

Decooda's Liquid Data Platform High Level Overview of Decooda VS Spark

Charles Wardell
Chief Technology Officer



ABSTRACT

Creating solutions using the [Decooda Liquid Data Platform](#) is a very different approach to processing data when compared to an Apache Storm implementation. This paper is just a highlights of the core differences between the approaches of the two technologies.

INTRODUCTION

Decooda Liquid Data Platform is a Massive Parallel Processing (MPP) platform designed to support the processing and analysis of extremely large data sets.

We categorize big data as any data that is either too large and/or too complex and/or too expensive to process using existing technologies and architectures.

Deployed on a cluster of commodity off the shelf hardware (COTS), Decooda Liquid Data Platform provides a cost effective and simple way to get massive parallelism with minimal effort.

USE CASES

- Real-time / Continuous / Stream Based Processing
- iSOA - Intelligent Parallelized SOA and RPC
- Big Data Processing

- Deep Dive AnalyticsMauris pharetra et purus mollis ornare.

A DIFFERENT APPROACH

The Decooda Liquid Data Platform can be considered a generalized utility when compared to the purpose-built solutions created with competing big data platforms on the market today.

When creating a solution for Storm, Spark, Hadoop, or HPCC, the developer is required to write a program, compile the program, and deploy the program as a purpose built job or solution. For example, if one were to deploy a streaming solution that performs a word count, that solution will always perform word counts. In order to add additional functionality (IE: Language detection), one either needs to:

1. Modify the word count program, compile it, and redeploy -OR-
2. Create a secondary solution, deploy it and customize the first solution to communicate with the second.

The first approach has the downside of a single monolithic program that lacks reusability. Since the workflow is hard coded into the solution itself, it has inherent dependencies.

In the second case, there is a benefit of having two stand-alone solutions but there is also a



drawback. One is forced to create a dependency by modifying the first solution to call or integrate with the second solution.

Decooda breaks this paradigm by being the first true **Big Data Platform as a Service. (BDPAAS)**. We separate the logic from the big data platform itself. This approach allows a single deploy of our big data platform to service requests from an unlimited number of vertical domains. A few notable benefits to this approach are:

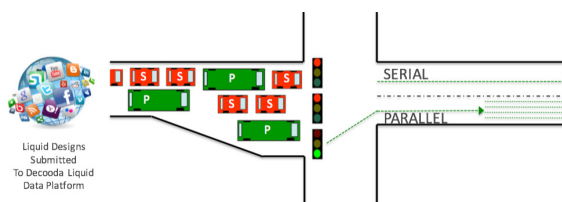
- A single investment in a BDPAAS can be leveraged across multiple domains and multiple companies. Decooda's Liquid Data Platform can essentially become one of the largest cloud-based big data processing platforms in the market today.
- Business logic is coded in the form of reusable plugins which we call droplets. These droplets can be shared and utilized across verticals.
- These individual droplets operate as atomic "black box" functional components. The droplets are designed to do one thing and one thing only. As a result, the droplets are highly reusable and shareable across domains. Additionally, the droplets can be monetizable in the form of an app store for big data processing.

There are only three components to designing atomic or complex multistep systems.

- **Decooda's Liquid Data Platform** operating as a big data platform as a service.(BDPAAS) -This is essentially the infrastructure that waits for jobs to be submitted in the form of Liquid Designs.
- **Droplets (AKA: Plug-ins):** Atomic "black box" functional components that are developed for a specific purpose. Essentially, droplets are the unit of logic. Droplets can be persistent in memory for faster execution or non-persistent and referenced as needed. Droplets are loosely coupled with

the Decooda Liquid Data Platform and are the basis for our component Flow-Based Programming paradigm.

- **Liquid Design Process Definition:** A Liquid Design is an XML representation of the data flow. The Liquid Design Process describes the sequential and parallel execution of how and when the droplets are invoked in the workflow.



Messages are sent to the Decooda Liquid Data Platform in the form of a Liquid Design Process Definition. The Process Definition contains not only the process definition but the payload on which the sequential and parallel steps operate. When the data is not present within the message ,a reference or pointer to the data is passed along with the Process Definition to minimize the movement of information packets throughout the system.

By approaching the architecture in this way, the Decooda Liquid Data Platform becomes an effective utility that allows a single grid to service multiple solutions, domains, and companies.

PERFORMANCE

Decooda's MPP approach is to automatically distribute and parallelize the execution of all Liquid Designs (process definitions) evenly across all available nodes in a cluster.

In an attempt to show that a true dynamic Big Data Platform as a Service (BDPAAS) can perform just as well as a purpose-built solution, we have

implemented a simple parallelized word count in both Decooda and Spark.

In order to compare apples to apples of raw processing performance, the comparison below utilizes a disk based file as its source.

	Decooda	Spark
100 MB File	15s	10s
500 MB File	32s	48s
1 GB File	61s	94s

Decooda -VS- Spark

Parallelized Word Count against Files

The above chart shows that the best of both worlds can be achieved. Decooda's Liquid Data Platform archives the performance of an optimized purpose-built solution while having flexibility that other Big Data solutions are unable to support. It should also be noted that the Decooda approach is unoptimized. We have not applied the optimization opportunities that our platform supports.

Likewise, in a simple comparison between Decooda and Hadoop, the below chart shows that the streaming approach that Decooda's Liquid Data Platform is far more efficient than the traditional Map/Reduce approach.

	Decooda	Hadoop
4 GB File	67s	199s

Decooda - VS - Hadoop

Single Node Word Count against 4GB File

FAULT TOLERANCE

Fault Tolerance on Spark is achieved only when it is set up on a fault tolerant underlying system like Hadoop Distributed File System (HDFS). Failure of a worker node or failure of a driver node will result in data loss.

Decooda's Liquid Data Platform is storage agnostic. Any node can fail within a Decooda Liquid Data Platform implementation without disrupting the process flow. Messages sent to the grid can be sent in a guaranteed messaging mode requiring an acknowledgement (ACK) upon completion. There is no single point of failure as found with Hadoop's name node and there is no dependency on an underlying system to ensure that fault tolerance is achieved.

COMPUTE INTENSIVE SCALE

Spark does not scale well on compute intensive jobs unless you know the intricate details and know exactly which knobs to turn and there are many.

Decooda's Liquid Data Platform can perform sequential as well as paralyzed tasks efficiently. Sometimes, one might want to avoid the distributed nature of computation because the network latency and overhead degrades the performance that might be better achieved on a single server with a simple parallel and multi-core approach. Decooda supports both approaches.

GENERAL SCALE

The Decooda Liquid Data Platform can scale horizontally across an unlimited number of commodity servers. There is no head node or named node that could be considered a single point of failure.

Furthermore, when operating in location transparency enabled through high-speed messaging, the Decooda Liquid Data Platform allows clusters in different parts of the world to operate as one seamless unified grid.

DECOODA LIQUID DATA PLATFORM VISION

Decooda's Liquid Data Platform is currently used by Decooda to glean insights from raw unstructured text. We have the foundation for something greater and to date, we have only extended our platform on a need by need basis. With that said, there is a much greater vision.

Imagine a single dynamic Big Data Platform as a Service (BDPAAS) where liquid designs from 1000s of companies across many domains are submitted for processing. One request can be processing sentiment analysis while another request is performing DNA blood protein folding.

Imagine an app store of reusable droplets (aka: plug-ins) where solutions can be built just by dragging and dropping the droplets in a Visio style workflow.

At Decooda, we are focused in creating a vertically integrated streaming cognitive analytics platform. In doing so, a byproduct of this development effort brought forth an innovation in how big data is processed. The Decooda Liquid Data Platform has merit and should be advanced as time and funding allow.

ROADMAP

- Droplet Designer - Eclipse based workbench so the community can develop, code, debug and deploy highly reusable units of logic.
- Liquid Designer - Eclipse based workbench so the community can develop workflows

that are consumed by our Big Data Platform as a Service (BDPAAS)

- LiquidDB - A dynamic in memory database that will increase the performance of batch processing exponentially.
- Droplet App Store - A place where developers can submit highly reusable units of logic. (Machine Learning, Text Mining, etc..) for monetization opportunities.
- Liquid GraphDB - A dynamic in-memory graph database. Leveraging an existing graph database, Decooda's Liquid Data Platform can populate the graph database on the fly. This will essentially bypass the limitation of nodes and edges as the graph is populated with only the relevant information when needed.

CONCLUSION

The core of our differentiation is our Liquid Design approach to creating vertical solutions. A Liquid Design is a paradigm that allows us to create "application systems" by networking and executing atomic and monolithic "black box" processes. Liquid Processing intelligently interprets Liquid Designs and optimizes the parallel execution of jobs and atomic level tasks - massive parallelism that exploits contemporary architectures.

Decooda's Liquid Data Platform is remarkably simple to deploy, maintain, and use. Our Liquid Design which has an inherent Flow-Based Programming Model (FBM) allows you to obtain high reusability across problem domains, while achieving better performance, higher flexibility and lower total cost of ownership (TCO).

About Decooda

Decooda delivers the facts that inspire every level of an organization to embrace a culture of aligning to the needs of the customer to understand each customer's state-of-mind in real-time.



© 2017 Decooda International Inc. All rights reserved. Decooda®, the Decooda logo, and the names and marks associated with Decooda's products are trademarks of Decooda and / or its affiliates.

DECOODA.COM