View Reviews

Paper ID
772

Paper Title
Using VDMS to Index and Search 100M Images

Track Name
Industry

Reviewer #1

Questions

2. Overall Rating
Weak Reject

4. Are there specific revisions that could raise your overall rating?
No

6. Summary of the paper (what is being proposed and in what context) and a brief justification of your overall recommendation. One solid paragraph.
The paper describes a visual data management system. The system is described in some detail, but it is not clear what the system is useful for since there are no applications in the paper.

7. Three (or more) strong points about the paper (Please be precise and explicit; clearly explain the value and nature of the contribution).
1. Real system

8. Three (or more) weak points about the paper (Please clearly indicate whether the paper has any mistakes, missing related work, or results that cannot be considered a contribution; write it so that the authors can understand what is seen as negative).
1. No application

2. Not clear how or whether the system is actually deployed.

3. There is just a synthetic benchmark for the system.

9. Novelty (Please give a high novelty ranking to papers on new topics, opening new fields, or proposing truly new ideas; assign medium ratings to delta papers and papers on well-known topics but still with some valuable contribution).
Novelty unclear

10. Significance
Improvement over existing work

11. Technical Depth and Quality of Content
Syntactically complete but with limited contribution

12. Experiments
OK, but certain claims are not covered by the experiments

13. Presentation
Reasonable: improvements needed

14. Detailed Evaluation (Contribution, Pros/Cons, Errors); please number each point.

I enjoyed reading this paper, but I am not clear with what I am left with. I completely lack an understanding what this system is good for -- it does not seem to be used in any large-scale search application? Is it used in data science workflows -- if so, in which ones? All the metrics (ingestion rates, concurrent queries, etc.) are not useful if I do not know what the system is used for in production.

As such, I have a hard time seeing that the paper would be very useful for the PVLDB audience.

15. If revision is required, list specific revisions you seek from the authors
Include a case study of the system in a production setting?

Reviewer #2

Questions

2. Overall Rating
Weak Reject

4. Are there specific revisions that could raise your overall rating?
No

6. Summary of the paper (what is being proposed and in what context) and a brief justification of your overall recommendation. One solid paragraph.
This paper is an experiment paper on VDMS -- a Visual Data Management System developed by Intel labs. In the experiments, the paper used a dataset that has 100 million images. It compared VDMS with a combination of off-the-shelf components including MySQL, Apache Web Server, and OpenCV. The comparison is pretty comprehensive and the results are very good in terms of image search and video search, e.g., it's orders of magnitude faster. Moreover, the system simplifies the data preparation process and provides useful functions that other systems don't have. However, this is not a system paper and the system design was published in another paper [20]. So, this paper is a pure experimental paper of an existing system VDMS, without providing sufficient insights and lessons in the system design tradeoffs.

7. Three (or more) strong points about the paper (Please be precise and explicit; clearly explain the value and nature of the contribution).
• S1: The paper studied an important problem. Visual data is increasingly important. VDMS is interesting, timely, and is an open-source system.
• S2: The experiments are somewhat comprehensive, although could be improved. The results are very impressive.
• S3: The paper is well structured and the presentation is great too.

8. Three (or more) weak points about the paper (Please clearly indicate whether the paper has any mistakes, missing related work, or results that cannot be considered a contribution; write it so that the authors can understand what is seen as negative).
• W1: This is an experiment paper, but failed to provide sufficient insights. It's not clear what factors contributed to the high performance of VDMS and the impact of each individual factor.
• W2: Some of the experiments (including setups) are not well justified and designed, see detailed comments.
• W3: The contributions are not clear.

9. Novelty (Please give a high novelty ranking to papers on new topics, opening new fields, or proposing truly new ideas; assign medium ratings to delta papers and papers on well-known topics but still with some valuable contribution).
Novelty unclear

10. Significance
No impact

11. Technical Depth and Quality of Content

Insignificant contribution

12. Experiments
OK, but certain claims are not covered by the experiments

13. Presentation
Reasonable: improvements needed

14. Detailed Evaluation (Contribution, Pros/Cons, Errors); please number each point.
• D1: This paper is not a system paper that I would expect. It's an experiment / benchmarking paper. But it only evaluated 1 visual data management system VDMS by comparing with MySQL, which is not fair since it's known that RDBMS is poor at handling images and videos. More importantly, it failed to provide sufficient insights and lessons in terms of how to build visual management systems. So, the contributions are not very clear to me.
• D2: It's not clear what factors contributed to the high performance of VDMS and the impact of each individual factor. For example, the paper mentioned I/O scheduling in Section 3.2.1. It's not clear what's the impact of the scheduling to the overall performance. There're many factors that can speedup or slow down the performance. The paper didn't clearly explain them.
• D3: The paper compared with MySQL for storing metadata, but it's not fair. For example, aototags are stored as in a table with primary-foreign key references, which can incur a lot of performance overhead for query processing.
• D4: In Section 3.2.2, the paper shows MySQL takes 3.72x longer to load data but without explaining why. So, what're the lessons for people to learn from the work?
• D5: In Section 3.4, the paper used only 4 artificial queries to evaluate image search. It would be more convincing if the paper can use more and real queries.
• D6: Also in Section 3.4, for q1, which finds metadata/images with one specific autotag. The best solution would be using inverted list for each tag, but why VDMS uses graph traversal?
• D7: There are some interesting parts, e.g., persistent memory on metadata, but "The benefits of Persistent Memory on metadata operations is left for another paper, and outside the scope of this evaluation."
• D8: The paper spent 3 pages in Section 3.4 and 3.5, but didn't compare with other solutions. It's not clear why Section 3.4 and 3.5 are necessary.

Reviewer #3

Questions

2. Overall Rating
Reject

4. Are there specific revisions that could raise your overall rating?
No

6. Summary of the paper (what is being proposed and in what context) and a brief justification of your overall recommendation. One solid paragraph.
This paper presents VDMS (visual data management system) to process and manage a large set of images and videos. VDMS is constructed by composing a few open source components for processing images (using OpenCV), videos (OpenCV), feature vectors (using TileDB), and property graph database for metadata management.

7. Three (or more) strong points about the paper (Please be precise and explicit; clearly explain the value and nature of the contribution).
1) The paper is well written and easy to understand.
2) Extensive experiment evaluation using YFCC100M dataset.

8. Three (or more) weak points about the paper (Please clearly indicate whether the paper has any mistakes, missing related work, or results that cannot be considered a contribution; write it so that the authors can

understand what is seen as negative).

1) The details of the VDMS is not described in the paper.

2) The pre-processing and indexing of the visual data is not explained in the paper.

3) The system is a combination of several open source components, as a result, the fundamental technical novelty is lacking.

9. Novelty (Please give a high novelty ranking to papers on new topics, opening new fields, or proposing truly new ideas; assign medium ratings to delta papers and papers on well-known topics but still with some valuable contribution).

Novelty unclear

10. Significance

Improvement over existing work

11. Technical Depth and Quality of Content

Syntactically complete but with limited contribution

12. Experiments

Very nicely support the claims made in the paper

13. Presentation

Reasonable: improvements needed

14. Detailed Evaluation (Contribution, Pros/Cons, Errors); please number each point.

This paper is well written and presents a useful system for managing large scale image and video data. The proposed overall system architecture seems quite reasonable and relies on open source systems to build. Unfortunately, the details of the proposed VDMS is missing from the paper, as a result, how exactly image and video data is pre-processed and indexed is not explained. The fundamental technical challenges that need to be overcome is also not clear, hence, the technical novelty of the system also needs further justification.