

TECHNICAL REPORT BD-TR-2026-02 * BASIS DESK RESEARCH DESK

Cross-Venue Entity Resolution, Regulatory Filing Ingestion, and On-Chain Whale Flow Attribution at the Edge

Deterministic Knowledge-Graph Linking and Multi-RPC Block Scanning for Institutional Market Intelligence

Authors: Julian Mercer (ORCID: 0009-0003-1162-170X) and Basis Desk Research Desk

Published: 2026-10-10 | JEL: G12, G28, K22 | License: CC BY 4.0

Canonical HTML: <https://basisdesk.news/research/bd-tr-2026-02> | PDF: <https://basisdesk.news/papers/BasisDesk-TR-2026-02.pdf>

ABSTRACT

Tracking institutional and regulatory developments in digital asset markets requires linking off-chain legal documents (such as SEC Form 8-K filings, MiCA supervisory guidelines, and court orders) with real-time on-chain capital flows and spot market microstructure. This study details the zero-LLM runtime entity resolution and on-chain telemetry architecture deployed at Basis Desk (<https://basisdesk.news>). We describe (1) a deterministic entity-matching grammar that maps news coverage and primary citations onto canonical regulator, statute, corporate treasury, and topic hubs with Wikidata alignment, and (2) a fault-tolerant, multi-RPC Ethereum and Bitcoin whale scanner that detects and labels transfers $\geq \$5,000,000$ USD across exchange hot/cold wallets, custodians, and treasury addresses.

Keywords: entity resolution; on-chain whale tracking; SEC EDGAR 8-K; regulatory intelligence; blockchain analytics; knowledge graph

1. Deterministic Entity Hubs and Primary Document Graphs

To eliminate runtime hallucination and latency in topic aggregation, Basis Desk maintains a curated registry of canonical entities spanning regulators (SEC, CFTC, ESMA, FCA, U.S. Treasury, Federal Reserve), statutory frameworks (MiCA, GENIUS Act), public corporate treasuries (Strategy, BlackRock, Coinbase, Tether, Circle), and market themes (crypto ETFs, tokenization, DeFi, prediction markets).

Each hub at <https://basisdesk.news/entities/<slug>> resolves coverage via SQL pre-filtering coupled with context-sensitive regular expressions (e.g., disambiguating Strategy's STRK perpetual preferred stock from Starknet's STRK token) and automatically extracts primary documents originating from the entity's official domain.

2. Fault-Tolerant Multi-RPC On-Chain Whale Telemetry

Public blockchain RPC endpoints exhibit high variance in rate limits, head lag, and full-block availability when accessed from edge worker egress IPs. Our Ethereum scanner rotates across a health-scored pool of public JSON-RPC nodes with per-isolate 5-minute cooldowns, capability-aware routing for full-transaction block fetches vs ERC-20 Transfer log scans, and a 3-block finality buffer.

Transfers of native ETH, BTC (≥ 100 BTC per output), and major ERC-20 assets (USDT, USDC, DAI, WBTC, stETH, LINK, UNI, AAVE, PEPE) exceeding \$5,000,000 USD are matched against an address-label registry to classify directional flow: `exchange_in` (potential sell supply), `exchange_out` (custody accumulation), internal treasury sweeps, or unlabelled whale movements.

3. Automated Market-Close Synthesis and Spot Microstructure

Every 60 seconds, the market engine aggregates spot quotes across Coinbase, Kraken, and

DefiLlama with Coinlore/CoinGecko rank synthesis, Ethereum EIP-1559 fee-history percentiles, Bitcoin mempool sat/vB targets, and the Crypto Fear & Greed Index.

At 00:05 UTC daily, the deterministic Market Close job compiles a zero-LLM data report (<https://basisdesk.news/crypto-market-today>) combining 24-hour asset returns, top-50 market breadth, and net exchange whale flows.

4. Data Availability and Reproducibility

Live whale telemetry is published at <https://basisdesk.news/data/whales> and exposed programmatically via the whale_transfers MCP tool at <https://basisdesk.news/mcp>. Historical article-to-source mappings are included in the public research exports at <https://basisdesk.news/api/research/dataset.csv>.

References and Live Reproducibility Endpoints

- [1] Basis Desk Topics, Regulators & Primary Document Hubs (2026).
<https://basisdesk.news/entities>
- [2] Basis Desk On-Chain Whale Tracker: \$5M+ Labelled Transfers (2026).
<https://basisdesk.news/data/whales>
- [3] Basis Desk Crypto Market Today & Daily Close Archive (2026).
<https://basisdesk.news/crypto-market-today>
- [4] Basis Desk Live Markets & Methodology (2026). <https://basisdesk.news/markets>

Appendix: Linked Basis Desk Verification & Data Endpoints

- * Basis Desk Home & Live Newsroom: <https://basisdesk.news/>
- * Research Portal & Longitudinal Datasets: <https://basisdesk.news/research>
- * Source-Verified Crypto News Feed: <https://basisdesk.news/news>
- * Research-Desk Deep-Dive Analysis: <https://basisdesk.news/analysis>
- * Regulator, Statute & Company Hubs: <https://basisdesk.news/entities>
- * Evergreen Crypto Explainers (300 Topics): <https://basisdesk.news/learn>
- * On-Chain Whale Tracker (\$5M+ Transfers): <https://basisdesk.news/data/whales>
- * Basis Desk Verify (Chrome & Firefox Extension): <https://basisdesk.news/tools/verify>
- * Smart-Contract Token Safety Checker: <https://basisdesk.news/tools/token-checker>
- * Model Context Protocol (MCP) Server for AI Agents: <https://basisdesk.news/developers>
- * Machine-Readable Dataset (CSV): <https://basisdesk.news/api/research/dataset.csv>
- * Machine-Readable Dataset (JSONL): <https://basisdesk.news/api/research/dataset.jsonl>
- * Frictionless Data Package (JSON): <https://basisdesk.news/api/research/datapackage.json>
- * BibTeX Citation Record: <https://basisdesk.news/api/research/cite.bib>