

# Data & Analytics Practice

# Data Ingestion Framework



## HIGHLY SCALABLE, DISTRIBUTED, SECURE AND FAULT TOLERANT

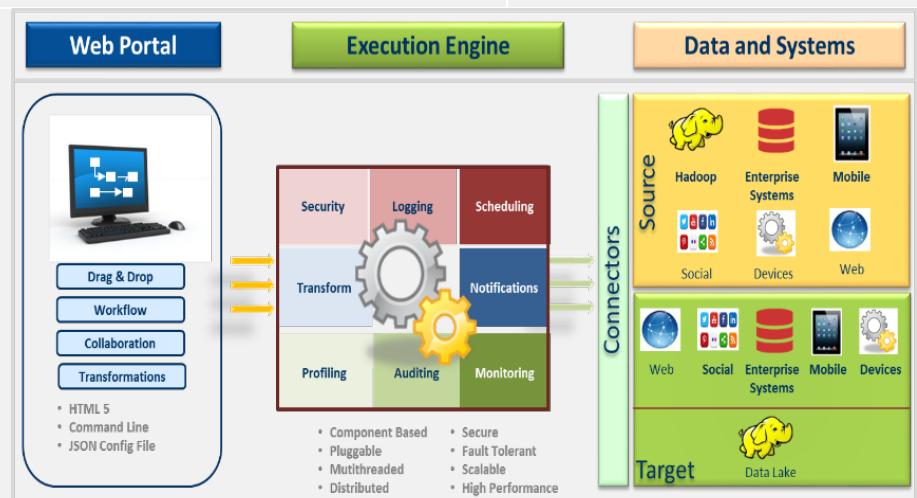
The RCG Data Ingestion Framework is a **fully integrated, highly scalable, distributed and secure solution** for managing, preparing and delivering data from a vast array of sources including: social media, mobile devices, smart devices and enterprise systems. The framework **eliminates the need** for IT professionals to become experts in Hadoop eco-system technologies and languages, and **speeds time to delivery at reduced costs** by simplifying and standardizing data management and data work flows.

The framework supports data sources (structured, semi-structured and unstructured) and targets in traditional enterprise systems, external systems and the Hadoop eco-system.

RCG's Data Ingestion Framework Features

- A **single consistent method** for capturing data
- The ability to **quickly add new data sources and targets**
- A foundation of open source technologies
- **A Highly scalable, distributed, secure and fault tolerant architecture**
- A component based architecture that **enables plug & play** of new connectors, transformers, etc.

RCG's Web Portal provides graphical drag and drop capabilities allowing developers to specify workflow and data transformations. Support is provided for **Generic Formats** (Database results, Delimited, JSON, XLS, XML Thrift, Avro, PDF), **Hadoop Formats** (ORC, RCFile, Sequence File, Parquet) and **Industry Specific Formats** (Accord, LAS, FIX)



# Execution Engine

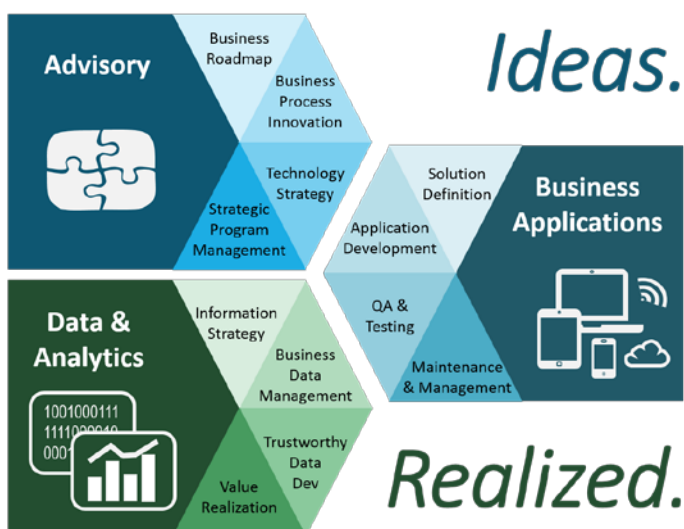
The execution engine captures, transforms and moves data. It can be initiated through a command line or the Web Portal or can be scheduled to run automatically. Features include:

- Parallel and distributed
- Visual Drag & Drop interface
- Portable, lightweight and secure
- Support for batch, micro-batch, near real-time and real-time delivery modes
- Metadata and lineage aware
- Secure end-to-end data routing encryption and compression
- 100% Open Source
- Over 100 supported endpoints
- Flexible Transformations
- Aggregations - Filters and Mappings
- Plugin architecture
- Distribution Agnostic
- Supports for structured, semi-structured and unstructured data

# Source and Target Data and Systems

Source and target data and systems are supported through connectors. Connectors are available for many types of data and systems and new ones can be added to the framework quickly and economically. Supported connectors include:

- Hadoop – HDFS, Hive, Hbase
- Database – Teradata, Netezza, SQLServer, MySQL, Oracle, DB2
- NoSQL – DataStax, Cassandra, MongoDB, Couchbase
- Social Media – GNIP, Datasift
- Search – Solr, ElasticSearch
- Messaging – Tibco, MQ, Kafka, Active MQ, RabbitMQ
- File Systems – Generic OS, FTP/SFTP, Splunk
- Streaming – Storm, Sockets
- Integration with Camel to support hundreds of different endpoints



## OUR BRAND PROMISE

Our *reputation* is built upon the premise that we are a company that *listens*.

We bring a *creative view* to your business initiative.

We are *collaborative* and *accountable* as we jointly create your solution.

We *continuously innovate* from concept to result and help you affect *business change*.

There will be *no surprises*.

For more information on any of RCG's solutions, check out <http://www.rcgglobalservices.com> or contact [solutions@rcggs.com](mailto:solutions@rcggs.com)

