
AWS
re:Invent

ANT314

Turbocharging query execution on Amazon EMR

Joseph Marques
Principal Engineer
AWS

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What will you learn?

- Performance philosophy
- General focus areas
- Optimization techniques
- Benchmarking results
- Beyond the query engine



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Performance philosophy

FOCUSED EFFORT TO PROVIDE…

- Cost savings



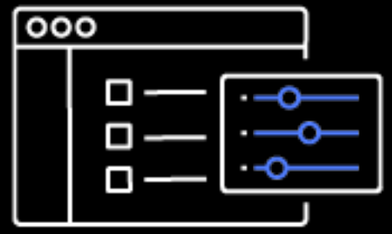
- Ease of use



- OSS compatibility



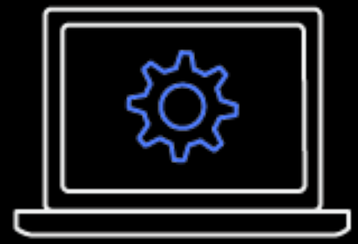
Focus areas



Configuration



Planning/Optimization



Query Execution

Optimization techniques



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

How to pull less data?

BEYOND DATA MODELING AND FILE FORMAT

Optimize metadata fetch



Eliminate redundant scans



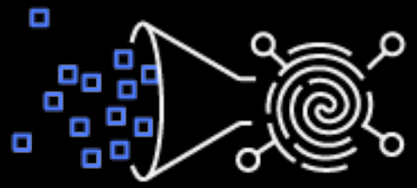
Dynamic partition pruning



How to shuffle less data?

BEYOND DATA BUCKETING AND QUERY REWRITE

Intelligent filtering



Broadcast join
w/o statistics



Join reordering
w/o statistics



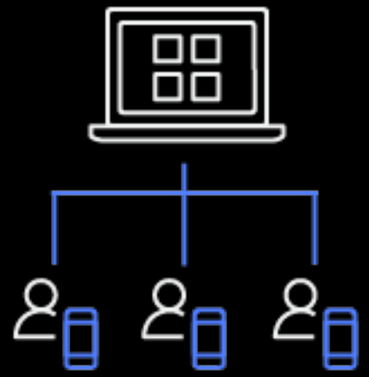
How to start queries faster?

BEYOND APPLICATION WARMING

Parallel/Async initialization



Early worker allocation



Data pre-fetch



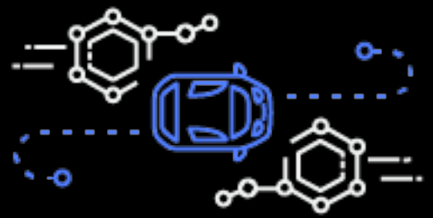
How to run queries faster?

BEYOND SCALING HARDWARE

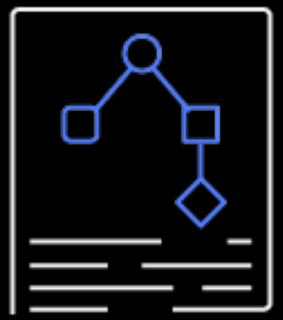
Operator optimization



Adaptive Query Execution



Stats inference

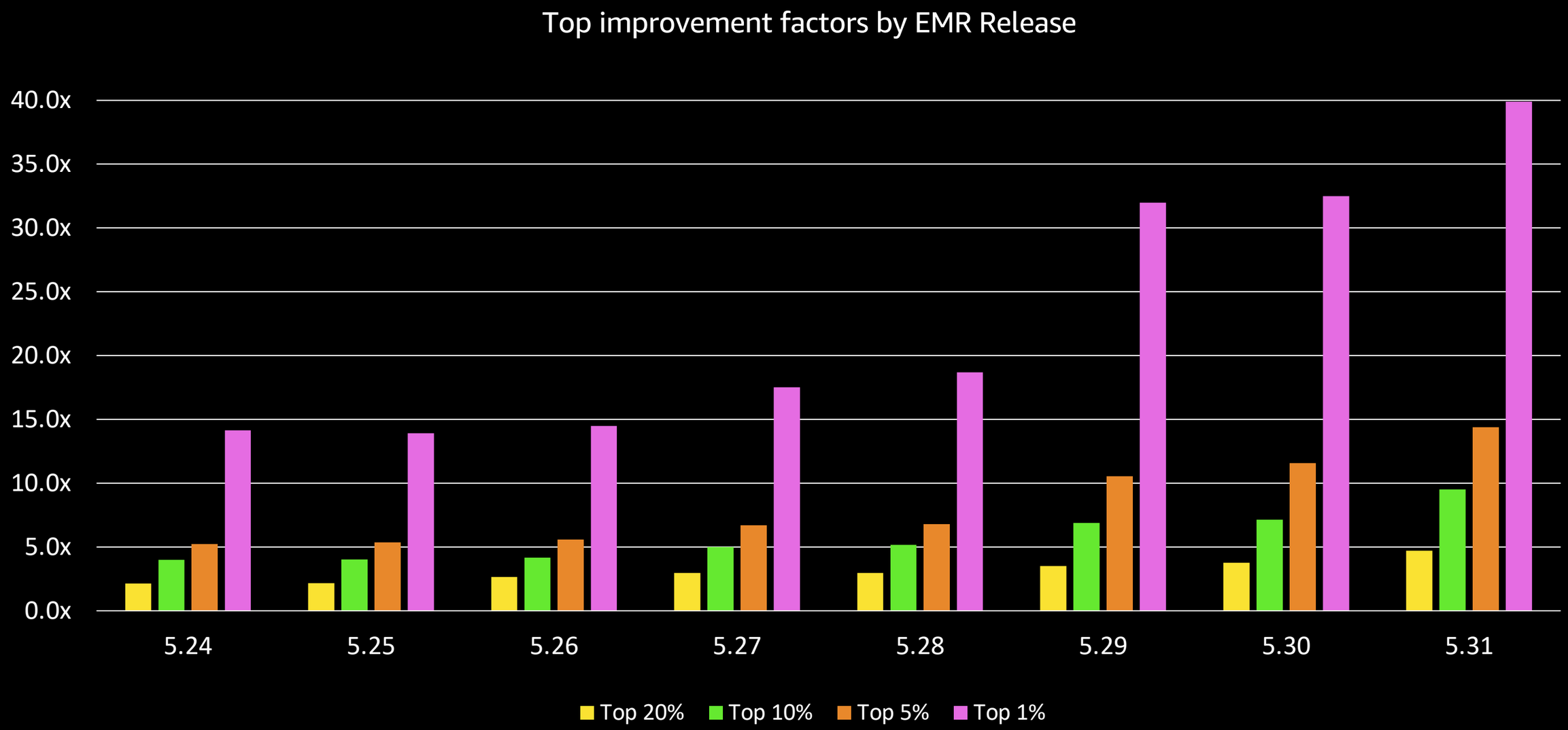


Benchmarking results

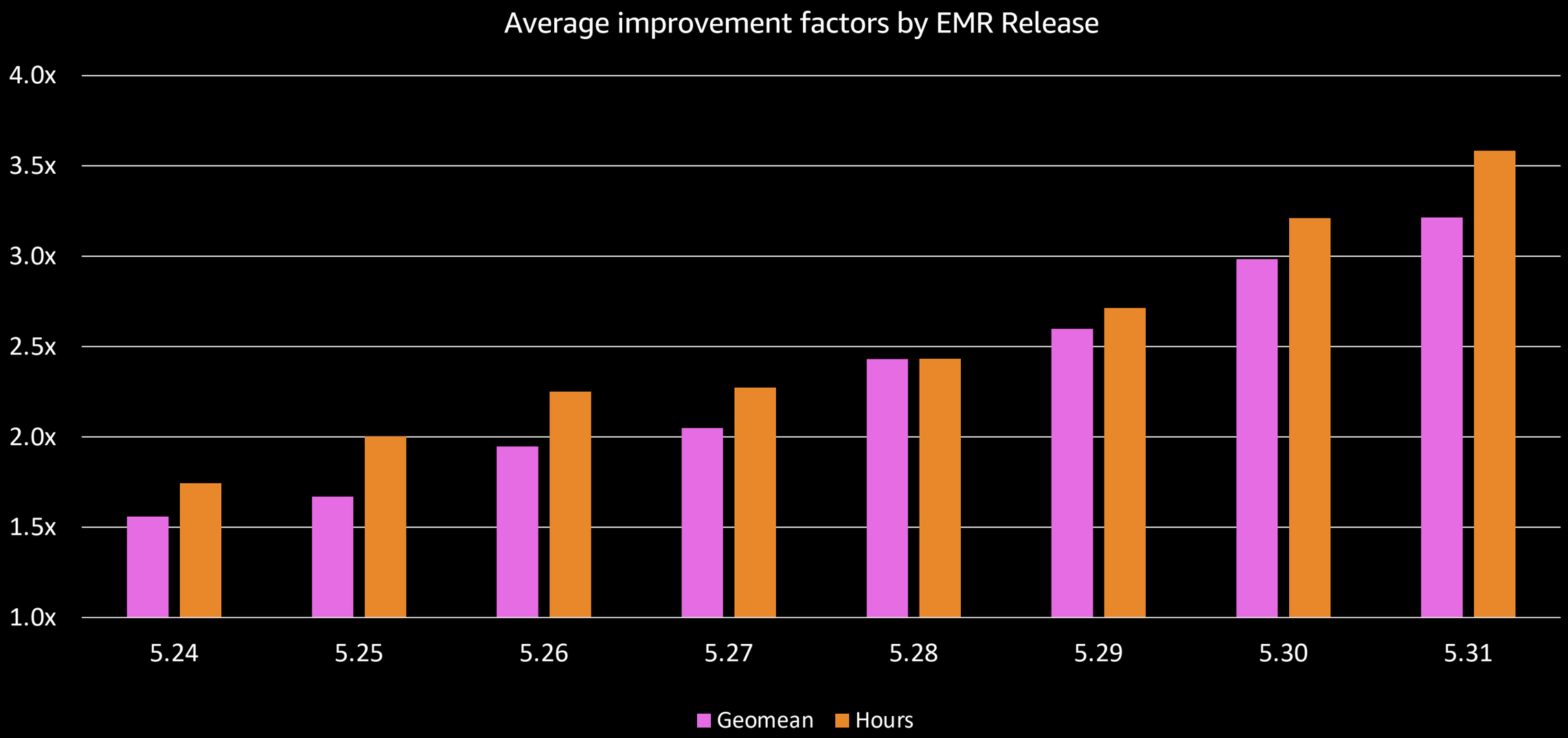


© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

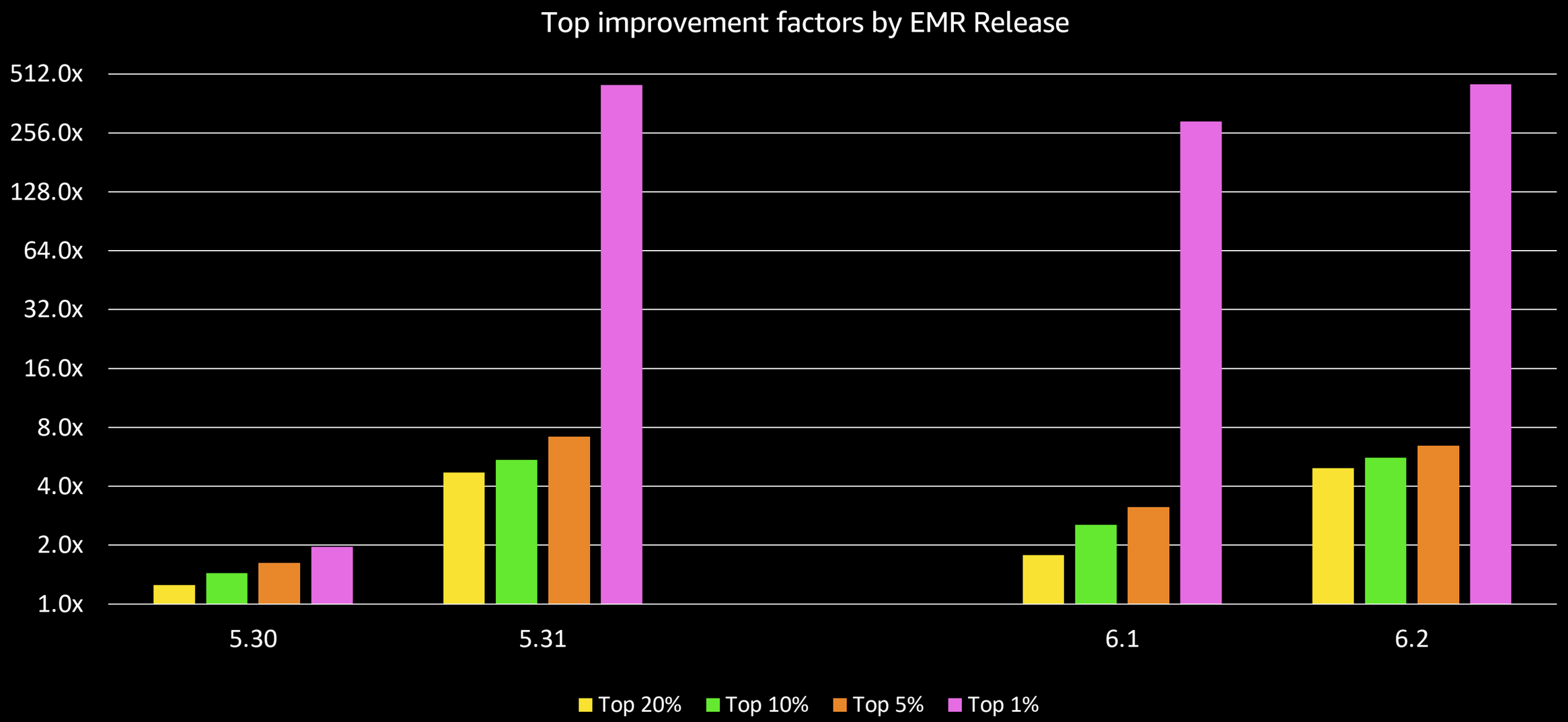
Spark benchmark results



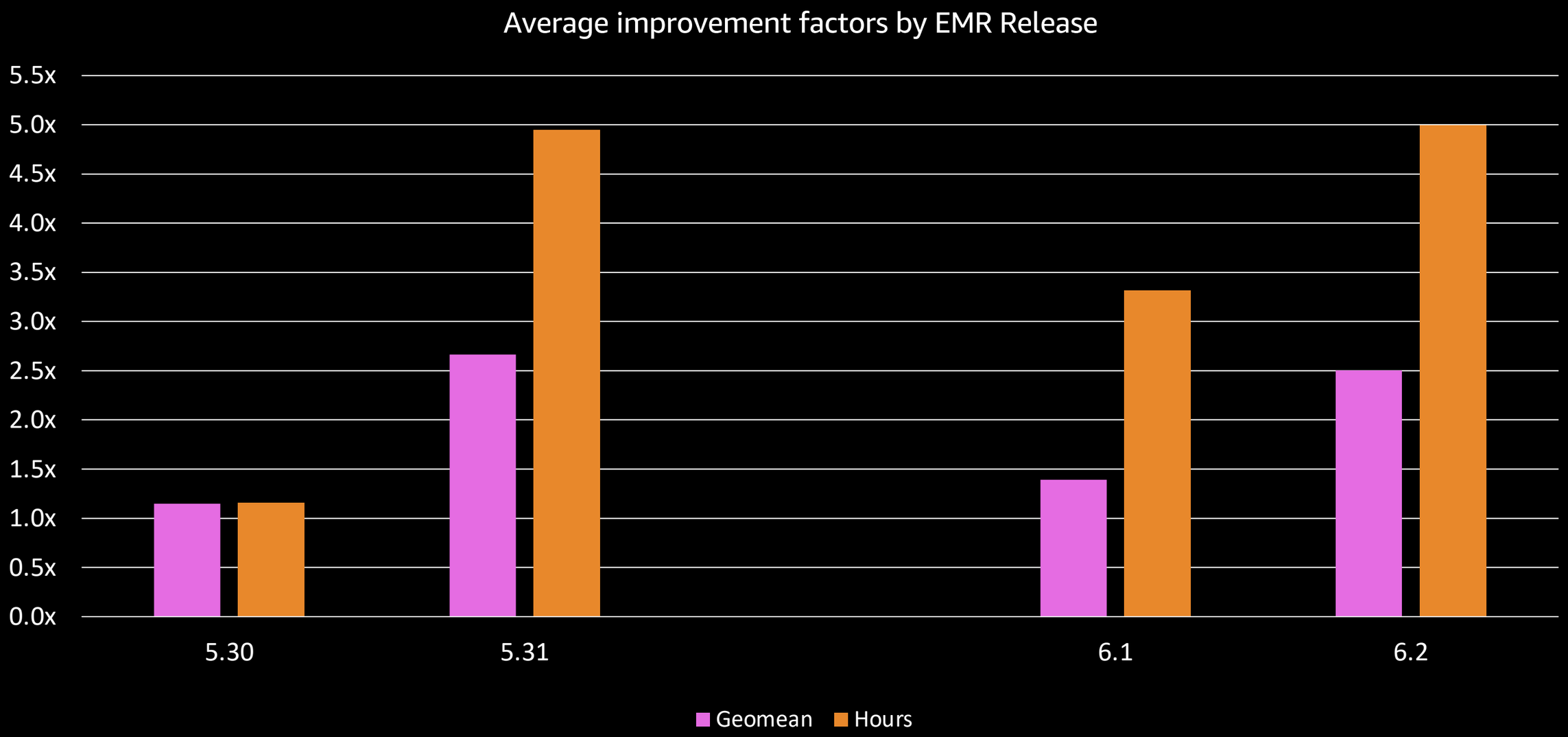
Spark benchmark results



Presto benchmark results



Presto benchmark results



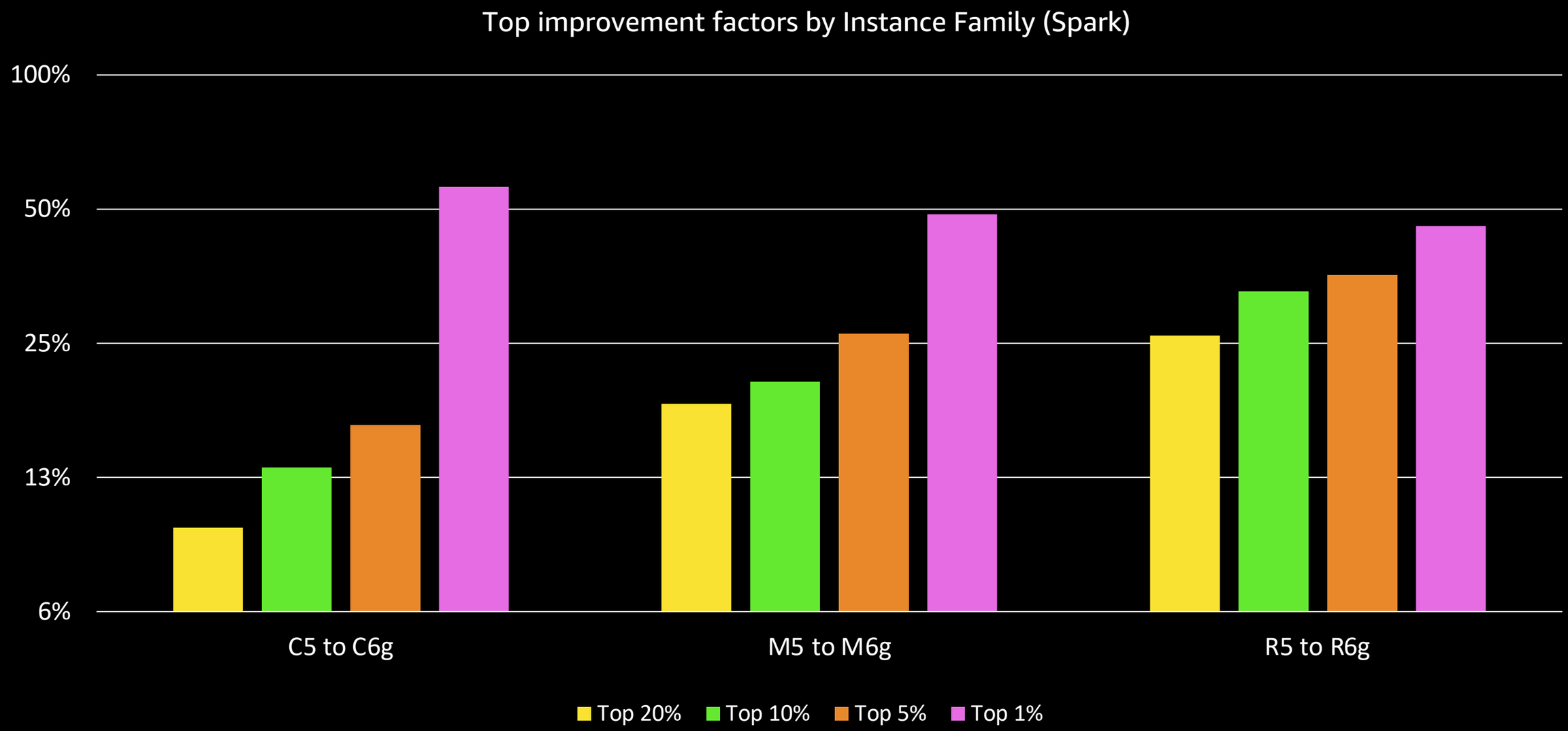
Beyond the query engine



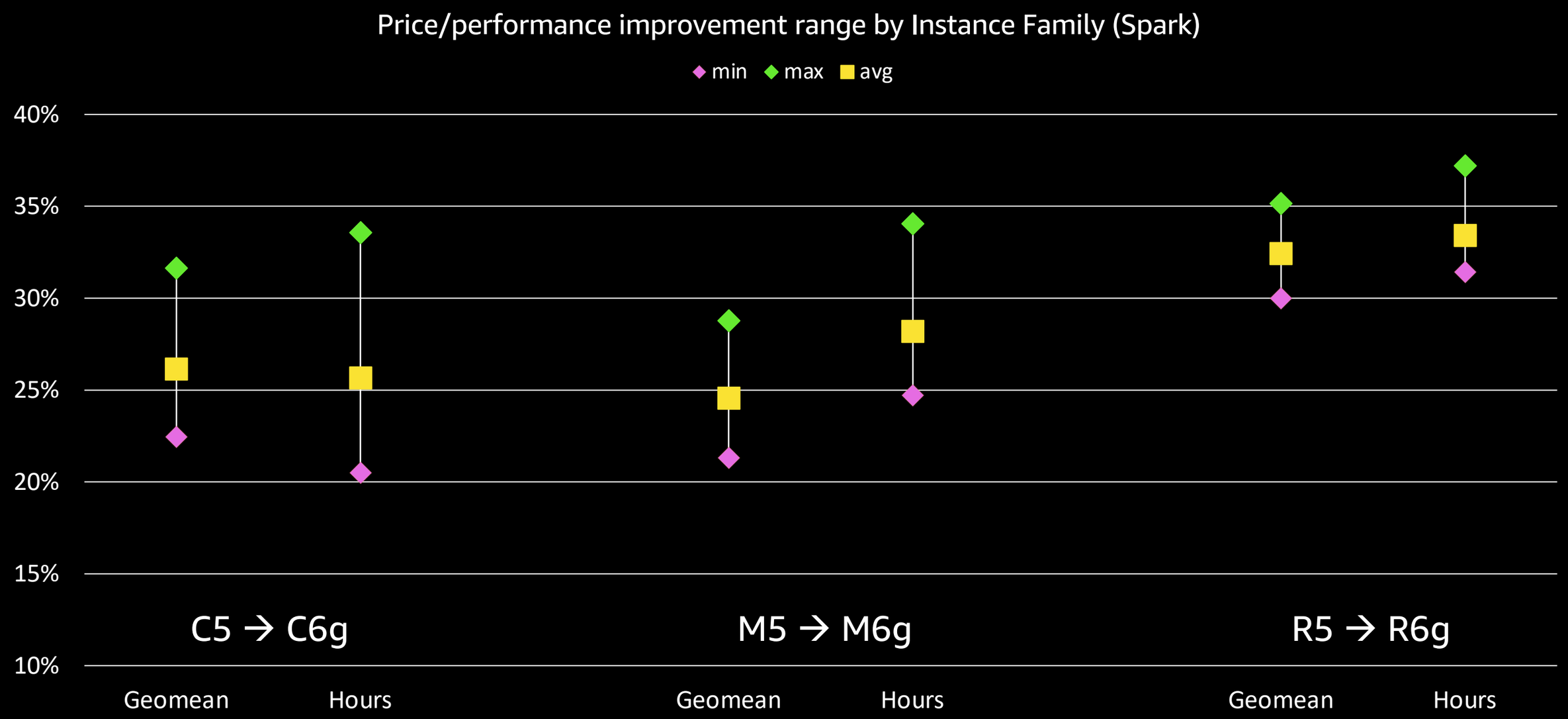
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

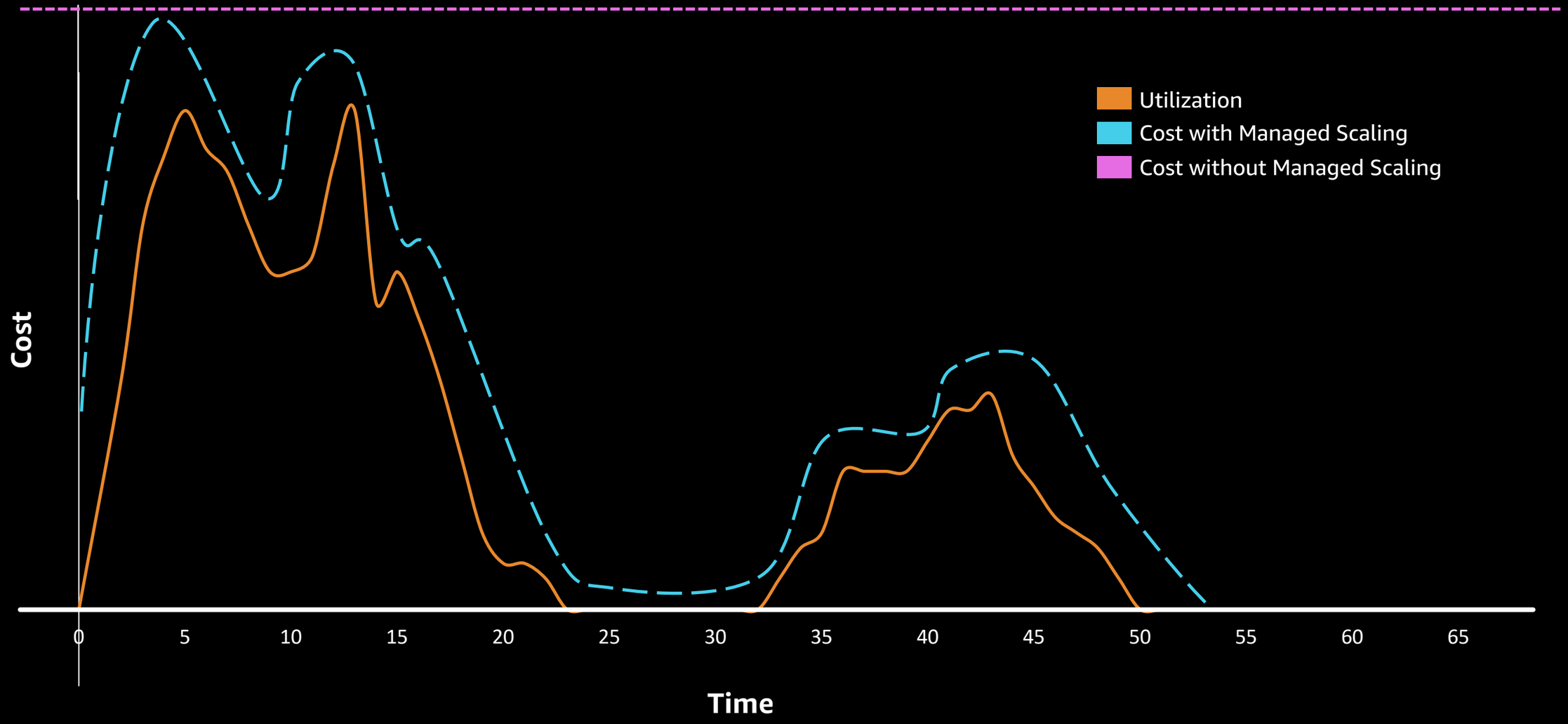
Graviton2 benchmark results



Graviton2 benchmark results



Managed Scaling overview



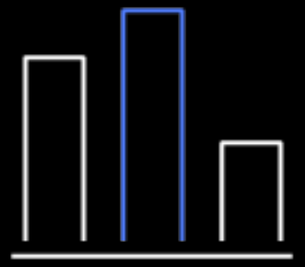
Managed Scaling benchmark results



Key Takeaways

SAVE TIME AND MONEY BY…

Upgrading EMR



Leveraging Graviton2



Enabling Managed Scaling



Thank you!

Joseph Marques

jmarques@amazon.com

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.