Thema 2.3: Apache Spark Skalierbare verteilte Datenanalyse

Lukas Wappler

May 11, 2017

Lukas Wappler Thema 2.3: Apache Spark May 11, 2017 1 / 19

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

Inhaltsverzeichnis

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

Lukas Wappler

Thema 2.3: Apache Spark

Einleitung
Apache Spark

Mehrere Komponenten im Verbund
Performance
Nutzung und Verbreitung
Fazit
Ausblick und Weiterentwicklung

May 11, 2017

2 / 19

Einleitung

Lukas Wappler

Thema 2.3: Apache Spark

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

Ausblick und Weiterentwicklung

Thema 2.3: Apache Spark
Mern-Bibliotheken / Komponenten
SQL-Abfragen
Verarbeitung von Datenstrmen (Spark-Streaming)
Berechnungen auf Graphen (GraphX)
Maschinelles Lernen (MLlib)
Skalierung von R Programmen (SparkR)

3 / 19

Einleitung

Lukas Wappler	Thema 2.3:	Apache Spark	May 11, 2017	4 / 19
Nutzu	Einleitung Apache Spark nenten im Verbund Performance ng und Verbreitung Fazit Weiterentwicklung	Berechnungen auf Maschinelles Lerne	Datenstrmen (Spark-Streaming) Graphen (GraphX)	

Lukas Wappler	Thema 2.3:	Apache Spark	May 11, 2017	5 / 19
Nutzu	Einleitung Apache Spark nenten im Verbund Performance ng und Verbreitung Fazit Weiterentwicklung	Berechnungen auf Maschinelles Lerne	, Datenstrmen (Spark-Streaming) Graphen (GraphX)	

Lukas Wappler

Thema 2.3: Apache Spark

Apache Spark

Mehrere Komponenten im Verbund
Performance
Nutzung und Verbreitung
Fazit
Ausblick und Weiterentwicklung

Mehrere Komponenten im Verbund
Performance
Nutzung und Verbreitung
Fazit
Ausblick und Weiterentwicklung

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

Lukas Wappler	Thema 2.3:	Apache Spark	May 11, 2017	7 / 19
Nutzu	Einleitung Apache Spark nenten im Verbund Performance ng und Verbreitung Fazit Weiterentwicklung	Berechnungen auf Maschinelles Lerne	, Datenstrmen (Spark-Streaming) Graphen (GraphX)	

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

Lukas Wappler

Einleitung Apache Spark

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

Lukas Wappler

Thema 2.3: Apache Spark

May 11, 2017 9 / 19

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

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

Lukas Wappler	Thema 2.3:	Apache Spark	May 11, 2017	10 / 19
Nutzu	Einleitung Apache Spark menten im Verbund Performance ng und Verbreitung Fazit Weiterentwicklung	Berechnungen auf Maschinelles Lerne	Datenstrmen (Spark-Streaming) Graphen (GraphX)	

Einleitung
Apache Spark

Mehrere Komponenten im Verbund
Performance
Nutzung und Verbreitung
Fazit
Ausblick und Weiterentwicklung

Lukas Wappler	Thema 2.3:	Apache Spark	May 11, 2017	12 / 19
Nutzu	Einleitung Apache Spark nenten im Verbund Performance ng und Verbreitung Fazit Weiterentwicklung	Berechnungen auf Maschinelles Lerne	, Datenstrmen (Spark-Streaming) Graphen (GraphX)	

Lukas Wappler Thema 2.3: Apache Spark May 11, 2017 13 / 19 Einleitung Apache Spark Mehrere Komponenten im Verbund Nutzung und Verbreitung Fazit Ausblick und Weiterentwicklung

Lukas Wappler	Thema 2.3:	Apache Spark	May 11, 2017	14 / 19
Nutzu	Einleitung Apache Spark nenten im Verbund Performance ng und Verbreitung Fazit Weiterentwicklung	Besonderheiten bei Netzwerk und I/O-	i der Speichernutzung Traffic	

Lukas Wappler

Thema 2.3: Apache Spark

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

May 11, 2017 15 / 19

Besonderheiten bei der Speichernutzung
Netzwerk und I/O-Traffic

Lukas Wappler	Thema 2.3: Apache Spark	May 11, 2017	17 / 19
Nutzu	Einleitung Apache Spark onenten im Verbund Performance ing und Verbreitung Fazit I Weiterentwicklung		

Lukas Wappler	Thema 2.3: Apache Spark	May 11, 2017 18 / 19
Nutzu	Einleitung Apache Spark nenten im Verbund Performance ng und Verbreitung Fazit Weiterentwicklung	

Lukas Wappler

Thema 2.3: Apache Spark

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

May 11, 2017 19 / 19