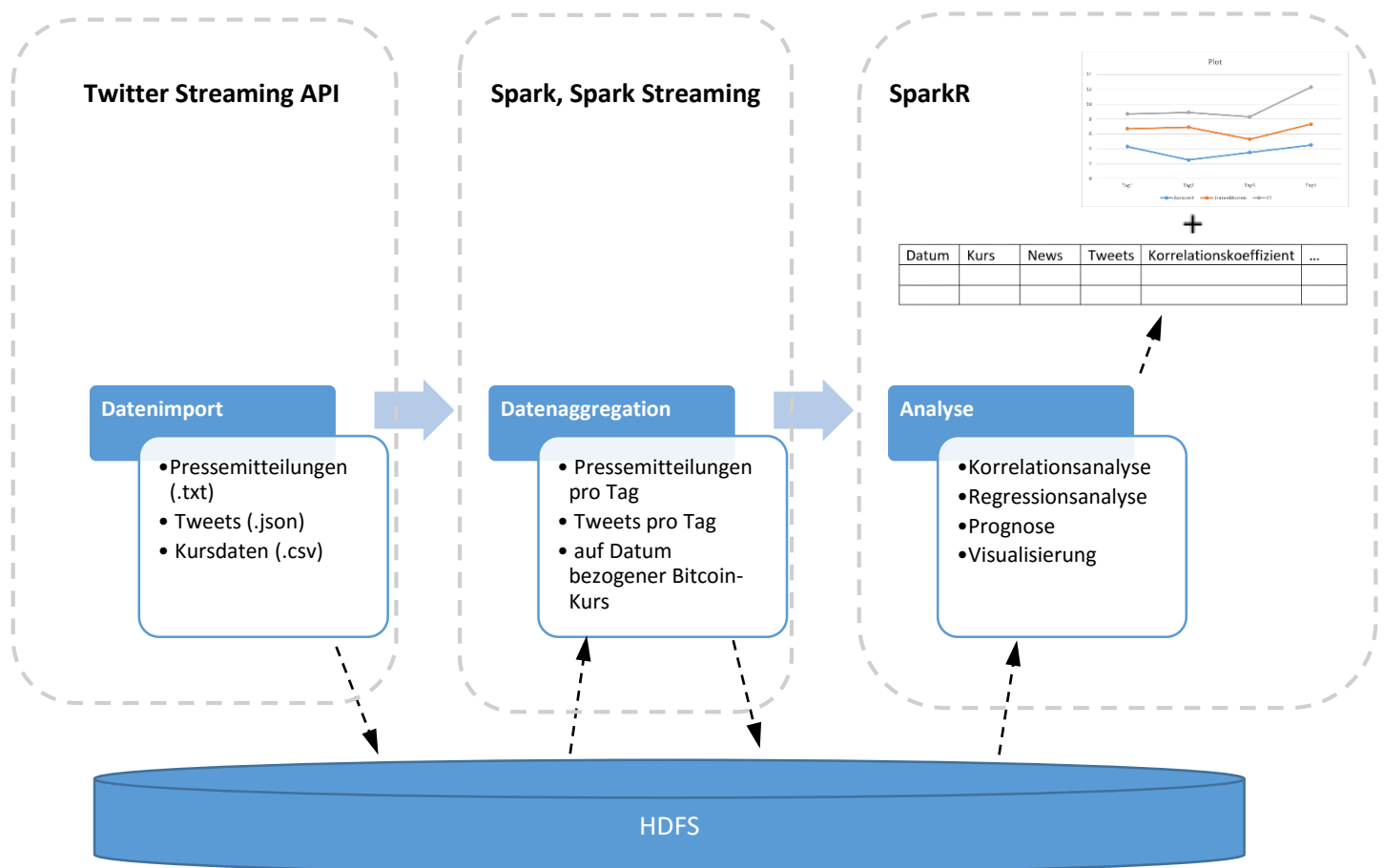


# Lösungsskizze

Temporale Analyse von News-Daten und Kursentwicklung



## System

- Linux Ubuntu 14.04 LTS und Arch Linux, Kernel: 4.6.2.1 (x86\_64)
- Entwicklungsumgebung: Eclipse Mars (inkl. Apache Maven 3.3.9-1 und Git)
- Hadoop 2.7.2
- Apache Spark 1.6.1 (inkl. Spark Streaming und SparkR)
- Twitter App (Erstellung auf apps.twitter.com)

## Datenimport

- Pressemitteilungen zum Thema Bitcoin (.txt), von cryptocurrenciesnews.com und Reuters
- Tweets mit #Bitcoin (.json), Streaming mit Twitter Streaming API (Query-Generierung mit OAuth Tool)
- Bitcoin-Kursdaten (.csv), von Quandl
- Import in HDFS

## Aggregation

- Pressemitteilungen, Tweets und Kursdaten
- nach dem Datum (dd-mm-yyyy)
- Anzahl Pressemitteilungen der verschiedenen Quellen (Cryptocoins-News, Reuters)
- Anzahl Tweets mit #Bitcoin
- Kurswert, Transaktionen, Bitcoin-Volumen

## Analyse des Einflusses von Pressemitteilungen und Tweets auf den Bitcoin-Kurs

- Korrelationsanalyse (News-Quellen untereinander, mit dem Kurswert, Transaktionen, Bitcoin-Volumen)
- Regressionsanalyse
- Prognose (Input: Daten eines bestimmten Zeitraums, Output: z. B. prognostizierter Kurswert des Folgetages)
- Visualisierung