung .....	22
9.2.	Lizenzbedingungen.....	23
10.	Zusätzliche Anforderungen .....	24
11.	Details zur Angebotserstellung .....	24
11.1.	Angebotsanforderungen .....	24
11.2.	Einstellen des Angebots .....	25
11.3.	Aufbau des Angebots.....	26
11.4.	Zusätzliche Vereinbarungen .....	27
12.	Ansprechpartner zur Ausschreibung .....	28
12.1.	Fragen zum Angebot .....	28
13.	Abkürzungsverzeichnis .....	28
14.	Tabellenverzeichnis.....	30



## 1. Einleitung

Im Frühjahr 2015 wurde in der IT der MAN Truck & Bus AG (MAN) ein Competence Center für Big Data gegründet. In diesem wurde eine Big Data Strategie entwickelt und so die Entscheidung vorbereitet, dass die IT der MAN eine eigene Big Data Plattform (Big Data Plattform) sowie zugehörige Prozesse und Knowhow erarbeiten und anschließend betreiben soll.

2016 erfolgte daraufhin der Aufbau und ab Mitte 2017 die Stabilisierung der internen Plattform. Parallel wurden seit Mitte 2016 die ersten MAN-internen Data Scientists und ab Mitte 2017 Data Engineers eingestellt. So steht der MAN derzeit ein Team aus einem Plattformarchitekten, zwei Data Engineers und sechs Data Scientists zur Verfügung. Im Juni 2018 kamen hierzu noch zehn weiteren Experten aus dem Umfeld von Business Intelligence.

Zeitgleich mit der Einstellung der internen Mitarbeiter und der Entwicklung der Plattform wurden die Arbeiten an ersten analytischen Vorhaben<sup>1</sup> begonnen. Dabei sind das Portfolio gesamtheitlich sowie die Größe der einzelnen Vorhaben stetig gewachsen, so dass die Vorhaben allein mit internen Mitarbeitern nicht mehr zu leisten sind. Zudem sind mitunter spezielle Experten erforderlich, die bei MAN nicht verfügbar sind.

Im Zuge dieser Ausschreibung sollen mittels eines Rahmenvertrags ein oder mehrere externe Dienstleister verpflichtet werden, auf Abruf die Umsetzung analytischer Vorhaben zu unterstützen.

---

<sup>1</sup> Analytische Vorhaben meint im Rahmen dieser Ausschreibung Data Science Use Cases oder sonstige Projekte mit analytischer Relevanz mit unterschiedlichen Reifegraden (z.B. Proof-of-concept, Pilot, Prototyp oder produktives Datenprodukt). Die Inhalte dieser können stark variieren und nicht pauschalisiert werden.



## 2. Notwendigkeit der Beauftragung

Ziel dieser Beauftragung ist es, das fachliche Wissen und die Erfahrungen der externen Dienstleister zu nutzen. Dabei ist die schnelle Reaktionsfähigkeit des Dienstleisters ein wichtiges Entscheidungsmerkmal. Somit soll die Möglichkeit entstehen, in analytischen Vorhaben spezielles Wissen zügig von externen Dienstleistern abrufen zu können.

Um interne Mitarbeiter jedoch für die Fortführung der Vorhaben zu befähigen ist angestrebt, einen Wissenstransfer von den externen Kollegen auf die MAN-internen Mitarbeiter zu ermöglichen. Dazu sind die Dokumentations-Vorgaben der MAN-internen Mitarbeiter zu befolgen.

## 3. Standortwahl

Die **Leistung soll überwiegen aber nicht zwingend ausschließlich onsite** erbracht werden. Die konkrete Wahl ist dabei auch von der Projektmethodik (agil/klassisch) des jeweiligen analytischen Vorhabens abhängig.

Reisekosten zum Standort München<sup>2</sup> werden als Inklusiv-Leistung erwartet. Reisekosten die für Reisen zu anderen Standorten der MAN entstehen können nach Aufwand abgerechnet werden, bedürfen jedoch der vorherigen Rücksprache mit dem Auftraggeber.

Als Sprache in der **verbalen Kommunikation werden Englisch und Deutsch** erwartet. Die schriftliche Dokumentation kann wahlweise **deutsch- oder englischsprachig** in **Abstimmung mit dem Auftraggeber** erfolgen.

---

<sup>2</sup> MAN Truck & Bus AG, Dachauer Straße 667 und 651, 80995 München,  
MAN Truck & Bus AG, Standorte im Gewerbegebiet Karlsfeld, 85757 Karlsfeld  
MAN Truck & Bus AG, Oskar-Schlemmer-Straße 19, 80807 München

## **4. Ausschreibungsgegenstand**

Innerhalb der MTB werden derzeit zahlreiche potentielle analytische Vorhaben und damit einhergehend relevante Datenquellen identifiziert. Diese in Zusammenarbeit mit den MAN-internen Experten aus den jeweiligen Fachbereichen, der IT und den Data Scientists und Data Engineers umzusetzen ist dabei Gegenstand der vorliegenden Ausschreibung.

### **4.1. Ablauf der analytischen Vorhaben**

Analytische Vorhaben durchlaufen bei MAN verschiedene Reifegrade vom Proof-of-concept über die Pilot- und Prototypenphase bis letztlich hin zum produktiven Datenprodukt. Diese Phasen können je nach analytischem Vorhaben in Gänze oder einzeln umgesetzt werden.

Der Arbeitsmodus der analytischen Vorhaben entspricht in der Regel agiler Projektmethodik, kann jedoch auch im Ausnahmefall in klassischen Projektstrukturen gem. IT PEP der MTB erfolgen. Konkrete Festlegungen erfolgen nach Vorgabe von MAN für jedes analytische Vorhaben individuell.

Grundsätzlich haben in analytischen Vorhaben die Data Scientists und Data Engineers der MAN die fachliche Führung. Ihre Vorgaben sind einzuhalten, allerdings können und sollen selbstredend eigene alternative Vorschläge mit eingebracht werden. Das MAN-interne Mitarbeiter den externen Kollegen unterstellt sind ist allerdings nicht anzunehmen.

Die Analysen der Daten dürfen nur auf/mit von der MAN zur Verfügung gestellten Hard-/Software gespeichert und verarbeitet werden. Vom Dienstleister mitgebrachte Entwicklungsumgebungen sind somit nicht zulässig.

### **4.2. Inhalt der analytischen Vorhaben**

Hauptbestandteil der analytischen Vorhaben sind zunächst deskriptive Datenanalysen für unterschiedliche Fachbereiche der MAN. Darüber hinaus kommen dann auch komplexere Algorithmen aus dem Bereich Machine Learning zum Einsatz.

Bei den zugrundeliegenden Daten handelt es sich hauptsächlich um relationale Tabellen<sup>3</sup>. Dies können beispielsweise Telematikdaten, Werkstatt-informationen oder Vertriebsdaten aus unterschiedlichen Quellsystemen (DB2 HOST, SAP, ORACLE DB, CSV/XLS, etc.) sein. Bei Bedarf müssen aber auch semi- und unstrukturierte Informationen bearbeitet werden können.

---

<sup>3</sup> Daten werden derzeit im Format AVRO in Hadoop gespeichert und sind teilweise als Tabellen in Hive und Impala verfügbar

Diese Daten sind in den meisten Fällen noch zu erschließen, aufzubereiten und zu persistieren. Die Entwicklung permanenter Schnittstellen in Realtime oder im Batch kann somit auch Bestandteil eines analytischen Vorhabens sein. Dies inkludiert nicht nur die Programmierung der Schnittstellen oder die Konfiguration von Pipelines sondern ebenso die Dokumentation der Schnittstellen sowie die technische Abstimmung mit allen betroffenen Bereichen (z.B. IT Ansprechpartner des Quellsystems, Verantwortliche für Backend, IT Security, etc.).

Nachdem es Ziel der analytischen Vorhaben ist, diese im Arbeitsalltag der Fachbereiche produktiv einzusetzen, ist die Produktivmachung ein weiteres wichtiges Themenfeld, welches in den Vorhaben umzusetzen ist. Dabei wird besonderes Augenmerk daraufgelegt, dass...

- die Entwicklung des Quellcodes abgeschlossen ist
- eine Bewertung durch IT-Security, Datenschutz, Legal und Betriebsrat erfolgt ist
- erforderliche Daten aus Schnittstellen kommen
- relevante Schnittstellen dokumentiert sind
- das Vorhaben in Betriebs- und Anwenderhandbuch beschrieben ist
- eine Architekturbeschreibung und –bewertung erfolgt ist
- eine CI-Strecke implementiert ist
- eine ausreichende Test-Coverage wie auch Code Qualität sichergestellt ist
- eine Betrachtung der Performance-Bedarfe erfolgt ist
- Abhängigkeiten zu anderen Systemen / Packages und Prerequisites erfasst sind
- der Betrieb bei den Betriebspartnern angemeldet und übergeben wurde
- Monitoring und Logging implementiert wurden
- Code von Schnittstelle und Applikation auf der Produktiven Umgebung laufen

Zum Zwecke der Nachhaltigkeit, Pflege der produktiven Datenprodukte und Betreibbarkeit wird großen Wert auf eine nützliche Dokumentation nach Vorgaben von MAN und die Einhaltung definierter Qualitätsstandards gelegt.

Auch ist es erforderlich, dass die Entwicklung im Rahmen analytischer Vorhaben auf der von MAN genutzten und vorgegeben Entwicklungsumgebung erfolgt, damit die Ergebnisse später wart- und veränderbar sind. Sollten Erweiterungen dieser Entwicklungsumgebung erforderlich sein, sind diese mit MAN abzustimmen. Die Umsetzung dieser Änderung kann in der Folge ein eigenes analytisches Vorhaben werden.

### 4.3. Abgrenzung

**Explizit nicht im Umfang dieser Ausschreibung sind nachfolgend aufgelistete Themen,** die dennoch in engem Verhältnis zum Gegenstand der Ausschreibung stehen:

- Betrieb der Big Data Plattform (Hardware und Betriebssystem)
- Betrieb von Datenbanken die für Hadoop erforderlich sind
- Betrieb weiterer Software der Big Data Plattform
- Betrieb externer Daten-Schnittstellen zu Quellsystemen
- Betrieb des Datenproduktes aus den analytischen Vorhaben

All diese Komponenten werden durch separate Provider betrieben. Der Auftraggeber geht jedoch davon aus, dass mit diesen Providern eng und im Sinne des Auftraggebers zusammengearbeitet wird.

### 4.4. Menge/Umfang

Die analytischen Vorhaben unterscheiden sich sehr stark in Ihrer Komplexität. Daher kann keine generische Aussage zu den einzelnen Abrufen getätigt werden. Indikativ kann jedoch von folgenden maximalen Mengen pro Jahr ausgegangen werden:

Nr.	Tätigkeitsprofile	Junior	Senior
1	Data Scientist	25	50
2	Data Engineer	50	200
3	Softwareentwickler	50	200
4	Visualization	25	25
5	Architekt	25	25

**Tabelle 1: Tätigkeitsprofile und jährliche Mengengerüste**

Einmalig wird eine Einarbeitung in die gegebene Big Data Plattform angeboten, für die je Tätigkeitsprofil ein Aufwand von 5 Tagen<sup>4</sup> verrechnet werden kann (in Summe also 30 PT). Weitere Einarbeitungsaufwände gehen zu Lasten des externen Dienstleisters.

<sup>4</sup> Mittelwert zwischen Junior- und Senior-Stufe

#### **4.5. Laufzeit**

Beginn der Leistungserbringung ist **ab 01.01.2019**. Die Leistungserbringung soll **bis 31.12.2021 erfolgen**. Die Laufzeit beträgt damit zunächst **36 Monate**.

#### **4.6. Mitwirkungspflichten Auftraggeber**

Der Auftraggeber verpflichtet sich im Rahmen der genannten Dienstleistungen die nachfolgend aufgeführten Zuleistungen zu erbringen:

- Projektleitung und regelmäßige Statustermine
- Bereitstellung eines angemessenen Projektraums und Zugang zu diesem (sofern relevant)
- Verfügbarkeit eines MAN-interner Mitarbeiters als direkter Ansprechpartner
- Unterstützung bei der Vermittlung von Ansprechpartnern für Anliegen des Auftragnehmers
- Treffen von für die Arbeit erforderlichen Entscheidungen und Bereitstellung von Standards
- Unterstützung bei der Einrichtung von Remote-Zugriff (sofern relevant)
- Bereitstellung von MAN Laptops sowie erforderlichen User Accounts und Zugriffsrechten
- Rechtzeitige Bereitstellung von allen zu untersuchenden Daten

#### **4.7. Material**

MAN unterstützt die Arbeit durch Bereitstellung MAN-spezifischer Materialien (z.B. Laptops), die nicht anders oder nur mit unverhältnismäßig großen Aufwänden beschafft werden können (z.B. Testfahrzeuge) und die für die Arbeit erforderlich sind.

Besondere Materialien, die für die Aufgabe erforderlich aber bei der MAN nicht verfügbar sind müssen vom Auftragnehmer auf eigene Kosten bereitgestellt werden.

#### **4.8. Umwelt/ Energie/ Verbrauchswerte**

Die gesetzlichen Anforderungen sind einzuhalten.

#### **4.9. Verfügbarkeit und Abruf**

Der Auftragnehmer hat sicherzustellen, dass die Bearbeitung der übertragenen Arbeitspakete durch ein konstantes Team erfolgt und somit Verzögerungen aufgrund von Personalschwankungen nicht stattfinden.

Abwesenheiten sowie weitere Ausfallzeiten sind vom Auftragnehmer frühzeitig mit einem Vorlauf von min. sechs Wochen zu benennen und zu kompensieren. Im Falle von



Personalwechseln sind mit ausreichend Vorlauf Übergaben zu gewährleisten, so dass die Erreichung der spezifizierten User stories bzw. Arbeitspakete sichergestellt ist. Die hierdurch entstehenden Mehrkosten sind durch den Dienstleister zu tragen.

Ferner wird darauf Wert gelegt, dass die **Ressourcen des Dienstleisters kurzfristig verfügbar sind**, da die Umsetzung der analytischen Vorhaben zügig nach Anfrage erfolgen soll. Die Dauer zwischen Anfrage und Arbeitsbeginn sollte dabei sechs Wochen nicht überschreiten.

Die Abrufe erfolgen spezifisch für jedes analytische Vorhaben. Dabei wird das analytische Vorhaben dem Dienstleister zusammen mit einem Personalbedarf vorgestellt. Der Dienstleister ist dann aufgefordert anonymisierte Personalvorschläge zu machen. MTB behält sich vor, im Rahmen von Telefonkonferenzen die fachliche Expertise zu validieren.

Je nach analytischem Vorhaben werden unterschiedliche Tätigkeitsprofile oder Kombinationen darauf in verschiedenen Mengen. Diese sind in der Regel keine kompletten Teams, da sie die MAN-internen Kollegen ergänzen und unterstützen sollen.

Für die Umsetzung eines analytischen Vorhabens werden dediziert verfügbare Mitarbeiter vorausgesetzt, es müssen jedoch nicht zwingend alle Vorhaben durch die gleichen Personen umgesetzt werden.

#### **4.10. Tätigkeitsprofile**

Die Tätigkeit lässt sich dabei in 6 verschiedene Tätigkeitsprofile zusammenfassen welche auch Grundlage des angefragten Angebots sein soll (s. Tabelle 1)

Die Stufe „Junior“ beschreibt dabei einen Mitarbeiter mit einer Berufserfahrung bis zu drei Jahren. Bei einer Berufserfahrung von mehr als drei Jahren kann der Mitarbeiter als Senior eingestuft werden.

#### **4.11. Beschreibung Profile**

Folgende Aufgaben werden von den zuvor genannten Tätigkeitsprofilen im Rahmen der Beauftragung erwartet:

## Data Scientist

### Kenntnisse

- Sicherer Umgang mit den Programmiersprachen Apache Spark, Python, R, Scala, Java
- Verständnis statistischer Methoden und lernender Systeme
- Tiefe Kenntnisse von Data Mining und Machine Learning, Deep Learning, Clustering, Recommendation, Classification (Senior)
- Mehrjährige Erfahrung im Bereich Datenanalyse und multivariate Statistik (Senior)
- Erfahrung in der Datenaufbereitung für spezifische Anwendungsfälle
- Grundlegende Expertise im Bereich Software Engineering für die Serienreifmachung produktiver Datenprodukte (inkl. Testen, Deployment, Code Qualität Logging, Monitoring, Betrieb)
- Kenntnisse des Coudera Hadoop Stack
- Detaillierte Kenntnisse agiler Softwareentwicklung nach Scrum
- Fachliche Automotive Branchen Kenntnisse

### Aufgaben

- Laden von Daten auf die Big Data Plattform
- Zielgerichtete Aufbereitung von Daten für spezifische Use Cases und Erstellen von Daten-Profilen
- Explorative Analyse der Daten
- Zielgerichtete Datenanalysen hinsichtlich definierter Fragestellungen
- Projektmanagement und Beratung der Fachbereiche hinsichtlich Umsetzbarkeit und Interpretation spezieller Fragestellungen
- Serienreifmachung von Use Case für die stabile produktive Nutzung
- Koordination der Aktivitäten des Dienstleisters (nur Senior)

### Relevante Tools

- Apache Spark (vertiefte Kenntnisse)
- R Studio (vertiefte Kenntnisse)
- Hue, Hive, Impala (vertiefte Kenntnisse)
- Jupyter Notebooks (vertiefte Kenntnisse)
- Gitlab (vertiefte Kenntnisse)
- Tableau, Shiny, Bokeh (Basiswissen)
- AirFlow (Basiswissen)
- Docker (Basiswissen)
- Redhat Linux (Basiswissen)

Tabelle 2: Beschreibung Tätigkeitsprofil Data Scientist



## Data Engineer

### Kenntnisse

- Sicherer Umgang mit den Programmiersprachen Apache Spark, Python, R, Scala, Java und Kotlin
- Verständnis des Cloudera Hadoop Stacks und der AWS Cloud Ingestion Lösungen
- Verständnis für effiziente Datenverarbeitung und Haltung in HDFS / S3 und Datenbanken (On-Prem und Cloud)
- Kenntnis relevanter Quellsysteme und deren Datentypen und – Strukturen
- Kenntnis in der Umsetzung von Batch- und Realtime-Schnittstellen
- Erweiterte Expertise im Bereich Software Engineering für die Serienreifmachung produktiver Datenprodukte (inkl. Testen, Deployment, Code Qualität Logging, Monitoring, Betrieb)
- Erfahrung in der Entwicklung verteilt-arbeitender Systemen
- Erfahrung in Docker basierendem Deployment und bzgl. Resource Manager
- Erfahrung in CI/CD Pipelines und Tools
- Kenntnis und praktische Erfahrung mit Workflow Managers
- Detaillierte Kenntnisse agiler Softwareentwicklung nach Scrum
- Fachliche Automotive Branchen Kenntnisse

### Aufgaben

- Laden von Daten auf die Big Data Plattform
- Zielgerichtete Aufbereitung von Daten für spezifische Use Cases und Erstellen von Daten-Profilen
- Entwicklung von betriebsreifen Datenpipelines
- Koordination der Aktivitäten des Dienstleisters (nur Senior)

### Relevante Tools

- Redhat Linux (vertiefte Kenntnisse)
- Docker (vertiefte Kenntnisse)
- Mesos, Marathon (vertiefte Kenntnisse)
- Yarn (vertiefte Kenntnisse)
- Kubernetes (vertiefte Kenntnisse)
- Spark (vertiefte Kenntnisse)
- Hadoop, Hadoop Command Line Tools (vertiefte Kenntnisse)
- HDFS (vertiefte Kenntnisse)
- AWS S3 (vertiefte Kenntnisse)
- Hive (vertiefte Kenntnisse)



• Sqoop	(vertiefte Kenntnisse)
• Airflow	(vertiefte Kenntnisse)
• AWS Glue	(vertiefte Kenntnisse)
• AWS Data Pipelines	(vertiefte Kenntnisse)
• Kafka, Confluent Connectors	(vertiefte Kenntnisse)
• AWS Kinesis	(vertiefte Kenntnisse)
• ELK und TICK-Stack	(vertiefte Kenntnisse)
• Oozie	(Basiswissen)
• GitLab or Jenkins	(Basiswissen)
• Artifactory	(Basiswissen)
• IntelliJ / Eclipse	(Basiswissen)

**Tabelle 3: Beschreibung Tätigkeitsprofil Data Engineer**

## Softwareentwickler

Kenntnisse	<ul style="list-style-type: none"> <li>• Sicherer Umgang mit den Programmiersprachen Apache Spark, Python, R, Scala, Java und Kotlin (JVM)</li> <li>• Verständnis von analytischen Vorhaben und deren Umsetzung</li> <li>• Kenntnis von Programmier-Paradigmen und deren praktischen Umsetzung</li> <li>• Hohes Qualitätsbewusstsein</li> <li>• Kenntnisse in Code Dokumentation und Testing</li> <li>• Erfahrung in der Entwicklung leicht-betreibbarer Software (inkl. Monitoring, Logging)</li> <li>• Erfahrung in Deployment und Test (automatisiert) von Code</li> <li>• Detaillierte Kenntnisse agiler Softwareentwicklung nach Scrum</li> <li>• Fachliche Automotive Branchen Kenntnisse</li> </ul>
Aufgaben	<ul style="list-style-type: none"> <li>• Umsetzung von analytischem Code in produktive Datenprodukte</li> <li>• Umsetzung von betriebs- und leistungsoptimierter Software</li> <li>• Übergabe von produktiven Datenprodukten an den Betrieb</li> <li>• Dokumentation von Code und Fehler-Handling</li> <li>• Test des entwickelten Codes</li> <li>• Koordination der Aktivitäten des Dienstleisters (nur Senior)</li> </ul>
Relevante Tools	<ul style="list-style-type: none"> <li>• Redhat Linux (vertiefte Kenntnisse)</li> <li>• Docker (vertiefte Kenntnisse)</li> <li>• ELK und TICK-Stack (vertiefte Kenntnisse)</li> <li>• Mesos, Marathon (Basiswissen)</li> <li>• Yarn (Basiswissen)</li> <li>• Kubernetes (Basiswissen)</li> <li>• Spark (Basiswissen)</li> <li>• HDFS (Basiswissen)</li> <li>• AWS S3 (Basiswissen)</li> <li>• Airflow (Basiswissen)</li> <li>• GitLab or Jenkins (Basiswissen)</li> <li>• Artifactory (Basiswissen)</li> <li>• IntelliJ / Eclipse (Basiswissen)</li> </ul>

Tabelle 4: Beschreibung Tätigkeitsprofil Softwareentwickler

## Visualization

Kenntnisse	<ul style="list-style-type: none"> <li>• Erfahren im UX-Design zur Anpassung der Darstellung an die Bedarfe des Auftraggebers</li> <li>• Fähigkeit durch verschiedene Arten der Visualisierung Verständnis und Begeisterung bei Angehörigen verschiedenster Fachbereiche mit unterschiedlichen Hintergrundwissen zu wecken</li> <li>• Erfahrung im Bereich interaktiver Datenvisualisierung (z. B. mit Java Script, D3...) für Webseiten</li> <li>• Detaillierte Kenntnisse agiler Softwareentwicklung nach Scrum</li> <li>• Fachliche Automotive Branchen Kenntnisse</li> </ul>
Aufgaben	<ul style="list-style-type: none"> <li>• Abstimmung von Storyboards mit den Auftraggebern</li> <li>• Darstellung komplexer Analyseergebnisse in leichtverständliche und interaktive Grafiken</li> <li>• Erstellung aussagekräftige Datenvisualisierungen</li> <li>• Aufbereitung erzielter Ergebnisse in Management-tauglicher CI-konformer Powerpoint- und PDF Präsentationen und Dokumentationen</li> </ul>
Relevante Tools	<ul style="list-style-type: none"> <li>• Shiny (vertiefte Kenntnisse)</li> <li>• Bokeh (vertiefte Kenntnisse)</li> <li>• Tableau (vertiefte Kenntnisse)</li> <li>• Java Script (vertiefte Kenntnisse)</li> <li>• SAP Lumira (Basiswissen)</li> <li>• SAP BI (Basiswissen)</li> </ul>

Tabelle 5: Beschreibung Tätigkeitsprofil Visualization

## Architekt

Kenntnisse	<ul style="list-style-type: none"> <li>• Erfahrung in der Entwicklung/Umsetzung von Big Data-/BI-Architekturen</li> <li>• Kenntnisse der Cloudera Hadoop und AWS-spezifischen Applikationen und Persistenzlösungen</li> <li>• Entwicklung von On-premise und Cloud-Infrastrukturen (AWS)</li> <li>• Kompetenz in Datenbanken und der Entwicklung verteilter Systeme</li> <li>• Bewusstsein für Anforderungen der IT-Sicherheit</li> <li>• Kenntnisse in Projektmanagement und Aufbau und Management von Betrieb (Applikation und Plattform)</li> <li>• Detaillierte Kenntnisse agiler Softwareentwicklung nach Scrum</li> <li>• Fachliche Automotive Branchen Kenntnisse</li> </ul>
Aufgaben	<ul style="list-style-type: none"> <li>• Entwicklung von Architekturen analytischer Vorhaben</li> <li>• Abstimmung der Architekturen von Applikation und Plattform</li> <li>• Ableiten von Anpassungsbedarfen der Plattform-Architektur und der deren Umsetzung</li> <li>• Planen von Projektvorhaben und deren Koordination</li> </ul>
Relevante Tools	<ul style="list-style-type: none"> <li>• LDAP, Kerberos (vertiefte Kenntnisse)</li> <li>• IT Security protocols like TLS, SSL, etc. (vertiefte Kenntnisse)</li> <li>• Redhat Linux (vertiefte Kenntnisse)</li> <li>• AWS, Azure (vertiefte Kenntnisse)</li> <li>• Mesos, Marathon (vertiefte Kenntnisse)</li> <li>• Oracle, SQL (vertiefte Kenntnisse)</li> <li>• Cloudera Hadoop Stack (vertiefte Kenntnisse) (incl. Yarn, HDFS, Impala, ...)</li> <li>• AWS S3 (vertiefte Kenntnisse)</li> <li>• Dynamo DB (vertiefte Kenntnisse)</li> <li>• AWS Glue (vertiefte Kenntnisse)</li> <li>• AWS EMR (vertiefte Kenntnisse)</li> <li>• AWS Data Pipeline (vertiefte Kenntnisse)</li> <li>• Microsoft Visio, Powerpoint (Basiswissen)</li> <li>• Docker, VMWare (Basiswissen)</li> <li>• Kafka (Basiswissen)</li> <li>• Sqoop (Basiswissen)</li> </ul>

Tabelle 6: Beschreibung Tätigkeitsprofil Architekt

Diese Übersicht der zuvor genannten Tätigkeitsprofile ist dabei indikativ zu verstehen und erhebt nicht den Anspruch auf Vollständigkeit.

#### **4.12. Allgemeine Anforderungen**

Der Auftraggeber legt besonderen Wert darauf, dass die eingesetzten Experten sich mit den folgenden Grundwerten identifizieren können bzw. folgende Erfahrungen mitbringen:

- Erfahrung in agilen Projekten und Arbeiten nach Scrum (Zertifizierung und praktische Erfahrung vorhanden)
- Ordentliche Dokumentation von Prozessen und Plattform
- Erfahrungen in der Steuerung und Leitung von Projekten
- Erfahrungen im Umgang mit Risiko/Unsicherheiten
- Einbringen eigener Ideen/Lösungsalternativen

Es wird somit erwartet, dass der Auftragnehmer über Lösungserfahrung für die Umsetzung von analytischen Vorhaben verfügt sowie die Bereitschaft sich VW- und MAN-Standards anzupassen (z.B. MAN-Projektstandards) und sich in etwaige bestehende Vorarbeiten einzudenken.

### **5. Dokumentation**

Hinsichtlich Dokumentation sind folgende Dokumente im Laufe der Tätigkeit anzufertigen und aktuell zu halten:

- Quellcode und dessen Beschreibung
- Beschreibung der Umsetzung des Vorhabens
- Beschreibung und Begründung eingesetzter Algorithmen
- Schematische Architekturdarstellungen
- Betriebsdokumentation inkl. Arbeitsanweisungen für Fehler-Handling
- Übersicht eigener Mitarbeiter die im Projekt eingesetzt sind (inkl. Einsatzdauer falls zeitlich befristet)
- Leistungsnachweise
- Sowie Vorgaben gem. Kriterien der User Stories





Sämtliche Dokumente sind in digitaler Form im nativen Format der jeweiligen Autorensysteme<sup>5</sup> und als PDF-Dokument dem Auftraggeber regelmäßig und bei Beauftragungsende zur Verfügung zu stellen.

In diesem Zusammenhang sei auf die IT-Sicherheitsrichtlinien der MTB hingewiesen, welche beim Einkauf erfragt werden können.

## **6. Logistik**

Nicht relevant

## **7. Garantie / Gewährleistung**

Die gesetzlichen Anforderungen sind einzuhalten.

## **8. Umwelt**

Die gesetzlichen Anforderungen sind einzuhalten.

---

<sup>5</sup> Es sind nur Autorensysteme zulässig, die bei MTB im Einsatz sind.

## 9. Technische Vorgaben

Die Architektur der Big Data Plattform liegt eine technische Beschreibung vor. Diese **finden Sie als Anlage** zu dieser Ausschreibung. Ebenso einen Auszug der Qualitätskriterien und der der Anwendungs-Guidelines der Big Data Plattform. Die detaillierte Beschreibung wird aus Gründen der Geheimhaltung nur nach Vertragsvergabe zur Verfügung gestellt.

Eine Zusammenfassung finden Sie nachfolgend. Die hier beschriebenen Entwicklungs-umgebung und –vorgaben werden als verbindlich vorausgesetzt.

### 9.1. Big Data Plattform-Umgebung

Die Big Data Plattform besteht im Wesentlichen aus folgenden zwei Komponenten:

- Storage und Ingestion (Data Lake und Data Connect)  
(Cloudera Hadoop, Confluent Kafka und Sqoop)
- Sandbox  
(Entwicklungsumgebung für analytische Vorhaben – s. a. Kap. 0)

Im Detail ist aktuell folgende Cloudera-nahe Software im Einsatz die als Grundlage der Tätigkeit dienen soll:

Nr	Software	Utility Node	Master Node	Worker Node
1	NameNode		X	
2	YARN ResourceManager + YARN JobHistory-Services		X	
3	ZooKeeper Node		X	
4	JournalNode		X	
5	FailoverController		X	
6	StateStore-Server		X	
7	Cloudera Manager	X		
8	Cloudera Management Services	X	X	X
9	Hive Metastore Server	X		
10	HiveServer2			X
11	Catalog-Server			X
12	Spark		X	X
13	Hue	X		

Nr	Software	Utility Node	Master Node	Worker Node
14	Oozie	X		X
15	Flume-Agent			X
16	Relationale Datenbank	X		
17	Gateway (Client-)Konfiguration	X		X
18	DataNode			X
19	YARN-NodeManager-Services			X
20	Impala		X	X
21	Sqoop	X		
22	Sentry		X	

Tabelle 7: Software Cloudera Hadoop

Deployment und Scheduling unserer Jobs läuft über eine dockerisierte Umgebung mit einem Mesos Demon. Dafür steht ein Mesos Cluster mit den folgenden Komponenten zur Verfügung:

Nr	Software	Utility Node	Master Node	Worker Node
1	Mesos Master		X	
2	Mesos Slave			X
3	ZooKeeper Node (=Cloudera Zookeeper)		X	
4	Marathon		X	
5	Metronome		X	
6	Consul	X		
7	Mesos-Consul		X	
8	Docker Registry	X		
9	Docker Demon	X		
10	Calico	X		
11	etcd	X		
12	CNTLM	X		
13	MANTL-Ui	X		

Tabelle 8: Software Deployment und Job Scheduling

Für alle persistierende Docker basierende Dienste (Sandbox 1.0, SFTP Servers, kleine DB Instanzen wie Postgres, Redis) wird ein **GlusterFS** Cluster benutzt.

Zusätzlich ist für Streaming Data ein Kafka Cluster (Open Source Confluent) in unserer Umgebung installiert mit Kafka Connect, laufend auf Mesos.

Nr	Software	Utility Node	Master Node	Worker Node
1	Apache Kafka		X	
2	Kafka Connect			X
3	Schema Registry	X		

Tabelle 9: Software Streaming

Für Data Pipelines und Batch Jobs, MAN nutzt eine Apache Airflow Instanz mit den folgenden Komponenten:

Nr	Software	Utility Node	Master Node	Worker Node
1	Scheduler		X	
2	Celery Worker			X
3	Flower (monitoring tool)		X	
4	Airflow WebUI (with LDAP authentication enabled)		X	
5	Redis (broker)	X		
6	Postgres (airflow states + result backend for Celery)	X		

Tabelle 10: Software Data Pipelines

Für CI Pipelines und Artifactory wird eine Gitlab Installation auf Mesos / Marathon mit den folgenden Komponenten genutzt:

Nr	Software	Utility Node	Master Node	Worker Node
1	Gitlab		X	
2	PostgreSQL		X	
3	Redis		X	

Tabelle 11: Software CI

Für Monitoring & Alerting ist ein TICK Stack installiert, mit dem alle physischen, OS-, Mesos-, und Docker-Ressourcen sowie Applikationen überwacht werden können.

Nr	Software	Utility Node	Master Node	Worker Node
1	Telegraf (agents)	X		
2	InfluxDB		X	
3	Chronograf		X	
4	Kapacitor		X	
5	HA Proxy		X	

Tabelle 12: Software Monitoring und Alerting

Für Logging im Speziellen, wurde ein Elastisearch Cluster aufgebaut:

Nr	Software	Utility Node	Master Node	Worker Node
1	Elasticsearch Master		X	
2	Elasticsearch Master Data Node		X	
3	Elasticsearch Data Node			X
4	Kibana		X	
5	Filebeat	X		

Tabelle 13: Software Logging

## 9.1. Entwicklungsumgebung

Es gibt 2 Versionen von Sandboxes die auf der Plattform kontinuierlich laufen:

Sandbox 1.0 – Docker basierend, AD Authentifizierte Maschinen die über GlusterFS persistiert sind. Die folgenden Clients laufen auf der Sandbox 1.0 vorkonfiguriert (und fallen unter Betrieb):

Nr	Software
1	Jupyter
2	RStudio
3	Shiny
4	R
5	Spark
6	Impala (Impala-Shell)
7	sssd
8	Java (JDBC, ODBC)
9	Cloudera Client Mounting

**Tabelle 14: Software Sandbox 1.0**

Sandbox 2.0 – große Edge Nodes (einem Server groß) wo man selber unterschiedlichen Diensten und Lösungen ausprobieren kann oder Ressource-konsumierende Daten-Initialisierungen durchführen kann.

Nr	Software
1	Jupyter
2	RStudio
3	Shiny
4	R
5	Spark
6	Impala (Impala-Shell)
7	sssd
8	Java (JDBC, ODBC)
9	Cloudera Client Mounting

**Tabelle 15: Software Sandbox 2.0**

Um die AD Directory Synchronisierung und DNS Forwarding / Routing zu ermöglichen gibt es ein paar angepasste/individualisierte Dienste (die dockerisiert laufen).

Diese Liste spiegelt den aktuellen Projektstand dar und ist damit nicht final.

Die Nachfolgenden beiden Abbildungen beschreiben die Entwicklungsumgebung, welche die Grundlage jeglicher Umsetzung von analytischen Vorhaben bei MTB darstellen.



Abbildung 1: Übersicht Big Data Plattform (Stand 04-2018)

## 9.2. Lizenzbedingungen

Die allgemeinen Einkaufsbedingungen für die IT ebenso wie die IT-Sicherheitsrichtlinien sind selbstredend auch für analytische Vorhaben verbindlich und auf [www.vwgroupsupply.de](http://www.vwgroupsupply.de) in Englisch und Deutsch veröffentlicht. Diese inkludieren u.a. auch die aktuell gültigen Lizenzbedingungen.

## 10. Zusätzliche Anforderungen

Zur Wahrung der gesetzlichen Vorgaben hinsichtlich Arbeitnehmer-überlassung benennen **Auftragnehmer und Auftraggeber benennen einen dedizierten Ansprechpartner** (sog. Single point of contact), welche als Eskalationsinstanz kontaktiert werden kann.

Darüber hinaus sei darauf hingewiesen, dass die **Beschäftigung von Unterauftragnehmern** dem Auftraggeber vorab angezeigt und von diesem genehmigt werden muss.

## 11. Details zur Angebotserstellung

Nachfolgend sind Anforderungen an Form und Inhalt des Angebots formuliert.

### 11.1. Angebotsanforderungen

Wir bitten Sie ausdrücklich darum, folgende Anforderungen bei der Erstellung Ihres Angebots zu berücksichtigen:

- Ihr Angebot muss vollständig sein. Unvollständige Angebote können nicht berücksichtigt werden.
- Ihr Angebot muss insbesondere **alle in den Ausschreibungsunterlagen beschriebenen Anforderungen erfüllen** und unseren Allgemeinen Geschäftsbedingungen entsprechen.
- Ihr Angebot ist nur gültig, wenn es in schriftlicher Form und in deutscher Sprache vorliegt.
- Das Angebot ist als PDF-Dokument abzugeben.
- Das abzugebende Angebot soll sich inhaltlich auf das Angebot der notwendigen Ressourcen konzentrieren.
- Die Kosten des Angebots trägt der Auftragnehmer bzw. Anbieter.
- Um die Aufwände bei Auftraggeber und Anbieter gering zu halten wird darum gebeten, die angegebenen Angebote kompakt zu halten, ein **Umfang von 30 Seiten** sollte hierfür ausreichend sein.





- **Zwingend erforderlich für eine positive fachliche Bewertung** sind überzeugende Darstellungen folgender Aspekte:
  - Erfahrung in der Anwendung der eingesetzten Applikationen
  - Programmiererfahrung für die genannten Programmiersprachen
  - Tiefe Kenntnisse in mathematischen Methoden
  - Lange Erfahrung im Anbinden und Aufbereiten von Daten
  - Verständnis für die Nutzung von Werkzeugen von Cloudera Hadoop und des AWS Eco-Systems
  - Verständliche und proaktive Kommunikationsfähigkeiten
  - Strukturierte und transparente Vorgehensweise

## **11.2. Einstellen des Angebots**

Nach Angebotsabgabe werden diese gesichtet und ggf. die priorisierten Anbieter zu einer Angebotspräsentation eingeladen. Es ist dabei darauf zu achten, dass die **Anbieter für die Angebotspräsentation die erforderlichen Mitarbeiter (insb. Den potentiellen SpoC sowie je einen exemplarischen Data Scientisten, Data Engineer und Softwareentwickler) vorhalten**. Diese Präsentation ist Teil des Angebotsprozesses und geht damit zu Lasten des Anbieters. Im Rahmen der Angebotspräsentation wird Ihnen möglicherweise eine ihnen fremde exemplarische Aufgabe gestellt, deren Bearbeitung in die Angebotsbewertung mit einfließt.

Bitte stellen Sie Ihr Angebot auf der B2B-Plattform [www.vwgroupsupply.com](http://www.vwgroupsupply.com) ein.

Sollten Sie Fragen zu der Benutzung der B2B-Plattform haben wenden Sie sich bitte direkt an unser Supplier Intergration Team.

Telefon: +49 5361 9 33099

E-Mail: [b2bteam@vwgroupsupply.com](mailto:b2bteam@vwgroupsupply.com)

Servicezeiten: Montag bis Freitag von 8:00 bis 18:00 Uhr (CET)

### 11.3. Aufbau des Angebots

Strukturieren Sie das Angebot bitte nach folgendem Schema:

- Beschreibung Leistungsumfang (Ausgangssituation, Aufgabe, Zielsetzung, Lösung, Liefer- und Leistungsumfang)
- Darstellung der **Ressourcen und deren Tagessätze** (Tätigkeitsprofile Kap. 4.11)
- **Explizit ausgewiesene Kosten für Einarbeitung** in die Big Data Plattform
- Nachweis der **kurzfristigen Verfügbarkeit** von Ressourcen und konkreter **Prozessvorschlag**
- **Aufzeigen exemplarischer Code-Beispiele für mathematische Algorithmen, Data Pipelines und Software Quellcode**
- Kenntnisse in **agiler Projektmethodik**

Darüber hinaus sollte Ihr Angebot folgende Informationen enthalten:

- **Kurze Darstellung Unternehmensprofil** Anbieter (Detailinformationen, Unternehmenskennzahlen, etc.), Lieferantendarstellung, Ansprechpartner
- Max. 2 anonymisierte **exemplarische Profile je Tätigkeitsprofil**
- Ihr Angebot sollte einen Umfang von 25 Seiten nicht überschreiten
- Allgemeine **Unternehmensreferenzen** (jeweils konkret nachzuweisen)
  - Erfahrung/Expertise auf dem für die Anfrage relevanten Fachgebiet
  - Nennung konkreter Referenzkunden und der dortigen Leistung
  - Engagement innerhalb der MTB, der MAN Gruppe und des VW Konzerns und seiner Marken
  - Eine Bewertung der inhaltlichen Kenntnisse über die Konzerndatenquellen von MTB und der MAN Gruppe mit den Abstufungen „Keine Kenntnisse“, „Grundkenntnisse“ und „Fundierte Kenntnisse“.



- Hinsichtlich der Vorstellung des Auftragnehmers sind folgende Fragen zu beantworten:
  - Wie sieht der prozessuale Ablauf aus, um einen neuen Mitarbeiter in die Firma/das konkrete Programm/Projekt zu integrieren?
  - Wie schaffen Sie es, eine Grundlast in Ihrem Unternehmen zu gewährleisten? Sind Sie ausschließlich von Projekten und deren Laufzeiten abhängig?
  - Wie schaffen Sie es, Auslastungsspitzen und –täler abzufangen?
  - Wie hoch war die Fluktuation in den Jahren 2016 und 2017?
  - Wie sieht die Projektabwicklungskette bei der Realisierung von Data Analytics Leistungen bei Ihnen aus (vom Lesen der Ausschreibungsunterlage, über die Entwicklung bis zum GoLive)?
  - Wie sieht das Projektcontrolling und –reporting gegenüber dem Auftraggeber bei Ihnen aus?

#### **11.4. Zusätzliche Vereinbarungen**

Es gelten die gesetzlichen Vorgaben sowie die Richtlinien der MTB bzgl. Datenschutz, IT-Sicherheit und Qualitätsmanagement. Darüber hinaus sind die Richtlinien, Normen und Arbeitsanweisungen der MTB zu befolgen.

Insbesondere dürfen mitgebrachte Rechner nicht im MTB Netzwerk angeschlossen werden. Des Weiteren sind die Vorgaben für den Datenaustausch zu beachten.

- **Datensicherheit und intellektuelles Eigentum**

Alle Pläne, Schriftstücke, Konstruktionszeichnungen, Modelle und andere Dokumente, die dem Auftragnehmer zur Verfügung gestellt werden, bleiben ausschließlich Eigentum der MTB.

Sie sind jederzeit auf Verlangen, spätestens jedoch nach Beendigung des Auftrages unaufgefordert an MTB zurückzugeben. Die Geltendmachung eines Zurückbehaltungsrechtes ist ausgeschlossen.

Alle vom Auftragnehmer unmittelbar erzielten Arbeitsergebnisse werden ausschließlich und exklusiv für MTB geschaffen und sind uneingeschränktes Eigentum der MTB.

Der Auftragnehmer wird über alle Vorgänge, Betriebseinrichtungen, betriebliche Anlagen usw., die im Zusammenhang mit seiner Tätigkeit für MTB bekannt werden, auch nach Erledigung seines Auftrages Dritten gegenüber Stillschweigen bewahren.

Falls bei der Tätigkeit Erfindungen entstehen, die zum Schutzrecht angemeldet werden können, behält sich die MTB jegliche Schutzrechte vor. Die Erteilung von Nutzungsrechten für den Auftragnehmer bedarf der separaten Klärung.

## 12. Ansprechpartner zur Ausschreibung

Ansprechpartner Fachbereich	Ansprechpartner des Einkaufs
Hr. Stefan Strehlow Abteilungsleiter (FIEB) <a href="mailto:stefan.strehlow@man.eu">stefan.strehlow@man.eu</a>	Fr. Delila Sigmund Einkäuferin (BAI) <a href="mailto:delila.sigmund@man.eu">delila.sigmund@man.eu</a>
Dr. Andreas Wittmann Teamleiter Analytics & Reporting (FIEBA) <a href="mailto:norbert-csaba.gergely@man.eu">norbert-csaba.gergely@man.eu</a>	

Tabelle 16: Ansprechpartner und deren Kontaktdaten

### 12.1. Fragen zum Angebot

Sämtliche Anfragen das Angebot betreffend sind ausschließlich in schriftlicher Form per Email an Dr. Andreas Wittmann und in Kopie Fr. Delila Sigmund zu stellen. Dabei ist die in der Zeitplanung angegebene Frist zur Fragestellung zu beachten.

Bitte beachten Sie, dass **Fragen die auf anderen Wegen oder außerhalb dieser Frist eingehen grundsätzlich nicht bearbeitet** werden.

## 13. Abkürzungsverzeichnis

Abkürzung	Bedeutung
-----------	-----------



B2B (-Plattform)	Business to Business (Zulieferer Portal von VW)
CET	Central European Time
Etc.	Et cetera
Ggf.	Gegebenenfalls
Inkl.	Inklusive
Insb	Insbesondere
IT	Informationstechnologie
MAN	MAN Gruppe / MAN SE
MTB	MAN Truck & Bus AG, Teil der MAN SE
Nr.	Nummer
PDF	Portable Document Format
PT	Personentage
sog.	So genannt
UrhG	Urheberrechtsgesetz
VW (Gruppe)	Volkswagen (Gruppe)
z.B.	Zum Beispiel

**Tabelle 17: Abkürzungsverzeichnis**



## 14. Tabellenverzeichnis

Tabelle 1: Tätigkeitsprofile und jährliche Mengengerüste .....	7
Tabelle 2: Beschreibung Tätigkeitsprofil Data Scientist.....	10
Tabelle 3: Beschreibung Tätigkeitsprofil Data Engineer .....	12
Tabelle 4: Beschreibung Tätigkeitsprofil Softwareentwickler.....	13
Tabelle 5: Beschreibung Tätigkeitsprofil Visualization.....	14
Tabelle 6: Beschreibung Tätigkeitsprofil Architect.....	15
Tabelle 7: Software Cloudera Hadoop .....	19
Tabelle 8: Software Deployment und Job Scheduling .....	19
Tabelle 9: Software Streaming .....	20
Tabelle 10: Software Data Pipelines .....	20
Tabelle 11: Software CI.....	20
Tabelle 12: Software Monitoring und Alerting .....	21
Tabelle 13: Software Logging.....	21
Tabelle 14: Software Sandbox 1.0 .....	22
Tabelle 15: Software Sandbox 2.0 .....	22
Tabelle 16: Ansprechpartner und deren Kontaktdaten .....	28
Tabelle 17: Mitgeltende Unterlagen.....	<b>Fehler! Textmarke nicht definiert.</b>
Tabelle 18: Abkürzungsverzeichnis.....	29