Show me all records which have been accessed by Peter Suspicious between January 10 and February 15.

Data Warehouse
- Long time before first result (schema creation, data loading)
- Manual preparation work
- Difficult to distribute
+ Fast query execution (milliseconds to seconds)
+ Support for complex queries

Hadoop
- New and unfamiliar programming model
- Slow query execution (minutes)
+ No preparation time required
+ Inherent distribution and parallelization
+ Scalable

Xadoop

Experiments
Our experiments show that the system scales out when adding more nodes to the cluster, and performance does not degrade when processing bigger data sets.

<table>
<thead>
<tr>
<th></th>
<th>30 Gb</th>
<th>30 Gb</th>
<th>30 Gb</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Workers</td>
<td>19m 00s</td>
<td>40m 30s</td>
<td>59m 20s</td>
</tr>
<tr>
<td>5 Workers</td>
<td>11m 05s</td>
<td>26m 20s</td>
<td>38m 10s</td>
</tr>
</tbody>
</table>

Worker specification: 2x2 cores AMD Operton 2.2GHz, 4g RAM, 160g HD

References
- XQuery 1.0: An XML Query Language. W3C Recommendation 23 January 2007
- MapReduce: Simplified data processing on large clusters, J. Dean and S. Ghemawat OSDI 2004