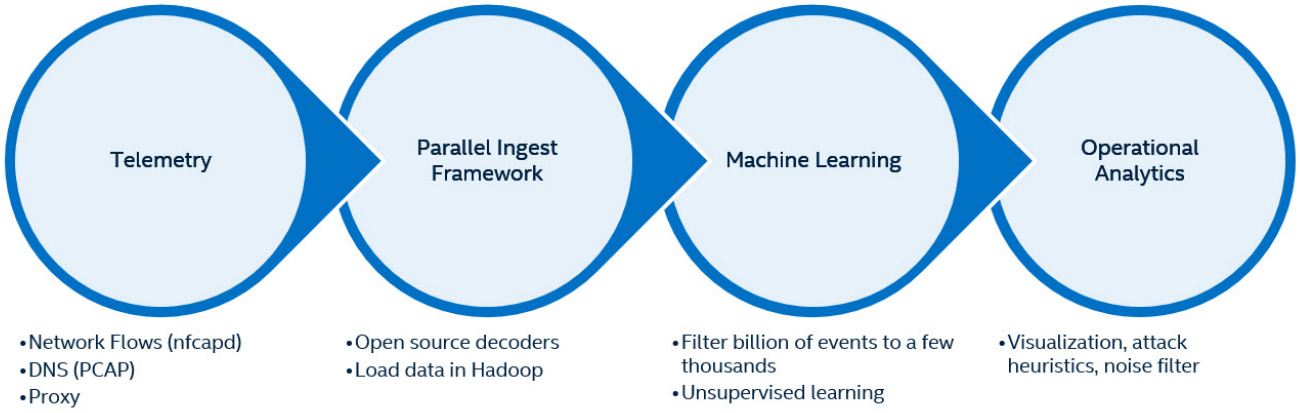
Technical Overview

Apache Spot is a solution built to leverage strong technology in both “big data” and scientific computing disciplines. While the solution solves problems end-to-end, components may be leveraged individually or integrated into other solutions. All components can output data in CSV format, maximizing interoperability.



Parallel Ingest Framework. The system uses decoders optimized from open source, that decodes binary flow and packet data, then loading the data in HDFS and data structures inside Hadoop. The decoded data is stored in multiple formats so it is available for searching, used by machine learning, transfer to law enforcement, or inputs to other systems.

Machine Learning. The system uses a combination of Apache Spark and optimized C code to run scalable machine learning algorithms. The machine learning component works not only as a filter for separating bad traffic from benign, but also as a way to characterize the unique behavior of network traffic in an organization.

Operational Analytics. In addition to machine learning, a proven process of context enrichment, noise filtering, whitelisting, and heuristics are applied to network data to produce a short list of the most likely patterns, which may be security threats.