

## Thema 2.3: Apache Spark

Skalierbare verteilte Datenanalyse

Lukas Wappler

May 11, 2017

Einleitung  
Apache Spark  
Mehrere Komponenten im Verbund  
Performance  
Nutzung und Verbreitung  
Fazit  
Ausblick und Weiterentwicklung

# Inhaltsverzeichnis

- 1 Einleitung
- 2 Apache Spark
- 3 Mehrere Komponenten im Verbund
- 4 Performance
- 5 Nutzung und Verbreitung
- 6 Fazit
- 7 Ausblick und Weiterentwicklung

**Einleitung**

Apache Spark

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

# Einleitung

Einleitung  
**Apache Spark**  
Mehrere Komponenten im Verbund  
Performance  
Nutzung und Verbreitung  
Fazit  
Ausblick und Weiterentwicklung

Kern-Bibliotheken / Komponenten  
SQL-Abfragen  
Verarbeitung von Datenströmen (Spark-Streaming)  
Berechnungen auf Graphen (GraphX)  
Maschinelles Lernen (MLlib)  
Skalierung von R Programmen (SparkR)

# Einleitung

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

**Kern-Bibliotheken / Komponenten**

SQL-Abfragen

Verarbeitung von Datenströmen (Spark-Streaming)

Berechnungen auf Graphen (GraphX)

Maschinelles Lernen (MLlib)

Skalierung von R Programmen (SparkR)

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

**Kern-Bibliotheken / Komponenten**

SQL-Abfragen

Verarbeitung von Datenströmen (Spark-Streaming)

Berechnungen auf Graphen (GraphX)

Maschinelles Lernen (MLlib)

Skalierung von R Programmen (SparkR)

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

**Kern-Bibliotheken / Komponenten**

SQL-Abfragen

Verarbeitung von Datenströmen (Spark-Streaming)

Berechnungen auf Graphen (GraphX)

Maschinelles Lernen (MLlib)

Skalierung von R Programmen (SparkR)

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

Kern-Bibliotheken / Komponenten

**SQL-Abfragen**

Verarbeitung von Datenströmen (Spark-Streaming)

Berechnungen auf Graphen (GraphX)

Maschinelles Lernen (MLlib)

Skalierung von R Programmen (SparkR)



Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

Kern-Bibliotheken / Komponenten

**SQL-Abfragen**

Verarbeitung von Datenströmen (Spark-Streaming)

Berechnungen auf Graphen (GraphX)

Maschinelles Lernen (MLlib)

Skalierung von R Programmen (SparkR)

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

Kern-Bibliotheken / Komponenten

SQL-Abfragen

**Verarbeitung von Datenströmen (Spark-Streaming)**

Berechnungen auf Graphen (GraphX)

Maschinelles Lernen (MLlib)

Skalierung von R Programmen (SparkR)

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

Kern-Bibliotheken / Komponenten

SQL-Abfragen

Verarbeitung von Datenströmen (Spark-Streaming)

**Berechnungen auf Graphen (GraphX)**

Maschinelles Lernen (MLlib)

Skalierung von R Programmen (SparkR)

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

Kern-Bibliotheken / Komponenten

SQL-Abfragen

Verarbeitung von Datenströmen (Spark-Streaming)

Berechnungen auf Graphen (GraphX)

**Maschinelles Lernen (MLlib)**

Skalierung von R Programmen (SparkR)

Einleitung

**Apache Spark**

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

Kern-Bibliotheken / Komponenten

SQL-Abfragen

Verarbeitung von Datenströmen (Spark-Streaming)

Berechnungen auf Graphen (GraphX)

Maschinelles Lernen (MLlib)

**Skalierung von R Programmen (SparkR)**

Einleitung

Apache Spark

**Mehrere Komponenten im Verbund**

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung

Einleitung  
Apache Spark  
Mehrere Komponenten im Verbund  
**Performance**  
Nutzung und Verbreitung  
Fazit  
Ausblick und Weiterentwicklung

Besonderheiten bei der Speichernutzung  
Netzwerk und I/O-Traffic

Einleitung  
Apache Spark  
Mehrere Komponenten im Verbund  
**Performance**  
Nutzung und Verbreitung  
Fazit  
Ausblick und Weiterentwicklung

**Besonderheiten bei der Speichernutzung**  
Netzwerk und I/O-Traffic



Einleitung  
Apache Spark  
Mehrere Komponenten im Verbund  
**Performance**  
Nutzung und Verbreitung  
Fazit  
Ausblick und Weiterentwicklung

Besonderheiten bei der Speichernutzung  
Netzwerk und I/O-Traffic

Einleitung

Apache Spark

Mehrere Komponenten im Verbund

Performance

**Nutzung und Verbreitung**

Fazit

Ausblick und Weiterentwicklung

Einleitung
Apache Spark
Mehrere Komponenten im Verbund
Performance
Nutzung und Verbreitung
<b>Fazit</b>
Ausblick und Weiterentwicklung

Einleitung

Apache Spark

Mehrere Komponenten im Verbund

Performance

Nutzung und Verbreitung

Fazit

Ausblick und Weiterentwicklung