RegEx 100B™

Product Brief

The DRC RegEx 100B™ engine delivers 100,000,000,000 (100 billion) character matches per second. Character patterns are specified using standard POSIX regular expressions. Actionable intelligence is delivered in realtime on unstructed, unindexed text.

DRC RegEx 100B engine

DRC RegEx 100B appliance with upto 4 engines

- Most advanced text analyzer
- Innovative design
- Uses proven algorithms
- Full Linux Support
- User friendly and configurable

DRC COMPUTER CORP
3375 Scott Blvd, Suite 206
Santa Clara, CA 95054

PHONE
+1. 408.562.0000

WEB
drccomputer.com

Background. With over 10 trillion SMS messages sent in 2012 alone people are rapidly shifting P2P communications from conventional semi-structured mechanisms, like email, to unstructured, casual messaging. With social messaging shorthand and misspellings are common. Specifying character patterns using regular expressions is well known and heavily used to detect key phrases and patterns.

Solution. The DRC RegEx 100B analyzer solves this dilemma by performing up to 100 billion character matches / second on real-time data. Up to 1,000 character patterns are compared in parallel on the input text that can be streamed in at 100M Characters / second. Patterns are specified using standard POSIC Regular Expressions. Users can also update the RegEx engine with new patterns in real-time with no perceivable delay or data loss. Thus, an unlimited number of character patterns can be used on the same data set. Matched patterns are output to the user in real-time as they are found. Indexing is not required.

Dense packaging. Each RegEx100B server contains up to four RegEx engines packaged in a 1U rack mountable configuration.

Highly flexible expression engine. Each analyzer engine can be individually configured to handle hundreds of search strings of variable length simultaneously.

Very low latency. By analyzing incoming data in realtime microsecond response times are achieved.

Multi-byte support. Each expression can consist of multi-byte characters enabling all languages to be supported.

Massive scalability. Clustered RegEx analyzers enable tens of thousands of expressions, petabytes of data and thousands of users to be supported simultaneously.

Ultra-low energy consumption. Each RegEx server with 4 analyzer engines requires less than 400 watts of power.

Cloud ready. Analyzers can be cloud based.

Specification (per analyzer engine):

Performance	
Character patterns /second	100,000,000,000
Detection latency	10^{-5}s (10us)
Character patterns	
Patterns executed in parallel	1,000
Maximum pattern length	64K char
Number of patterns	unlimited
Input Data	
Character encoding	8, 16, 32-bit
Data throughput	100 MB/s
Maximum length	unlimited

1U server contains up to 4 analyzer engines. 1Us can be clustered.

The DRC Difference

With over 200 man-years of experience in developing low latency, high capacity solutions, DRC has a unique talent in big data applications.

By utilizing a task based architecture DRC has optimized data management/data analysis balance. The key to the ultra-high performance of Novara is distributing processing capacity so that's its available where it's needed rather than centralizing it. Moves the processing to the data versus moving the data to the processor.

DRC Complete Services

In addition to providing the standard Novara analyzer DRC also delivers a complete design and implementation service for those customers who require assistance with designing, configuring and customizing their Novara environment.

DRC programs include data security, indexing, biometrics, image processing, genomics and financial services applications.

In many cases DRC engineers are used to supplement customer's internal resources as consultants to assist with design and performance optimization.

Made in America

DRC is a US based company with all staff US citizens, and all its engineering and manufacturing conducted in the US.