ZALONI

BEDROCK

Capitalize on Big Data for Competitive Advantage with Bedrock™, an integrated Management Platform for Hadoop Data Lakes

Highly competitive enterprises are increasingly finding ways to maximize and accelerate the value that can be derived from Big Data. A Hadoop Data Lake is a cost-effective, highly-scalable architecture for collecting and processing virtually any data format from any source. It enables new business insights and data use cases that were previously unachievable.

For example, companies can now experience a 360° view of each of their customers. They can extend the life of their Enterprise Data Warehouse investments, which have grown unwieldy, or in some cases, simply too expensive to maintain. They can also achieve a new level of operational intelligence with visibility across complex business systems.

In order to obtain business value from an enterprise Data Lake and the powerful, but ever-changing ecosystem of Hadoop, a robust, enterprise-grade data management platform is required. Zaloni's Bedrock™ is the industry's only fully integrated Hadoop data management platform. It is a unified solution for the managed ingestion, organization, enrichment, and extraction of data in Hadoop. Bedrock makes it easy.

Bedrock simplifies and automates common data management activities allowing the enterprise to focus its time and resources building the insights and analytics that drive its business. The result is improved time to market and a competitive advantage.

BEDROCK Managed Data Pipeline

**Ingest** **Organize** **Enrich** **Extract**
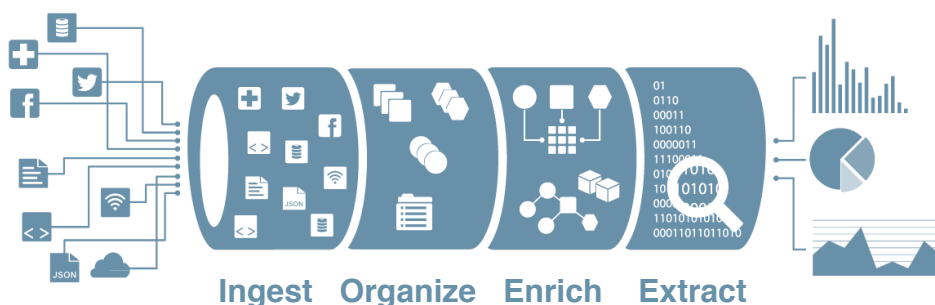
Benefits of Bedrock

- **Unified Data Management:** The only fully integrated, easy-to-use solution that offers data ingestion, organization, enrichment and extraction with metadata management and workflow.

- **Data Reliability:** Gives the user confidence that their analytics are always running on the right data, with the right quality.

- **Data Visibility:** Metadata management capabilities allow the user to keep track of what data is in Hadoop, its source, its format and its lineage.

- **Native to Hadoop:** Designed, developed and tested to work with Hadoop from the ground up, Bedrock enables users to leverage the full power of Hadoop and its ecosystem.

- **Platform Agnostic:** Compatible with all major Hadoop distributions, most data processing engines, and applicable to both on-premise and cloud based deployment models.

Ingest: Set the foundation for a well-defined data pipeline

While there are many ways to get data into Hadoop, doing so in a managed way translates to control over how data is ingested, where it comes from, when it arrives, and where it lands in the Data Lake. With Bedrock, all steps of the data ingestion pipeline are defined in advance, tracked and logged. The process is repeatable and scalable.

A well-defined data ingestion process provides operational benefits such as the ability for IT to troubleshoot and diagnose ingestion issues. It simplifies the onboarding of new data sets and therefore simplifies the development of new use cases and applications for the Data Lake.
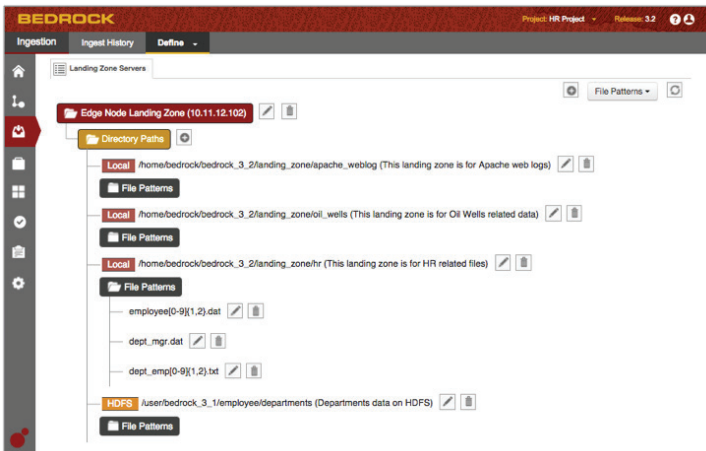

Automate your data pipeline: Configure Bedrock to automatically pick up incoming files, tag them with metadata, and run transformations. Streaming ingest is also supported.

Organize: Catalog data to reduce risk and time to insight

Bedrock makes it easy to organize and manage data, regardless of volume. It captures operational, technical and business metadata, so that consumers can search, browse, and find the data they need for analytics, reducing the time to insight for new analytics projects.

With file and record level watermarking, Bedrock provides the ability to see data lineage, where data moves and how it is used. This safeguards data and reduces risk, as the data manager will always know where data has come from, where it is, and how it is being used.
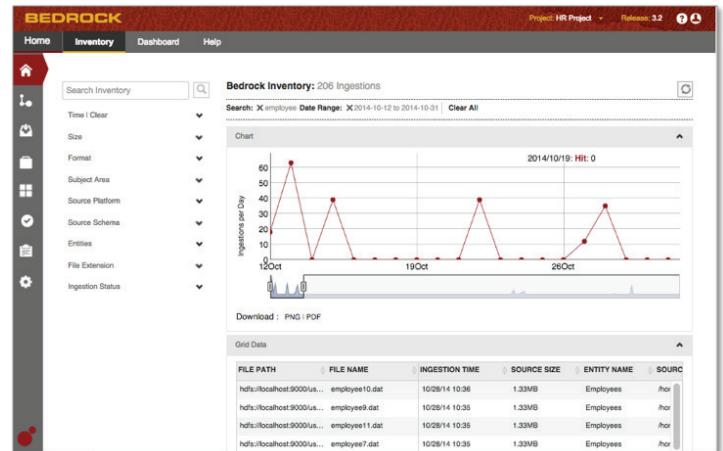

Know what is in your Data Lake: Bedrock's Inventory Dashboard enables Data Lake exploration via a faceted search.

Enrich: Handle data consistently for increased confidence in the Data Lake

Enabling users to access and leverage data that resides in the Hadoop Data Lake is critical to meeting business goals that the Data Lake was created to solve. This requires adequately preparing the data for use, which may include many things:

- Ensuring that sensitive data is masked or tokenized appropriately
- Evaluating the quality of data as it comes in
- Executing a complex workflow to reconcile updated or changed data
- Converting data formats to handle international characters
- Tagging data to indicate its lineage

All of these tasks can be orchestrated by Bedrock using a combination of built-in and user defined transformations. Each workflow can run on a schedule or be triggered by an event such as new data ingestion.
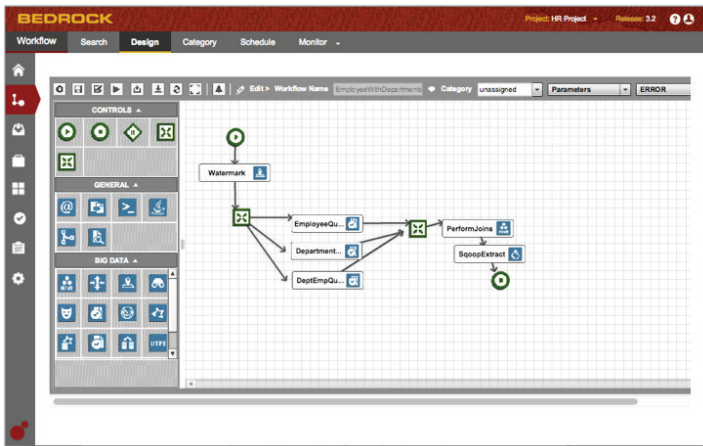

Create a repeatable transformation process: Stitch complex workflows together, including Hadoop and non-Hadoop workloads. Schedule workflows to run on a recurring basis, and set notifications for specific events, such as workflow failures.

Extract: Enable self service to democratize the Data Lake

The business value of Hadoop, is predicated on giving business experts power. Build extraction as part of a repeatable, managed workflow, or allow the Data Lake consumers to query the data and create their own extracts.

The Query Builder allows business analysts with little to no SQL knowledge to explore the data that they have visibility into, and extract the data so they can use the visualization tool of their choice. Alternatively, data scientists can create their extracts that can then be used for analytics.
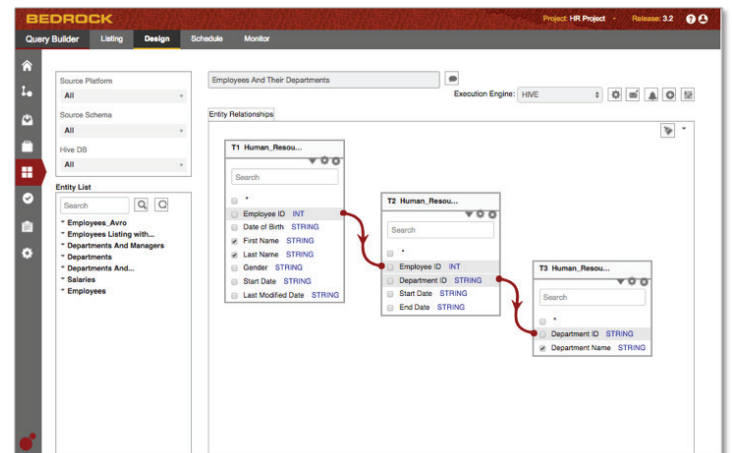
Empower business analysts to extract data: The Bedrock Query Builder enables business analysts to create their own extracts without an IT request and without SQL knowledge.

Bedrock 3.2 Feature Overview

Ingest

Ingest Streaming Data Ingestion	Ingestion is possible as Files, Streaming via Flume, and Relational via Sqoop. A wide variety of file formats are supported. E.g. Fixed, Delimited, XML, JSON, and domain specific formats like HL7 and X12.
Managed Landing Zones	Landing Zone agents watch for new files to arrive in the edge node or HDFS, associate them with metadata Entity types, initiate the capturing of operational metadata and kick off the data transformation process.
Scalable Ingestion	Landing zones may be distributed across multiple servers to maximize ingestion throughput and scale.

Organize

Inventory Dashboard	An inventory dashboard allows exploration of ingested data via faceted and free text searches.
Entity Definition	Capture business descriptions, field formats, characteristics and tags through metadata Entities.
Entity Versioning	Support for changing data formats and ensuring that applications are accessing the expected version of the data.
HCatalog Integration	Automatically update metadata in HCatalog so that Hadoop data query tools are always working from up-to-date and accurate information.
Change Data Capture	Automatically reconcile incremental updates with data in Hive to ensure that the data consumers have access to the most current active view of the data as well as a historical view of previous changes.
Data Quality	Incoming data is checked by applying the associated rules so that Data consumers are assured that the data they are using meets the expected quality standards
Data Lineage	Show where data came from, when it arrived, where it went to in Hadoop.
Data Catalog	Search for data by business name, source, and creation date. Browse data for defined Entity types, their fields and their ingestion history.
Bulk Import	Bulk import of Metadata for integration with relational database schemas and metadata management systems.

Enrich

Workflow & Orchestration	Graphical workflow designer to orchestrate preparation and transformations steps. Workflow execution triggered by data arrival, on a schedule, or on demand.
Built-in Data Preparation Functions	Drag and drop data preparation and transformation actions like Watermarking, Masking, Tokenization, Data Quality, and Change Data Capture. Avro entities can also be created to provide a single view across different versions of the data.
Big Data Preparation Functions	Support for data transformation and preparation functions that leverage Hadoop-based toolsets. Examples include MapReduce, Hive, Spark, SparkSQL.
User Defined Data Preparation Functions	Support to build your own data transformation and preparation functions, and to deliver integrations to third party systems. Examples include Java, and Script actions.

Bedrock 3.2 Feature Overview

Enrich	Query Builder	• Easy to use, drag and drop interface to develop and execute queries on the data in a Bedrock-managed Hadoop Data Lake. • Choose either Hive, Spark, or Impala as the Execution Engine. • Enable quick and iterative data exploration and analysis by enabling users to create analytics ready data-sets, while controlling the visibility of data through projects. • Self-service access reduces the typical IT barriers to ad hoc data analysis and exploration, and enables business users and data scientists to develop rapid responses to data driven questions. • Allows ETL developers and technical users to use Query Builder queries as scheduled data transformation steps or in workflows for data preparation.
Extract	Self Service	Data consumers can download the resulting data after running queries created using the Query Builder.
	Orchestrated extract	Leverage Sqoop as part of a workflow to export data to a relational system.
Operations & Deployment	Support for all Major Hadoop Distributions	Compatible with all major Hadoop distributions and with most data processing engines including MapReduce, Hive, Pig, Impala, and Spark.
	Reporting	Reporting framework for user-defined operational dashboards and reports. Examples include: Data Ingestion volume, Workflow execution time, and Data Quality.
	Hadoop Edge Node Application	Bedrock is a Java application that runs on a Hadoop edge node. It operates as a data management or control plane and does not interfere with the data processing plane or performance.
	High Availability Support	May be deployed in redundant and highly available architectures.
	Logging and Notification	All Bedrock actions are logged for operational troubleshooting and auditing activity. Customizable notifications can be sent when errors or alarms occur.
	Role / Project Based Access Control	Manage Bedrock users locally or via integration with an enterprise LDAP server. Associate Users with Roles to grant or restrict privileges. Limit visibility to approved members of a project.
	Web Interface and API	All functions are available via Web based user interface, and REST API.
	Deployment	Deployable on premise or in the cloud.
	Hadoop 2.0 / YARN support	Bedrock is able to detect the YARN resource queues defined and allowed for Hadoop jobs to be submitted against the appropriate queues.
	Import & Export	Stagger the deployment of your application from development to test, and finally to production using the Import and Export feature to move your artifacts and related resources.

ZALONI

BEDROCK

Capitalize on Big Data for Competitive Advantage with Bedrock™, an integrated Management Platform for Hadoop Data Lakes

About Zaloni

Zaloni provides enterprise data management solutions for Hadoop. Our software solutions enable customers to gain competitive advantage through organized, actionable Hadoop data lakes. Serving the Fortune 500, Zaloni has helped its customers build production Hadoop implementations at many of the world's leading telecommunications, financial services and healthcare companies.

Services and Training

Zaloni Professional Services offer expert consulting and training services to help you reduce risk, accelerate adoption, and improve business performance. Through the use of these services, delivered globally, you can move quickly from pre-production to post-production to maximize the value of your investment.

Customer Support

We provide a comprehensive customer support experience that includes online portal access to software and documentation, and a robust knowledge base.

To Learn More

For more information about Bedrock, the industry's only integrated Data Management Platform for Hadoop.
Call us: +1 919.323.4050
E-mail: info@zaloni.com
Visit: http://www.zaloni.com

Find us on social media

Twitter: https://twitter.com/zaloni/ (Twitter handle @zaloni)
LinkedIn: https://www.linkedin.com/company/Zaloni
Facebook: https://www.facebook.com/pages/zaloni/131938830174886

"One of the greatest challenges with managing massive amounts of data, located across multiple sources, is managing cost, with Zaloni's Bedrock Platform in place we quickly realized cost savings, and those savings accompanied a scalable, highly available and reliable platform."

Pradeep Varadan
Associate Director
Verizon Enterprise Solutions