Open your Big Data solutions to gain insights with IBM

Christophe Menichetti
IBM

Revolutionizing the Datacenter
Power is designed for Big Data and Analytics

**4X**
threads per core vs. x86 (up to 1536 threads per system)

**4X**
memory bandwidth vs. x86 (up to 16TB of memory)

**5X**
more cache vs. x86 (up to 224MB cache per socket)

**Processors**
flexible, fast execution of analytics algorithms

**Memory**
large, fast workspace to maximize business insight

**Cache**
ensure continuous data load for fast responses

Supports growth of users, reports and complex queries

Delivers fast analytics results for real-time decision-making

Handles large volumes of data for better response times

Optimized for a broad range of big data & analytics workloads:

82X is based on IBM internal tests as of April 17, 2014 comparing IBM DB2 with BLU Acceleration on Power with a comparably tuned competitor row store database server on x86 executing a materially identical 2.6TB BI workload in a controlled laboratory environment. Test measured 60 concurrent user report throughput executing identical Cognos report workloads. Competitor configuration: HP DL380p, 24 cores, 256GB RAM, Competitor row-store database, SuSE Linux 11SP3 (Database) and HP DL380p, 16 cores, 384GB RAM, Cognos 10.2.1.1, SuSE Linux 11SP3 (Cognos). IBM configuration: IBM S824, 24 cores, 256GB RAM, DB2 10.5, AIX 7.1 TL2 (Database) and IBM S822L, 16 of 20 cores activated, 384GB RAM, Cognos 10.2.1.1, SuSE Linux 11SP3 (Cognos). Results may not be typical and will vary based on actual workload, configuration, applications, queries and other variables in a production environment.
IBM Solution for BLU Acceleration – Power Systems Edition

- Accelerate complex queries & reports a mean of 82X

Integrated platform solution with DB2 BLU

IBM Data Engine for NoSQL – Power Systems Edition

- Lower cost, greater workload density for 24:1 server consolidation and up to 3X lower TCA

Integrated solution for NoSQL

IBM Solution for Analytics – Power Systems Edition

- Deliver 2-3X faster scoring results for predictive actions

Integrated solution with Cognos BI and/or SPSS

IBM Data Engine for Analytics – Power Systems Edition

- Less than half storage infrastructure with only 1 copy of data*
- Competitive performance with up to 50% less servers*

Integrated platform for Hadoop and Spark
New systems approach for big data
Game-changing innovation for fast access NoSQL data stores

IBM Data Engine for NoSQL
Lower Total Cost of Ownership for NoSQL infrastructure with no performance degradation

- Reduce server footprint with DRAM-FLASH consolidation
- 100% Redis compliant (no application changes)

24:1 infrastructure consolidation
3X cost savings
6X less rack space
2U server+2U FlashSystem vs. typical deployment

Source: for 24:1 system consolidation ratio (12:1 rack density improvement) based on a single IBM S824, (24 cores, POWER8 3.5 GHz), 256GB RAM, AIX 7.1 with 40 TB memory based Flash replacing 24 HP DL380p, 24 cores, E5-2697 v2 2.7 GHz), 256GB RAM, SuSE Linux 11SP3. Inbound network limits performance to 1M IOPs in both scenarios, equal capacity (#user, data) in both cases. x86 cost includes 10k$ for 2x 1U switches
Power Gzip Compression Accelerator

Human Whole Genome 3GB (hg19, GRCh37) 2/2009

Original file size: 3199905909

Compression ( % of orig. )

Hardware Compression

- Power Gzip Compression Accelerator
  - compression hardware accelerated
  - compression cpu seconds: 2.112
  - compressed file size: 1040892182
  - compression ratio: 0.32528837147130003315

Gzip/Gunzip

- compression cpu seconds: 103.492
- compressed file size: 1046714150
- compression ratio: 0.32710778996845810068

Compression ( % of orig. )

IBM

50x

15x

D. Jamsek, A. Martin, K. Agarwal e.a. IBM

© 2014 IBM Corporation
Acceleration Use Case Example:

- CAPI Flash and RDMA Leveraged Transparently to Spark Applications
- Coming…. HDFS CAPI FPGA Erasure Code Acceleration, CAPI FPGA Compression Acceleration, …. 

**x Degrees of Separation on Spark**

**CAPI Flash for RDD Cache = 4X memory reduction at equal performance**

**RDMA for Spark Shuffle = 30% Better Response Time, Lower CPU Utilization**

![Graph showing improvement in TPC-H 100 GB - Q1/Q2/Q18 with CAPI Flash and RDMA](chart.png)
Draw me an Hadoop appliance

With just one provider, with Open solutions, no lock in, accelerated analytics

**Simplicity** of optimized Expert Integrated Systems

**Flexibility** of open source choice and multi-workload capabilities

**Quality** of service that Power systems uniquely delivers over +20 years

Proprietary Intel CPU strategy

Proprietary Cloudera Hadoop Strategy

No Hardware accelerations
Join the conversation at #OpenPOWERSummit

cristophe.menichetti@fr.ibm.com