
MAIN POINTS

Lesson 6 & 7 – Apache Pig

Do less and accomplish more

WHOLENESS OF THE LESSON

Apache Pig, a data flow language, is developed on top of MapReduce by Yahoo. Virtually all parts of the processing path in Pig are customizable: loading, storing, filtering, grouping, and joining can all be altered by user-defined functions (UDFs). This abstract layer of Pig on top of MR adds ease of development. **Science & Technology of Consciousness: Physics** describes creation as built up of layers, from the innermost subtle levels to the outer levels of matters. Beneath the subtlest layer is the abstract, unmanifest field of pure Being.

MAIN POINTS

1. Pig has a rich collection of operators in order to perform operations like join, filter, and sort. Using these direct operators, Pig extracts the huge data set, performs operations on them and dumps the data in the required format in HDFS (ETL). **Science & Technology of Consciousness: This is exactly in accord with Maharishi's principle of "Do less and accomplish more" which only applies when the action is taken from the level of transcendental consciousness.**
2. In Pig, no operation is evaluated until the transformed data is required. The actual processing of data starts when we submit the dump or store commands. Traditionally, while churning data we like to see transformed data at every step but while dealing with big data a lot of time would be wasted in repeated calculations to transform the data. This Lazy evaluation in pig reduces the repeated calculations, hence improves performance. **Science & Technology of Consciousness: Rest and activity are the steps of progress.**
3. It is possible to develop algorithms in MapReduce that cannot be done easily in Pig. And the developer gives up a level of control. So, for less common algorithms or extremely performance-sensitive ones, MapReduce is still the right choice. **Science & Technology of Consciousness: Deeper levels of intelligence enable one to have greater insight and greater mastery over the more expressed values of life.**

UNITY CHART

CONNECTING THE PARTS OF KNOWLEDGE WITH THE WHOLENESS OF KNOWLEDGE:

Building High-Level Dataflows over Map-Reduce

1. Development cycle of MR is very long and so for productivity, we need higher level tools like Pig.
 2. Virtually all parts of the processing path in Pig are customizable: loading, storing, filtering, grouping, and joining can all be altered by user-defined functions (UDFs).
-
-

3. Transcendental consciousness: is the experience of the simplest and most abstract state of awareness which underlies all states of greater excitation.
4. Impulses within the Transcendental Field: Nature accomplishes what it needs by having its impulses in the transcendental field be as efficient as possible.
5. Wholeness moving within itself: In unity consciousness one experiences that all layers of the universe are only different expressions of the same infinite field of pure consciousness.

