

HeatWave Lakehouse

HeatWave Lakehouse는 탁월한 성능 및 가성비를 바탕으로 객체 스토리지, MySQL 데이터베이스 또는 양쪽 모두에 대한 데이터 쿼리를 지원하는 서비스입니다.

도전 과제: 오늘날 기업의 데이터 수집 속도는 기하급수적인 속도로 증가 중입니다. 기업은 일상적 운영에 사용하는 미션 크리티컬 데이터 쿼리 뿐만 아니라, 객체 스토리지에 저장된 대용량 데이터(예: IoT 데이터, 웹 콘텐츠, 로그 파일)에서도 인사이트를 확보해야 합니다. 그러나 기업이 이와 같이 다양한 데이터 유형에 대한 분석을 수행하기 위해 객체 스토리지에서 데이터베이스로 대규모 데이터를 복사하고 변환하는 데 소요되는 복잡성, 시간, 비용은 반드시 해결해야 하는 문제입니다.

해결책: HeatWave Lakehouse는 객체 스토리지에서 최대 0.5 페타바이트에 달하는 CSV, Parquet, Avro, JSON 및 다른 데이터베이스로부터의 내보내기(예: Aurora, Redshift) 등의 다양한 파일 형식의 데이터를 HeatWave로 로드하고, 필요한 경우 로드한 데이터를 단일 SQL 쿼리로 MySQL 데이터베이스 내의 트랜잭션 데이터와 결합하여 사용 편의성을 극대화할 수 있도록 지원합니다. 또한 HeatWave Lakehouse는 고객이 데이터베이스 내의 데이터를 쿼리하는 것처럼 **빠르게** 객체 스토리지의 데이터를 쿼리할 수 있는 업계 최초의 서비스입니다.

HeatWave 엔진이 모든 쿼리를 처리하므로 MySQL 호환 워크로드 뿐만 아니라 비MySQL 워크로드에도 HeatWave를 함께 활용할 수 있습니다. 또한 HeatWave Lakehouse에서 HeatWave AutoML을 실행하여 머신러닝 모델을 자동으로 학습시키고, 객체 스토리지에 저장된 파일에 대한 추론 및 설명을 진행할 수 있습니다. 또한 AWS 고객은 하나의 서비스로 최대 5개의 AWS 서비스를 대체함으로써 복잡성은 줄이고, 업계 최고의 데이터 분석 가성비를 누릴 수 있게 되었습니다.

머신러닝 기반 최적화

MySQL Autopilot은 HeatWave Lakehouse에서 대량의 데이터를 효율적으로 쿼리하고 관리하기 위한 핵심적인 역할을 수행합니다. 프로비저닝, 병렬 데이터 로딩, 쿼리 실행 등 대규모 쿼리의 신속한 수행에 필수적이고, 종종 까다로운 문제를 발생시키는 여러 요소를 자동화해 줍니다.

객체 스토리지 내 데이터 지원, 사용 편의성 개선, 사용자의 시간 절약을 위해 기존의 여러 MySQL Autopilot 기능이 HeatWave Lakehouse에 맞게 개선되었습니다. 예를 들어, 워크로드를 처리할 수 있을 만큼 큰 클러스터를 원하면서도, 과도한 프로비저닝으로 필요 이상의 비용을 지출하고 싶지는 않은 사용자의 경우 적절한 클러스터 크기 예측이 어려울 수 있습니다. 또한 프로비저닝, 쿼리, 데이터 로드 시간이 길어지면 실시간 인사이트 확보 및 데이터 분석이 지연됩니다.



500TB TPC-H
벤치마크를 통해
입증된 바와 같이,
HeatWave
Lakehouse의 쿼리
성능은 Redshift보다
15배, Snowflake보다
18배,
Databricks보다
18배, Google
BigQuery보다 35배
빠릅니다.

HeatWave
클러스터는 최대
512개 노드까지 확장
가능하고, 최대
0.5페타바이트의
데이터를 메모리에서
처리할 수 있습니다.

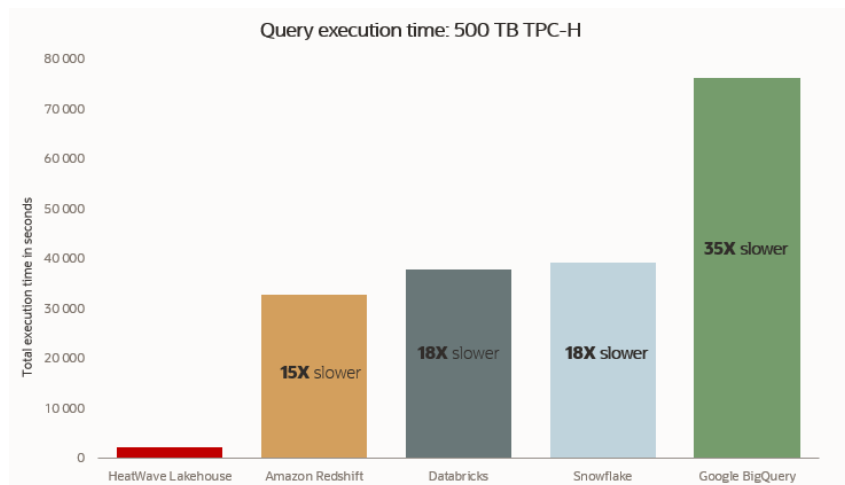
이제 **자동 프로비저닝(Auto-provisioning)** 기능은 워크로드 실행에 필요한 HeatWave 노드 수를 예측하고 객체 저장소의 파일을 지원합니다. **자동 쿼리 계획 향상(Auto query plan improvement)** 기능은 쿼리 실행 과정에서 다양한 런타임 통계를 학습하여 향후 다른 쿼리의 실행 계획을 개선합니다. **자동 병렬 로딩(Auto parallel loading)**은 객체 저장소에서 HeatWave로 데이터를 로드하는 시간을 예측하여 고객이 데이터에 대한 쿼리 실행 시기를 계획할 수 있도록 지원합니다.

HeatWave Lakehouse에 도입된 새로운 MySQL Autopilot 기능은 다음과 같습니다.

- **자동 스키마 추론(Auto-schema inference)**은 지원되는 모든 파일 유형의 해당하는 스키마 정의에 대한 파일 데이터 매핑을 자동으로 추론합니다. CSV 등의 메타데이터가 포함되지 않은 파일의 경우에도 마찬가지입니다. 결과적으로 고객이 파일의 스키마 매핑을 직접 정의 및 업데이트할 필요가 없어지므로 그만큼의 시간 및 노력을 절약할 수 있습니다.
- **적응형 데이터 샘플링(Adaptive data sampling)**은 지능적으로 파일을 샘플링하여 예측에 필요한 데이터에 대한 정보를 도출합니다. 이를 통해 MySQL Autopilot은 400TB 데이터 세트를 1분 이내에 스캔하고 그에 대한 스키마 예측을 수행할 수 있습니다.
- **적응형 데이터 흐름(Adaptive data flow)** 기능은 HeatWave Lakehouse가 모든 리전의 기반 객체 저장소 성능에 동적으로 적응할 수 있게 해 줍니다. 이를 통해 전반적인 성능, 가격 대비 성능, 가용성이 개선됩니다.

업계 최고의 성능

아래는 객체 저장소에 저장된 500TB 워크로드에 대한 HeatWave Lakehouse 및 타사 서비스의 로드 및 쿼리 성능 비교 그래프입니다.



사용자는 표준 SQL 구문을 사용하여 쿼리를 실행하고, 객체 스토리지와 데이터베이스에서 동일한 쿼리 성능을 누릴 수 있습니다.

HeatWave Lakehouse는 업계 최고의 성능 및 가성비를 바탕으로 객체 저장소에 저장된 모든 형식의 파일에 대한 데이터 쿼리 및 로드를 지원합니다.

추가 정보

- [HeatWave MySQL](#)
- [HeatWave Lakehouse](#)
- [기술 개요](#)
- [HeatWave Lakehouse 무료 체험하기](#)

문의처

한국 오라클 대표번호 **02-2194-8000**, 또는 **oracle.com/kr** 웹사이트를 통해 Oracle 담당자에게 연락하실 수 있습니다. 북미 외 지역의 경우, **oracle.com/contact**에서 국가별 지사의 연락처를 확인하실 수 있습니다.

 blogs.oracle.com

 facebook.com/oracle

 twitter.com/oracle

저작권 ©2024, Oracle 및/또는 그 계열사. 무단 전재와 무단 복제를 금합니다. 본 문서는 참고용으로만 제공되며, 문서의 내용은 사전 통지 없이 변경될 수 있습니다. Oracle은 본 문서의 무오류성을 보증하지 않습니다. 또한 본 문서에는 상업성 또는 특정 용도 수행을 위한 적합성과 관련된 암시적 보증 및 조건을 비롯한 구두상의 표현 또는 법 규정에 의한 어떠한 보증 또는 조건도 포함되어 있지 않습니다. Oracle은 본 문서에 관한 법적 책임을 일체 지지 않으며, 본 문서로 인한 직접 또는 간접적 계약 구속력 역시 일체 발생하지 않습니다. 본 문서는 Oracle의 사전 서면 승인 없이 전자적, 기계적 및 기타 어떠한 형태나 수단으로도 복제되거나 전송될 수 없습니다.

Oracle, Java, MySQL은 Oracle 및/또는 그 계열사의 등록 상표입니다. 기타 명칭들은 각 명칭을 소유한 기업의 상표일 수 있습니다.

벤치마크에 사용된 쿼리의 출처는 TPC-H 벤치마크이지만, 벤치마크 결과를 공개된 TPC-H 벤치마크 결과와 비교할 수 없는 이유는 TPC-H 사양을 준수하지 않았기 때문입니다.