



Easily Leverage the Power of Hadoop with AppSymphony

The Promise of Hadoop

The potential benefits of the Hadoop ecosystem are well known and include improved information timeliness, improved data processing robustness and the ability to scale processing as data volumes increase. Companies such as Cloudera, Hortonworks, and MapR, as well as the Apache Foundation have done an excellent job educating companies about these benefits. They have also made it extremely easy to install and manage a Hadoop environment within a corporate data center. For organizations that do not want to manage their own Hadoop environment, Amazon's Elastic Map Reduce and Azure's HDInsight provide easy access to a Hadoop environment in a public cloud offering. Unfortunately, none of these Hadoop platform providers have made it easy for a company to develop and run Hadoop analytics. In fact, many organizations that have stood up Hadoop environments are failing to see a positive return on investment because they don't have a suite of tools to effectively leverage their initial investment in the Hadoop platform. The companies that are successfully extracting value from the Hadoop ecosystem are usually giant technology companies like Facebook and Google who have thousands of software developers. To successfully leverage Hadoop, companies typically attempt to hire software developers that are also competent data scientists as well as experts in the modern programming paradigms that Hadoop supports. There are simply not enough software developers with these skills and the competition for them is too great. A different approach and a different set of tools are needed to better extract value from an initial Hadoop platform investment.

AppSymphony and Hadoop

AppSymphony enables companies to realize the benefits of Hadoop with their current workforce by empowering analysts and data scientists to build new Hadoop analytics without writing a single line of new computer code. They can leverage their company's existing proprietary algorithms, even if these algorithms were not written to be used in the Hadoop ecosystem. AppSymphony provides them with the ability to access and process data from a wide array of the open-source high-performance compute and storage ecosystem including Hadoop HDFS, Hadoop Map Reduce, Apache Accumulo, MongoDB, Neo4j and others.

Graphical Composition

Without AppSymphony, developers must develop code to implement mapper and reducer classes, write custom software to access files in HDFS or the variety of other compatible Hadoop data sources, and integrate that new, custom software with the larger Hadoop framework. With AppSymphony, users employ an intuitive graphical composition canvas to connect existing processing components into executable analytics. This capability enables users to rapidly compose simple but powerful new analytic capabilities without requiring software development skills. Figure 1 shows a workflow where the user describes her data set location, executes a map reduce job against that data set, and displays the result. In this case the end result is a plot showing the Twitter users who have had the biggest increase in followers in recent days.

Optensity

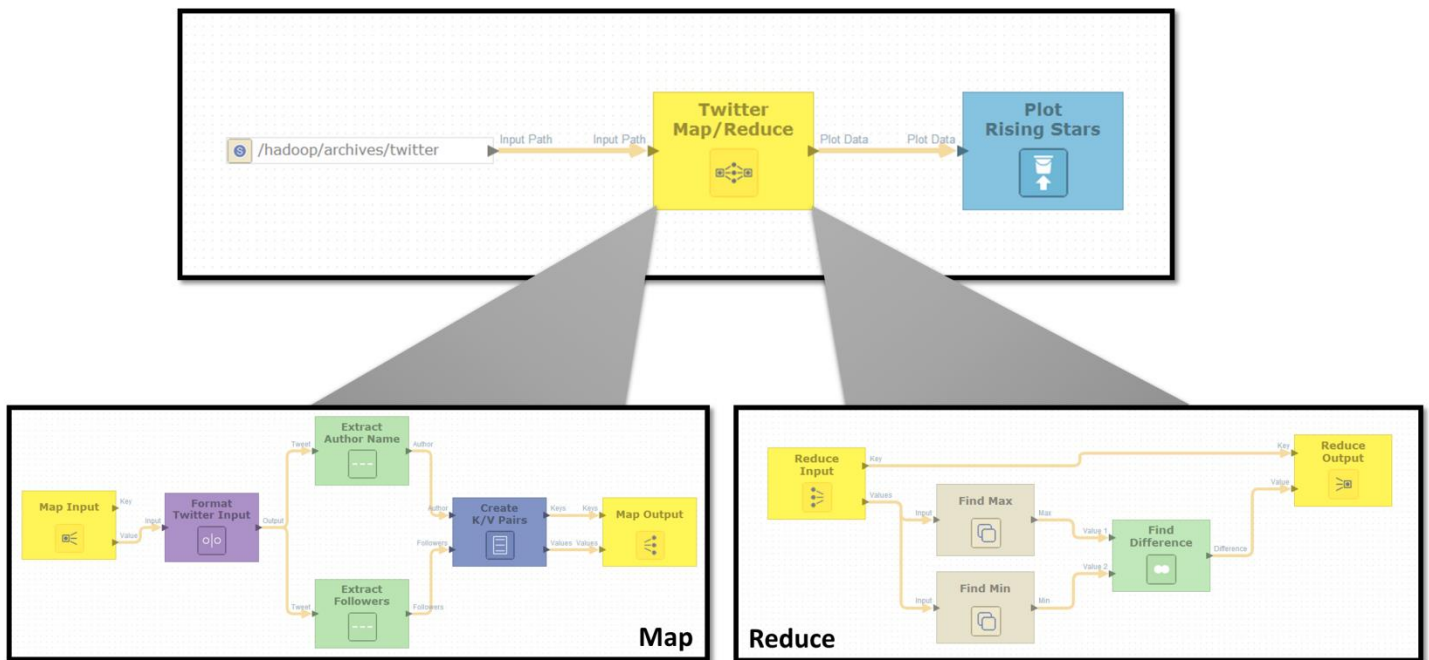


Figure 1- Twitter Rising Star Map Reduce App

Inside the Map Reduce component, AppSymphony presents the analyst with a separate composition canvas for the Map and Reduce phases of execution. These canvases are pre-populated with input and output components that provide adapters to HDFS, Accumulo, and many other high-capacity data stores that support Hadoop Map Reduce processing. The analyst or data scientist is free to place any components between these adapters to process the data without worrying if the components are built specifically for Hadoop. The data scientist does not have to write Mapper and Reducer classes in Java or any other computer code, she simply graphically composes how the data should be processed. This approach substantially enhances agility by reducing the time needed to validate new algorithms and gain new insights from existing algorithms that were previously unable to deal with large data volumes.

Leverage existing analytics

AppSymphony comes with a large set of components so analysts can immediately begin doing useful work. Because most companies have made extensive investments in proprietary data processing algorithms, AppSymphony allows those to be incorporated as components as well. AppSymphony has a component development kit which is an Eclipse-based wizard that walks data scientists and programmers through the steps needed to create a component from an existing algorithm. The existing algorithm can be written in a variety of languages including JAVA, C, C++, Python, Perl, and others. The entire process to create a new component can be completed in minutes by a user with no Hadoop skills. Once the component is created, it can be leveraged to compose AppSymphony Hadoop apps even if the original algorithm was not built to be used in a Hadoop environment.

Full Hadoop Ecosystem

AppSymphony provides highly efficient access to the Hadoop ecosystem through our adapters for a variety of Hadoop environments. AppSymphony users can copy data into and out of HDFS as well as a variety of Hadoop-capable databases including Hive, Accumulo, HBase, and MongoDB. For example, analysts only need to replace a single graphical component in order to run their Map Reduce analytic on

data stored in Accumulo instead of HDFS. AppSymphony also allows users to execute Map Reduce analytics against commercially available data stores such as Amazon's S3 and Azure's Blob storage.

Demonstration

Please contact Scott Zimmer via email szimmer@optensity.com or phone (512)299-7218.