Evolution of Analytics: SAP and HP Breaking Analytic Performance Barriers with SAP HANA and HP AppSystems

An ENTERPRISE MANAGEMENT ASSOCIATES® (EMA™) White Paper
Prepared for SAP

May 2012

EMA™
IT & DATA MANAGEMENT RESEARCH,
INDUSTRY ANALYSIS & CONSULTING

Evolution of Analytics: SAP and HP Breaking Analytic Performance Barriers with SAP HANA and HP AppSystems

Table of Contents

Evolution of Analytics: SAP and HP Breaking Analytic Performance Barriers with SAP HANA and HP AppSystems

Executive Summary

In May 2012 Hewlett-Packard (HP) became the first SAP Partner to receive the Benchmark Certificate for the execution of the SAP NetWeaver Business Warehouse Enhanced Mixed Load (BW-EML) Standard Application Benchmark. HP has 20 years of experience with SAP applications as both a hardware environment vendor and a professional services partner. Part of this relationship includes HP's long-term history with the SAP Benchmarking program. This experience and expertise make HP a well-positioned candidate for the initial run of the SAP BW-EML Benchmark.

Evolution of Analytics

One of the main drivers for additional processing power in analytical platforms is for analysts to explore new business areas at faster speeds. Linked with an increase in demand for greater analytic maturity is the demand for the utilization of the wider array of available detailed data sources. Often in the past, concessions at the database or analytical platform level were made to make analytics processes practical. These compromises are no longer being accepted by business analysts or required by Information Technology (IT) teams.

SAP Standard Application Benchmarks

Openly published and distributed, the SAP Standard Application Benchmarks provide insight for customers on strengths and limitations of particular software configurations and hardware environments. These performance testing specifications of hardware, system software components, and database management systems place a substantial load upon a given implementation of an SAP application under a carefully constructed set of technical requirements. With this framework and associated metrics, SAP technology partners have the opportunity to show customers the value of their configurations.

SAP NetWeaver Business Warehouse Benchmark

Like its predecessor, the SAP NetWeaver Business Warehouse Mixed Load Benchmark (BW-MXL), the SAP BW-EML Benchmark focuses on a mix of multi-user database query demand and incremental data updates. The SAP BW-EML benchmark specification updates continue to keep pace with analytical user demands.

Intended Audience

This document is intended for business unit and IT managers; database and application system administrators; and experienced users wishing to learn more about the capabilities of SAP NetWeaver Business Warehouse 7.3 and the HP AppSystems for SAP HANA™ appliance. This paper also outlines how the SAP Benchmarking program tests the specifications of the changing analytical platform marketplace.

EMA Key Observations

- **Neutral Advocate:** The SAP Benchmark Group, with the definition of its standardized application benchmarks, maintains a neutral status and valued advisor position for customers and partners.
- **Business Demands Driving Change:** Evolution of the SAP BW-EML Benchmark shows maturity of customer requirements in analytics to a more flexible, ad-hoc analytical query structure.
- **Total Cost of Ownership (TCO):** Using SAP NetWeaver BW with HP AppSystem appliances serve to provide the enablers for lower TCO with integrated performance of an in-memory database.
- **Details:** Go to the following for more detailed information on:
  - **SAP HANA database** http://www.sap.com/hana
  - **HP AppSystems for SAP HANA appliance** http://www.hp.com/go/sap/hana

Breaking Analytics Performance Barriers

Before Chuck Yeager and the first flight to successfully break the sound barrier, there were compromises in aircraft design because of failings in the technology associated with aircraft engines, wings or airframes. However, there was always the quest for speed.

Whether it was the "original" goal to get into the sky and achieve flight in the case of the Wright Brothers or a more "modern" goal such as moving cargo from one continent to another in the case of DHL, speed has been an important aspect of both the technology of aircraft design and use of aircraft. The evolution of aircraft design technology has enabled new ways of using planes and the effort to use them in new ways has driven the development of innovative aeronautical technologies.

Just like aircraft design, the analytics of business intelligence is evolving. And just like aircraft design, one of the main drivers of this evolution has been the quest for speed and new ways to leverage it.

In the early days of business intelligence, standardized, aggregated reports were the norm. Often this was the case due to limitations in the underlying analytical environments. However as organizations matured, more data became available and advanced analytical techniques have left the "laboratory" of the analytical back office and pushed for more processing power to be built into those analytical applications.

However as organizations matured, more data became available and advanced analytical techniques have left the "laboratory" of the analytical back office and pushed for more processing power to be built into those analytical applications.

One of the main drivers for additional processing power in analytical platforms is to explore new domains and at faster speeds. Quantitative analysts, like their aircraft pilot counterparts, are pressing for continually more sophisticated analytic requirements. Linked with increase in demand for analytic maturity is the demand for the utilization of the wider array of detailed or atomic level data sources. Often in the past, because of the deficiencies of database management, system platforms and the processing power associated with analytical environments, concessions were made to make analytics processes more practical. Aggregated data was used to speed analytics. Pre-determined dimensions were established and processed to enable analysis to be available in the timeframes expected by the user community. However, just as analysts have pushed higher and higher with analytical demands, analysts have been pushing to utilize finer levels of detail data. No longer are aggregations by region or time acceptable for analysts to determine the best path for competitive advantage.

It was once acceptable to define "detail data" at a region, or store, level when the focus was on internal corporate perspectives. However, now detail level data is defined at the line-item level transactions because the focus is on the customer transaction and how those customers roll up to a store and/or a regional level. As analysts combine fine-grained Web analytics techniques, like click-stream analysis, with "pre"-transaction information, such as sensor based, in-store RFID information, detail data will again be redefined. Detail data may soon be described as not only the act of purchase, but the steps that lead to the transaction. This will provide understanding of not only what is purchased, but what is not purchased.

Powering Analytical Change

With the above demands for more advanced analytics and finer detail data definitions, technical environments have evolved to meet these challenges of flexible analytical dimensions, expected analyst response and processing power.

Applying columnar techniques to structured datasets has allowed analytic environments to remove some of the pre-established dimensions on analytics. By using columnar approaches, IT teams can speed the overall time to implementation by not pre-configuring the dimensions that the analysts will use. In the past, with row-based approaches, the cost of including all the dimensional columns in the analytical datasets had an overwhelming burden at execution time. With columnar data stores, decisions can be made at execution time with the use of select queries to avoid some of the penalties associated with row-based approaches.

The development of Massive Parallel Processing (MPP) for database management systems has made it possible to process the type of data associated with modern detailed data loads. No longer are legacy Symmetric MultiProcessing (SMP) models dictating compromises to IT teams that must in turn be placed upon the analyst community. MPP allows for datasets that meet not only Big Data requirements for ingesting, processing and delivery, but also enables future growth of detailed data sets when organizations move to the next definition of detail.

In-memory technology further solves the latency issue associated with performing analytics on detailed data. With falling memory prices, the ability to place larger and larger data sets in memory makes analytics much faster and less expensive. When an analytical platform and/or data store can avoid the Input/Output (I/O) costs associated with a rotating disk drive, primary access architectures, the performance of that analysis increases dramatically.

Putting these types of technologies together produces a synergistic effect to improve the response time of analytics to meet the expectations of analysts and business users. These technologies also expand the realm of possibilities of data usage in analytics. The limitations from row-based, SMP, rotating disk architectures were much like the "sound barrier" in early aeronautics. When the appropriate combination of wing, engine and airframe were put together, that barrier no long existed and aeronautics became aerospace. This is similar to the combination of MPP, in-memory and columnar for analytics. With these properly aligned, analytics can now expand beyond its past limitations.

Putting these types of technologies together produces a synergistic effect to improve the response time of analytics to meet the expectations of analysts and business users. These technologies also expand the realm of possibilities of data usage in analytics.

Pushing the Envelope

Similar to the advances in the technologies of aeronautics that allowed for supersonic flight and space launches, there needed to be tests that proved out the technologies. For aircraft, this is often the test flight and pushing the "envelope" of performance characteristics. In information technology, "test flight" is a technology load test or benchmark. There are internal testing benchmarks that organizations use to test their internal software and technology. Many of these benchmarks never saw the light of day because they represented proprietary information for specific internal purposes.

There were also independent industry benchmarks. These benchmark details were published and promoted in the technology industry. However, these benchmarks were often more of a generalized benchmark rarely directly applicable to particular use cases. By becoming independent and generalized

in nature, industry benchmarks represent more of a theoretical use case rather than a specific application use case. Transaction Processing Performance Council (TPC)[1] is the most well-known of the independent application/database benchmark development standards body.

Both of these types of benchmarks have evolved over time. Initially for applications using database management systems, the performance benchmarks tested the state of the art for operational processing. However, as analytics became more critical to organizations and the combination of technology and technique raised the analytical use case to a critical component, benchmarks incorporated the techniques and technologies of early OLAP. At each stage it is necessary to adjust benchmarks to meet the appropriate use case and technology being presented.

Combination Benchmarks

Initiated in 1995 and supported by the SAP Benchmark Council, the SAP Standard Application Benchmarks are a combination of internal load tests and industry benchmarks. With the evolution of SAP as a near industry standard vendor in many application areas, SAP developed its benchmarking program to help customers and partners load test various hardware environment and database management system configurations for SAP solutions. Working together with its hardware and database management system partner organizations, SAP developed the SAP Standard Application Benchmarks to test the hardware and database performance of SAP applications and components just as internal load tests and industry benchmarks would. However, in contrast with industry benchmarks, specific SAP application use cases and application configurations are used to specifically represent customer relevant scenarios.

SAP developed its benchmarking program to help customers and partners load test various hardware environment and database management system configurations for SAP solutions.

Openly published and distributed among SAP's technology partners, the SAP benchmarks provide environment configuration guidelines for customers. During the benchmark execution, load testing of hardware, system software components, and database management systems place substantial load upon a given implementation of an SAP application. Performance data applicable to overall technical environment, user interaction, and business application performance are monitored during a benchmark run. The end metric for a given benchmark and can be used to compare platforms.

Evolving to Match Customer Requirements

The new SAP NetWeaver Business Warehouse Enhanced Mixed Load Benchmark (BW-EML) is the latest member in the family of Standard Application Benchmarks from SAP.[2] These benchmarks include server power utilization trials; enterprise resource planning, supply chain management, customer relationship management, product lifecycle management, business intelligence, planning and optimization application load tests; and SAP industry measures for banking, utilities and retail environments.

Like its predecessor, the SAP NetWeaver Business Warehouse Mixed Load Benchmark (BW-MXL), the SAP BW-EML Benchmark focuses on a mix of multi-user database query demand and incremental data ("delta data") updates. "Delta data" is loaded into the database management system simultaneously with the query execution. The SAP BW-EML benchmark specification updates continue to keep pace with the evolution of user demands. These benchmark specifications simulate the loads associated with modern analytics and the analyst tasks associated with that analysis.

[1] "TPC Benchmarks", Transaction Processing Performance Council, http://www.tpc.org/information/benchmarks.asp
[2] "SAP Standard Application Benchmarks", www.sap.com, http://www.sap.com/solutions/benchmark/index.epx

SAP NetWeaver Business Warehouse Enhanced Mixed Load Benchmark

The SAP BW-EML Benchmark consists of simulated user capacity and data load tests on a multi-layered data warehouse environment. This includes three InfoCubes and seven DataStore objects. Each of these objects holds data of one particular year. The three InfoCubes, which provide the physical data model responsible for multidimensional analytics, hold the same data as the corresponding DataStore objects for three fiscal years.

The InfoCubes have a set of 16 dimensions that comprise a total of 63 characteristics, having cardinalities of up to 1 million different values and one complex hierarchy. With its 30 different key figures, including key figures requiring exception aggregation, the InfoCube data model has been defined in close accordance with typical customer data models. In the DataStore Object data model, the high cardinality characteristics have been defined as key members, while other characteristics have been modeled as part of the data members.

SAP BW-EML Benchmark can be executed with various data volumes. In its smallest configuration, the benchmark rules require an initial load of a total of 500 million records (i.e., 50 million records per InfoCube/DataStore Object) coming from ASCII flat files. Further possible configurations include initial load volumes of:

- 1000 million records
- 2000 million records

A single benchmark run is designed to last at least one hour. During this time, "delta data" is loaded in increments every five minutes.

Simulated User Process Description

For the SAP BW-EML Benchmark, each of the simulated users logs on to the system and executes a series of analytical queries and their associated follow-up investigation queries. Unlike actual analysts who spend time reviewing and comparing results, the simulated user analytic processes are run without delay to simulate a greater application load. These simulated analytical queries pick information for one particular year at random. An appropriate SAP BW InfoCube or DataStore object containing the appropriate data is accessed. Within any one of the required analytical queries, further exploration actions are executed. These exploration actions are referred to as **ad-hoc navigation steps**.

Each simulated user executes eight analytical queries and their associated filtering and drill-down actions, to comprise a total of 40 ad-hoc navigation steps. Each of these ad-hoc navigation steps result in an individual query and access to the underlying database management system. Although analytical queries follow similar navigation patterns, the filter and drill-down operations have been randomized to address the demand for ad-hoc type queries associated with the demands of a typical analyst community. To maintain the ad-hoc nature of the navigation steps, random values for filter parameters ensure that different partitions of data are accessed. Random selections of dimensional characteristics are used for drill downs, and other slice-and-dice operations. The user then logs off completing a particular cycle.

The above actions are simulated in the benchmark environment beginning with the initial load of 500 million, 1000 million or 2000 million rows. During the benchmark test run (a total 1 hour), a delta data package is uploaded 120 times. The total size of these packages is 1/1000th of the total records of the initial load or between 500k and 2 million rows per hour.

HP Runs Initial SAP BW-EML Benchmark

In May 2012, HP became the first SAP Partner to receive the SAP Benchmark Certificate for the execution of the SAP BW-EML Standard Application Benchmark. HP has 20 years of experience with SAP applications as both a hardware environment vendor and a professional services partner. Part of this long-term relationship has included HP's continued participation with the SAP Benchmarking program and associated workgroups. This experience and its expertise with business intelligence and data warehousing professional services make HP a well-positioned candidate for the initial run of the SAP BW-EML Benchmark.

In addition to their technical and domain experience, HP and SAP share thousands of joint customer implementations. Certification with the SAP BW-EML Benchmark would provide guidance on options for SAP customers on technical implementation configurations for SAP NetWeaver Business Warehouse.

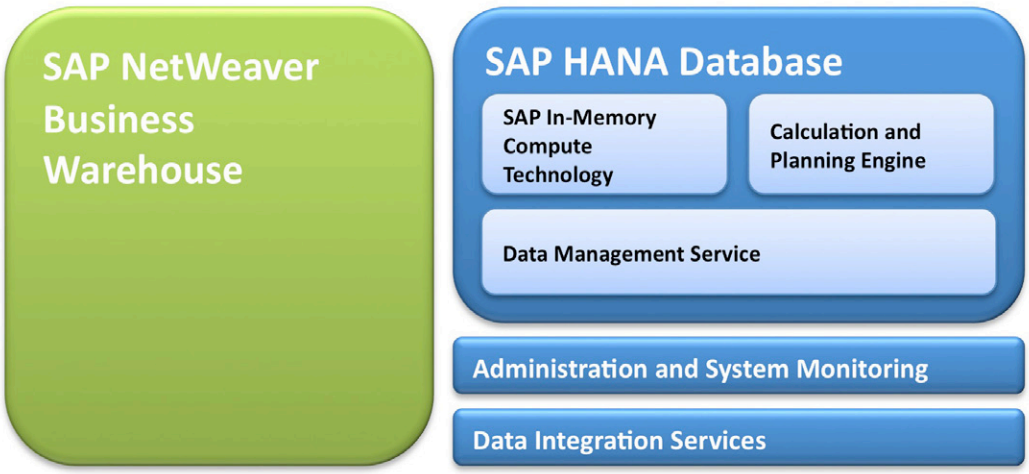
HP has 20 years of experience with SAP applications as both a hardware environment vendor and a professional services partner. Part of this long-term relationship has included HP's continued participation with the SAP Benchmarking program.

Software Architecture

For the SAP BW-EML Benchmark execution, HP was challenged with making the proper environment choices for the best benchmark run performance. With HP's wealth of knowledge associated with business intelligence/data warehousing requirements assessment and software architecture, HP had a wide selection of choices for the underlying database management system. The SAP BW-EML Benchmark specifications place a high emphasis on ad-hoc queries and user response time. This parallels changes in the analytics domain and the expectations of the high-level analyst community.

With these constraints in mind, HP made the architectural decision to utilize the SAP HANA in-memory database as the database management system. SAP HANA's strengths in detailed data management and query response were the main drivers for this decision.

The configuration of the SAP NetWeaver Business Warehouse environment for the benchmark test is configured as follows:



Technical Implementation

Again, leveraging HP's experience with technical architecture and hardware environment implementation, HP made choices associated with the hardware implementation for SAP NetWeaver Business Warehouse as the application and SAP HANA for the database management system. Employing HP's long history developing infrastructures supporting in-memory technologies, the HP AppSystems for SAP HANA Appliance[3] was selected as the best solution for the underlying database server. Integrating the SAP HANA functionality with a pre-configured server appliance has benefits for both performance and time to implementation.

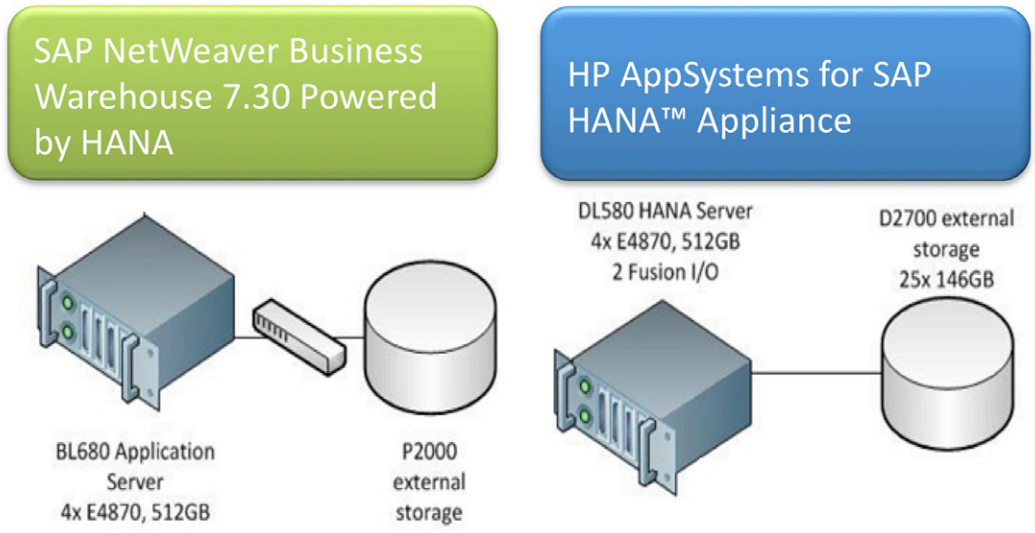
HP AppSystems for SAP HANA Appliance was selected as the best solution for the underlying database/ server combination. Integrating the SAP HANA functionality with a pre-configured server appliance has benefits for both performance and time to implementation.

For this particular environment and the specifications of the SAP BW-EML Benchmark, the HP AppSystems for SAP HANA "medium" appliance was selected. The "medium" SAP HANA AppSystem is a single node environment running with the following configuration:

ATTRIBUTE	SPECIFICATION
Server	ProLiant DL580 G7
CPUs	Four (4) Intel x86-E7
Operating System	SUSE Linux Enterprise Server 11
Persistent Storage	Two (2) 300GB SAS drives expandable with 24 146GB SAS drives
Storage Capacity	Up to 1.28TB (depending on compression

The "medium"-level AppSystem appliance, used for the SAP BW-EML Benchmark run, is one of a continuum of single-node and scale-out appliances available for the SAP HANA database.[4]

The combined SAP NetWeaver Business Warehouse and SAP HANA environment is configured with the following specifications:



[3] "HP AppSystems for SAP HANA™ Appliance Solution Brief", www.hp.com, http://h20195.www2.hp.com/v2/GetPDF.aspx/4AA3-3037ENW.pdf

[4] "HP AppSystems for SAP HANA", www.hp.com, http://www.hp.com/go/sap/hana

Benchmark Results

The SAP BW-EML Benchmark run on May 13, 2012, at the HP Test Labs in Houston, Texas, with a total of 1 billion (1000 million) records achieved the following performance results:

METRIC	VALUE
Ad-Hoc Navigation Steps Per Hour	65,990
Database Management System	SAP HANA 1.0
Application Software Release	SAP NetWeaver Business Warehouse 7.30 powered by SAP HANA
Operating System Release	SUSE Linux Enterprise Server 11

EMA Perspective

SAP has created a modern benchmark to keep up with the demands of today's analytic OLAP market and developing a general framework for hardware and software partners. The establishment of the SAP Benchmark program allows partners to test the performance of software and hardware configurations. These load tests reflect the direct feedback and use cases of the SAP customer base. This saves time and effort generally associated with individual internal load tests against various SAP environments by customers who want to make a purchase decision. The SAP Benchmark program metrics also give a much better sense of comparison between vendor hardware and database configurations. In addition, SAP has done a good job being a neutral observer when customers want to evaluate hardware and database options.

The establishment of the SAP Benchmark program allows partners to test the performance of software and hardware configurations. These load tests reflect the direct feedback and use cases of the SAP customer base.

As discussed above, the evolution of the specific SAP NetWeaver Business Warehouse Benchmark with the Enhanced Mixed Load analytical requirements mirrors the maturity and increased sophistication of the analytical community. The inclusion of ad-hoc navigation steps in the benchmark process and a depth of detailed data and near continual loading of information into the database management system represents the new realities of business requirements. The **ad-hoc navigation steps per hour** measurement also reflects the industry paradigm shift away from simple queries per hour to a more complex measurement.

With the changes to analytical business requirements, HP has made the decision to incorporate the SAP HANA database, implemented on an HP AppSystems for SAP HANA appliance, as the underlying data store for SAP NetWeaver BW. SAP HANA brings with it the benefits of columnar structure, MPP processing and in-memory calculation speed. By using the flexibility of the columnar architecture, analysts should enjoy the promised flexibility and performance gains associated with ad-hoc select attribute queries. The MPP processing will positively impact the ability to load data into the data store on an ongoing basis. Finally, the in-memory processing will match the expectations of the user community for returning fast results.

In addition to the use of the HP AppSystem appliance in the Benchmark run configuration, HP has displayed the value of its experience in business intelligence and data warehousing professional services. One of the major selling points of the SAP HANA database is lowering Total Cost of Ownership

(TCO). This is accomplished by reducing time to implementation and the complexity of database design. For the SAP BW-EML Benchmark execution, the HP implementation team decided to run SAP HANA "out of the box." In this configuration, advanced caching and indexes, usually required to meet general SAP BW Benchmark performance standards, were not implemented. Also, the SAP BW InfoCubes and DataStore objects were implemented using a simplified configuration rather than a more complex one generally used with other database platforms. In both of these cases, the short-term (time to implementation) and long-term (table configuration and maintenance) TCO arguments appear to be made.

Relating to the actual performance of the SAP NetWeaver BW/AppSystems for SAP HANA environment, the results of 65,990 ad-hoc navigation steps per hour provides a good baseline for an initial benchmark execution run. It should be noted that this absolute benchmark measurement was accomplished in a simulated user environment without the typical "think time" associated with a normal analyst's work process. The benchmarked number ad-hoc navigation steps roughly equates into a user community of approximately 825 concurrent analysts accessing the benchmarked SAP NetWeaver environment for a given hour. This assumes an average "think time" of 45 seconds per ad-hoc navigation step, which typically requires more analyst involvement than the analysis associated with an average standard report.

Overall, the measured performance of the SAP NetWeaver BW environment using the HP AppSystems for SAP HANA appliance shows the value of combining pre-configured hardware and tuned software platforms together. These systems allow for customers to lower the overall time to implementation of a SAP NetWeaver Business Warehouse with the delivery of a purpose-built appliance architecture. Linked with the SAP BW-EML Benchmark results, customers will not only have a strong sense of the performance potential of the "medium" AppSystem; they will have ability to implement in a short timeframe as well.

Overall, the measured performance of the SAP NetWeaver BW environment using the HP AppSystems for SAP HANA appliance shows the value of combining pre-configured hardware and tuned software platforms together.

About Enterprise Management Associates, Inc.

Founded in 1996, Enterprise Management Associates (EMA) is a leading industry analyst firm that provides deep insight across the full spectrum of IT and data management technologies. EMA analysts leverage a unique combination of practical experience, insight into industry best practices, and in-depth knowledge of current and planned vendor solutions to help its clients achieve their goals. Learn more about EMA research, analysis, and consulting services for enterprise line of business users, IT professionals and IT vendors at www.enterprisemanagement.com or blogs.enterprisemanagement.com. You can also follow EMA on Twitter or Facebook.

Corporate Headquarters:
5777 Central Avenue, Suite 105
Boulder, CO 80301
Phone: +1 303.543.9500
Fax: +1 303.543.7687
www.enterprisemanagement.com

2472.051512