

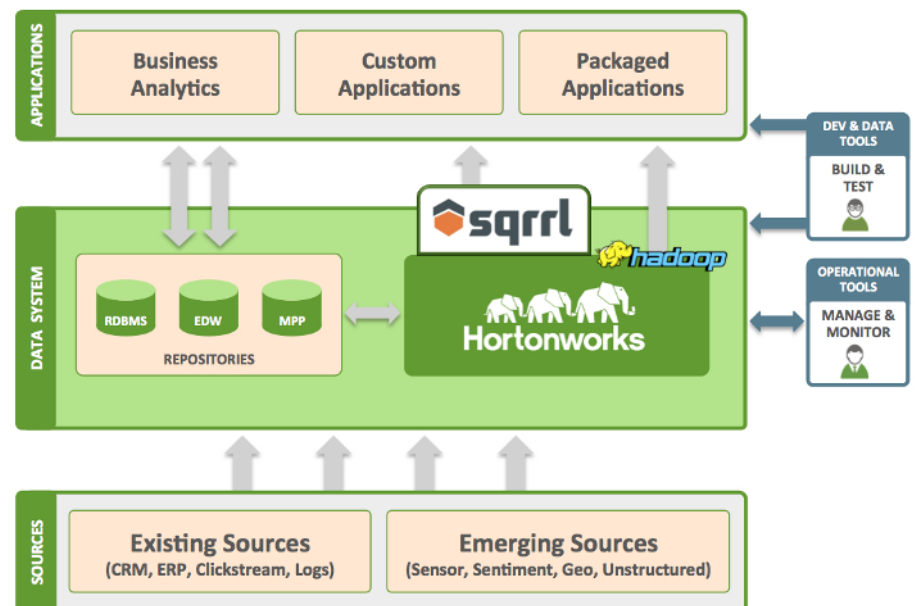
## Reference Architecture

### Hortonworks Data Platform and Sqrrl Enterprise

#### Executive Overview

Sqrrl Enterprise and the Hortonworks Data Platform (HDP) are complementary technologies that provide advanced secure data processing capabilities for Apache Hadoop. HDP provides the Hadoop platform and Sqrrl Enterprise provides a secure, interactive data layer. Apache Accumulo is an alternative to HBase in HDP. The benefits of this solution are better security, better performance, and richer data storage and query capabilities.

The diagram shows how Sqrrl and Hortonworks fit into a Modern Data Architecture with existing data systems and applications while enabling new types of data to be processed.



#### Company Overviews

**Hortonworks** - Hortonworks develops, distributes and supports the only 100% open source distribution of Apache Hadoop explicitly architected, built and tested for enterprise grade deployments. Formed by the original architects, builders and operators of Hadoop, Hortonworks stewards the core and delivers the critical services required by the enterprise to reliably and effectively run Apache Hadoop at scale.

**Sqrrl** - Sqrrl is the Big Data company that develops Sqrrl Enterprise, which is the most secure, massively scalable NoSQL database that powers real-time applications and is built with Apache Accumulo™ and Hadoop. Sqrrl Enterprise extends the capabilities of Accumulo with additional data ingest, security, and analytical features that help unlock the power of Big Data.

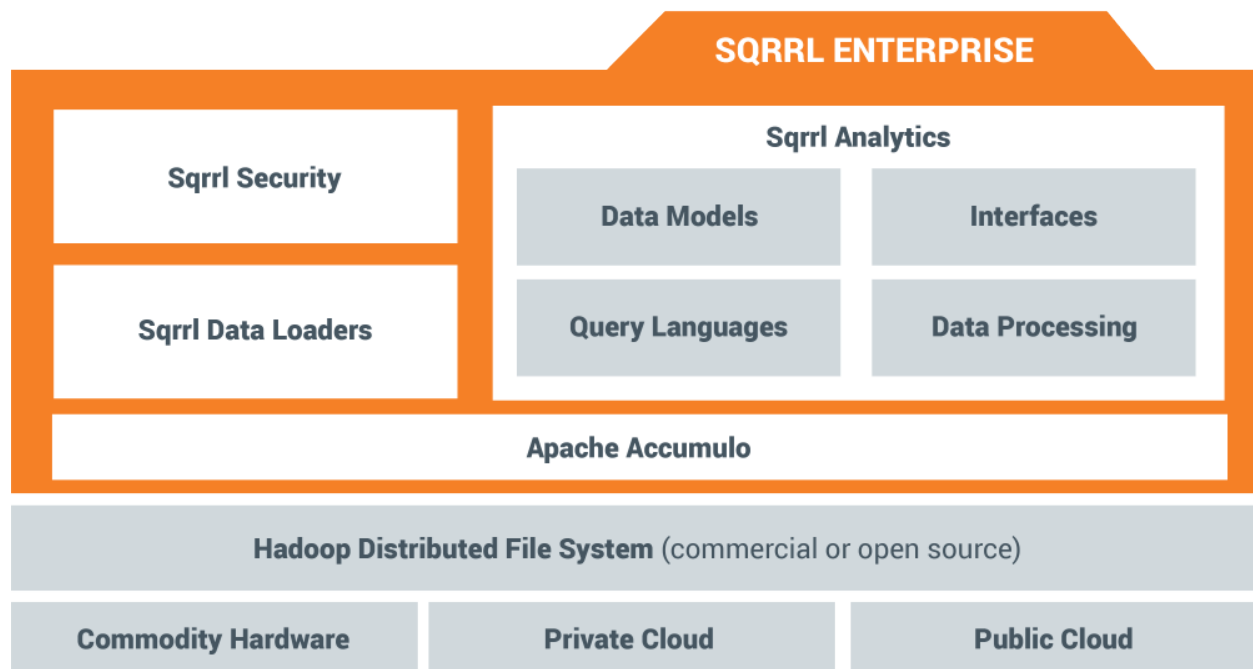
Sqrrl was founded in 2012 by creators of Apache Accumulo. Although originally developed by the National Security Agency (NSA), Accumulo is currently used throughout the federal government (in particular in the defense, intelligence, homeland security, and law enforcement communities) and in various regulated sectors, such as healthcare, finance, and telecommunications. With their roots in the U.S. Intelligence Community, Sqrrl's founders have deep experience working with and building applications for complex petabyte-scale datasets. Sqrrl is headquartered in Cambridge, MA and is a venture-backed company with investors from Matrix Partners and Atlas Venture.

## Sqrrl Enterprise

The figure below depicts the architecture for Sqrrl Enterprise. Sqrrl Enterprise can sit on top of any open source or commercial distribution of the Hadoop Distributed File system, including the HDP. It can also be deployed in a cloud environment or on bare metal. Sqrrl Enterprise consists of the following items:

**Apache Accumulo:** Sqrrl Enterprise utilizes the latest open source version of Apache Accumulo, preventing vendor lock-in and promoting interoperability.

**Sqrrl Data Loaders:** Integration with Apache Flume provides Sqrrl Enterprise with the ability to ingest streaming data feeds, such as log files. Sqrrl Enterprise can also easily bulk load JSON, CSV, and other file formats.



**Sqrrl Security:** Sqrrl features best-in-class data-centric security and is the only Big Data solution that can provide security at the “cell-level” (i.e., each individual key/value pair is tagged with a security label). The components of Sqrrl Security include the following:

- **Encryption-at-Rest:** File system-level encryption of all data held in Sqrrl Enterprise (i.e., both the keys and values associated with each key/value pair); an optional key management service stores the private keys associated with the encryption on a separate server; compatible with any encryption algorithm.
- **Encryption-in-Motion:** SSL/TLS encryption of all data in flight, including data flowing between HDFS and Accumulo, between Accumulo and Sqrrl Server, and between Sqrrl Server and end user applications.
- **Audit:** Sqrrl Enterprise maintains a record of all transactions taking place within the platform, including who is performing queries and what types of queries are being performed. This data can be used for compliance and forensics purposes.
- **Labeling Engine:** The Labeling Engine automates the tagging of key/value pairs with security labels using policy-based heuristics. These labeling heuristics are derived from an organization’s existing information security policies, and they are loaded into the Labeling Engine to apply security labels to data at the time of ingest. For example, a labeling heuristic could require that any piece of data in the format of “xxx-xxx-xxxx” receive a specific type of security label.

## Partner Reference Architecture

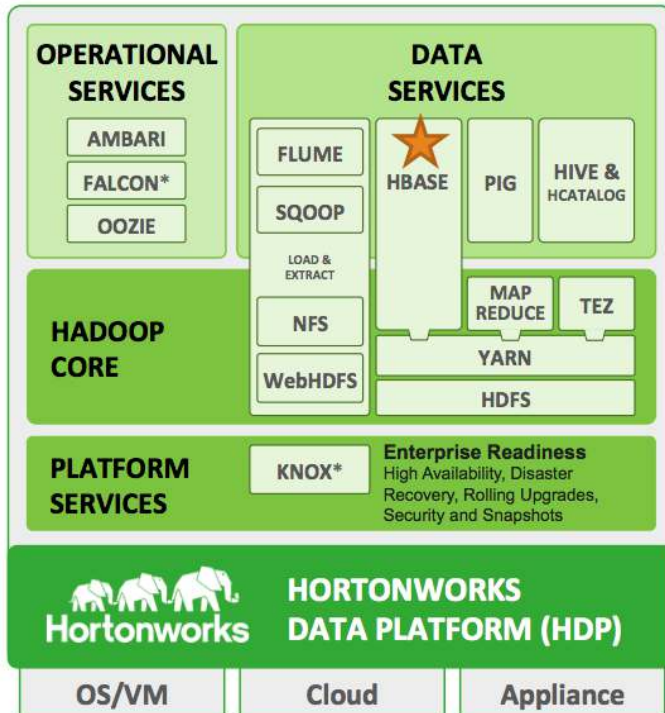
- **Policy Engine:** The Policy Engine enables both Role-Based and Attribute-Based Access Controls. It allows an organization to transform identity and environmental attributes into policy rules that dictate who can access certain types of data. For example, the Policy Engine could support a rule that data tagged with a certain security label can only be accessed by current employees during the hours of 9-5 and who are located within the United States.
- **Identity Management Integrations:** Sqrrl Enterprise has integrations with enterprise authentication and authorization systems, such as Active Directory, LDAP, and Siteminder.

**Sqrrl Analytics:** Sqrrl Enterprise has powerful analytic capabilities that enable organizations to build a variety of analytical applications on top of Sqrrl Enterprise. Organizations can also plug existing applications and tools into Sqrrl Enterprise. These capabilities are described below.

- **Query Languages:** Sqrrl Enterprise supports a variety of query languages and data models. Each of these languages and data models are fully integrated with cell-level security concepts.
  - **SqrrlQL:** Sqrrl Enterprise's SQL (SqrrlQL) capabilities include functions such as SELECT, FROM, WHERE, LIMIT, and GROUP BY and provide users with real-time statistical capabilities via a variety of scalar and aggregation functions. Users can also construct SQL queries that utilize keywords (instead of fields) as the predicate of the query.
  - **Lucene:** Sqrrl Enterprise features full-text key word search utilizing the Lucene syntax and includes phrase search, wildcards, and range queries.
- **Data Processing:** Sqrrl Enterprise's real-time processing capabilities are powered by custom iterators and an extensible indexing framework. Sqrrl Enterprise can also serve as a source or sink for Pig/MapReduce jobs via custom connectors.
  - **Iterators:** Iterators are Accumulo's server-side programming framework that enables operations that are akin to "continuous MapReduce." Sqrrl Enterprise includes various custom iterators that power its real-time analytics.
  - **Indexing:** Sqrrl Enterprise indexes all data upon ingest, including extraction of graph relationships and document-distributed indices.
  - **MapReduce and Pig:** Sqrrl Enterprise can connect to MapReduce jobs written in native MapReduce formats or via Apache Pig.
- **Data Models:** Sqrrl Enterprise provides users with a 3 databases-in-1 capability by enabling storage of data as key/value pairs (native Accumulo data format), graph data, and JavaScript Object Notation (JSON) data.
  - **Graph:** Sqrrl Enterprise supports storage of data as graph data (i.e., nodes and edges) and allows for graph search capabilities (similar to Facebook Graph Search).
  - **JSON:** Sqrrl Enterprise supports JavaScript Object Notation, which is a lightweight data interchange format that is human-readable. JSON is utilized in a variety of NoSQL document stores, such as MongoDB and Couchbase.
- **Interfaces:** Sqrrl Enterprise's API is provided through Apache Thrift and includes bindings for various popular web programming languages, including Java, Python, Ruby, and C#. Developers can utilize Sqrrl Enterprise's sample demo apps, as starter kits for custom user interfaces. The demo apps are built using the open source D3 visualization framework.

## Sqrrl Enterprise and Hortonworks Data Platform

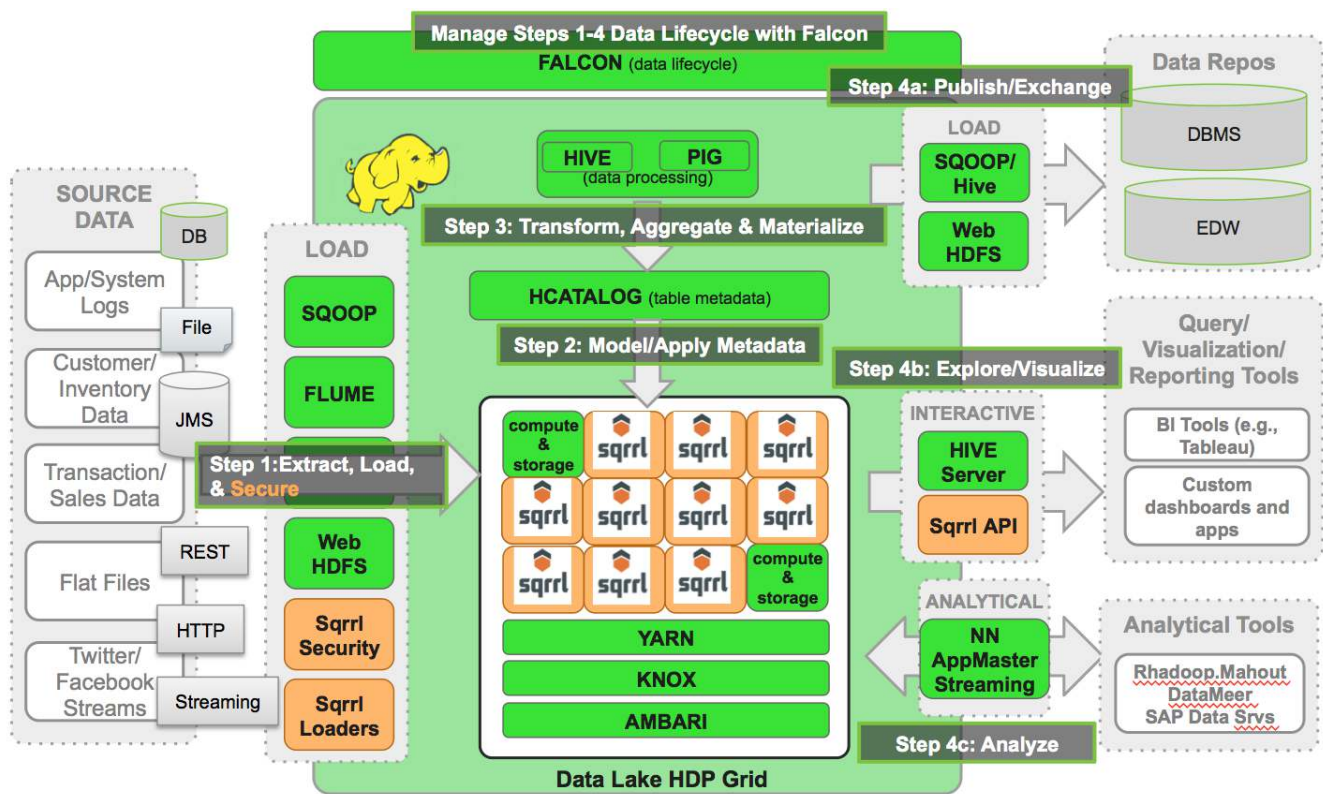
Sqrrl Enterprise natively uses the Hadoop Distributed File System (HDFS) as its storage layer. As depicted in the figure below, Sqrrl Enterprise replaces HBase in the Data Services layer of HDP, and Sqrrl Enterprise is deployed on each data node in the HDP cluster. Deployment of Sqrrl Enterprise includes installation of both Accumulo and Sqrrl Server on each data node. Sqrrl Server includes all Sqrrl-developed add-ons to Accumulo, including Sqrrl Security, Sqrrl Data Loaders, and Sqrrl Analytics.



- ★ Sqrrl Enterprise is an alternative to HBase in the HDP technology stack
- Sqrrl Enterprise (powered by Accumulo) advantages include cell-level security, better performance metrics, massive scalability, and additional data models including JSON and graph support

## Use Cases

Sqrrl expands Hortonworks' "data lake" use case to become a "secure data lake" use case, as depicted in the figure below. As data is ingested into Sqrrl Enterprise via Sqrrl Data Loaders, it is both tagged with security labels and encrypted via Sqrrl Security tools. These security labels dictate who can access each piece of data at the application layer. Sqrrl Enterprise also indexes all the data ingested into it in a variety of ways and stores the data as key/value pairs, JSON documents, and graph data. This data is made available for query through the Sqrrl API, which enables SQL, full-text key word search, and graph-style queries. Users can plug existing or new custom apps or dashboards into Sqrrl Enterprise. An upcoming release will support full ODBC/JDBC driver to enable integration with various Business Intelligence tools.



Specific analytical use cases, include searching for anomalies in cybersecurity datasets or searching for patterns in healthcare data (all while ensuring security of sensitive pieces of information).

## Customer Benefits

There are many benefits in using Sqrrl Enterprise (and Accumulo) with HDP. Some of these benefits include the following:

- Best-in-class security, including cell-level security access controls and encryption
- Superior performance, including massive scalability to 10s of petabytes, very low latency reads (i.e. “real-time”) and half a million writes per second per node
- Automatic secondary indexing and more diverse data models, including document and graph storage

Sqrrl Enterprise is particularly strong whenever there are regulated security and privacy requirements and there is a desire to mix datasets together in the “Data Lake” that require differing levels of protection. Sqrrl’s cell-level security can easily handle these complex security requirements.

## Contact Information

To schedule a demonstration or request additional information, please contact:

Ely Kahn  
VP, Business Development  
Sqrrl  
275 3<sup>rd</sup> St.  
Cambridge, MA 02142  
617.902.0784 or [info@sqrrl.com](mailto:info@sqrrl.com)