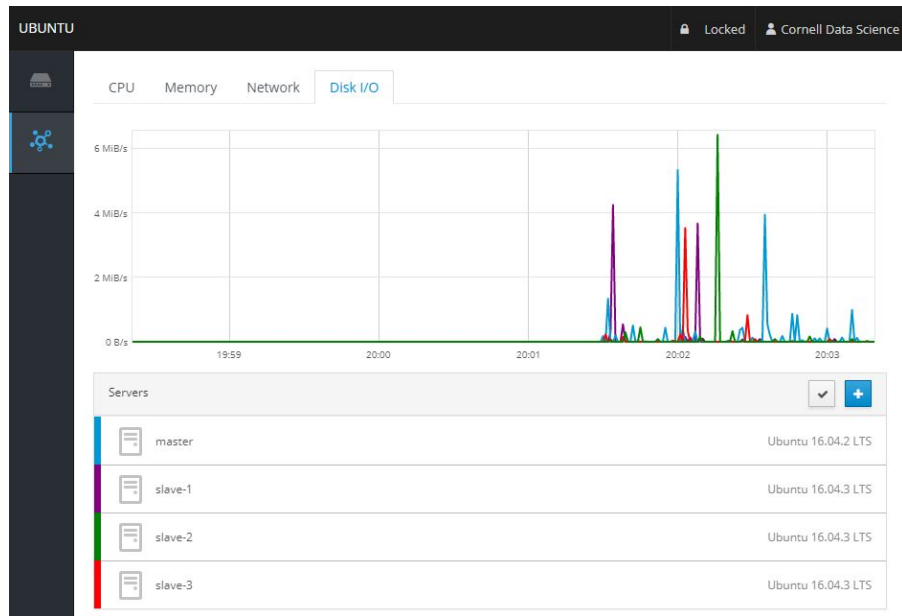


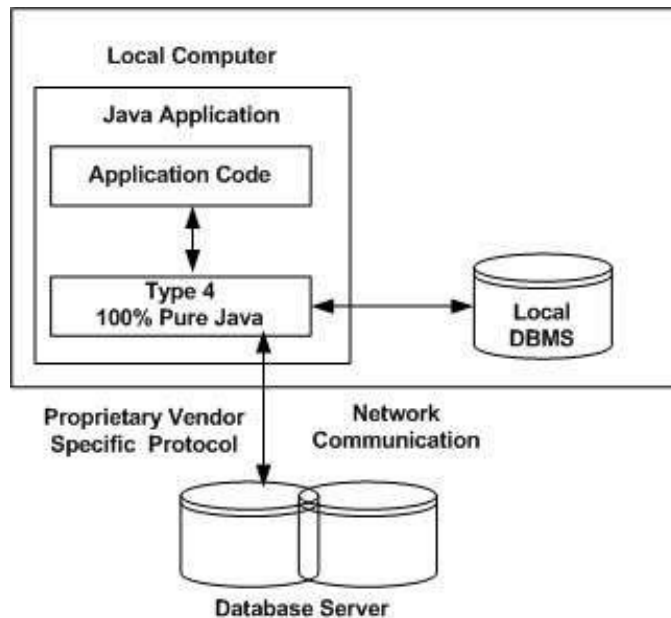
Server Administration

- Cgroups
 - Resource limits for a group of processes
 - Granular control
 - Set up to prioritize Spark processes
- Cockpit to monitor servers (CPU, Memory, Disk I/O, Logs)
 - Integratable in website



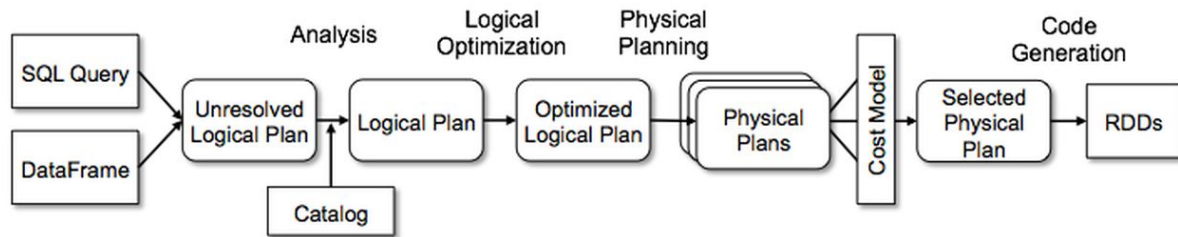
MySQL on Spark & JDBC

- Determined how to link mySQL to Spark
 - 4 Types of JDBC (Java Database Connectivity) Driver
 - I - JDBC-ODBC Bridge
 - II - Native API
 - III - Net-Pure Java
 - IV - Pure Java (Direct) - Best for connecting to single type of database
- Query caching greatly improves performance for SELECT operations
- 3 Cache levels
 - Spark query caching
 - MySQL query caching
 - Database engine caching



Catalyst Optimizer

- Underlying optimizer for data transformations
- Pattern matching based rules and strategies
- Can implement new rules/strategies to speed up transformations



Steps of the Catalyst Optimizer