

PRODUCT INTRODUCTION

Starburst Enterprise

제품 소개

Trino 기반 엔터프라이즈 레이크하우스 분석 플랫폼 — 데이터를 옮기지 않고, 있는 자리에서 분석합니다.

오픈소스 Trino 엔진 위에 보안 · 거버넌스 · 성능 가속 · 운영 자동화 · 24x7 지원을 더한 상용 배포판

제품군

Starburst Enterprise(SEP) ·
Starburst Galaxy

쿼리 엔진

Trino — 오픈소스 분산 MPP SQL

핵심 가치

페더레이션 · Iceberg 네이티브 · 보안/거
버넌스 · Warp Speed

문서 기준일

2026-07-17

Data Dynamics Inc · Starburst 파트너

무엇이고, 왜 필요하고, 무엇을 얻는가

무엇

Starburst Enterprise(SEP)는 Trino의 상용 배포판입니다. 엔진은 오픈소스 Trino와 동일하며, 그 위에 보안·거버넌스·성능 가속(Warp Speed)·운영 자동화·지원 SLA 레이어를 지원합니다.

왜

데이터가 레이크·DW·RDBMS·SaaS에 흩어져 있고, 레거시 Hadoop/Hive는 인터랙티브 분석에 구조적으로 느립니다. 규제 산업은 세밀한 접근 제어·감사·패치 SLA까지 요구합니다.

얻는 것

데이터 이동 없는 단일 SQL 분석(50+ 커넥터), Hive 대비 최대 22배 쿼리 성능, Warp Speed로 추가 5배 가속, 그리고 컴플라이언스 리스크와 운영 부담을 구조적으로 감소시킬 수 있습니다.

최대 22배

TPC-DS 쿼리 성능 (vs Hive)

50+

데이터 소스 커넥터

84%

2024년 Trino 커밋 기여

24×7

지원 · 30분 SLA · 전담 TAM

※ 성능 수치는 Starburst 공개 벤치마크 기준이며 워크로드에 따라 달라집니다 — 도입 전 PoC 검증을 권장합니다.

목차

Part 1	개요와 아키텍처	Trino와 Starburst의 관계, 플랫폼 아키텍처, 핵심 가치, 커넥터	04 - 07
Part 2	레이크하우스 코어	Apache Iceberg 소개와 Starburst의 Iceberg 지원, MinIO 연동	08 - 11
Part 3	성능	Warp Speed 가속 원리와 효과, Hive·클라우드 DW 대비 벤치마크	12 - 15
Part 4	보안 · 거버넌스 · 운영	BIAC/RBAC·마스킹·감사, Data Products, HA·오토스케일링, Stargate	16 - 19
Part 5	제품 구성과 버전 정책	OSS Trino vs SEP 매트릭스, LTS/STS, Enterprise vs Galaxy	20 - 22
Part 6	활용 시나리오와 사례	Hadoop/Hive 현대화, 글로벌 금융권 도입 사례, Starburst AI	23 - 25
Part 7	Data Dynamics 서비스	라이선스 · 기술지원 · 교육/PoC, 문의 및 Next Steps	26 - 27

Trino와 Starburst Enterprise의 관계

SEP는 Trino를 다시 만든 것이 아니라, Trino 위에 엔터프라이즈 레이어를 쌓은 인증 배포판입니다. 엔진은 동일합니다.

- **Trino** — 오픈소스 분산 MPP SQL 엔진(이전 PrestoSQL). 상시 코디네이터+워커 구조로 페타바이트급 데이터에 대화형 응답을 제공하며, 약 2주마다 새 버전을 릴리스합니다.
- **Starburst** — Trino 창시자들이 설립한 회사로, 2024년 Trino 코드 커밋의 약 84%를 기여한 최대 Trino 전문가 조직입니다.
- **동일한 엔진** — SEP N-e는 업스트림 Trino N과 동일한 엔진 — SQL 시맨틱, 비용 기반 옵티마이저(CBO), 커넥터 SPI가 모두 같습니다.
- **확장** — 그 위에 보안(BIAC·Ranger), 전용 커넥터, Warp Speed 가속, 관리 UI·Insights, 지원되는 배포 경로, 24x7 지원 계약이 추가됩니다.

SEP = Trino 엔진 + 엔터프라이즈 레이어

지원 · 릴리스 24x7 지원 · 30분 SLA · 전담 TAM · LTS/STS

운영 · 배포 Insights · HA · 오토스케일링 · K8s/OpenShift/Ansible

보안 · 거버넌스 BIAC · RBAC/ABAC · 마스크 · 감사 · Data Products

성능 · 커넥터 Warp Speed · 자동 테이블 유지관리 · 전용 커넥터

Trino 엔진 (동일) ANSI SQL · CBO · MPP 코디네이터+워커

※ 멘탈 모델: Trino는 엔진, Starburst는 그 엔진을 둘러싼 플랫폼. 순수 엔진 성능은 같고, 차이는 엔진을 둘러싼 레이어에서 나옵니다.

단일 쿼리 엔진, 다양한 데이터 소스

데이터를 옮기지 않고 Starburst가 직접 각 소스에 접근해 분산 쿼리를 수행합니다. 한 번의 SQL로 레이크하우스·데이터베이스·SaaS를 JOIN합니다.



Why Starburst — 4가지 핵심 가치

데이터 페더레이션

데이터를 이동·복제하지 않고 여러 소스를 단일 SQL로 JOIN합니다. ETL로 미리 옮겨 모으는 구조 자체가 필요 없어집니다.

50+ 커넥터 · Stargate 클러스터 간 페더레이션

Iceberg 네이티브

히든 파티셔닝·파티션 진화·타임트래블·ACID를 완전 지원하고, SEP는 컴팩션·스냅샷·만료 등 테이블 유지관리를 자동화합니다.

레이크하우스의 결정적 SQL 엔진

성능 가속 · 안정성

Warp Speed 스마트 인덱싱·캐싱으로 쿼리 최대 5배·CPU 10배 개선. 내결함 실행(FTE)은 최대 60TB 규모까지 하드닝되어 있습니다.

TPC-DS 기준, 워크로드에 따라 상이

세밀한 보안 정책

행/열 단위 RBAC·ABAC, 동적 컬럼 마스킹, 로우 필터링, Apache Ranger 통합, 쿼리 감사까지 — 규제 산업 요건을 플랫폼에서 충족합니다.

금융 · 의료 · 공공에서 결정적

커넥터 — 엔터프라이즈 데이터의 긴 꼬리까지

일반적인 엔터프라이즈의 데이터는 Oracle·Teradata·SAP·Salesforce 같은 곳에 있습니다. SEP 전용 커넥터가 라이선스 가치의 상당 부분을 차지하는 이유입니다.

SEP 전용 커넥터 (업스트림 Trino에 없음)

Snowflake · Teradata · IBM Db2 · SAP HANA · Salesforce · Azure Synapse · Greenplum · Netezza · Cosmos DB · DynamoDB · Splunk · Neo4j · MaxCompute · Generic JDBC · Stargate(클러스터 간 페더레이션)

업스트림 커넥터의 개선판 (푸시다운 강화 · Warp Speed 통합 · 추가 인증)

Hive(Warp Speed) · Iceberg(Warp Speed) · Delta Lake · Oracle · BigQuery · SQL Server · PostgreSQL · MySQL · MongoDB · Kafka · Elasticsearch · Redshift · ClickHouse · Vertica 외

업스트림 Trino와 동일

DuckDB · Druid · Pinot · Prometheus · Redis · OpenSearch · Google Sheets · TPC-H/DS 등 커뮤니티 커넥터 전체

※ 판단 기준: "S3 + PostgreSQL 위의 Trino"만으로 요건이 커버되면 OSS로 충분합니다. 아키텍처에 Snowflake·Oracle·Teradata·SAP·Salesforce가 등장하는 순간 SEP가 필요합니다.

Apache Iceberg — 레이크하우스의 개방형 테이블 포맷

Iceberg는 오브젝트 스토리지 위의 파일 더미를 데이터베이스 테이블처럼 다루게 하는 오픈 테이블 포맷입니다. 특정 엔진·벤더에 종속되지 않는 개방형 표준으로, 레이크하우스의 사실상 표준이 되었습니다.

Iceberg 테이블의 메타데이터 구조

카탈로그 (Catalog)

현재 metadata 파일 포인터 — HMS · REST · Glue



metadata.json

스키마 · 파티션 스펙 · 스냅샷 목록



Manifest List

스냅샷별 매니페스트 목록 + 파티션 범위 통계



Manifest File

데이터 파일 목록 + 컬럼 min/max 통계 → 파일 단위 프루닝



Data Files

Parquet · ORC · Avro — 오브젝트 스토리지(S3 · MinIO)

- **ACID 트랜잭션** — 스냅샷 기반 원자적 커밋 — 쓰기 도중에도 읽기가 일관되고, 동시 쓰기가 안전합니다.
- **히든 파티셔닝 · 파티션 진화** — 파티션을 물리 디렉터리가 아닌 컬럼 transform으로 정의 — 원본 컬럼 필터만으로 자동 프루닝, 파티션 전략 변경도 재작성 0.
- **스키마 진화** — 컬럼 추가·삭제·이름 변경·타입 확장을 데이터 재작성 없이 안전하게.
- **타임트래블 · 롤백** — 과거 스냅샷 시점으로 조회(FOR VERSION/TIMESTAMP AS OF)하고, 잘못된 쓰기는 스냅샷 롤백으로 복구.
- **엔진 중립 · 개방형 스펙** — Trino/Starburst · Spark · Flink 등 여러 엔진이 같은 테이블을 공유 — 포맷 스펙 (v1/v2/v3)이 공개되어 벤더 종속이 없습니다.

※ Hive 테이블 포맷의 구조적 한계(디렉터리 파티셔닝 · 비원자적 쓰기 · 느린 파티션 나열)를 테이블 포맷 차원에서 해결한 것이 Iceberg입니다.

Starburst의 Iceberg 지원 — 파티션의 벽을 넘고, 운영까지 자동화

히든 파티셔닝·파티션 진화는 Trino + Iceberg의 공통 이점입니다. SEP는 그 이후 남은 운영 숙제(스몰파일·스냅샷·통계)를 자동화합니다.

Trino + Iceberg 공통 이점

- **히든 파티셔닝** — 파티션을 디렉터리가 아닌 컬럼 transform(day·bucket·truncate)으로 정의 — 분석가가 원본 컬럼으로 필터만 걸어도 자동 프루닝, 무심코 풀스캔이 사라집니다.
- **파티션 진화** — 데이터 재작성 없이 파티션 전략 변경 — 구·신 스펙이 혼재해도 한 쿼리가 투명하게 처리.
- **매니페스트 프루닝** — 파일 단위 min/max 통계로 프루닝 — RDBMS 메타데이터 병목 없이 수백만 파일로 확장.
- **ACID · 타임트래블** — 스냅샷 기반 트랜잭션과 시점 조회, 스키마 진화.

SEP가 더 지원하는 것 — 자동 테이블 유지관리

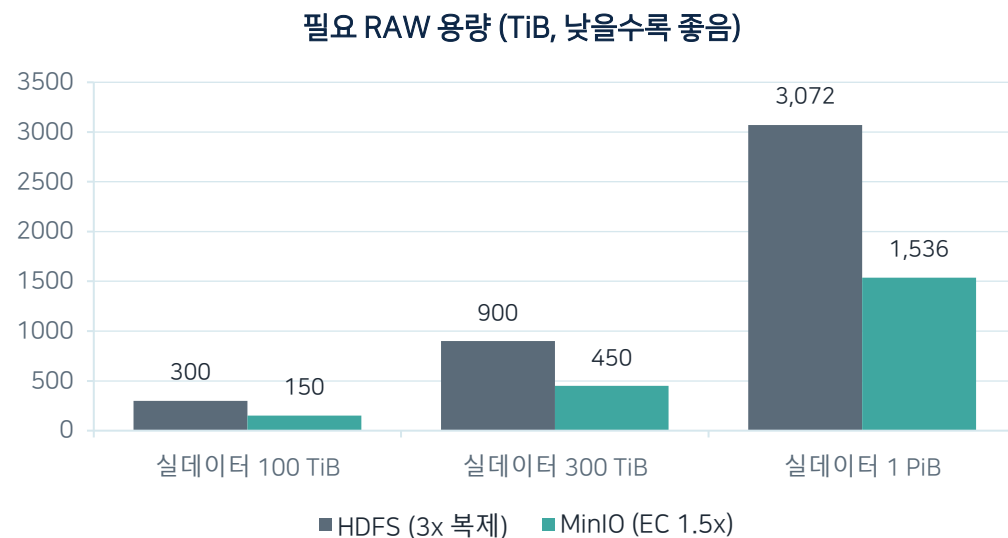
- **자동 컴팩션** — optimize(스몰파일 병합)를 스케줄 기반으로 자동 실행.
- **스냅샷 · 오편 파일 정리** — expire_snapshots · remove_orphan_files 자동화로 메타데이터 비대화 방지.
- **통계 자동 수집** — CBO가 쓰는 테이블 통계를 최신으로 유지.
- **스트리밍 인제스트 · 파일 로더** — Kafka 등에서 Iceberg 테이블로의 연속 적재를 관리형으로 제공 — 별도 Spark/Flink 파이프라인 불필요.

※ OSS Trino에서도 같은 명령을 수동/스케줄로 실행할 수 있지만, 운영 책임은 플랫폼 팀의 몫입니다. SEP는 이를 관리형·자동화로 제공합니다.

MinIO 연동 — 오브젝트 스토리지 레이크하우스

MinIO는 S3 API 100% 호환 고성능 오브젝트 스토리지입니다. 온프레미스에서 Starburst + Iceberg 레이크하우스의 스토리지 레이어를 담당하며, HDFS의 구조적 병목을 제거합니다.

구분	HDFS	MinIO (S3 호환)
위치 조회	NameNode 중앙 관리 — 단일 병목	객체 키로 직접 접근 — 중앙 마스터 없음
읽기 방식	DataNode 블록 읽기(로컬리티 의존)	HTTP GET · Range Request 병렬 읽기
스토리지 효율	3x 복제 — RAW의 33%만 사용 가능	Erasure Coding — RAW의 67% 사용 가능
확장 단위	Compute + Storage 결합 증설	Storage 독립 수평 확장
스물파일	NameNode 메모리 부하 · GC	객체 단위 관리 — 내성 확보



- **스토리지 50% 회수** — 같은 데이터를 절반의 RAW 용량으로 — 예: 900 TiB HDFS(실데이터 300 TiB)를 EC 기준 약 450 TiB로 전환.
- **"로컬이 원격보다 빠르다"는 오해** — 로컬리티 이득 < NameNode 병목·강결합 손실 — 매니페스트 프루닝 + Range read + (SEP) Warp Speed 캐시로 원격 IO를 상쇄합니다.

※ Kubernetes 네이티브 · Active-Active 멀티사이트 복제(DR) · Bitrot 감지 — 전담 Hadoop 운영팀 없이 자동 복구·확장이 가능합니다.

MinIO 연동 구성 — 코드 변경 없이 스토리지만 교체

표준 S3 API라 별도 플러그인이 없습니다. Starburst 카탈로그에 엔드포인트를 등록하고, 기존 Hadoop 잡은 S3A 커넥터로 경로(hdfs:// → s3a://)만 바꾸면 됩니다.

구성 항목	설정 포인트	핵심 값
Starburst Iceberg/Hive 카탈로그	카탈로그 properties에 S3 엔드포인트·키 등록	hive.s3.path-style-access=true
Hadoop S3A 커넥터	core-site.xml — 기존 잡 코드 변경 없이 연결	fs.s3a.endpoint · fs.s3a.committer.name=magic
데이터 이관	Trino CTAS · DistCp · s5cmd 병렬 복사	external_location = 's3a://datalake/...'
보안 통합	Ranger → MinIO IAM 정책 이전, TLS·SSE 암호화	AD/LDAP · OIDC 직접 연동 (Kerberos 불필요)

예시 — Trino CTAS로 HDFS 테이블을 MinIO로 이관

```
CREATE TABLE iceberg.analytics.transactions
WITH (format = 'PARQUET',
      external_location = 's3a://datalake/transactions/')
AS SELECT * FROM hive.legacy_hdfs.transactions;
```

AIStor Tables — Iceberg 카탈로그 내장

MinIO AIStor는 Iceberg 카탈로그를 스토리지에 내장해 기존 4레이어(메타 DB·카탈로그 서비스·스토리지·엔진) 스택을 2레이어로 단순화 — Starburst가 REST 카탈로그로 바로 연동합니다.

→ 권장 전환 순서: 병렬 클러스터 구축 → 데이터 이전(distcp·이중 수집) → 보안 전환 → HDFS 노드 순차 해제 — 무중단 단계적 전환이 가능합니다.

Warp Speed — 스마트 인덱싱 · 캐싱의 동작 원리

Hive·Iceberg·Delta 커넥터 위에 투명하게 얹히는 가속 레이어입니다. 쿼리 패턴을 학습해 인덱스와 캐시를 자율적으로 만들고 유지합니다 — 수동 인덱스 설계가 필요 없습니다.

Hot 각 워커의 NVMe SSD

인덱스·캐시 히트 — 원격 IO 없이 로컬에서 서빙

Warm 워커가 공유하는 오브젝트 스토리지 버킷

스케일업 시 재인덱싱 최소화 · 복원력 확보

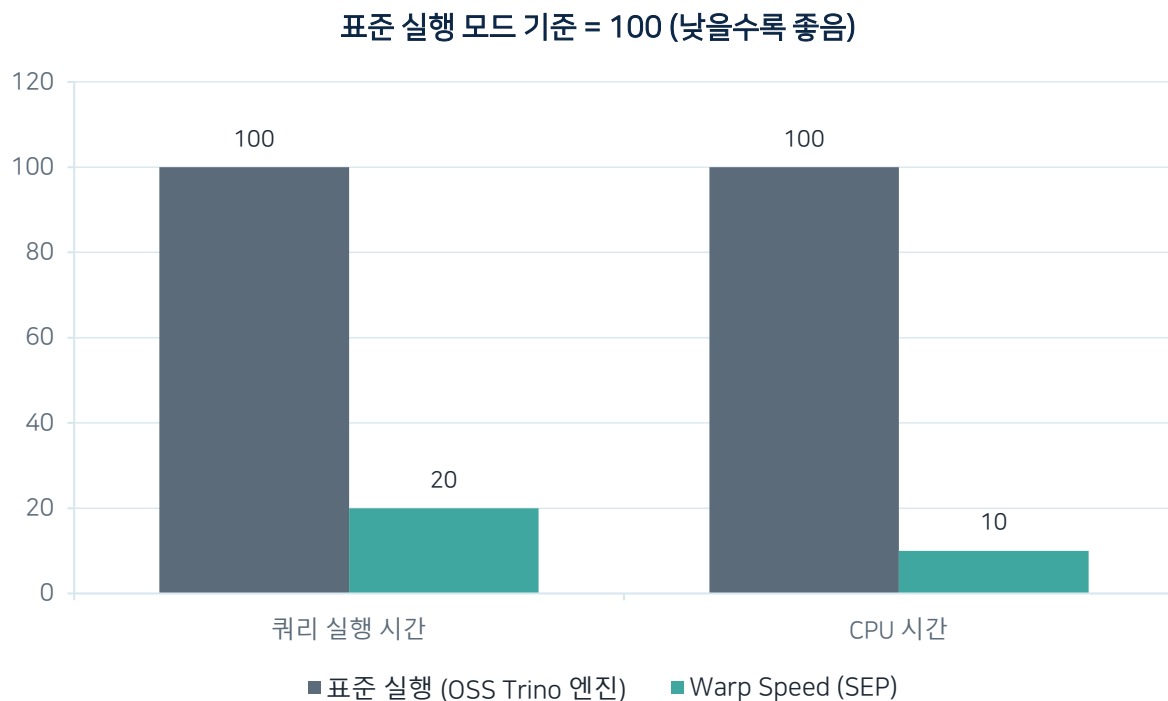
Cold 원본 S3 · MinIO · ADLS

첫 접근 시 여기서 읽고, 인덱스·캐시는 비동기 생성

- **블록 단위 스마트 인덱싱** — 데이터를 블록으로 나누고 블록별 최적 인덱스 타입(bitmap · dictionary · tree)을 자동 선택 — 조건에 맞지 않는 블록·행을 통째로 스킵합니다.
- **자율 캐싱** — 자주 접근하는 컬럼·범위를 로컬 SSD에 캐싱하고, 워크로드 빈도·우선순위에 따라 동적으로 갱신합니다.
- **왜 빨라지는가** — 인덱스 프루닝이 처리할 일 자체를 줄여 CPU 시간을, 로컬 SSD가 원격 IO를 없애 체감 실행 시간을 각각 줄입니다 — "CPU 10배 · 시간 5배" 비대칭의 이유.
- **한계** — 콜드(첫 접근) 쿼리는 이득이 없고, 이미 파티션이 잘 잡힌 대형 풀스캔은 이득이 제한적입니다. 선택적(selective)·반복적 워크로드일수록 효과가 큼니다.

Warp Speed 효과 — SEP 전용 가속의 실측치

엔진이 같아서 가속을 끄면 OSS Trino와 숫자가 동일합니다. 따라서 의미 있는 성능 차이는 곧 Warp Speed의 효과입니다. (TPC-DS SF1000 · Iceberg)



5배

빠른 쿼리 실행 시간 — 표준 실행 대비

10배

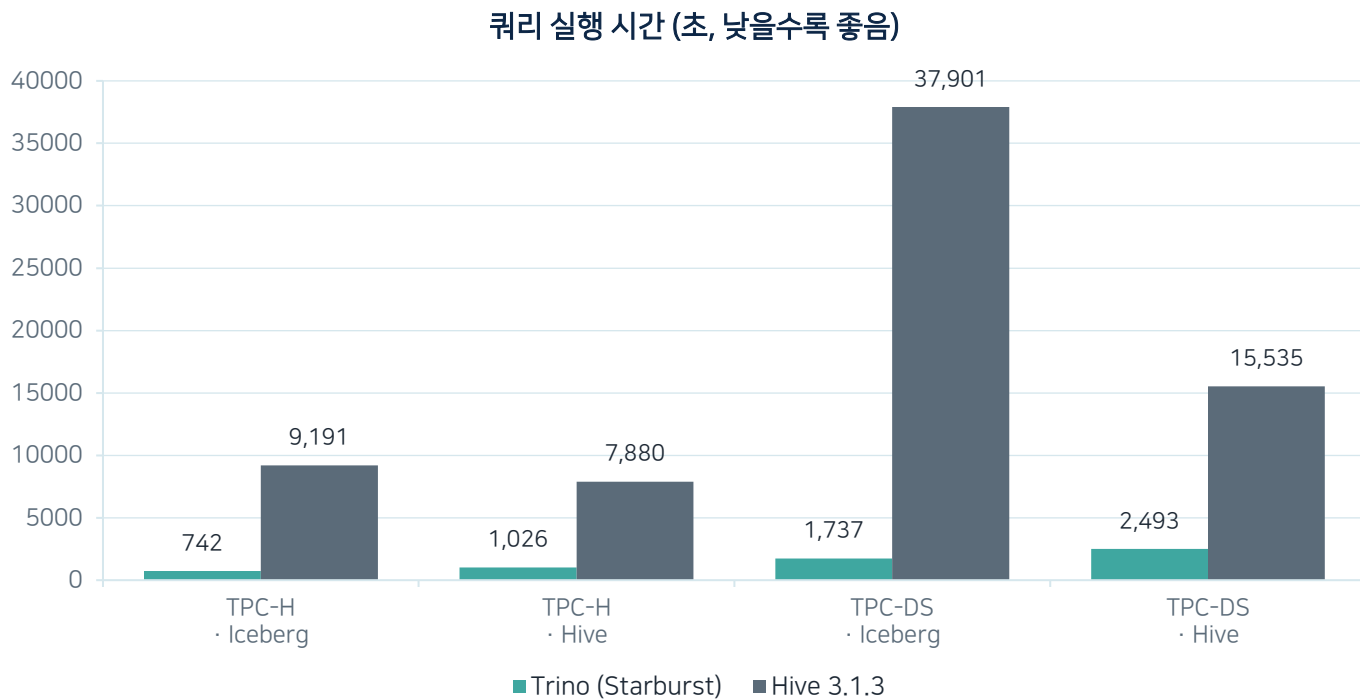
절감된 CPU 시간 — 표준 실행 대비

- 인덱싱 가속 — 쿼리 최대 6~7배(최대 700%) 향상
- 반복 유사 서브쿼리 — 최대 62% 향상
- 인터랙티브 — 평균 40% 가속 · 비용 최대 40% 절감
- 실사용 — El Toro 300% · Jungle Games 30% 향상

※ 모두 Starburst 자사 벤치마크이며 워크로드 성격에 좌우됩니다. 반복적·선택적 adhoc/BI 조회에서 효과가 크고, 절대 수치는 PoC 검증을 권장합니다.

벤치마크 — Hive 대비 최대 22배

같은 데이터를 Hive 엔진과 Starburst Enterprise로 각각 실행한 벤치마크입니다. 1TB 원본(S3 · Parquet), Starburst 476-e vs Hive 3.1.3.



가장 극적인 사례

10.5시간 → 29분

TPC-DS · Iceberg — 37,901초 → 1,737초 (약 22배)

벤치마크	배수
TPC-H · Iceberg	약 12배
TPC-H · Hive 테이블	약 8배
TPC-DS · Iceberg	약 22배
TPC-DS · Hive 테이블	약 6배

- 점진 전환 논리 — 엔진만 바뀌어도 6~8배, 이후 Iceberg로 이관하면 12~22배. 빅뱅 없이 단계적으로 이득을 키울 수 있습니다.

※ 테스트 조건: TPC-H 9노드 / TPC-DS 17노드(각 32 vCPU·128GB), Apache JMeter 4회 순차 루프. 출처: Concurrency Labs 벤치마크 리포트(2025-09).

벤치마크 — 클라우드 데이터 웨어하우스 대비 가격·성능

오픈 테이블 포맷 위 SQL 실행 기준, Starburst는 최대 6.3배 빠른 쿼리와 최대 12.7배 비용 절감을 주장합니다. Runtime·Cost 모두 낮을수록 좋습니다.

엔진	Runtime (분)	Cost (USD)
Amazon Redshift	325	470
Google BigQuery	125	400
Snowflake	55	185
Amazon Athena	150	65
Databricks SQL Serverless	38	60
Databricks SQL Pro	55	70
Starburst Galaxy	50	30

읽는 법

Runtime과 Cost가 둘 다 최하위권인 것은 Starburst뿐입니다. Databricks SQL Serverless가 Runtime은 근소하게 빠르지만, 비용까지 함께 보면 Starburst가 가장 균형 잡힌 조합입니다.

레이크하우스(오픈 포맷) 전략이라면, 데이터를 DW로 복제하지 않고도 DW급 성능·비용 효율을 확보한다는 의미입니다.

※ 테스트 조건: TPC-DS sf1000 · 동시성 5 · 전 벤더 Iceberg(단 Databricks는 Delta Lake) · 2024-06 기준. 표 값은 원본 차트에서 읽은 근사치입니다.

보안과 접근 제어 — 규제 산업의 기본기

인증은 OSS Trino도 충분합니다. 갈리는 곳은 세밀한 인가(Authorization)와 감사 — 규제 산업(금융·의료·공공)에서 SEP가 결정적인 선택의 이유입니다.

Built-in Access Control (BIAC)

SEP 네이티브 인가 모델 — 카탈로그/스키마/테이블/컬럼/로우 단위 역할 기반(RBAC)·속성 기반(ABAC) 권한을 UI로 관리.

Apache Ranger 통합

글로벌/카탈로그/컬럼/로우 레벨 — 기존 Hive/HDFS에서 쓰던 Ranger 정책을 그대로 공유.

인증 통합

LDAP · OAuth2 · JWT · Kerberos에 더해 Okta 통합, 사용자 임퍼소네이션 일반화.

컬럼 마스킹 · 로우 필터링

민감 컬럼 동적 마스킹과 행 단위 필터를 기본 제공 — 개인정보·거래정보 보호 요건 충족.

Credential Passthrough

Kerberos·패스워드 자격 증명 전달(캐싱 포함) — 워커가 장수명 서비스 계정 없이 사용자 신원으로 소스에 접근.

Query Audit

누가, 언제, 어떤 카탈로그에, 어떤 쿼리를 실행했는지 영속 기록 — 감사 추적 요건에 즉시 대응.

※ 실무 효과: 정책이 한 곳(BIAC 또는 Ranger)에 모이고, SEP가 노출하는 모든 커넥터에 균일하게 적용됩니다.

데이터 거버넌스 · 디스커버리 · 관측성

순수 쿼리 엔진(Trino)과 데이터 플랫폼(SEP)의 차이가 가장 크게 드러나는 영역 — OSS에서는 전부 별도 도구를 붙여 직접 구축해야 하는 항목입니다.

Data Products

카탈로그 위에 오너·문서·SLO·액세스 정책이 부여된 큐레이션 데이터셋을 게시 — 데이터 메시 아키텍처의 셀프서비스 마켓플레이스.

리니지 · 메타데이터

데이터 리니지 추적과 Apache Atlas 통합 — 어디서 와서 어디로 가는지 추적 가능.

품질 · 분류

데이터 품질 규칙, 스키마 변경 알림, 자동 데이터 분류(민감정보 탐지), 소스 전반 통합 검색.

Starburst Insights

쿼리 패턴·자원 사용량·슬로우 쿼리 분석 대시보드 — "클러스터가 어제 새벽 3시에 뭘 했는가"에 답합니다. CloudWatch·Stackdriver 통합.

※ OSS Trino 대비: 통합 검색·스키마 자동 발견·Data Products·리니지·품질 규칙·변경 알림·자동 분류 — 7개 항목 모두 SEP 제공, OSS 미제공.

운영 · 배포

- **High Availability** — 코디네이터와 주변 백엔드 서비스의 고가용성 구성.
- **Graceful Scaledown 오토스케일링** — 진행 중인 쿼리를 죽이지 않고 워커를 drain하며 축소 — 비용과 안정성의 균형.
- **비용 자동화** — 클러스터 자동 일시중지·유휴 시 자동 종료·클러스터 스케줄링.
- **내결함 실행(FTE) 하드닝** — 최대 60TB 규모 배치·ETL까지 엔터프라이즈급 검증 — 장시간 쿼리가 통째로 실패하지 않습니다.
- **웹 UI · 쿼리 편집기** — 쿼리 작성·이력·운영자 친화적 클러스터 뷰.

지원되는 배포 경로

Kubernetes 기본 제공 Helm 차트 · 매니페스트

Red Hat OpenShift 공식 지원

Starburst Admin 온프레미스 · VM용 Ansible 배포 도구

클라우드 Marketplace AWS · Azure · GCP 등록, Marketplace 빌링

관측 통합 CloudWatch · Google Stackdriver 로그/대시보드

※ OSS는 JMX·로그·기본 UI까지 — 그 위의 쿼리 이력 저장소, 쿼리를 존중하는 오토스케일러, 메트릭 파이프라인은 직접 만들어야 합니다. 엔지니어링 시간을 비용으로 환산하면 "직접 만들기"가 우세한 경우는 드뭅니다.

Stargate — 클러스터를 넘고, 국경을 넘는 페더레이션

한 Trino/SEP 클러스터가 다른 클러스터를 하나의 소스처럼 쿼리합니다. 데이터를 복제하지 않고 클러스터 사이로 쿼리를 라우팅하는 SEP 전용 커넥터입니다.

글로벌 ↔ 리전

글로벌 클러스터가 리전별 SEP 클러스터를 페더레이트 — 데이터는 각 소스 시스템 가까이에서 둔 채 통합 분석.

중앙 ↔ 테넌트

별도 보안이 적용된 테넌트 클러스터들 위에 중앙 클러스터가 큐레이션된 뷰를 노출.

마이그레이션 브릿지

새 클러스터에서 레거시 클러스터로 Stargate를 걸어, 전환 기간 동안 양쪽 데이터가 동일하게 보이도록 유지.

데이터 주권(Data Sovereignty) 사례

62개국에서 운영되는 글로벌 은행그룹은 Stargate로 데이터 이동 없이 국경·클라우드를 넘는 분석을 구현해 GDPR·CCPA를 준수하면서 쿼리 성능을 20배 개선했습니다. 다른 글로벌 투자은행은 100+개국의 페타바이트 데이터를 원위치(in-place)에서 질의해 AML 분석과 데이터 반출 규제를 동시에 충족했습니다.

OSS Trino vs Starburst Enterprise — 기능 매트릭스

카테고리	OSS Trino	Starburst Enterprise (SEP)	판정
핵심 쿼리 엔진	완전한 Trino 엔진 — 최신 기능 최선 반영	인증·검증된 동일 엔진 + 선제 패치·백포트	공통
보안 · 인가	파일/OPA 수동 구성, RBAC UI 없음	BIAC(행/열 RBAC·ABAC) · 마스킹 · Ranger · 감사 UI	SEP 우위
거버넌스 · 카탈로그	외부 도구 의존(OpenMetadata 등)	Data Products · 리니지 · 품질 규칙 · 탐색 포털	SEP 우위
성능 가속	CBO · dynamic filtering	+ Warp Speed(자동 인덱싱·캐싱) · 워크로드 관리	SEP 우위
테이블 유지관리	optimize 등 직접 스케줄링	컴팩션 · 스냅샷 만료 · 통계 수집 자동화	SEP 우위
커넥터	50+ 커뮤니티 커넥터(품질 편차)	인증 커넥터 + 전용(Snowflake·SAP·Salesforce 등)	SEP 우위
배포 · 운영	Helm 자체 관리	HA/DR · 오토스케일 · OpenShift · Marketplace	SEP 우위
지원	커뮤니티 Slack · GitHub	24×7 · 30분 SLA · 전담 TAM · LTS 12개월 패치	SEP 우위
비용 · 라이선스	Apache 2.0 무료 — 벤더 종속 없음	상용 구독(노드/컴퓨트 기반)	OSS 우위

※ 요약: 엔진의 최신성·개방성·비용은 OSS Trino, 보안·거버넌스·성능 가속·운영·지원 SLA는 SEP 우위 — 규제 산업과 엔터프라이즈 규모일수록 SEP가 유리합니다.

버전 정책 — LTS / STS, 1년 단위로 계획 가능한 플랫폼

Trino는 약 2주마다 릴리스됩니다. SEP는 그 위에 예측 가능한 릴리스 주기와 LTS 패치 라이프사이클을 얹어, 엔터프라이즈가 따라갈 수 있는 속도로 바꿉니다.

버전 넘버링

Trino 릴리스 → 479

SEP STS 릴리스 → 480-e

SEP LTS 릴리스 → 479-e.3

앞의 정수는 언제나 상위 Trino 버전 — 패치 접미사(.x)가 있으면 LTS, 없으면 STS입니다.

항목	LTS	STS
지원 기간	릴리스로부터 12개월	다음 STS까지
패치 릴리스	제공 (-e.0 → -e.1 → ...)	없음(다음 릴리스에 반영)
릴리스 주기	분기당 최대 1개(2·5·8·11월)	Trino를 따라감(빈번)
보안(CVE) 수정	패치로 백포트	다음 STS에서만
운영 적합도	운영 환경의 기본값	기능 주도 · 업스트림 근접

- **권장 플레이북** — 1년에 한 번 LTS → 다음 LTS로 통제된 업그레이드. 남은 지원 기간 9~12개월인 라인을 타깃으로, breaking changes를 라인 단위로 처리합니다.
- **EOL은 하드 데드라인** — 지원 종료 후에는 CVE 패치가 제공되지 않습니다 — 어느 시점이든 약 4개의 LTS 라인이 겹쳐 지원되어 업그레이드 활주로가 확보됩니다.

제품군 — Starburst Enterprise vs Starburst Galaxy

같은 엔진 패밀리, 다른 운영 소유권 경계 — 운영 경계를 어디에 두고 싶은지로 선택합니다.

Starburst Enterprise (SEP)

고객 인프라(온프레미스 · Private VPC)에 직접 배포하는 자가관리형 상용 배포판. 데이터 통제권을 유지하면서 엔터프라이즈 보안·운영 기능을 활용합니다.

- ✓ Self-managed — K8s · OpenShift · Ansible
- ✓ LTS/STS 버전 정책, 업그레이드는 운영자 소유
- ✓ 데이터 종력 · 망분리 · 규제 거주 요건에 적합
- ✓ 심층 커스터마이징 가능

Starburst Galaxy

AWS · Azure · GCP에서 동작하는 완전 관리형 SaaS. 클러스터 자동 스케일링과 페이-퍼-쿼리, 크로스클라우드 페더레이션으로 운영 부담 없이 즉시 시작합니다.

- ✓ Fully Managed — Starburst가 운영·업그레이드
- ✓ 오토스케일링 · Pay-per-query
- ✓ 멀티 클라우드 페더레이션
- ✓ 운영 부담을 0에 가깝게

※ 온프레미스·프라이빗 환경 중심의 국내 엔터프라이즈(특히 규제 산업)에는 SEP가, Cloud First 조직에는 Galaxy가 일반적인 출발점입니다.

대표 시나리오 — 레거시 Hadoop / Hive 플랫폼 현대화

Hive의 느린 adhoc 쿼리는 튜닝 부족이 아니라, 배치 엔진을 인터랙티브 용도로 쓰는 구조적 미스매치가 원인입니다. Starburst + Iceberg + 오브젝트 스토리지는 이를 아키텍처 차원에서 제거합니다.

문제 (Hive + HDFS)	근본 원인	Starburst + Iceberg 해결
콜드 스타트	매 쿼리 MR/Tez 컴파일 + YARN 할당	상시 워커 즉시 실행 — 잡 기동 자체가 없음
파티션 프루닝 실패	파티션 = 물리 디렉터리, 명시 필터 필요	히든 파티셔닝 — 원본 컬럼 필터만으로 자동 프루닝
메타데이터 병목	HMS(RDBMS) 행 단위 파티션 나열	매니페스트 파일 단위 프루닝, RDBMS 병목 없음
경직된 확장 · 3배 복제	Compute·Storage 강결합(HDFS 3x)	분리 아키텍처 — 워커만 탄력 증감, EC 스토리지
EOL · CVE 리스크	레거시 배포판 지원 종료	LTS 12개월 CVE 백포트 · 24×7 벤더 SLA

1. 병행(Coexist)

기존 Hive 메타스토어에 연결, 읽기부터 — 데이터 이동 0

2. 오프로드(Offload)

무거운 adhoc/BI 쿼리 이관 — 즉시 성능 체감

3. 현대화(Modernize)

핵심 테이블 Iceberg 전환 — 이득 극대화

4. 정착(Consolidate)

레거시 축소, 보안·거버넌스·오토스케일 정착

→ 기존 Hive 데이터를 그대로 질의할 수 있어 빅뱅 없이 단계마다 KPI·롤백 포인트를 두고 전환합니다.

글로벌 도입 사례 — 금융권을 중심으로

Hive/Hadoop·Teradata·DB2 등 레거시에서 Trino 기반 Starburst로 전환·페더레이션한 실제 사례입니다.

HDFC Bank

인도 최대 민간은행 — 레거시 DB2 → 개방형 아키텍처. Azure AKS·온프레 OpenShift 병행.

클라우드·온프레 통합 거버넌스 · DB2 비용 절감

Crédit Mutuel Arkéa

프랑스 — 온프레미스를 유지한 채 Hadoop 위 페더레이션 쿼리 엔진 구축.

Hadoop 쿼리 10~20배 향상 · 1년 내 550+ 사용자

Bank Hapoalim

이스라엘 — 20년 Teradata를 데이터레이크로. Hive·Spark 벤치마크 후 Starburst 채택.

고동시성 ad-hoc + BI 쿼리 성능 최우수

Banco Inter

브라질 — Amazon Athena → Starburst. 탐색·거버넌스·셀프서비스 분석 단일 플랫폼화.

인사이트 도출 days → seconds · 월 \$100K 절감

글로벌 투자은행

Hive를 Starburst로 대체 — 데이터 이동 없이 100+개국 원위치 (in-place) 쿼리, AML 분석.

데이터 주권 준수 · 규제 과징금 리스크 감소

글로벌 은행그룹 (62개국)

Stargate 기반 국경 간 분석 — 데이터 이동 없이 GDPR·CCPA 준수.

쿼리 성능 20배 향상 · 데이터 주권 준수

※ 이 외 Starburst 공식 사례: El Toro 쿼리 성능 300% 향상 · Talkdesk 에러율 150배 감소 · Kovi 애드혹 쿼리 85% 단축. (출처: Starburst Customer Use Case)

Starburst AI — AI-Ready 데이터 플랫폼으로의 확장

거버넌스가 적용된 단일 액세스 포인트 위에서, 정형·비정형 데이터를 AI 워크로드에 바로 연결하는 방향의 로드맵입니다.

AI Search

정형·비정형 데이터를 벡터 임베딩으로 Iceberg에 저장·검색 — 별도 Vector DB 없이 레이크하우스에서 시맨틱 검색.

AI SQL Functions

prompt() · classify() · mask() 등 LLM 함수를 SQL 안에서 직접 호출 — 분석가가 쓰던 언어 그대로 AI를 활용.

AI Model Access Management

모델 접근 제어 · 사용량 모니터링 · 감사 로그 — 데이터와 같은 거버넌스 체계로 AI 모델을 관리.

성숙도 주의 Starburst AI 기능군은 상당수가 프라이빗 프리뷰 단계이며 SEP 버전·시점에 따라 가용성이 다릅니다. 즉시 프로덕션 기능이라기보다 "AI-Ready 기반 확보 + 로드맵" 관점으로 접근하고, PoC 시점에 GA 상태를 확인하는 것을 권장합니다.

Data Dynamics가 제공하는 서비스

Starburst 파트너로서 라이선스부터 구축·운영·교육까지, 도입 전 주기를 지원합니다.

라이선스 판매

Starburst Enterprise·Galaxy 라이선스 견적 및 발주 지원.

- ✓ 용량·노드 기반 산정
- ✓ PoC 라이선스 지원
- ✓ 갱신·확장 컨설팅

기술지원

구축부터 운영까지 Starburst 전반에 대한 기술지원.

- ✓ 설치·업그레이드 지원
- ✓ 쿼리 성능 튜닝
- ✓ 장애 대응·에스컬레이션

교육 · PoC

Trino·Starburst 기술 교육 및 맞춤 PoC 수행.

- ✓ 관리자 과정
- ✓ 데이터 분석가 교육
- ✓ 맞춤 PoC 설계·수행

■ **Trino 생태계 전문성** — Trino·Iceberg·오브젝트 스토리지 기반 레이크하우스 구축과 Hive/Impala 마이그레이션 실무 경험을 보유하고 있으며, 관련 기술 자료를 홈페이지 블로그에 지속 공개하고 있습니다.

☎ 문의: 유지관 · 010-2356-5921 · trendy.ryu@data-dynamics.io

NEXT STEPS

결론 및 다음 단계

Starburst Enterprise는 "더 빠른 쿼리 엔진"이 아니라, 데이터가 있는 자리에서 분석하는 개방형 레이크하우스 전략의 엔터프라이즈 완성형입니다.

엔진의 성능은 오픈소스 Trino와 같고 — 보안·거버넌스·가속·운영·지원이라는, 프로덕션을 가르는 다양한 기능이 추가됩니다.

1

요건 정의 · PoC 설계

워크로드 인벤토리, 사용자·동시성·SLA 기반 규모 산정, 1TB TPC-DS 벤치마크를 포함한 PoC 설계

2

PoC 수행 · 검증

고객 데이터 기반 성능·보안·거버넌스 검증 — Warp Speed 효과와 비용 모델을 실측

3

단계적 도입

병행 → 오프로드 → 현대화 → 정착의 점진 로드맵으로 리스크를 통제하며 전환

Data Dynamics Inc

문의: 유지관 · 010-2356-5921 · trendy.ryu@data-dynamics.io · www.data-dynamics.io