softtech group

BIG DATA Testing

Data Sheet

Big Data Testing

BigData testing is known as the testing BigData applications. Big data is a series of massive databases, which cannot be processed using conventional computing techniques. Testing of such datasets requires different computing methods, techniques, and frameworks. Big data refers to the production, storage, retrieval and analysis of data that is exceptional in terms of volume, variety and speed.

Big Data Testing Strategy

Instead of checking the individual features of the software product, Big Data application is more a verification of its data processing. The secret is accuracy and functional testing when it comes to broad data testing. In large data testing, QA engineers verify that data terabytes are successfully processed via commodity clusters and other supportive components. It needs a high degree of testing, as processing is very fast. There may be three forms of processing:

- Batch
- Real Time
- Interactive

How to test Hadoop Applications

Big data testing can be split widely into three stages

1. **Data Staging Validation**

   The first phase in big data testing often referred to as the pre-Hadoop stage includes the validation in processes.

   - ✓ To ensure that correct data is pulled into the framework, data from different sources such as RDBMS, weblogs, social media, etc. should be checked.
   - ✓ Comparison of the source data with the details forced into the Hadoop framework to ensure that it fits
   - ✓ Verify that the correct data is extracted and loaded into the correct position for HDFS
   - ✓ For data staging validation, tools such as Talend, Datameer, can be used.

2. **MapReduce Validation**

   The validation of "MapReduce" is the second step. At this point, the tester verifies the validation of business logic on each node and then validates it after running against multiple nodes, ensuring that the node is validated.

   - ✓ Chart Process Reduction works correctly
   - ✓ Data aggregation or segregation laws apply to the data
   - ✓ Key values are generated in pairs
   - ✓ Validation of the data after processing Map Reduce

3. **Output Validation Phase**

The process of performance validation is the final or third stage of Big Data research. Based on the requirement, the output data files are created and ready to be transferred to an EDW (Enterprise Data Warehouse) or any other device.

Third-stage activities include:

- ✓ To ensure the transformation rules are being applied correctly
- ✓ Checking the integrity of the data and the effective load of data into the target system
- ✓ Compare the target data with the HDFS file system data system to verify that there is no data corruption

Challenges in Big Data Testing

The Automation

Big data automation research needs someone who has professional knowledge. Automated methods are often not designed to deal with unforeseen issues that occur during research.

Virtualizing

It is one of the main stages of research. In real-time, big data research, virtual machine latency causes timing issues. Managing images in Big Data is also a concern.

Wide collection of data

- **You need to check more data and do so more quickly.**
- **Need to automate the effort for testing**
- **The need to be able to test through multiple platforms**