HADOOP CAREERS: GETTING HIRED

Hadoop hiring & career considerations

As you *may* have heard, Hadoop and other big data professionals are in demand. The results for a Google search of "Hadoop jobs" might be large enough to qualify as big data itself. Many young and mid-career IT professionals alike are trying to ride the Hadoop elephant to better opportunities. To support this ambition, numerous training and certification programs have been created – although it must be noted that no certification is available through the Apache Hadoop foundation. IT professionals are often left wondering which skills and certifications are most valuable, while organizations with little or no big data experience wonder what specific skills and other qualities they need from their candidates.

Hadoop falls under the category of data science, but being successful with it is really more of an art. Scientific discovery is about producing repeatable results by applying the scientific method. Hadoop projects don't always follow established processes because today Hadoop is often used to try things that haven't been done before. That is why Hadoop success is part art and part science. At a minimum you'll need the technical knowledge to extract and process data, but the art is in finding new sources to extract data from, and new ways to process and express it.

Hadoop job titles and roles

Hadoop has similarities with other environments in that people are needed to develop applications, interpret the output, and manage the environment. In the Hadoop world, these roles are designated by titles like:

- Big data analyst
- Big data software engineer
- Big data architect
- BI specialist
- Data engineer
- Data scientist
- ETL developer
- Hadoop architect
- Hadoop developer

Some of the most commonly sought-after skills and certifications for these positions include:

- Experience with the Apache Hadoop stack, which includes HDFS, YARN, HBase, Hive, Pig, Spark, etc.
- Cloudera Certified Professional (CCP) certifications
- Hortonworks certifications, for example, HDPCD, HDPCA, et al
- MapR certifications – (MCHA, MCHD, MCHBD, MCSD)
- Shell scripting experience (especially Python, R, Unix shell and PowerShell)
- SQL/NoSQL experience

Individual organizations will likely only require certification by one of the principle Hadoop distribution vendors (Cloudera, Hortonworks, or MapR).

The Hadoop certification landscape

Hadoop Certifications >

Considerations for non-technical skills

Finding a candidate that checks all the desired boxes for desired certifications, skills, and experience is no guarantee of success. Just like big data development itself, finding the right people for roles might be more art than science. Big data is about finding context and business value. It takes technical skill to create the Hadoop environment to process big data, but a different skillset to make it meaningful and actionable. If you are hiring, it is important to consider the business acumen of professionals on the big data team, even those in technical roles. If you are a candidate, you should be prepared to discuss how you can apply Hadoop/big data to help the business solve its problems and find new opportunities.

Flexibility is another important trait for Hadoop professionals. The Hadoop environment is growing and changing quickly. New tools and techniques are frequently introduced, and may enable Hadoop to do things it wasn't able to do before. Individuals that prefer working in a very structured, unchanging role may not be comfortable in this environment. On the other hand, individuals who do not need close supervision, and like being able to explore new technologies, and recommend new projects and approaches to their organization, may thrive in the fast-moving Hadoop world.

Here are some other observations about swimming in the Hadoop talent pool:

- It may be very difficult to find a candidate that has the specific skills an organization seeks and experience in its industry. Business acumen is more important for data scientist and advisory roles than for development and maintenance staff. Hadoop technical skills are transferrable among industries.
- Hadoop specialists that join organizations may know more about big data than the people who hired and will manage them. All parties should be aware of this possibility and be comfortable with it. For organizations with limited big data experience, it is important to hire candidates that communicate well and can present ideas and requests to make their case from a business, not overly technical, perspective.

Ideas for successful interviews

There is a lot of discussion and debate within the Hadoop community about the relative merits of the various components – for example, what's better – Pig, Spark or Hive? When should an organization use MapReduce and when should it consider alternatives? And so on. The interviewer and the candidate should each be prepared for such questions. In many cases there is no clear right or wrong answer, so an effective answer is one where the candidate can clearly and persuasively explain why he or she favors one option. Commenting on specific circumstances that favor each option demonstrates some understanding of business requirements and suggests the candidate is flexible. Similarly, be prepared to talk about different hardware and configuration options for the Hadoop infrastructure.

Candidates shouldn't be afraid to ask questions. Asking why the Hadoop infrastructure is being configured in a certain way, why different methods and technologies were selected, how workloads are being constructed, etc. will show that you will not just maintain a system, but are thinking about ways it could be made better to help the business.

Candidates should probe to see if the prospective employer will support their professional development, for example by sending the candidate to conferences, paying for certification courses,

etc. Need to get a list of questions prepared, or prepare for your Hadoop interview? Check out this list of top [50 Hadoop interview questions](#).