

Data Warehousing with Amazon Redshift



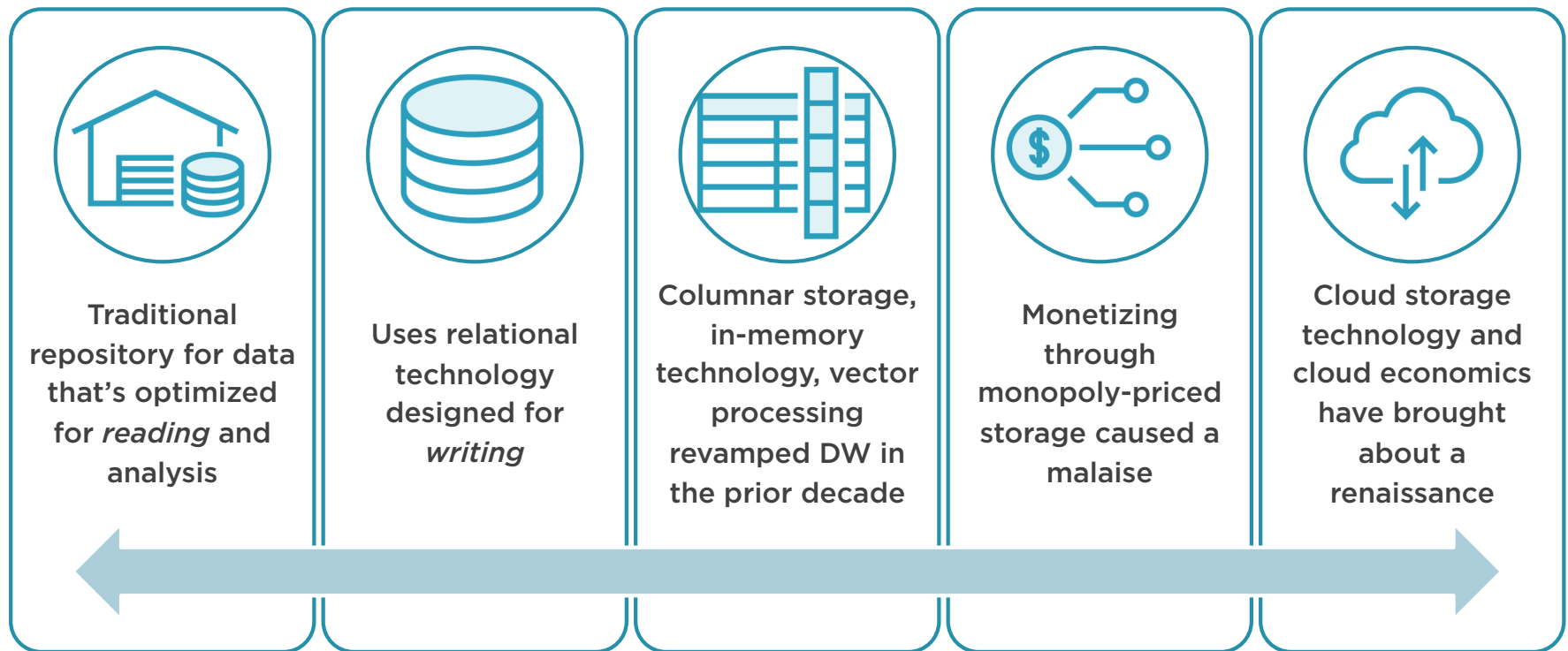
Andrew Brust

FOUNDER & CEO, BLUE BADGE INSIGHTS

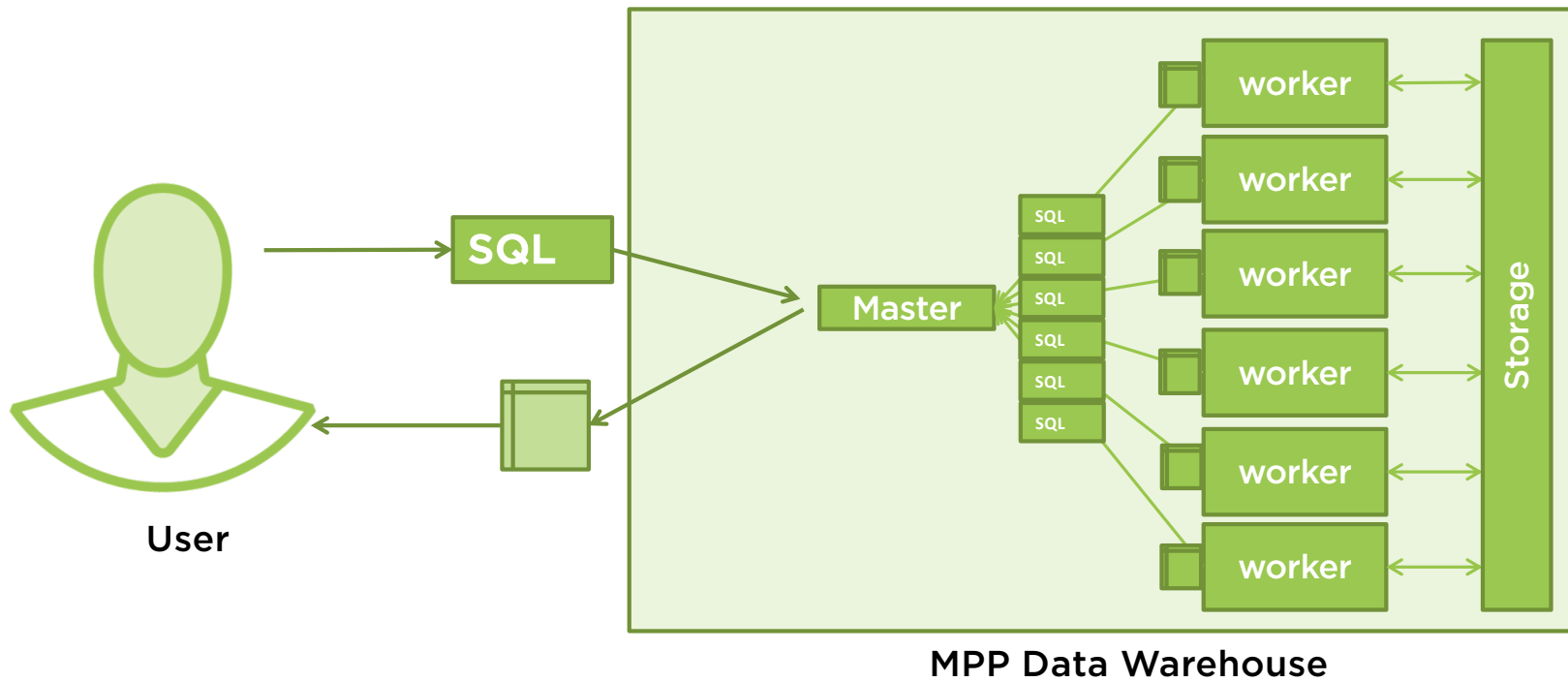
@andrewbrust www.bluebadgeinsights.com



Concepts: Data Warehouse



What Is MPP?



Redshift



Cloud-based MPP Data Warehouse; petabyte scale, sized elastically



Uses conventional storage (not S3)



Integrates with S3, DynamoDB, other services



Pioneer in cloud DW w/ very big ecosystem, but has lost ground to competitors



Redshift

Lineage in on-prem DW
ParAccel, in which Amazon
was an investor

Like many MPP DWs,
compatible with Postgres

Unlike competitors, doesn't
use object storage; can't
pause/resume

Fast, expensive



Redshift and Data Lakes



Redshift Spectrum allows direct querying of data files in S3 via EXTERNAL tables

Redshift integrates with Glue's data catalog, as does...

- Hive
- Spark SQL
- Presto (on EMR)
- Athena

