

# Verteiltes Rechnen

High Performance Computing für KMUs

# Herausforderungen

## 1. Datensicherheit

In hochsensiblen Bereichen wie z.B. Finanzdaten ist die Datensicherheit zentraler Bestandteil aller Prozesse. Auch wenn Cloud-Computing bei korrekter Konfiguration sicher genutzt werden kann, besteht grundsätzlich ein anderes Risikopotenzial als bei einem eigenem Rechenzentrum. Hinzu kommen erhöhte Aufwände im Risk Assessment, da dies eine externe Auftragsdatenverarbeitung darstellt. Die Anforderungen hierfür sind insbesondere im stark reguliertem Finanzbereich erheblich. Aus diesem Grund hat sich die PPA dazu entschieden jegliche Produktionsdaten ausschließlich in eigenen Rechenzentren zu verarbeiten.

## 2. Ausfallsicherheit

Bei lokalen Rechenzentren ist das Thema Ausfallsicherheit gerade für KMUs ein enormer Kostentreiber. Rechenkapazität redundant vorzuhalten kann abhängig von der Systemarchitektur einen deutlichen Einfluss auf die Wirtschaftlichkeit haben. Daher ist die PPA von einem monolithischem System zu einem hochflexiblen, modularen System gewechselt.

## 3. Verarbeitungsgeschwindigkeit

Gerade wenn kein Zugriff auf hochskalierbare Ressourcen des Cloud-Computings besteht, muss eine ausreichende Rechenkapazität im eigenen Rechenzentrum vorgehalten werden.

# Monolithisches System versus Cluster

## 1. Klassische Datenverarbeitung auf Servern

Die Serverumgebung der PPA besteht aus einem ESXi Verbund von drei physischen Servern. Einer der Server stand ausschließlich den HPC (High Performance Computing) Aufgaben zur Verfügung. Hierfür wurden 2 XEON CPUs mit je 20 Kernen und 40 Threads bereitgestellt. Im Falle eines Ausfalls konnten die Aufgaben auf einen anderen physischen Server verschoben werden. Hier steht dann allerdings nicht die gleiche Rechenleistung zur Verfügung.

## 2. HPC auf einem Cluster aus Desktop-Rechnern

Als Alternative wird nun einen Rechencluster aus 3 Intel® Core® i9-9900K und 3 AMD Ryzen® 9 3950X bereitgestellt. Hierbei wurde auf eine Kühlung geachtet, die es den CPU erlaubt dauerhaft den vollen Turbotakt bereitzustellen. Gleichzeitig wurde der Programmcode so optimiert, dass eine effiziente Stapelverarbeitung auf einem verteiltem System möglich wird. Fällt einer der Rechner aus, beeinflusst das zwar die Verarbeitungsgeschwindigkeit stört allerdings den Prozess als solchen nicht.

# Vergleich der Systemleistung

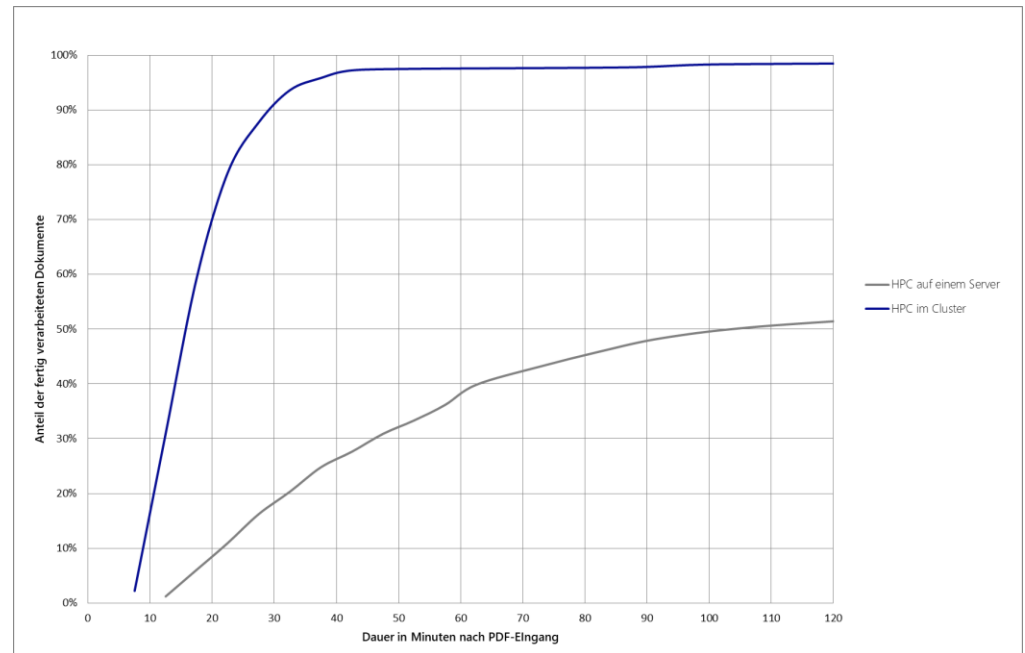
## 1. Theoretisch

Da die Kühlleistung in dem ThinkSystem SR630 nicht ausreicht um dauerhaft den Turbotakt bereitstellen zu können, kann der Server eine Rechenleistung von 96GHz bereitstellen. Diese stehen durch die Virtualisierung auch nicht vollständig zur Verfügung.

Jeder Rechner des Clusters hat eine ausreichende Kühlung um den Turbotakt dauerhaft bereitstellen zu können. Er hat eine Rechenleistung von 345,6GHz.

## 2. Am Beispiel der Verarbeitung von Jahresabschlussdokumenten

Die PPA verarbeitet in der Bilanzsaison im Schnitt 1000 Jahresabschlussdokumente mit in Summe 39.000 Seiten. Die Verarbeitung besteht aus einer Texterkennung, Dokumentaufbereitung und automatisierter Extraktion von Daten. Der Rechencluster hat 90% aller Dokumente bereits nach 30 Minuten verarbeitet. Der Server ist zu diesem Zeitpunkt erst bei 20%. Insbesondere große Dokumente haben im alten System zu einer enormen Verzögerung geführt.



# Wirtschaftlicher Vergleich und Fazit

## Kosten

- Hardwarekosten: Der Server hat ~13.000€ gekostet. Die insgesamt 6 Desktoprechner haben ~8.500€ gekostet
- Betriebssystem: Bei den Desktops war eine Windows 10 Lizenz inkludiert. Diese langt für die Anforderungen. Für den Server muss noch zusätzlich eine Windows Server Lizenz erworben werden.
- Sonstige Lizenzkosten: Die Lizenz der eingesetzten OCR Software muss auf eine Netzwerklizenz umgestellt werden.

## Vergleich mit Cloud Computing

Vorteile des Cloud Computing ist die enorme Flexibilität. Nicht genutzte Rechenleistung muss nicht gezahlt werden. Wenn man allerdings auf lokalen Systemen eine hohe Auslastung erzielen kann und die Wartung durch das bestehende Team abgedeckt werden kann, ist ein lokales System durchaus Konkurrenzfähig. So muss man für ein System mit 64 vCPUs mit 2\$/h (Linux) und über 5\$/h (Windows) kalkulieren. Wenn man dieses Systeme nur in 50 % der Zeit benötigt, werden Kosten von über 8000€/anno (Linux) oder über 20.000€/anno (Windows) verursacht.

## Fazit

Ein Cluster aus Desktops kann bei gleichen Kosten erheblich mehr Rechenleistung bereitstellen als ein Serversystem. Das System ist zusätzlich sehr viel ausfallsicherer. Wenn das lokale System ausreichend ausgelastet werden kann, braucht es den wirtschaftlichen Vergleich mit dem Cloud Computing nicht zu scheuen.



Felix Gratz

PPA Gesellschaft für Finanzanalyse  
und Benchmarks mbH  
Hilpertstraße 24b  
64293 Darmstadt

+49 6151 7804 500  
info@ppaworld.com  
www.ppaworld.com