As MapReduce is an important paradigm for solving big and costly problems today, there are really many use cases. Almost every problem that can be divided into small pieces and processed independent can be distributed and calculated in parallel by following the MapReduce paradigm. Not only scientific calculation, but also today’s companies are trying to solve their problems by using Map Reduce.

The following shows an exemplary list of problems that can be efficiently solved by MapReduce:

* Factorization of (big) Integers (Quelle: <https://code.google.com/p/mapreduce-integer-factorization/>)
* Matrix factorization (Quelle: <http://www.redaktion.tu-berlin.de/fileadmin/fg131/Publikation/Papers/sys024-schelter.pdf>)
* Multiplying terabit integers (Quelle: <https://people.apache.org/~szetszwo/ssmr20110430.pdf>)
* Fourier transformations
* Genetic analysis (Quelle: <http://genome.cshlp.org/content/20/9/1297.long>)
* Search engine indexing
* Big sort problems (Quelle: <http://courses.cs.washington.edu/courses/cse490h/08au/lectures/algorithms.pdf>)
* Users' behavior analytics
* Group text documents into topically related groups
* ETL and Data Mining
* Friend finder
* Defending E-Mail-Spam (Yahoo Mail: <http://readwrite.com/2010/05/24/map-reduce-yahoo-mail#awesm=~ozyRktADsuJsXT>)
* Shortest Path (Quelle: <http://courses.cs.washington.edu/courses/cse490h/08au/lectures/algorithms.pdf>)
* Word count (Quelle: <http://courses.cs.washington.edu/courses/cse490h/08au/lectures/algorithms.pdf>)
* Weather prediction

Some of these use cases are more scientific, others are for commercial use. Especially while data warehouses are rising, there is a need for calculating big data. When data is filled into the data ware house, there are often calculations to be done like calculating aggregations. This can be efficiently done by using Map Reduce. When data is already filled inside the data ware house, there are several techniques for data mining. As many parts of data mining can be split into pieces and calculated separately, it is very efficient to take Map Reduce algorithms to do so.

As MapReduce processing is computed by large clusters, nowadays with many cloud providers it is more a matter of being interested in the results than having the money to buy an own cluster. Almost all providers of cloud computing are having resources or hardware optimized solutions for doing MapReduce. E.g. Amazons Elastic Map Reduce (EMR) is between $0.015 and $0.50 per hour (Quelle: <http://aws.amazon.com/de/elasticmapreduce/pricing/>), which isn’t really expensive - especially if it is for commercial use.

Genetic analysis is an upcoming trend. Today the costs for a complete genetic analysis are below $1000. By doing this analysis on public cloud infrastructure and doing the analysis with the Map Reduce paradigm, these costs could be significantly reduced. Maybe in some years, it will be a normal thing to analyses the genomes of a person? - The technic is ready to do so. The bigger question is: Are the people willing to do that?

Factorization of big integers is a very interesting but also a dangerous use case. Many security and cryptographic methods are based on the problem that it is not possible to efficiently factorize big integers. Map Reduce is a way to solve some factorization problems. But if someone could find a really efficient way (much more efficient than Map Reduce), this would lead to new challenges in cryptography.

Search engine companies like Google do have a big need for indexing huge amount of data to give their users the best and most relevant search results in the shortest possible time. It is hardly surprising that Google as the market leader and inventor of MapReduce uses MapReduce to do that job while it is highly parallelizable. Google also uses the same techniques for indexing Emails in Gmail.

Friend finders were a revolutionary new thing for social networks. Companies. With Facebook’s friend finder leading the way big social networks implemented better search engines for people who are searching for their friends. A large amount of data has to be analyzed to give good suggestions for possible friends. Indexing profiles and searching for people can be highly parallelized. Facebook implemented its friend finder by using the MapReduce pattern.

Quellen:

<http://medcitynews.com/2012/12/genetic-analysis-company-nanostring-will-use-15-3m-for-first-launch-in-molecular-diagnostics/>

<http://courses.cs.washington.edu/courses/cse490h/08au/lectures/algorithms.pdf>

<http://highlyscalable.wordpress.com/2012/02/01/mapreduce-patterns/>

<http://www.infoq.com/news/2012/02/MapReducePatterns>

<http://stevekrenzel.com/finding-friends-with-mapreduce>

<https://mahout.apache.org/>