



aws **SUMMIT**

ATLANTA | MAY 18-19, 2022

ISV201

Event streaming with modern data pipelines in a SaaS architecture

Matthew Duren
Director of Reliability Engineering
KnowBe4

Ray Dickenson
ISV Solutions Architect
AWS



Agenda

About us

The problem

AWS Well-Architected

The old versus the new



About KnowBe4

- Founded 2010 in Tampa, FL
- Industry-leading security awareness training and phishing simulation service
- **Fortune's Best Places to Work in Technology**
- Ranked a best place to work for women, millennials, and in technology nationwide by Fortune Magazine three years in a row
- Ranked #1 or #2 in the Tampa Bay Top Workplaces Survey for the last six years

**Great
Place
To
Work**

Certified

AUG 2021–AUG 2022

USA

The problem

- Our internal warehouse needs data
- Relational databases have lots of stakeholders
- Our customers need better reporting

Unicorn – as envisioned



Unicorn – as implemented



AWS Well-Architected Framework

- Stop guessing our capacity needs
- Test systems at production scale
- Automate to make architectural experimentation easier
- Allow for evolutionary architectures
- Drive our architecture using data
- Improve with the help of Game Days

AWS Well-Architected Framework

- Stop guessing our capacity needs
- Drive our architecture using data
- Allow for evolutionary architectures

The old versus the new



Evolution of the architecture

The old

- Nightly, long-running Amazon ECS-based job
- Clone the entire database
- Anonymize the entire dataset
- Export to Amazon S3 (Parquet)
- Slow

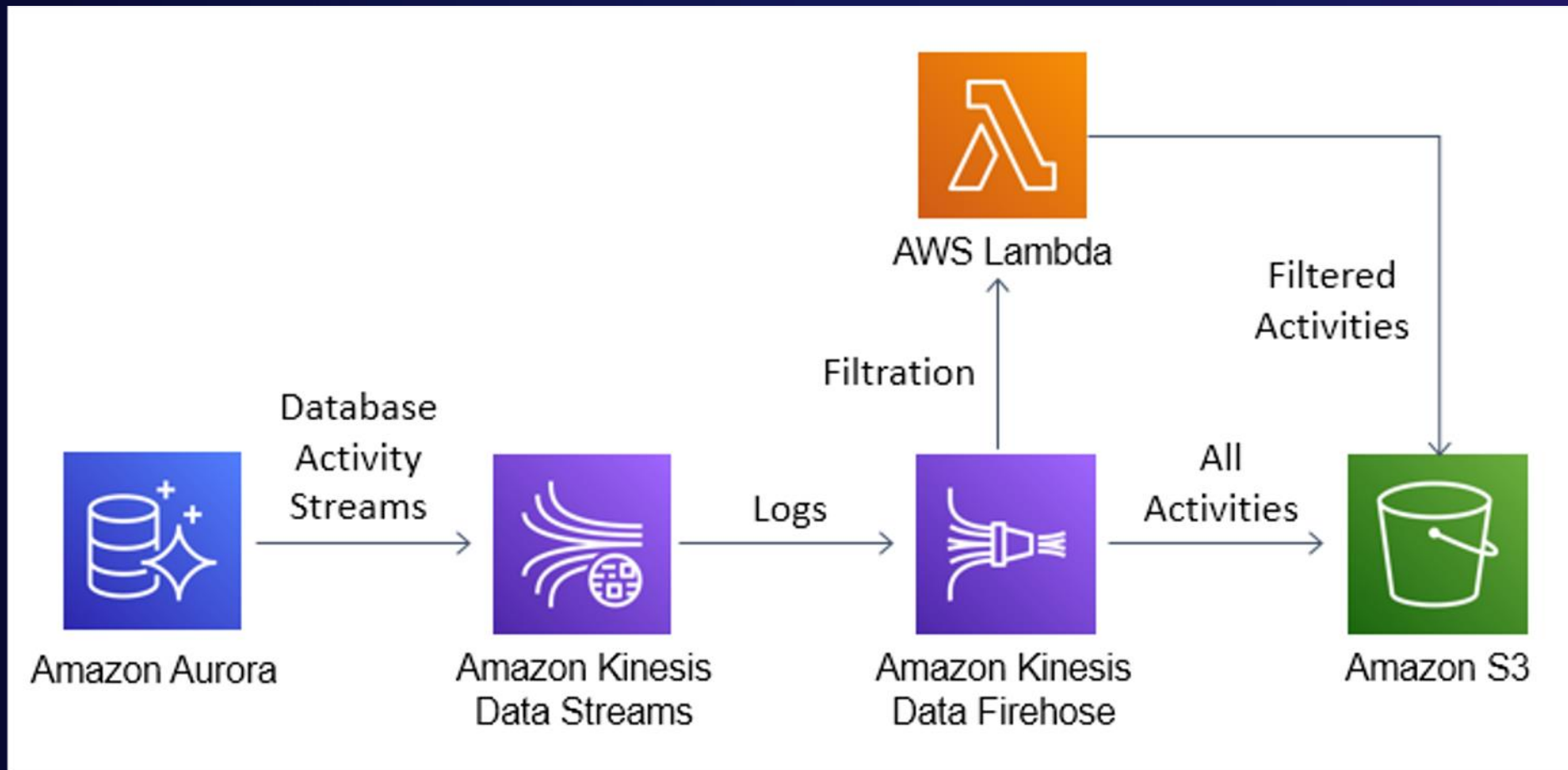
The new

- No waiting around (AWS Lambda)
- Work with new data only (CDC)
- Anonymize the data once
- Product to Amazon S3
- Fast!

Colonial – the pipeline

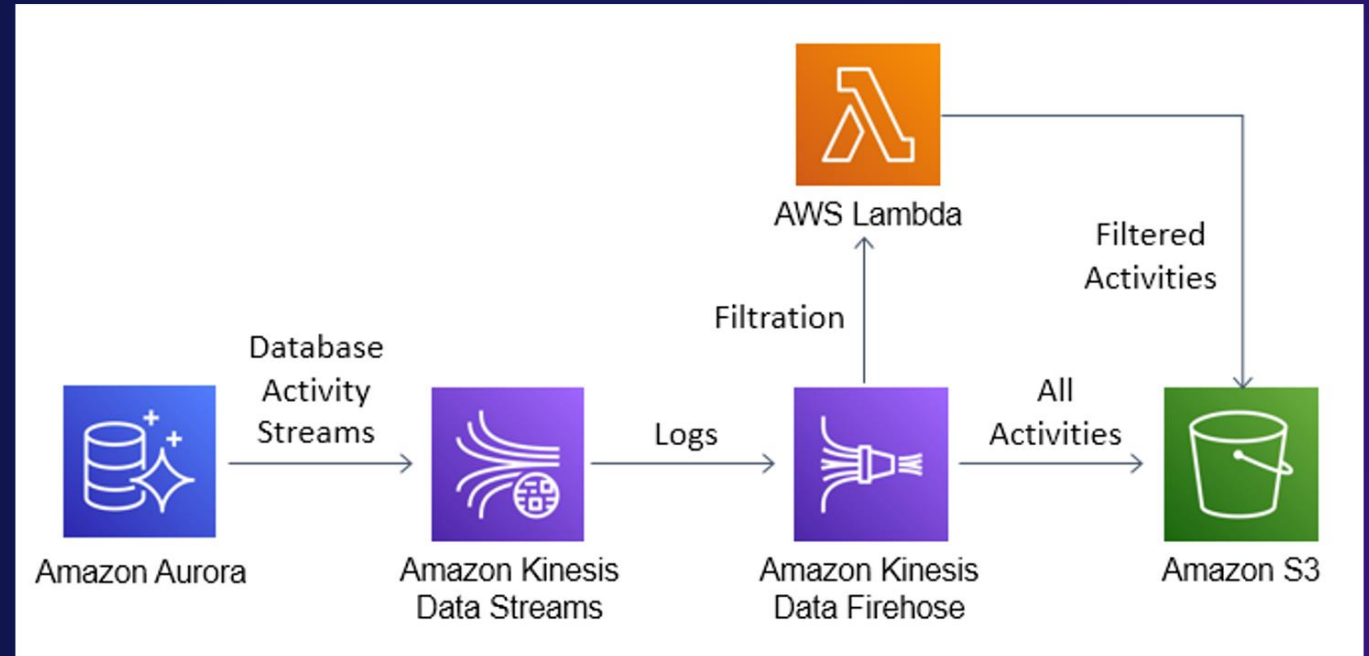


Database activity streams



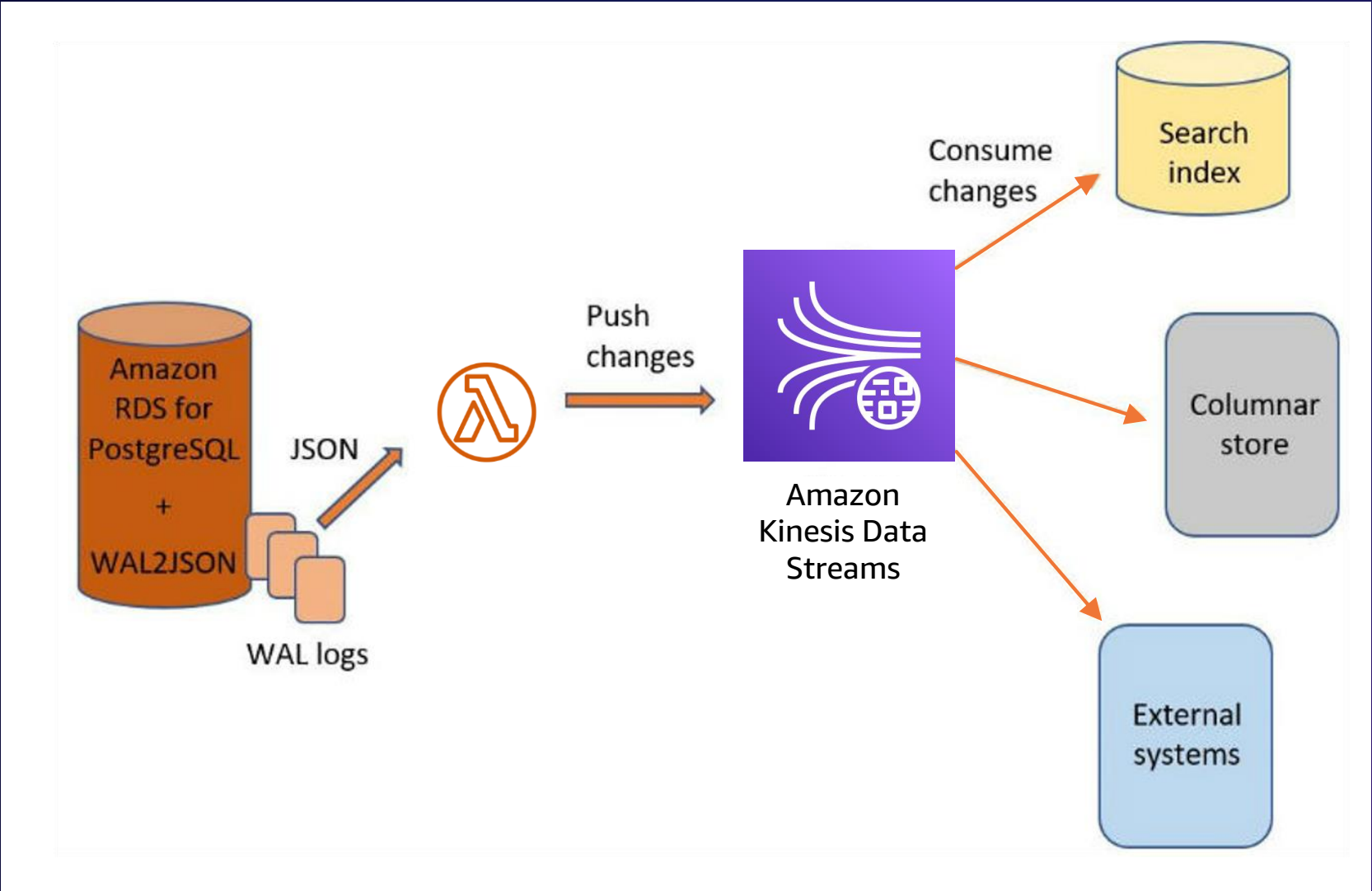
Database activity streams

- Free with Amazon Aurora
- No full load
- Not proper CDC
- No “middleman” features
 - Select schemas/tables
 - Stop/pause
 - Data store targeting



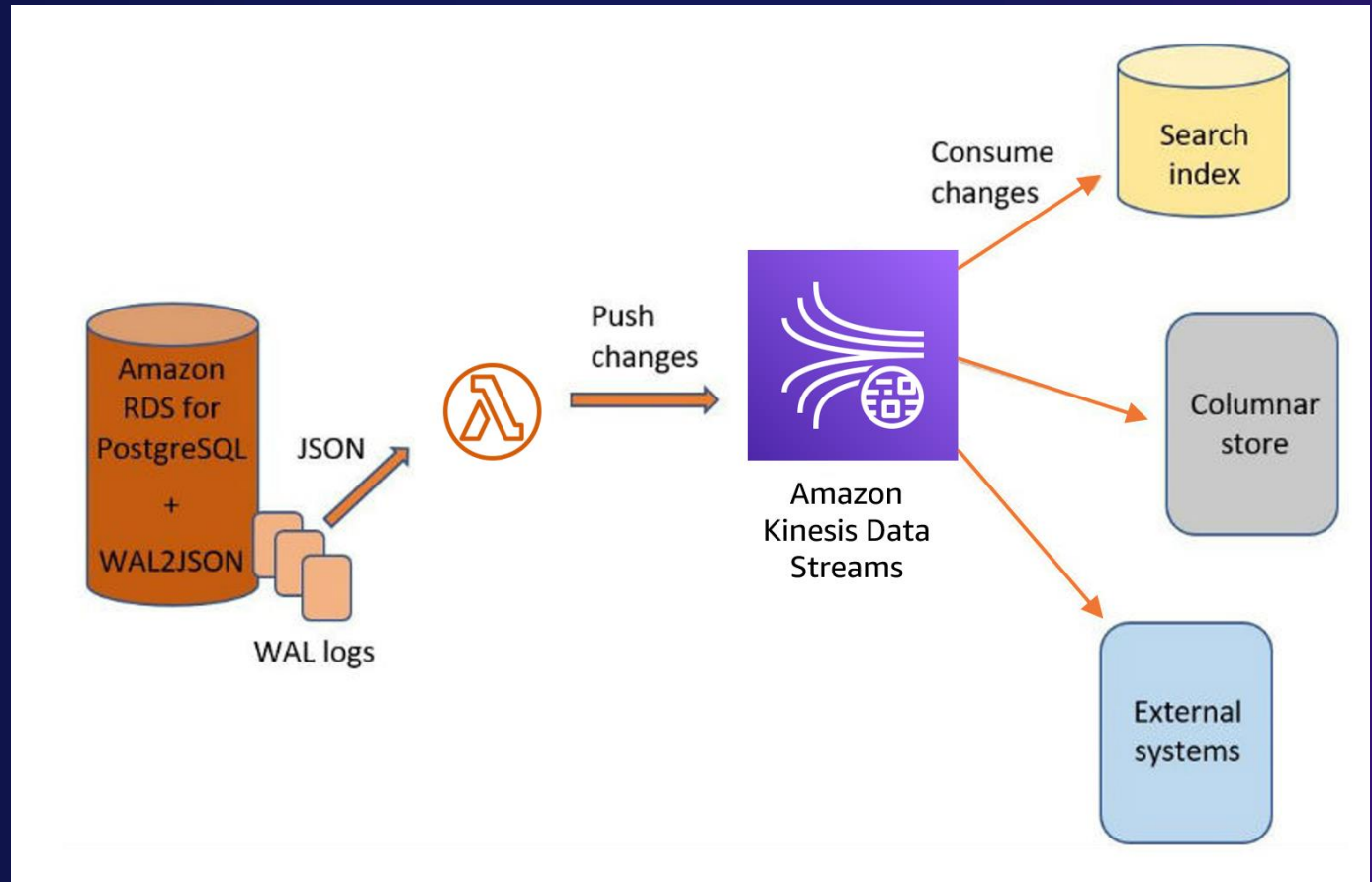
<https://aws.amazon.com/blogs/database/filter-amazon-aurora-database-activity-stream-data-for-segregation-and-monitoring>

Write-ahead logs (WAL)



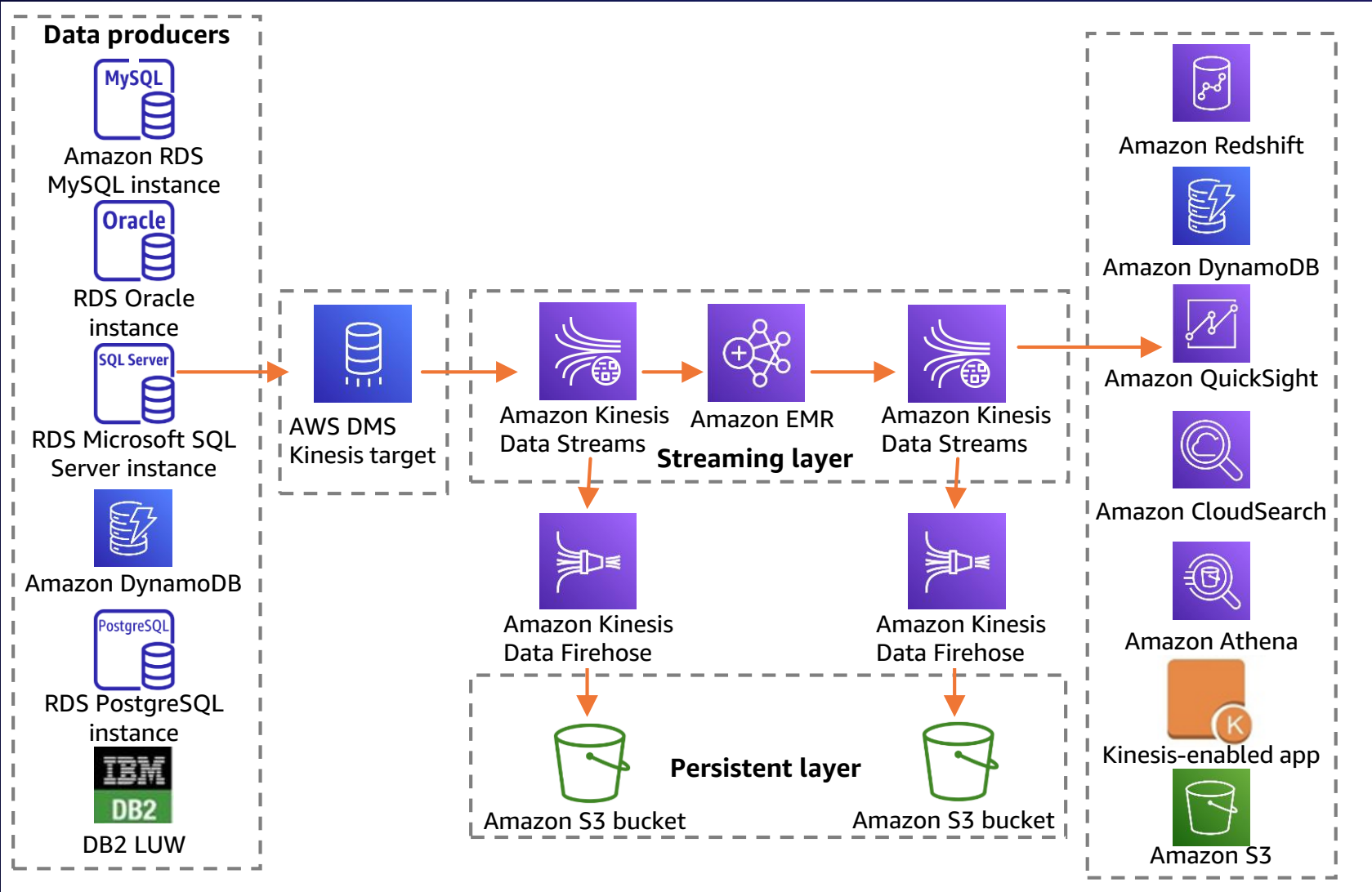
Write-ahead logs (WAL)

- Postgres only
- Delivery to Amazon CloudWatch



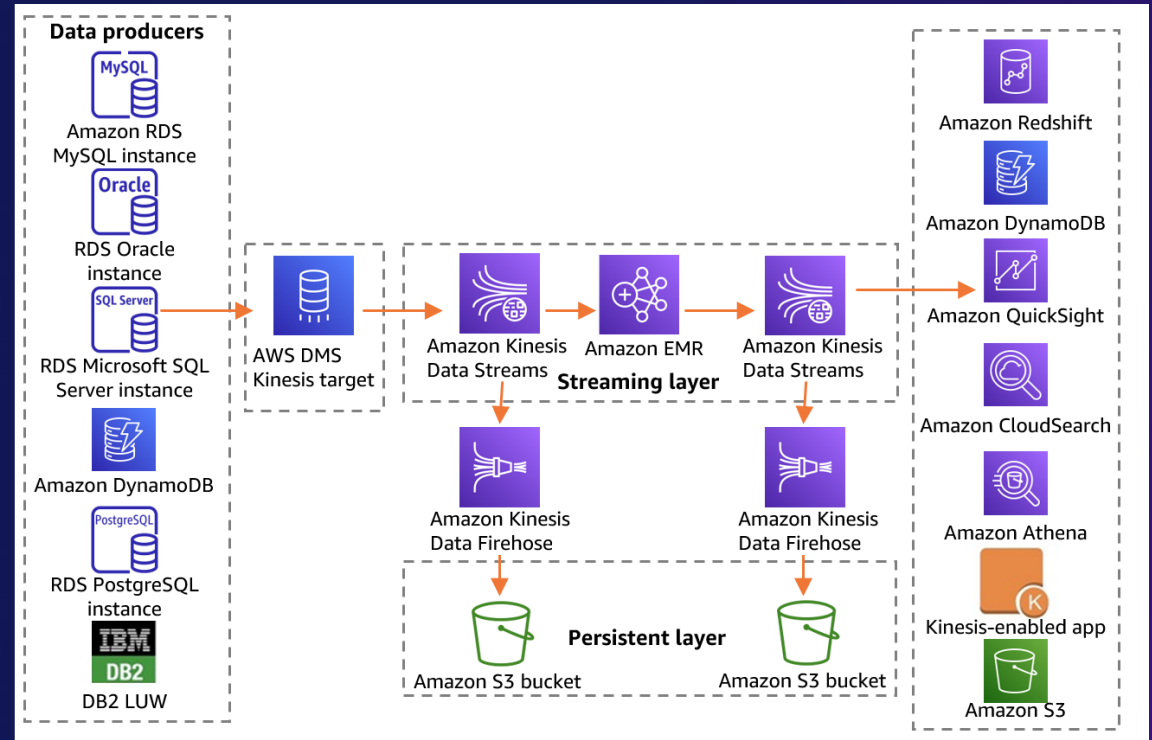
<https://aws.amazon.com/blogs/database/stream-changes-from-amazon-rds-for-postgresql-using-amazon-kinesis-data-streams-and-aws-lambda>

AWS Database Migration Service (AWS DMS)



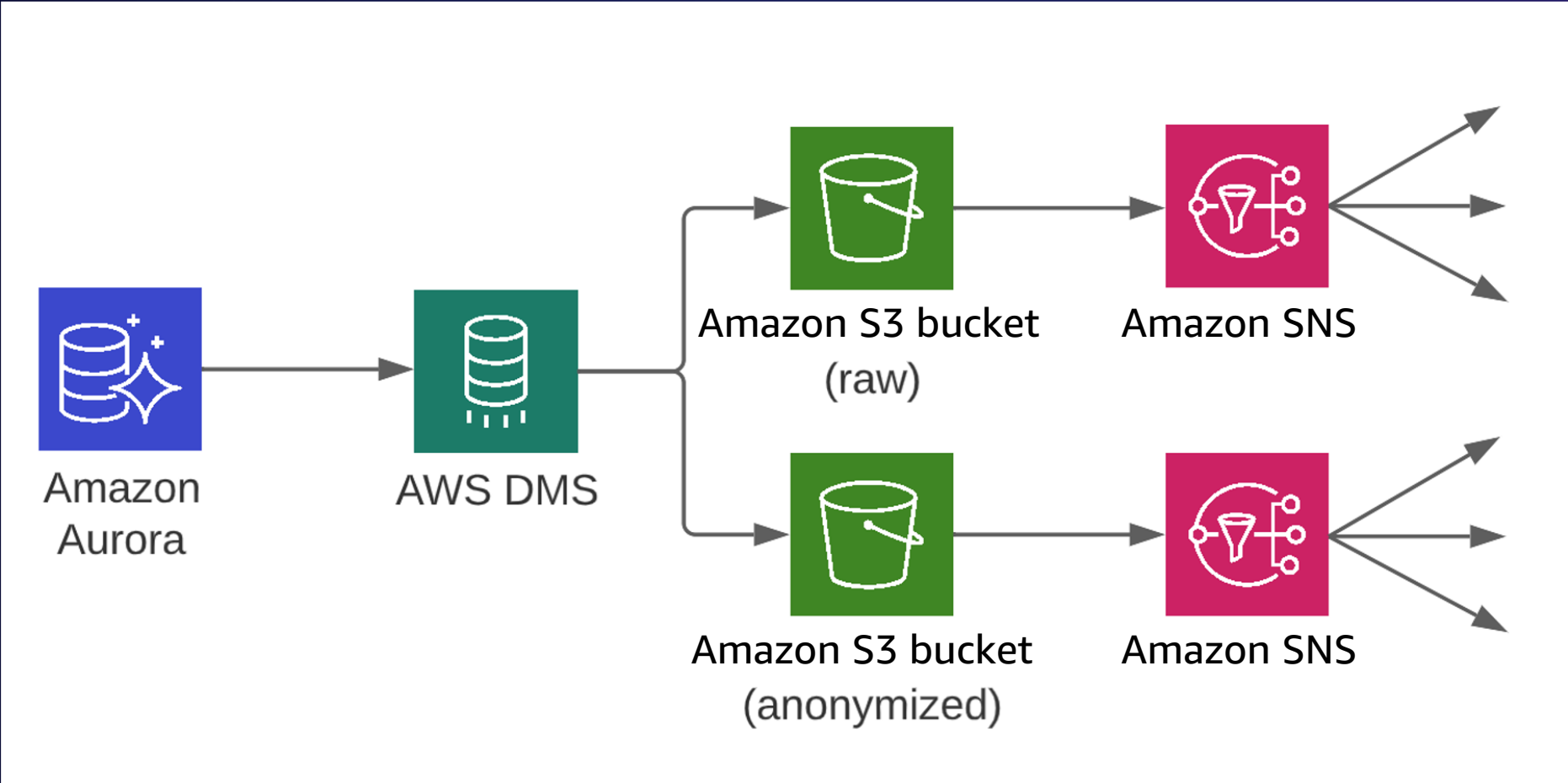
Database activity streams

- Full load capabilities
- Control over the export task (stop/start/pause)
- True CDC solution
- Many options for schema selection
- Native anonymization (bonus points!)



<https://aws.amazon.com/blogs/database/use-the-aws-database-migration-service-to-stream-change-data-to-amazon-kinesis-data-streams>

AWS DMS



How it's going so far

- A few minor issues with DDL pausing the DMS (e.g., adding columns)
- Limited anonymization capabilities
 - Will lead to AWS Glue some day
- “If you build it, they will come” has been successful
 - User sync
 - New reporting capabilities
 - Internal search tool
- Applied to 3 Amazon RDS databases and growing
- Overall, very successful!

Learn in-demand AWS Cloud skills



AWS Skill Builder

Access **500+ free** digital courses and Learning Plans

Explore resources with a variety of skill levels and **16+** languages to meet your learning needs

Deepen your skills with digital learning on demand



Train now



AWS Certifications

Earn an industry-recognized credential

Receive Foundational, Associate, Professional, and Specialty certifications

Join the **AWS Certified community** and get exclusive benefits



Access **new** exam guides

Thank you!

