A Comparison of Approaches to Large-Scale Data Analysis

Magnus Kirø

Norwegian University of Science and Technology

October 26, 2012

Presentation Goals

The purpose of paper is to consider MapReduce and a regular Database Management Systems for large-scale data analysis.

1. **Parallel DBMS** and **MR**, two approaches to large-scale data analysis.
2. Thee **Architectural Elements** of DBMS and MR.
3. Outline of **Benchmark Results** and **Best Practice** for large-scale data analysis.

MR vs DBMS

- MR
  - Only two Functions:
    - **Map**
    - **Reduce**
  - Data is stored in a distributed file system on the node.
- BDMS
  - Tables are partitioned across nodes
  - Query optimizer, that translates SQL to a query plan. execution of the query plan is divided among multiple nodes.
  - underlying storage details can be disregarded by the programmers.

Schema Support

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Indexing

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Programming Model

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Data Distribution

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Execution Strategy

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Flexibility

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Fault Tolerance

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Benchmark Results

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Best Practice

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

MR vs DBMS

The different setups have different strong points.

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Drawbacks

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

Related Topics

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

A more time specific tentative plan

1. ... (content omitted)
2. ... (content omitted)
3. ... (content omitted)
4. ... (content omitted)

This is the last slide.

Any questions?