

Hunk®: Splunk Analytics for Hadoop and NoSQL Data Stores

Explore, analyze and visualize raw big data

HIGHLIGHTS

- Lower the cost and risk of big data initiatives with a full-featured platform to rapidly explore, analyze and visualize data in Hadoop and NoSQL data stores
- Obtain proven business value for security, risk management, product analytics and the Internet of Things
- Explore and visualize raw big data without building fixed schemas or moving data to an in-memory store
- Download the free Hunk Sandbox for a step-by-step tutorial on how to search, analyze and visualize raw big data

Product Overview

Hunk is a full-featured platform for rapidly exploring, analyzing and visualizing data in Hadoop and NoSQL data stores. Based on years of experience building big data products deployed at thousands of Splunk customers, Hunk drives dramatic improvements in the speed and simplicity of getting insights from raw, semistructured or unstructured big data—all without building fixed schemas or moving data to a separate in-memory store. Hunk delivers proven value for security, risk management, product analytics, a 360-degree customer view and the Internet of Things.

While many big data initiatives take months and have high rates of failure, Hunk offers a unique approach. Hunk provides a single, fluid user experience designed to drive rapid insights from your big data. Hunk empowers self-service analytics for anyone in your department or organization to quickly and easily unlock actionable insights from raw big data, wherever it may reside.

One of the key innovations in Hunk is Splunk Virtual Index™ technology. This patent-pending capability decouples the storage tier from the data access and analytics tiers, enabling Hunk to transparently route requests to different data stores. By decoupling the tiers, Hunk incorporates data models, pivot, dashboarding, role-based access controls and the Splunk developer environment. The analytics tier is powered by Splunk's Search Processing Language (SPL™), the industry-leading method for interactive data exploration across large, diverse data sets.

A key Splunk technical differentiator, schema-on-the-fly, allows Hunk to deliver the flexible, interactive analytics experience that business and IT teams need in working with unstructured data. With schema-on-the-fly, you can store data in any format and the structure is applied at query time. This provides a more iterative workflow with major productivity advantages. Now you no longer need to plan the questions up front—you can immediately explore and analyze your data in Hadoop and NoSQL data stores through visual interactions and use SPL for deeper analysis.

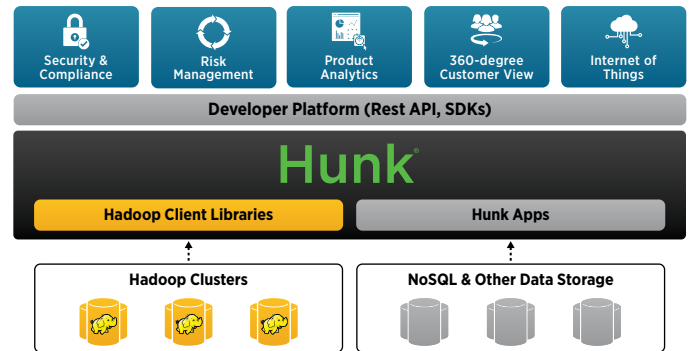


Figure 1: Extend the power of exploratory analytics to data in Hadoop, NoSQL, Amazon S3 and other data stores, on-premises or in the cloud.

Hunk Capabilities

Full-Featured, Integrated Analytics. Rapidly explore, analyze and visualize data; create dashboards and share reports from one integrated analytics platform that works with Apache Hadoop and NoSQL data stores. Drive deep analysis and find anomalies across terabytes or petabytes of raw data (see Figure 1). Instantly pivot from any search and automatically identify meaningful patterns in your data.

Fast to Deploy and Drive Value. Simply point Hunk at your Hadoop cluster and start exploring and analyzing data immediately. Using the Data Explorer wizard, select the most relevant data sets for analysis in Hunk to browse and prepare raw data in Hadoop. To connect Hunk to NoSQL data stores, download free apps at apps.splunk.com.

Interactive Search. Schema-on-the-fly technology means that you don't need to know anything about the data in advance. Hunk automatically adds structure and identifies fields of interest at search (query) time, such as keywords, patterns over time, top values and more. Event pattern detection automatically identifies meaningful patterns in your data.

Supported Data Formats. Hunk automates access to text files and sequence files, along with popular Apache Hive data formats including Record Columnar Files (RCFile), Optimized Row Columnar (ORC) files and Parquet columnar files. Hunk data can also be accessed using third-party Hadoop tools such as Hive or Pig.

Report Acceleration. Significantly improve reporting response times and cluster performance by transparently caching data with the check of a box.

Results Preview. When you start a query in Hunk, it streams back interim results immediately while the MapReduce job continues to run in the background. This delivers a faster, more interactive experience because you can pause and refine queries without having to wait for full MapReduce jobs to complete.

Drag-and-Drop Analytics. Empower business and IT teams to analyze raw data in Hadoop and NoSQL data stores. Data models describe relationships in the underlying raw data, making it more meaningful and usable. Quickly generate charts, visuals and dashboards using pivot.

Rich Developer Environment. The web framework makes building apps on top of the Hadoop Distributed File System (HDFS) or NoSQL data store look and feel like building any modern web application. The framework enables developers to integrate data and functionality from Hunk into enterprise big data applications using a standards-based web framework, documented REST API and software development kits (SDKs) for C#, Java, JavaScript, PHP, Python and Ruby. Developers can build Hunk apps with custom dashboards, flexible UI components and custom data visualizations using common development languages such as JavaScript, HTML5 and Python. For more information visit the [Developer Portal](#).

Custom Dashboards and Views. Business and technical colleagues can edit dashboards and change chart types with a simple interface. Pan and zoom to change perspectives, overlay charts and drill down from anywhere in a chart or dashboard to the underlying raw events. Save reports, integrate them into dashboards and view them all from your desktop or mobile device. Dashboard panels can be saved independently and shared across teams.

Archival from Splunk Enterprise. Archive historical data from Splunk Enterprise to commodity storage on HDFS and Amazon S3. Use unified queries to correlate real-time data from Splunk Enterprise with historical data in Hadoop.

Secure Access. Role-based access controls protect sensitive data. Pass-through authentication enables submission of MapReduce jobs and secure access to Hadoop clusters using the authenticated credentials and roles assigned by IT.

Hunk Apps. Enable the Hunk analytics platform to explore, analyze and visualize data in NoSQL and other data stores, including Apache Accumulo, Apache Cassandra, MongoDB and Neo4j. Download free Hunk apps at apps.splunk.com.

Start with the Hunk Sandbox. Learn about Hunk, Hadoop, interactive search and analytics in a single download that runs on Windows®, Apple® and Linux operating systems, without having to set up a Hadoop cluster. The Hunk Sandbox includes Apache Hadoop, Hunk software, a step-by-step tutorial and sample dashboards.

Hunk 6.3 for Production Deployments. Hunk will work with almost any Hadoop distribution running on 64-bit Linux, including any MapReduce 1.0 compatible distribution and YARN (MapReduce 2.0) distributions. Hunk works with your choice of Hadoop distribution, including Cloudera CDH, Hortonworks Data Platform, IBM InfoSphere BigInsights, MapR M-series and Pivotal HD.

Hunk on the AWS Cloud. In addition to the option to deploy Hunk for on-premises Hadoop clusters, Hunk is available as a preconfigured instance on the cloud from Amazon Web Services (AWS). AWS customers have the option to bring existing Hunk licenses or purchase Hunk by the hour from AWS. Visit the [Amazon EMR](#) webpage to sign up now.

Hunk runs on 64-bit Linux and virtual machines for Windows or Apple operating systems. For additional information on configuring Hunk, refer to the Hunk Tutorial at docs.splunk.com.

Splunk's Big Data Product Portfolio

Splunk's breakthrough products change how organizations use data. Splunk Enterprise is the industry-leading platform for operational intelligence, providing the ability to collect, index and analyze massive amounts of real-time and historical machine data. Hunk is your platform for exploring, analyzing and visualizing historical data in Hadoop. The products are complementary:

Features	Splunk Enterprise	Hunk
Indexing	Native	Virtual
Where Data is Stored and Read	Splunk Buckets on Local or SAN Disks	Any Hadoop Compatible File System (HDFS, MapR, Amazon S3)
60 Day Free Trial License	500MB/Day	Unlimited
Pricing Model	Data Ingested per Day	Number of TaskTrackers (Compute Nodes in YARN)
Real-time, Streaming Events	•	•
Data Model	•	•
Pivot	•	•
Rich Developer Environment	•	•
Event Breaking, Timestamp Extraction, Source Typing	Index Time	Search Time
Rare Term Search	Fast: Uses Index and Bloom Filters	Slow: Requires Full Data Scan Within Partitions
Report Acceleration	•	•
Access Control and Single Sign-on	•	•
Universal Forwarder	•	
Forwarder Management	•	
Splunk Apps	•	Limited
Premium Apps	•	
Standard Support	•	•
Enterprise Support	•	•

Free Download

To unlock the value of big data, start by [downloading the Hunk Sandbox](#) and the [installation instructions](#). To prepare a Hunk deployment, download Hunk 6.3 for a 60-day free trial of the production software. You can purchase a license for Hunk by contacting sales@splunk.com.

Try Hunk in the Cloud. Get started in minutes with Hunk on the AWS Cloud. [Visit the Amazon EMR webpage](#) to sign up now.