Project 1: **Big Data Workflows: Locality-Aware Orchestration Using Software Containers**

They have proposed architecture for managing and executing big data workflows is structured around three main layers, each with distinct responsibilities

(Control layer, Data layer , and Compute layer).

1. Control Layer - This layer Manages the execution of workflows based on their definitions. Ensures correct step sequencing, processes data correctly, and coordinates the execution of workflows.
2. Data Layer- This layer Handles all aspects of data storage, retrieval, and movement between hosts.
3. Compute Layer- This layer Contains the processing logic for workflow steps. like Multiple compute steps, each performing specific processing tasks within the workflow.

In Data Processing They have used the parallel data processing

Parallel Processing: Each data unit is processed independently, allowing multiple units to be processed simultaneously across different computer steps.

Also they have used Centralized Architecture which helped them in Data -Locality and Simplified Management.

Paper 2: **BDPS: An Efficient Spark-Based Big Data Processing Scheme for Cloud Fog-IoT Orchestration**

This paper introduces a novel framework BDPS (Big Data Processing Scheme), designed to enhance data processing efficiency in Cloud-Fog-IoT ecosystems using Apache Spark's Resilient Distributed Datasets (RDDs). The authors address the challenges of high latency and inefficient data management in IoT networks by optimizing data routing and processing through in-memory computing and advanced algorithms.

The authors present BDPS, a Spark-based framework tailored for decentralized data processing across cloud and fog networks, significantly improving data delivery times and reducing network overhead.

The use of depth-first search-based shortest path algorithms in BDPS for efficient data routing is an innovative approach that outperforms traditional algorithms like Bellman-Ford, Floyd-Marshall, Dijkstra, and Hadoop’s map-reduce in terms of latency and resource management.

The paper provides a detailed performance analysis, demonstrating BDPS’s superiority in handling real-time data processing needs within IoT networks compared to existing methods.

Paper 3: