# Optimizing query execution time in the LHCb Bookkeeping System

The LHCb Bookkeeping System is a metadata and provenance catalogue for files and jobs, which are used for defining datasets. It is based on an **Oracle** RDBMS warehouse.

## Partition pruning and partition wise joins

### Why?

- The number of datasets is large and rapidly growing at almost double size each year.
- The partitioning of big tables is essential for scalability, manageability and performance.

### What Is it?

**Partition pruning** is a technique used to improve the performance of the database, when the tables are partitioned. The optimizer analyzes the **FROM** and **WHERE** clauses of the SQL statement in order to eliminate the partitions which are not relevant to the statement.

**Partition-Wise joins** is a query optimization technique, which reduces the query response time by minimizing the data read from the cache/memory. It is used when **both tables** are partitioned on the **join key**.

## Current partition schema:

-Range partitioning of *jobs* and *files* tables based on the **jobid** column.

**Drawback**:

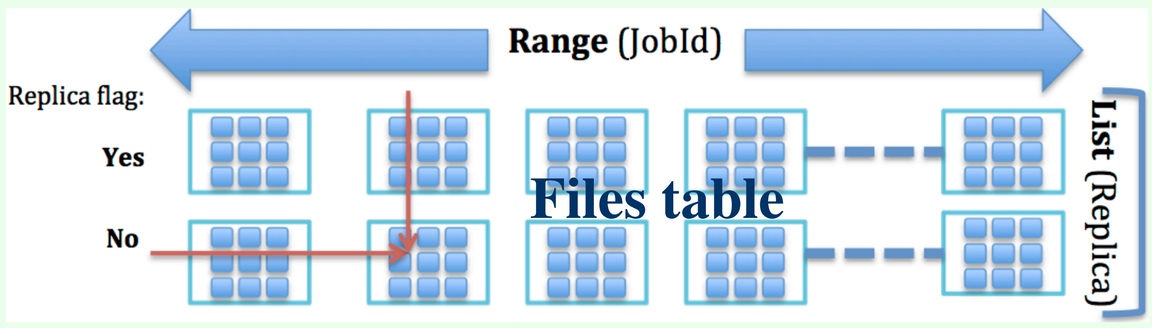
-If the **jobid** is not in the **FROM** and **WHERE** clauses, partition pruning is only used when the two tables are joined together.

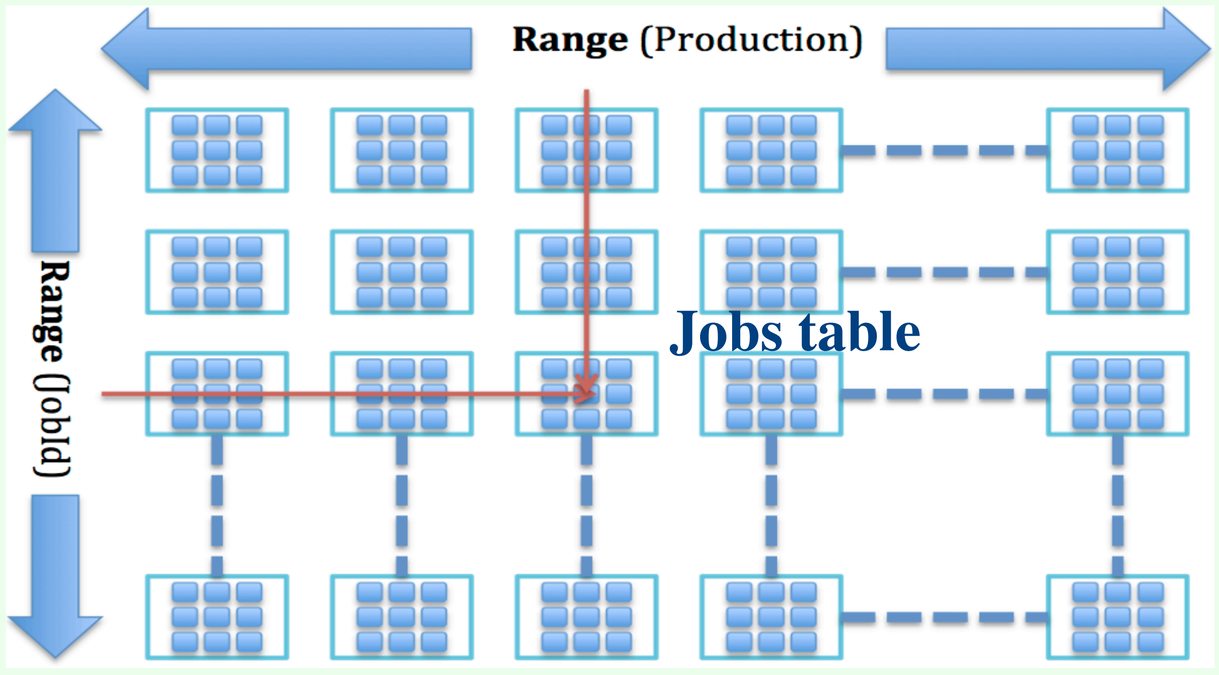

***jobs* produce and process *files* that may have a *replica* or not**

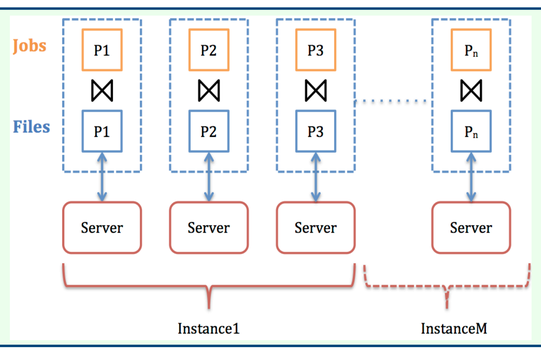
## New Partition schema:

**Composite** partitioning:

- Range-Range partitioning of the *jobs* table on the **production** (collection of jobs) and **jobid** columns;
- Range-List partition of the *files* table on the **jobid** and **replica** (flags physical existence of files) columns




## Full partition wise join parallel

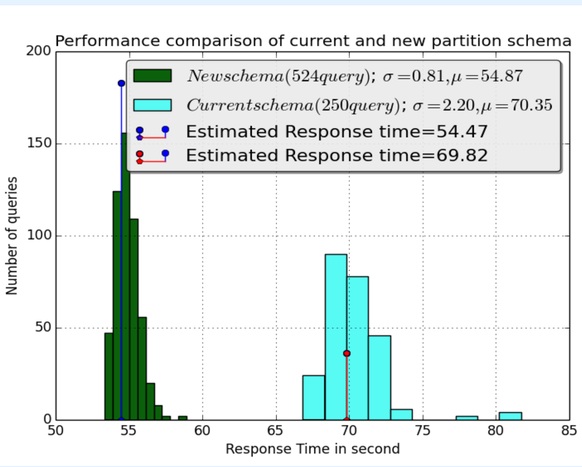
When the *jobs* and *files* tables are joined together and the parallel execution is allowed, then the large query divided into smaller joins. Each pair of **partition** ($P_1$,$P_2$, …$P_n$) from the two joined table are executed by different processes called **server**s.
Each server belongs to a database **instance.**

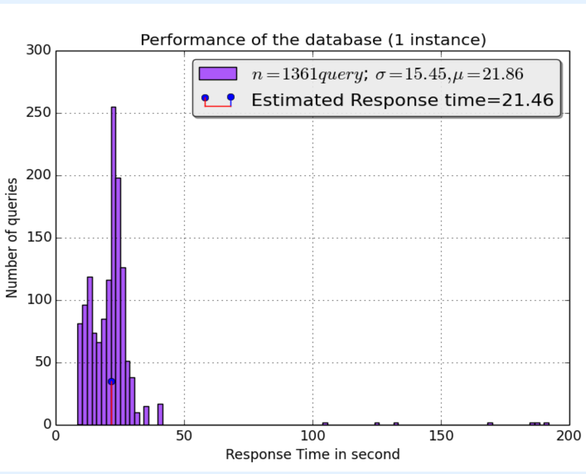
## Test Conditions:

- We used the same query in our tests. The query is executed **serial** and **parallel** during 1 hour.
- We measured the **Response Time** (Response Time = Latency + Execution Time) and number of queries using 1 and 2 database instances.
- 2 node cluster each node had 4 x Quad Core Intel E5630 @ 2.53 CPU, 48 GB of RAM and 36 SATA 7200 RPM 2 TB disks configured with Oracle ASM as RAID10

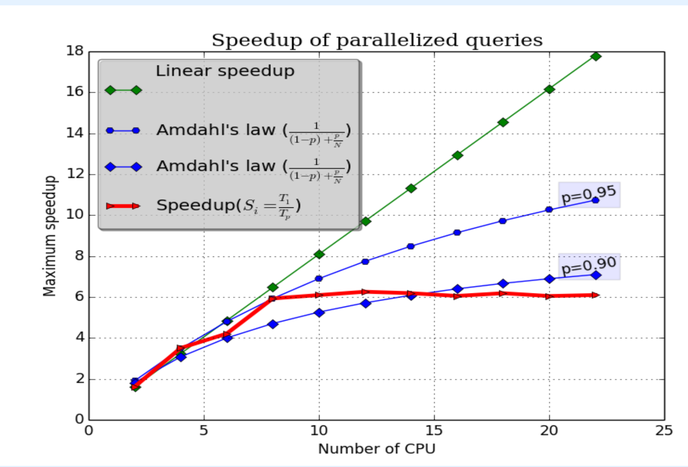
| Table Name | Number of rows |
|---|---|
| files | 154M |
| Jobs<br>productions | 69M<br>25k |

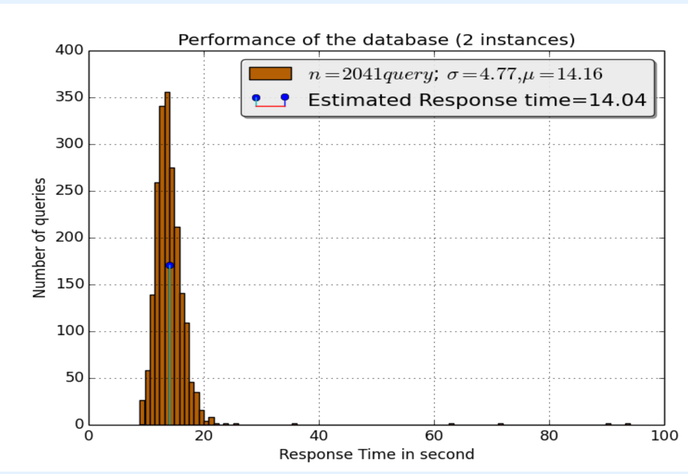
## Results: a. Serial execution




## b. Parallel execution (New Partition schema)




*Z.Mathe and P. Charpentier on behalf of the LHCb Distributed Computing Team*