



POWERING ELASTICSEARCH IN THE CLOUD:

Transform High Performance Analytics in the AWS Cloud for Fast, Data-Driven Decisions

Today's actionable insights are trapped inside compute platforms and analytics ecosystems that were, at the core, architected more than 70 years ago. Traditional von Neumann-based processors have ruled the IT infrastructure world, but with new types of data—digital images, video, geospatial and more—proliferating at increasingly breakneck speeds, these legacy architectures are doing more to hold back data-driven insights than to facilitate them. Thriving in the new era of data means being free from the prison of slow, sequential processors that have monopolized on-premises and cloud-based analytics infrastructures.

Newer, more specialized architectures, such as field-programmable gate arrays (FPGAs), are gaining traction with organizations that need fast insights and cannot wait for the latency associated with data preparation on quickly changing data. Initially considered tools for highly niche-use cases, FPGA-accelerated architectures have made the jump to more mainstream use cases in recent years. But full adoption has been stymied by the need for costly specialized hardware and specially trained team members to fully reap the benefits of FPGA technology. Not anymore. Ryft is delivering the performance and ease-of-use needed for FPGA-accelerated computing architectures in the cloud on Amazon's EC2 F1.



ELIMINATING BIG DATA GRIDLOCK WITH HIGH PERFORMANCE, HIGH-EFFICIENCY INFRASTRUCTURES

Ryft has specialized in delivering immediate insight from large and complex data sets using FPGA-accelerated systems for more than 15 years. By eliminating the need for data preparation including indexing, transformation and curation and by boosting the speed of analytics, Ryft’s massively parallel FPGA-accelerated architecture delivers 200X or more data analytics performance gains over traditional CPU-based approaches. Whereas traditional FPGA-based data analytics appliances proved too difficult and too costly for the average enterprise to deploy, program and maintain, Ryft’s cloud-based solutions make use of Ryft’s open API, providing a library of analytics algorithms and connectors that abstract away complexity. The result is a fast and simple cloud-based architecture harnessing the power of FPGA-based heterogeneous compute systems. With Ryft, all organizations can benefit from the FPGA’s performance gains without needing any specialized knowledge.

Organizations finally have the choice of storing and analyzing their data using Ryft’s technology with any combination of on-premise datacenters and cloud-based services running on F1 instances.

SUPERCHARGING THE CLOUD WITH THE AMAZON EC2 F1 INSTANCE

Amazon announced its F1 instance in November 2016 to help customers benefit from the performance gains of FPGA-accelerated heterogeneous compute. With two F1 instance offerings, including up to eight of the latest 16nm Xilinx UltraScale+ VU9P FPGAs per instance, organizations that rely on the AWS cloud to store and analyze data can now get the massive performance gains that hardware acceleration offers.

Being able to deploy FPGA acceleration in the cloud enables many businesses to achieve FPGA-accelerated gains without massive internal infrastructure deployment costs. Ryft realizes that no single solution is right for all enterprises. Some organizations are more comfortable with on-premises solutions, whereas others are migrating to the cloud or some hybrid combination of the local datacenter and cloud. Since no two organizations have the same infrastructure, expertise, budget or security and privacy issues, Ryft offers multiple deployment options to address the diversity as shown in Figure 1, below.

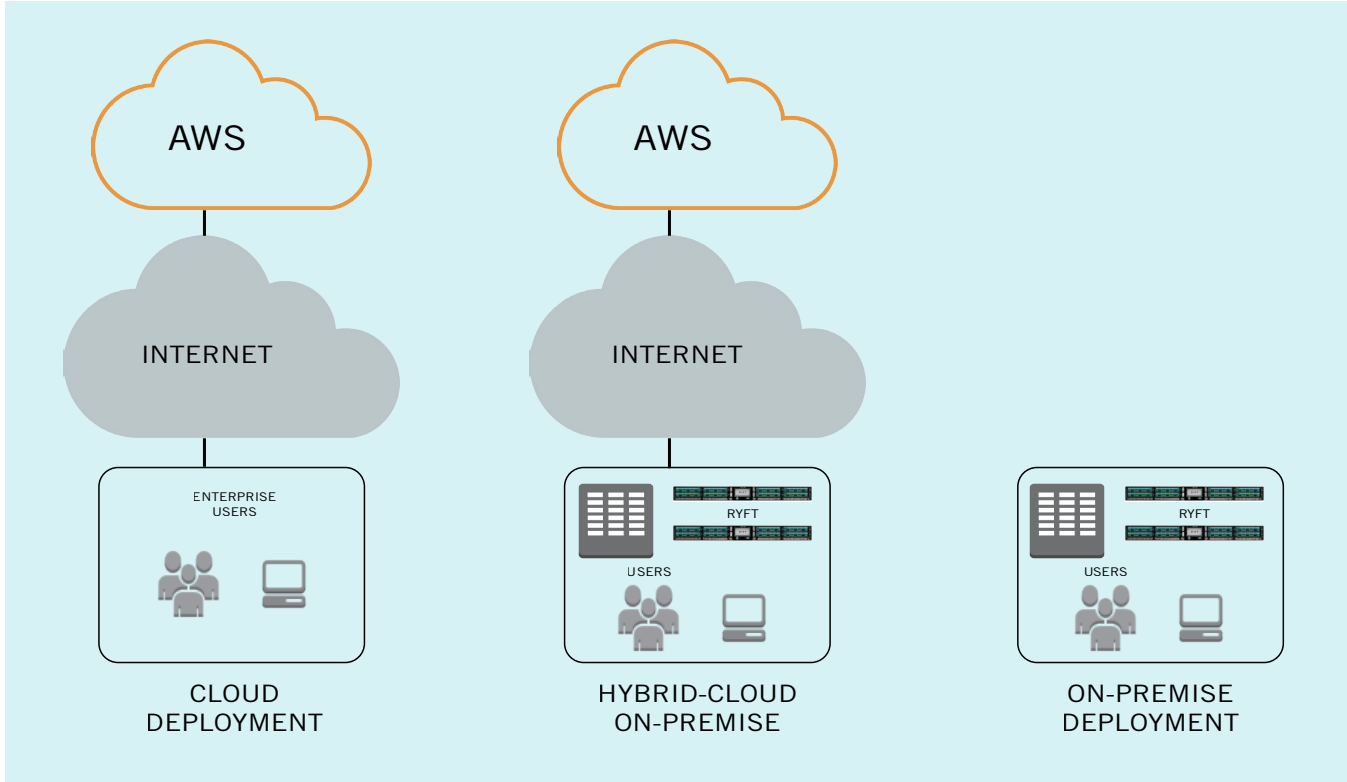


Figure 1: Ryft seamlessly operates in cloud, hybrid and on-premise deployments.

RYFT ON AWS HELPS BRING ACCELERATED COMPUTE BENEFITS TO ALL ORGANIZATIONS

The F1 instance helps expand Ryft’s existing solutions portfolio by adding cloud-based x86/FPGA accelerated compute to existing on-premise and hybrid cloud/on-premise deployment capabilities. This new deployment method enables organizations that have moved their analytics duties to the AWS cloud to:

- Gain **predictable performance improvements**, because FPGAs are not shared across instances, users or accounts, so customers know that their FPGAs are dedicated while in use.
- More **affordably use FPGAs** by enabling customers to spin up and down instances when needed, instead of letting specialized hardware sit idle in the datacenter.
- **Quickly deploy hardware acceleration** by spinning up an F1 instance alongside the necessary application from one of AWS’s dedicated F1 partners.

FPGA-accelerated performance greatly benefits an organization, whether it is quickly running IP analysis on incoming web traffic to ensure no security breaches, or trying to run large-distance Levenshtein or Hamming searches on difficult-to-index data or any number of compute-intensive workloads. And now, with the F1 instance, those same workloads can be analyzed using the cloud and not just in on-premise datacenters. The difficulty in programming FGPA-based applications is reduced with the F1 instance as Amazon allows partner developers such as Ryft to provide packaged solutions that are seamless and as simple to use as standard CPU-based instances.

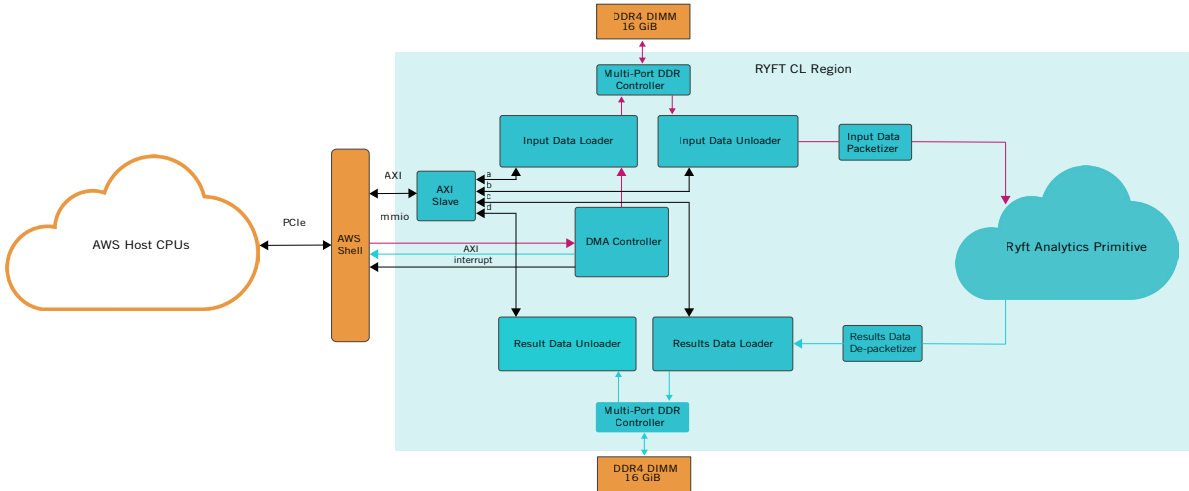
Ryft’s ongoing collaboration with AWS to deliver FPGA-enabled capabilities is a great step forward to putting the power of heterogeneous compute into the hands of every business.

ARCHITECTING SIMPLE FPGA USE IN THE CLOUD WITH RYFT

The Ryft Cloud technology on the F1 instance’s FPGA architecture is tailored for cloud-based heterogeneous computing platforms where data may come from a variety of sources, such as streaming data from Amazon Kinesis, Simple Storage Service (S3), Elastic Block Store (EBS) or instance local storage. Ryft’s architecture is designed to operate on data streams or complete data files with no reliance on special distributed file systems that chunk data for parallel computing and acceleration. Acceleration is provided using massive bitwise parallelism within the FPGA. Performance scalability is achievable with both scale-up and scale-out architectures using F1.

All F1 instance FPGAs are Xilinx UltraScale+ VU9P devices and are programmable. Each device has a dedicated x16 PCIe interface and 64 GiB of ECC-protected memory on 4x DDR4 channels. All FPGAs in the largest F1 instance can communicate and transfer data over a high-speed (400 Gbps) bidirectional ring.¹

The Ryft Cloud technology on the F1 instance is a custom accelerator used for the heterogeneous acceleration of data analytics and machine learning algorithms. The architecture utilizes two of the four DDR4 channels, one for buffering input data and one for buffering accelerator output data. A diagram of the Ryft Cloud on the F1 instance is shown in Figure 2, below.



¹ See <https://aws.amazon.com/ec2/instance-types/f1>.



RYFT ACCELERATES INSIGHTS FROM AWS-BASED DATA ANALYTICS

The Ryft Cloud interface provides an AXI bus master interface mapped to PCIe BAR space. This AXI bus is used for register-based command and status throughout the Ryft FPGA region. A separate AXI slave bus provides a high bandwidth DMA-based interface to the DDRs. The upper DDR, as depicted in Figure 2, serves as a staging cache for data destined for an arbitrary Ryft analytics primitive. Data is cached there so that it can be subsequently accessed without additional fetch operations from the host memory. Input data is moved into the upper DDR using a DMA-engine implemented through the module path (DMA Controller, Input Data Loader and Multiport DDR Controller). These modules are register-programmed to autonomously transfer the desired amount of data. A second and simultaneous DMA operation moves data from the upper DDR into the Ryft analytic primitive through the module path (Multi-Port DDR Controller, Input Data Unloader and Input Data Packetizer).

The result produced by the Ryft analytic primitive is transferred similarly to the output side DDR. All DMA data, both input and output, is multiplexed by the DMA Controller module. The Ryft implementation uses a custom DMA approach to transfer DDR data; however, alternative architectures are possible using built-in Xilinx and AWS functions, should the FPGA designer so desire.

OPEN, EASY-TO-USE INTERFACES WITH RYFT OPEN API AND ANALYTIC ALGORITHMS FOR F1

The Ryft analytic primitive section of the architecture implements data analytic and machine learning accelerators that are accessed by simple command line interface or software function calls to the Ryft Analytics Primitives library. Ryft's initial release on the F1 instance includes support for JSON alongside Levenshtein distance, to extend and enhance the popular ELK Stack (Elasticsearch/Logstash/Kibana). In the coming months, we will release support for all our other primitives including exact search, fuzzy Hamming search and detection and ranging for date, time, numeric, currency, IPv4 addresses and IPv6 addresses. Each of those primitives operates the same way in on the F1 instance as it does on deployed Ryft ONE FPGA-accelerated hardware, allowing for operation against any data type, including unstructured, semi-structured or structured data, to include not just JSON but also XML and comma separated value (CSV)

data. In fact, Ryft is committed to supporting all of its primitives and data types moving forward on both our own Ryft ONE hardware family as well as on the F1 instance.

Ryft is able to achieve parity between its hardware- and cloud-based offerings due to the development of an extensive design flow and methodology for managing and building heterogeneous analytic primitives, allowing for rapid build-out of our analytics library. The Ryft Cloud technology on the F1 was architected to maintain our current FPGA build methodology used on our own Ryft ONE hardware platform and not require any platform-specific library modifications when migrating these analytics to the F1 instance. Migrating this library to the F1 instance was a seamless effort since AWS supports the standard Xilinx Vivado (SDx) 2017.1 environment, supporting both local and cloud-based development via Amazon's FPGA Developer AMI. The Ryft Cloud on the F1 FPGA regions are implemented using a Tcl-based Non-Project Vivado flow initiated from GNU make. This Tcl-based scripting is based on the scripts found in the Vivado Partial Reconfiguration Tutorial (Xilinx UG947). Additionally, using GNU make helps resolve file dependencies within the Ryft design library and simplifies the process of rebuilding the entire library.

Ryft utilizes locally licensed Mentor QuestaSim HDL simulators during FPGA development, allowing for the verification of new FPGA code that was needed to interface to the F1. Ryft's Cloud + F1 FPGA bitstreams are generated automatically by submitting a fully routed design check point (DCP) from the Xilinx Vivado toolkit to AWS via AWS CLI commands, referred to as the ingestion process. After AWS performs a check of the DCP for validity, an Amazon FPGA Image (AFI) is generated, which can then be programmed into an F1 instance under the control of an Amazon Machine Image (AMI) that Ryft built based on the Ubuntu 16.04 LTS operating system. FPGA management and control is automatically handled by Ryft's control software included in the AMI that runs on the F1 instance, so that end users can focus purely on their own software applications, seamlessly running Ryft analytic primitives using the same software ecosystems that they currently use today. That allows for faster performance and enhanced features that are typically unavailable to CPU-only compute infrastructure.

² See www.ryft.com/api.

RYFT POWERS COMPLEX ANALYTICS APPLICATIONS TO SPEED BUSINESS INSIGHTS

The magic of the faster performance and enhanced features is in the power of abstracted heterogeneous compute afforded by Ryft's technology. Ryft is committed to ensuring that FPGA-based technology is seamlessly integrated into any ecosystem for widespread adoption by the masses, allowing them to achieve unparalleled performance gains while accessing previously unavailable features that only FPGAs can deliver.

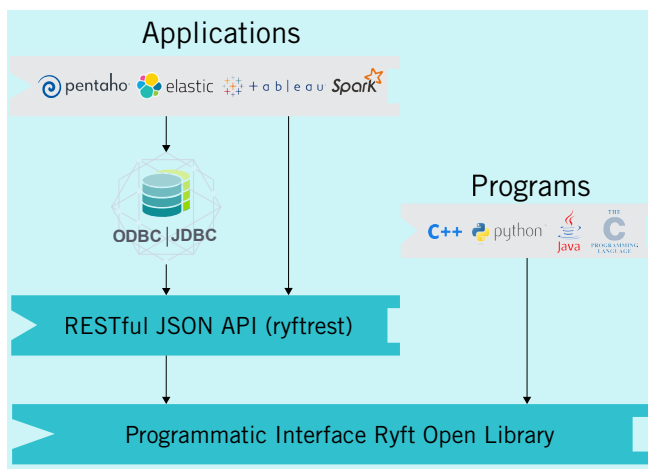


Figure 3: Ryft's Open API and Library of Connectors enable users to deploy Ryft accelerators in minutes

As seen in Figure 3, above, Ryft provides an array of interconnection options for any ecosystem, operating as low as programmatic language or as high as standards-based APIs. Programmatic interfaces (C, C++, Java and Python) are available out of the box using bindings to today's most popular development languages. A simple-to-use programmatic interface for developers is a great first step but by itself is not enough. The desire for increased performance with minimal impact to the existing data ecosystem is a challenge that all new successful technologies must address. Ryft addresses it by providing standards-based interfaces, such as ODBC and JDBC, which allows any application that supports those protocols to talk to the Ryft solution without requiring any changes to existing code or methodologies.

Ryft also provides a command line interface for scripting, as well as a RESTful JSON API that is extremely popular in today's IT world. It provides quick integration into an existing ecosystem using RESTful interfaces without users or developers needing to know anything about FPGAs. For example, various end-user applications like Elasticsearch can benefit from heterogeneous compute hardware and our RESTful API, providing for new functionality that is

not available in the traditional x86 implementation. A benefit for many customers is the ability to search various types of logs, protocol capture files and structured and unstructured data using Elasticsearch without spending the typical time on data indexing, transformation and curation. Integration with popular cluster-based tools such as Apache Spark is also available to accelerate various pieces of the data pipeline, as well as the ability to enable clustering of both cloud-based F1 instances and, if desired, locally deployed Ryft ONE servers.

For more information on the ODBC, JDBC and RESTful API, check out www.ryft.com/api, and for the Spark-Ryft Connector look at <https://spark-packages.org> (search for Ryft). Ryft layered products enhance data analysis using heterogeneous compute techniques alongside seamless integration into our customers' solutions.

With F1 and Ryft, it is easy to get started. In the near future, from the AWS Marketplace, just look for "Elasticsearch Powered by Ryft."

EXTENDING AND EXPANDING ELASTIC'S ELK STACK WITH RYFT ON THE F1 INSTANCE

Ryft's Elasticsearch is a preconfigured, ready-to-run image for the powerful open-source distributed real-time search and analytics engine Elasticsearch on the F1 instance, allowing for seamlessly obtaining the benefits of heterogeneous computing. Included is Elasticsearch (deep search and data analytics), Logstash (centralized logging, log enrichment and parsing), Beats family of interfaces (lightweight data agents including Packetbeat, Filebeat and Metricbeat that send data back to Elasticsearch) and Kibana (a data visualization tool that creates charts, graphs and dashboards of content indexed in Elasticsearch), as seen in Figure 4, below.

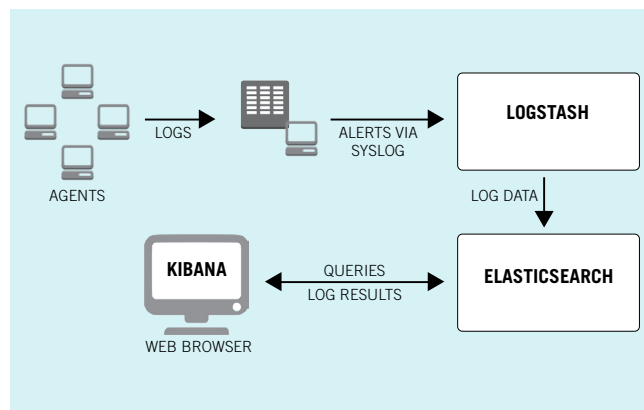


Figure 4: Ryft seamlessly expands and extends the Elastic (ELK) stack.

By using the modern heterogeneous compute architectures that F1 provides, users can benefit from enhanced functionality for both pre-index and post-index use cases. Capabilities include:

- Increasing the power of searching with user selectable changes to large (greater than two) fuzzy distance requests for Hamming or Levenshtein (edit distance) searches.
- Speeding search and analysis across unstructured data.
- Using JSON, XML, LOGs, CSV, TSV and other files with no transformation and indexing.
- Enhancing wildcard searches to include leading wildcard characters.
- Supercharge regular expressions and machine learning, coming later in 2017.

The heterogeneous computing capability allows users to accelerate workflows by deploying pre-index and post-index searches, searching their data when and how they want. Ryft's technology powers the next generation of tools that organizations need to analyze and understand their data immediately, with no delay, before indexes are created.

SET YOURSELF UP FOR DATA ANALYTICS SUCCESS

Enterprises today need to modernize their data analytics infrastructures, leveraging the right compute engine for the right job by incorporating new heterogeneous architectures to speed complex data workloads. Combining the powerful yet easy-to-use Ryft software ecosystem of analytics algorithms and connectors with Amazon's high performance FPGA-based F1 instance gives organizations a new and unprecedented ability to quickly and easily extract needed insights from all of their data.

Ryft's approach is simple, whether running in the cloud or on-premise. Ryft's seamless integration to any ecosystem automatically manages the underlying heterogeneous compute technology, allowing you to focus on what is truly important: getting the fastest possible answers to your important business questions.

Ryft is delivering the performance and ease of use needed for FPGA-accelerated compute architectures in the cloud on Amazon's EC2 F1 instance.

GETTING STARTED

When you hit the inevitable performance wall, it's time to look to Ryft. You can get started today by trying out Amazon's EC2 F1 instances, or you can contact us. Our team will work with you to match the perfect Ryft solution with your needs and budget. Visit ryft.com today for more information.

ABOUT RYFT

Ryft powers cloud-based and on-premise data analytics applications for faster, smarter data-driven decisions. By making heterogeneous compute fast and simple, Ryft delivers instant insights into any data by eliminating the latency associated with data indexing, transformation and curation. Today, industry leaders rely on Ryft to quickly and simply unlock the value hidden in their data for real-time actionable insight.

INFO@RYFT.COM
WWW.RYFT.COM
855.RYFT.ONE
855.793.8663

7362 CALHOUN PLACE
ROCKVILLE, MD 20855