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.