

# Dell's Data Science and Big Data Analytics v2 Builds Competencies in Latest Technologies

Dell Technologies Education Services  
 Best Advance in Competencies and Skill Development  
 September 2019



## Company Background

### DELL Technologies

<b>Company-at-a-Glance</b>	Dell Technologies is an U.S. multinational company that offers a wide range of IT products (hardware and software) and services for enterprise, government, small business and consumer markets.
<b>Headquarters</b>	Round Rock, TX
<b>Year Founded</b>	1984
<b>Revenue</b>	\$90.621 billion
<b>Employees</b>	More than 146,000 (2019)
<b>Global Scale</b>	Global
<b>Customers/Output, etc. (Key customers and services offered)</b>	Dell's products include personal computers, servers, smartphones, televisions, computer software, computer security and network security, as well as information security services.
<b>Industry</b>	Computer Hardware and Software
<b>Website</b>	<a href="http://www.dell.com">www.dell.com</a>

## Budget and Timeframe

<b>Overall budget</b>	Developed Internally
<b>Number of Operations or Subject Matter Expert employees involved with the implementation?</b>	Three
<b>Timeframe to implement</b>	Eight months

## Business Conditions and Business Needs

---

Data science has been an integral part of business for decades. It has been called by different names: industry analysis, business analytics, predictive modeling, technical analysis, quantitative analysis, decision sciences and so on. But at the core, the principles of all quantitative analysis remain the same. It is about collecting, organizing, analyzing the data, and making reliable predictions based on that data. While the processes for doing this have been in the making for centuries, the advent of modern computing has been a game changer. Modern computers allow everyone to look at the data and analyze it meaningfully to further business growth.

Lately, the growth in computing power has increased exponentially and so have the data analytics tasks. While data is recorded and analyzed continuously to make real-time decisions, there is a great demand for engineers and professionals who can make the most out of this information to make meaningful contributions in their organizations.

Data science itself became a coined term when *Harvard Business Review* published an article calling "Data Science ... The Sexiest Job of 21<sup>st</sup> Century." However, in more recent years, data sciences have transcended the capabilities that were expected, even in the previous decade. Skills such as Machine Learning, Artificial Intelligence, Deep Learning and Big Data Analytics have become necessary skills with very high demand.

This Data Science and Big Data Analytics (DSBDA) course was preceded by a business transformation course which was aimed at "teaching leaders and executives about Big Data and data science" as part of Dell EMC's data science curriculum. However, as interest built up about delivering actionable skills and developing capabilities in-house, an attempt was made to create a course which quickly evolved into a textbook which became an industry standard, an instructor-led intensive five-day course, a video-based course from top data scientists in the company and an online instructor-led course. All courses are supported by intensive lab exercises which cover all techniques and algorithms discussed in the course.

The course has been updated to stay relevant and up to date to deliver a strongly interconnected set of skills which help professionals to dive right in to solve pertinent business issues.

The certification has remained prestigious and is considered a true mark of skill across the industry.

## Overview

---

### Organizational Objectives/Goals from Dell's Perspective

DSBDA v2 is an open-curriculum course. In other words, it isn't product specific or specific to an industry. It is a general set of skills required by certain professionals who want to do more with the data available at their disposal. This includes simple outcomes such as reporting, dashboarding and monitoring metrics and complex outcomes such as predictive analytics, as well as running queries on large databases and handling end-to-end model building exercises that can be integrated into the decision-making process.

A high-level description of the program:

- The core of the program is the data analytics lifecycle. This provides a repeatable, reproducible template for any data analytics/predictive modeling/machine learning exercise that could come up in an industry.
- The program introduces basic concepts of statistics and hypothesis testing and builds on these to introduce advanced predictive modeling techniques such as regression, classifiers and time series analysis. Concepts such as accuracy, precision and model effectiveness are discussed in detail and put in perspective.
- The program introduces R programming and application for the algorithms discussed in the course. There also is a final exercise that helps students to apply the data analytics life cycle to a case study. The exercise tests every skill that is taught in the course along with an emphasis on strategic thinking.
- The program also places emphasis on creating a report for various stakeholders and the level of complexity that must be introduced in those presentations. For example, business leaders must be able to see the overall value, while the technical transition team needs to know the components of the model so that they can optimize the deployment.
- Hadoop/SQL
- Labs
- There is also a strong emphasis on data visualization and the standards to make accurate and effective data visualization.

## Design of the Program

The Data Science and Big Data Analytics is part of Dell Technologies Proven Professional path which focuses on continuously improving skill sets based on application and work experience.

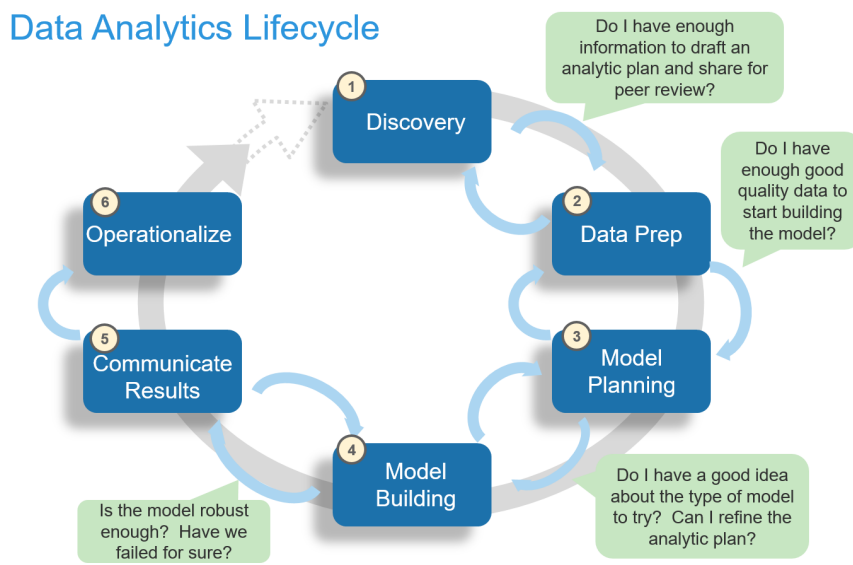
The same philosophy was included when designing the DSBDA course. The course is delivered in two levels:

- The Associate Program:** Intended for business analysts, database admins, people with at least two years of experience working on data reporting, dashboarding, running SQL queries, charting, etc. While a working knowledge of SQL is desired, it isn't assumed.
- The Advanced Methods in Data Science:** Intended for those who have successfully completed the associate course or have passed the certification exam. Additionally, the student should have worked as a data scientist for two or more years. The course expects the student to be well-versed in the project lifecycle, statistics and machine-learning algorithms.

As a result, the design of the course focuses on a well-rounded Data Scientist who is capable of picking up a project and delivering it autonomously.

The core of this entire program rests with the data analytics lifecycle. All problems are tackled taking the life cycle into consideration.

**Figure 1: A Description of the Data Analytics Lifecycle**



Source: Dell Technologies Education Services

The design also looked closely at the real problems faced by new data scientists. While there are plenty of resources on individual topics in data science, new data scientists find it hard to put together everything they have learned in a cohesive manner and apply their skills to the problem at hand. They often miss critical elements in steps such as data preparation and often need to revisit the step once they see the model isn't functioning satisfactorily.

These are skills that people build up over years of working on complex projects. As a result, the program leverages the decades of experience of the authors and instructors to help the students hit the ground running.

## Delivery of the Program

---

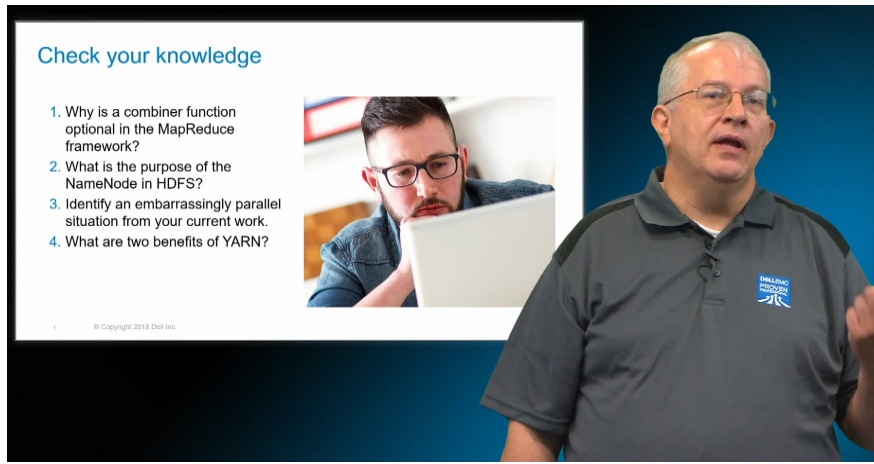
The program has an international demand and outreach. While the course is primarily marketed to enterprise users, there is a huge demand from consumers who identify data science as a key skill to be added to their CV. Given the demand, there are qualified data scientists who can work autonomously with extended teams of data engineers, database administrators and other stakeholders like business managers/sponsors who fund future projects within the organization.

The course can be taken in the following formats depending on the time and resources available:

1. **Instructor Led Training (ILT):** The course is delivered by a highly qualified Data Scientist with more than 15 years of experience in the field across various positions. The delivery is done in North America, Europe and Asia. The delivery is done over five days and covers five modules, including hands-on lab exercises and a case study to cover all concepts discussed. The program gives the students an intense, extended session where they are immersed in the content through discussion and scenario building. The students also can share their own domain expertise and the problems they are looking to solve in their industry to find optimal solutions.
2. **On-Demand Course:** The course is priced lower than the ILT and is available online for anyone wishing to take a self-paced course. The course has been built using cutting edge digital development technologies to visualize complex statistical algorithms. There is also a step-by-step instruction of the lab exercises, along with alternative packages that students can use to achieve the same result. The course videos and the lab videos are highly detailed to cover every aspect of the final assessment for certification.

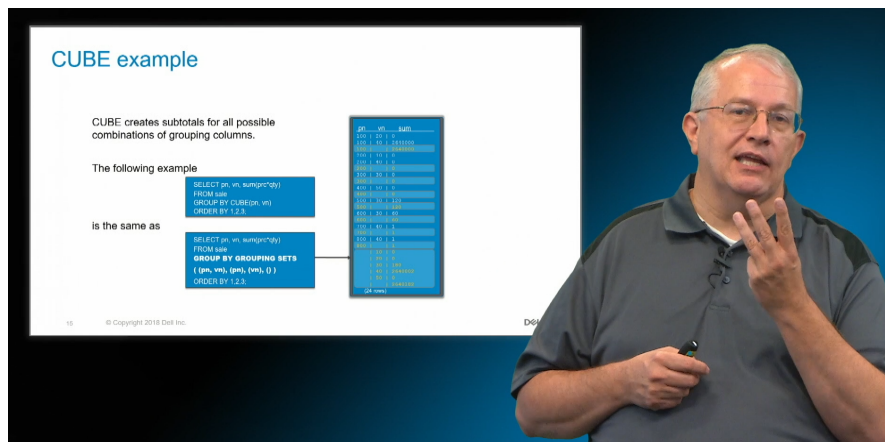
Some screen shots from the course are given below.

**Figure 2: Check Your Knowledge Section Explained by the Instructor**



Source: Dell Technologies Education Services

**Figure 3: Picture-in-Picture Format Used by the Digital Development Team**



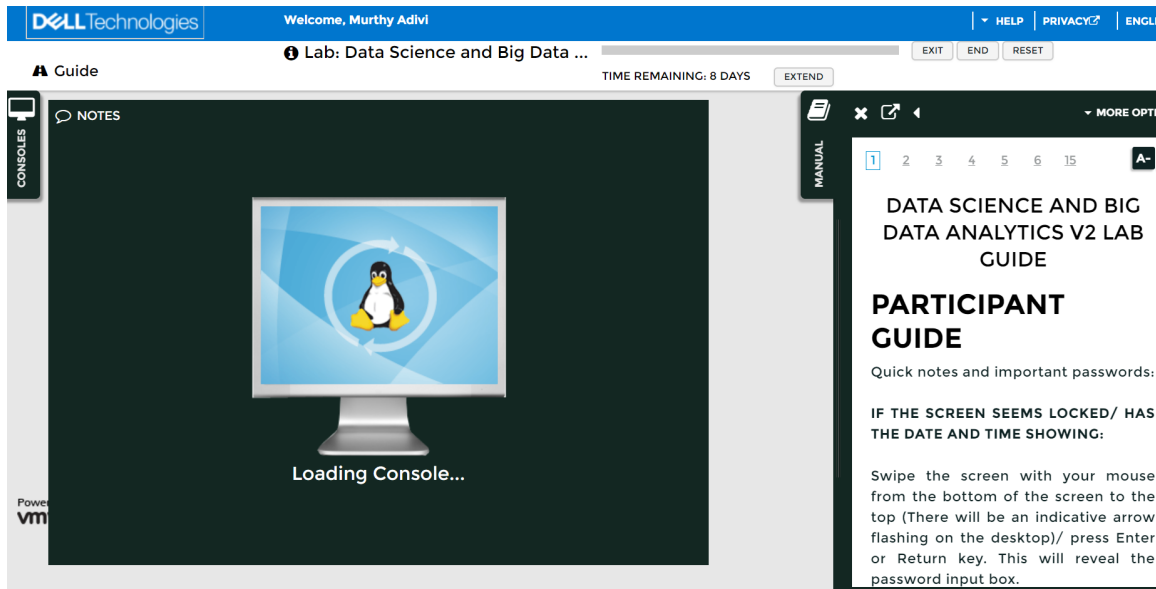
Source: Dell Technologies Education Services

3. **Virtual Instructor-Led (VILT):** The course can also be taken in the same format as ILT but via a remote instructor in a virtual classroom. The students will be taken through the course and lab exercises in the same fashion as the ILT.
4. **On-Demand Lab:** The course is primarily for engineers who have had vast exposure to statistics and who work with data analytics on a regular basis, but who wish to understand and apply advanced techniques such as text analytics, Naïve Bayes classifier, decision trees and time-series analysis along with an exposure to Hadoop environment. The On-Demand Lab provides a virtual environment hosted on VMware powered machines. The On-Demand Lab also gives the user access to the instructional videos from VILT to help the student

get up to speed. In addition, the student is permitted to write code independently and use it as a sandbox environment.

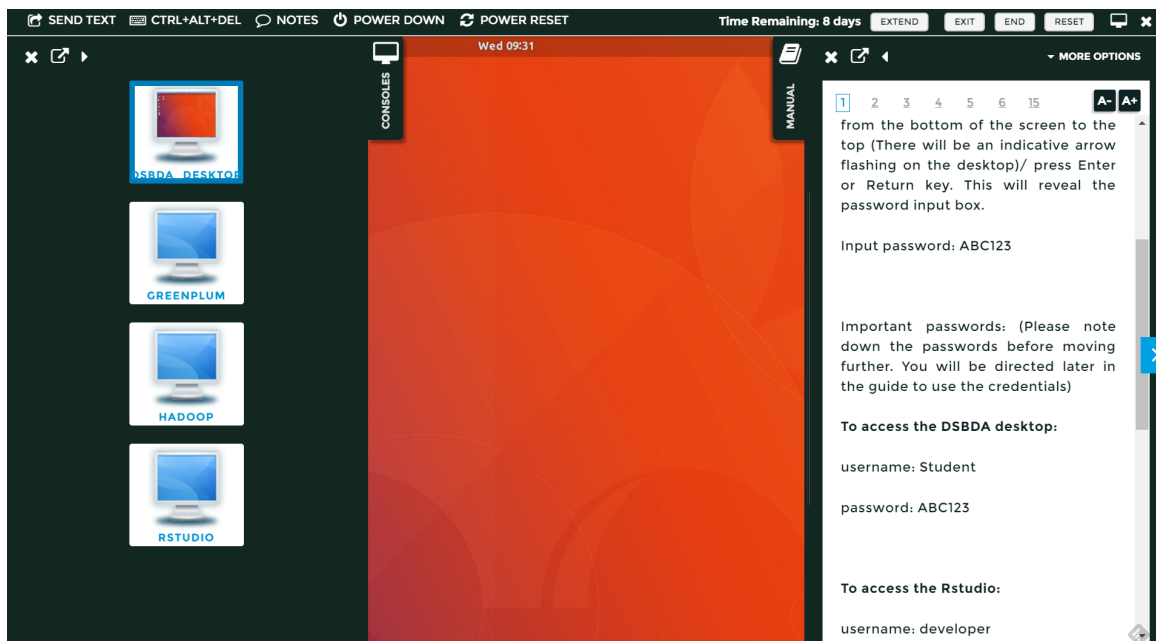
The following screen shots give a look and feel of the virtual environment where the On-Demand Lab is hosted.

**Figure 4: A Live Lab Guide Giving Step-by-Step Instructions is Included in the VLP**

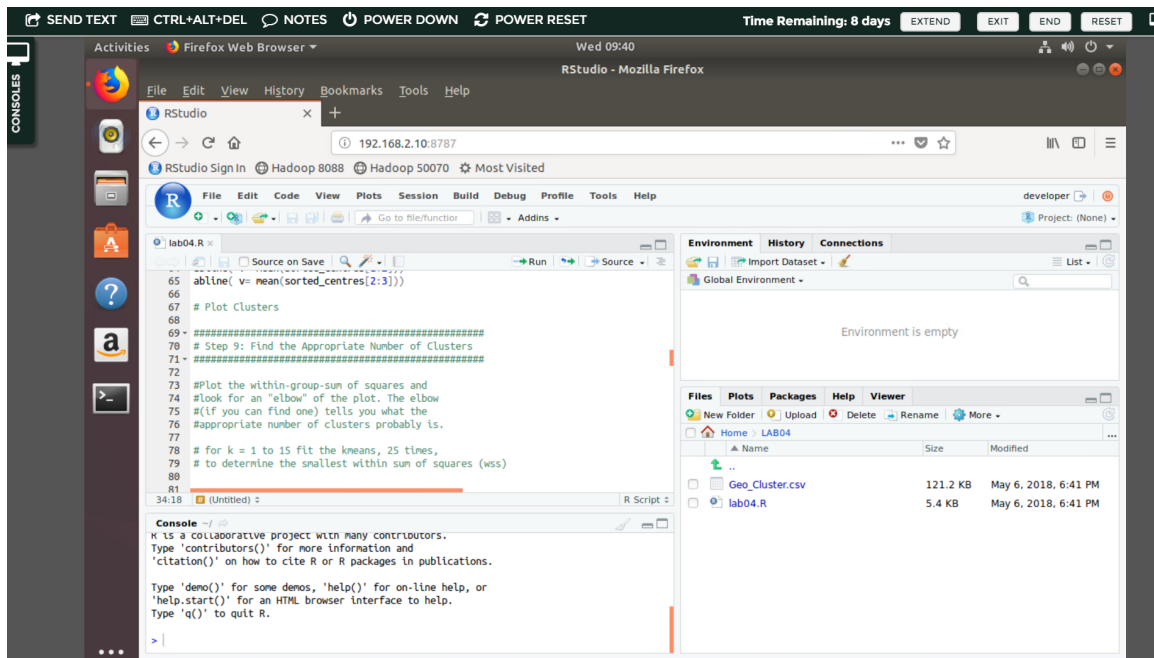


Source: Dell Technologies Education Services

**Figure 5: The Various Virtual Machines Used for the Exercises in the Lab**



Source: Dell Technologies Education Services

**Figure 6: Glimpses of the RStudio Working Environment**


Source: Dell Technologies Education Services

- 5. Certification Exam:** The certification exam is the minimum requirement on the part of a student to be certified by the “Dell Technologies Proven Professional Certification Path.” All above modalities allow the student to take a sample test and prepare themselves for the final certification. If a student is confident in their knowledge and expertise on the lab, the student is allowed to take the exam with Dell partner vendor Pearson, who will host the exam on Dell’s behalf. The exam is multiple choice, but also tests knowledge of coding. The students are also given access to a mock test to prepare them for the complexity and style of questions they could expect in the actual certification exam. The questions are updated on a regular basis to make sure some students do not have an undue advantage over others — keeping the value of certification high in the industry.

All modalities of the delivery of this course aim at teaching the student the importance of the data analytics lifecycle and that this is a core learning skill one can expect working in the field. It is necessary for the student to understand how to apply these skills in a real work environment while dealing with various stakeholders. The technical skills, however important, evolve over time and need upkeep by the students through practice and study. The core principles, statistics, algorithms and the data analytics lifecycle are tools that mold the work practices of the student with a long-term impact. That is the primary goal of this course.



## Change Management Efforts

---

### **Need to Keep Content Relevant**

Data science is a fast-changing field and the students do expect the most cutting-edge training on the same. There is a need to update and keep up with the fast-changing environment. Over the past year, several changes have been made to keep the course fresh and relevant.

### **Need to Manage Content Changes Across Modalities**

A use of Learning Content Management System (LCMS) was proposed throughout the organization and it has been implemented for DSBDA. Xyleme has been identified as the choice content management/authoring tool. It makes version control, editing of content and many other aspects of content management much easier.

### **Need for Creating Practical Application-based Learning**

There has been a push toward moving all lab exercises to a centralized virtual computing resource. A lab environment is allocated for each student to let them monitor their progress. Lab guides have been included directly into the environment, helping students to look up instructions in real-time.

The virtual environment is available to the student for up to 216 hours (or 60 days) to give them enough time to practice and get their hands-on experience with the various tools and technologies, while revising their theoretical content. These machines can also be reset to allow students to work out the exercises themselves.

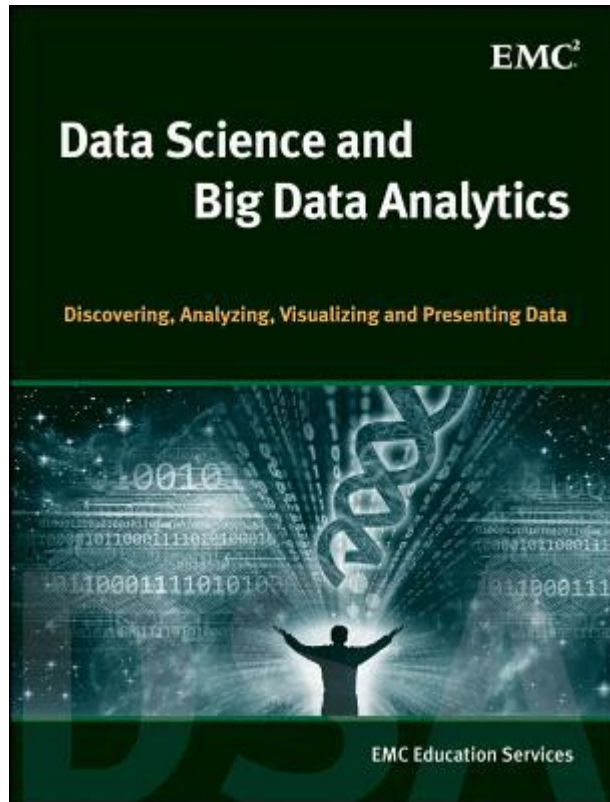
Similar exercises are done for both Associate and Advanced Methods courses every year.

## Measurable Benefits

---

The textbook that was published based on the course content has since become the industry standard for beginner and intermediate data scientists who would like to buy a reference textbook. The book comes up in most searches for data science textbooks.

Figure 7: The Course Content Published as a Leading Text/Reference Book



Source: Dell Technologies Education Services

The course is available for consumption for internal employees and provided to all employees at the beginning of their employment with Dell. There is a healthy participation from internal employees who see the value of data science and its application in their day-to-day work.

External customers have also been consistently increasing in number for all modalities of delivery. This provides a steady revenue stream for the company.

The course has been acting as a bridge for skill gap, both internally and externally, for Dell Technologies, in addition to the revenue generated on a steady basis.

As part of the Academic Alliance Initiative from what was EMC, the course, along with the virtual lab environment, is given as a package to more than 3,000 universities worldwide. The course was updated and released last fall in time for the start of college courses.

## Overall

---

During this course development, the team identified a few development opportunities:

- Importance of developing the learning content efficiently so that it takes less time for development, resulting in decreased time to market and improved pricing

- Effective use of LCMS features for content development and reusability
- Leverage the skills of developing interactive course elements to improve audience engagement
- Collaboration with industry to keep up with the fast-changing industry of data sciences and deliver skills that are most in demand

Future programs will continue to focus on specific skills needed by data engineers to do their day-to-day jobs. Dell is planning to create a highly interactive course which can be delivered in one day and will hone in on one particular skill at a time. Some of the skills Dell is focusing on are SQL queries, Hadoop management and Python programming.

## About Brandon Hall Group

---

Brandon Hall Group is an HCM research and advisory services firm that provides insights around key performance areas, including Learning and Development, Talent Management, Leadership Development, Talent Acquisition, and HR/Workforce Management. With more than 10,000 clients globally and more than 25 years of delivering world-class research and advisory services, Brandon Hall Group is focused on developing research that drives performance in emerging and large organizations, and provides strategic insights for executives and practitioners responsible for growth and business results.

### Inspiring a Better Workplace Experience

Our mission: Empower excellence in organizations around the world through our research and tools every day. At the core of our offerings is a Membership Program that combines research, benchmarking and unlimited access to data and analysts. The Membership Program offers insights and best practices to enable executives and practitioners to make the right decisions about people, processes, and systems, coalesced with analyst advisory services which aim to put the research into action in a way that is practical and efficient.

### Membership Offers Tailored Support

Our membership delivers much more than research. Membership provides you direct access to our seasoned team of thought leaders dedicated to your success, backed by a rich member community, and proactive support from our client services team.

### RESEARCH ACCESS & EVENTS

- Reports
- Case Studies, Frameworks & Tools
- DataNow® & TotalTech®
- Webinars and Research Spotlights
- Annual HCM Conference

### ADVISORY SUPPORT

- Ask the Expert
- 1 on 1 Consultations
- Research Briefings
- Benchmarking

## **CLIENT SUCCESS PLAN**

- Your Priorities
- Executive Sponsor
- Client Associate
- Monthly Meetings

## **Strategic Consulting Offers Expert Solution Development**

Our consulting draws on constantly updated research and hundreds of case studies from around the globe. We provide services that simplify and target efforts to produce business results.

## **BENCHMARKING**

- Competitive/Comparative
- Maturity Model
- Custom Research

## **STRATEGY**

- Business Case
- Planning
- Organization & Governance

## **TECHNOLOGY SELECTION**

- Vendor Selection
- Architecture Design
- Systems Evaluation

## **DEVELOPMENT & INTEGRATION**

- Program Design
- Assessment
- Survey
- Process Integration

For more information, contact us at [success@brandonhall.com](mailto:success@brandonhall.com).