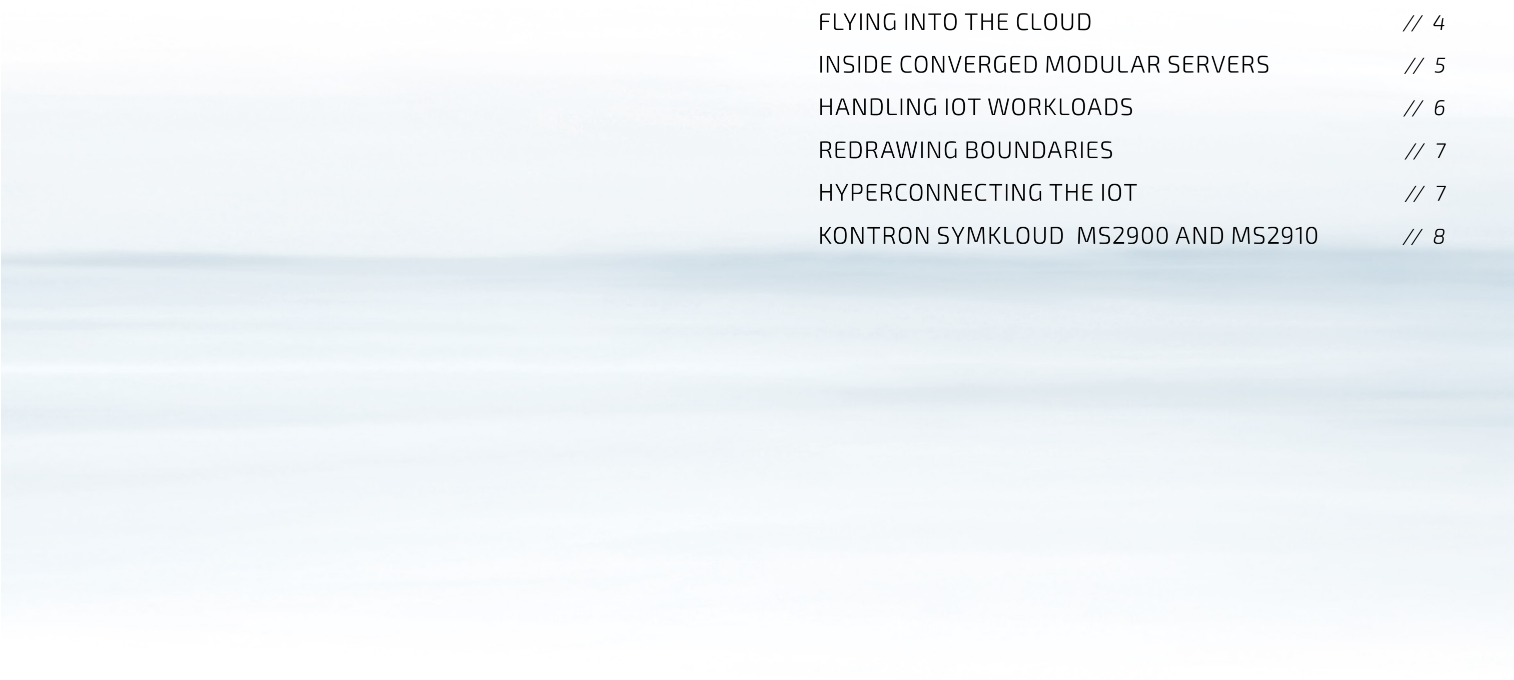
WHITE PAPER

HYPERCONNECTING THE IOT

REALIZING THE FULL POTENTIAL OF LARGE-SCALE END-TO-END IOT APPLICATIONS WITH PULIC AND PRIVATE CLOUD INFRASTRUCTURE

POSSIBILITIES START HERE kontron

CONTENT



WHERE DOES DATA ACQUIRED FROM SENSORS GO, AND WHAT IS THE RIGHT ARCHITECTURE TO ANALYZE IT? THE SIMPLE ANSWER WOULD BE "THE CLOUD."

THE END-TO-END CHALLENGE

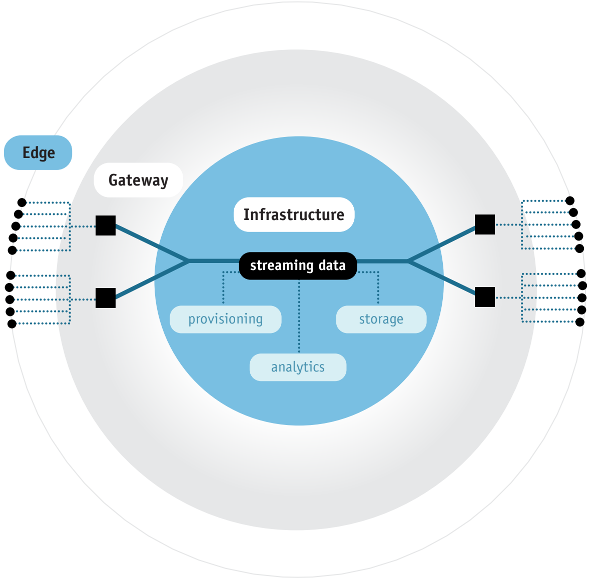
With projects on the Internet of Things (IoT) gaining speed, developers are now facing new territory in creating the infrastructure. Where does data acquired from sensors go, and what is the right architecture to analyze it?

The simple answer would be "the cloud." Scalability and connectivity found in the cloud are certainly desirable features for an IoT platform, but not enough. Successful IoT application design requires an end-to-end approach, drawing on expertise from both information technology (IT) and operational technology (OT) disciplines. Public cloud infrastructure provides low-cost processing and storage capability, but can pose many unforeseen risks in integration and operation as IoT applications grow. Security is one of the biggest concerns in a public cloud; safeguarding valuable IoT data streams and analytics is essential for trust of IoT applications.

Hybrid cloud environments offer a better, safer solution for end-to- end enablement of IoT deployments. Making the task of infrastructure design for the IoT simpler and more flexible are converged modular servers. This new breed of commercial-off-the-shelf (COTS) modular server hardware incorporates the best characteristics of rackmount and blade servers, integrating high-density compute, storage, and network switching into one platform. Walking through some of the aspects of IoT design from sensors to analytics shows how a converged modular server-based approach scales to meet the demands of critical infrastructure designed around hybrid clouds.

BASICS OF IOT TOPOLOGY

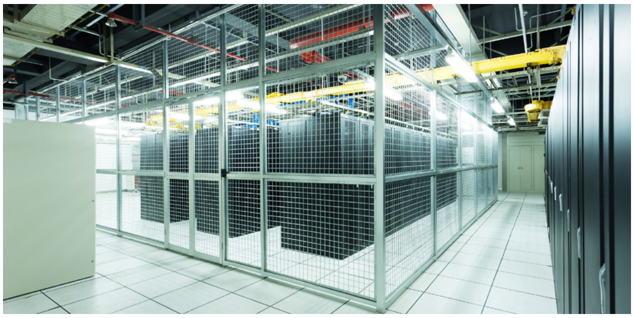
Regardless of the network topology connecting individual nodes, IoT implementations usually fall into a three-tier model. Architectural decisions within each tier establish how IoT applications integrate, scale, and deliver value.
At the outermost tier lies the edge, where end points of the IoT meet the real world. Here, a cluster of sensors and actuators translate physical characteristics to digital data. Sensor networks can be monolithic, with sensors of the same type and interface, capturing data from different physical locations. Many applications deploy multiple sensor types with the intent of fusing the data to derive context. For example, GPS, accelerometers, gyroscopes, and magnetometers combined with sensor fusion algorithms provide indoor navigation capability. Sensors often have long life cycles, and as networks grow, updated versions of sensors with better capability appear.



Gateways at the second tier provide protocol conversion, data aggregation and formatting, and real-time analysis capability. A key role of gateways is accepting a variety of highly optimized protocols for sensors, and converting them to Internet Protocol (IP) for sharing with the cloud and enterprise networks. One example of a gateway is a typical smartphone. With Bluetooth, Wi-Fi, and 3G or 4G connections, a smartphone can connect with sensors, gather data, and upload it to the cloud. Applications using smartphones typically have fewer sensors and lower sample rates, and less critical real-time requirements. For more advanced needs, small form factor embedded computers with specialized software serve as gateways. These platforms can scale, aggregating many sensor data streams and making decisions on data locally in real-time if needed.

AN INFRASTRUCTURE UNABLE TO KEEP UP WITH THE SUM TOTAL OF DATA FLOWING INTO IT RISKS SOFT FAILURE. REALIZING THE FULL POTENTIAL OF A LARGE-SCALE IOT APPLICATION DEPENDS ON INFRASTRUCTURE DESIGN CHOICES.

Application services come from the third tier, an IoT infrastructure in an enterprise data center, or distributed across the cloud. Many IoT applications run dark, capturing and processing data and executing analytics automatically. This implies a higher level of trust for sensors and software in IoT deployments. Asset management services help provision sensors, making sure they are installed and operating correctly, and preventing unauthorized tampering in the sensor/ gateway tiers. Storage services can record both raw data streams and processing results. Advanced analytics capability can examine data, looking for trends and anomalies and even predicting events, and distribute the information across the enterprise network.



FLYING INTO THE CLOUD

Most embedded computing engineers are familiar with designing solutions in the outer two IoT tiers – sensors and gateways. The majority of value in IoT applications comes from analytics, powered in the infrastructure tier. Realizing the full potential of a large-scale IoT application depends on infrastructure design choices.

For example, a popular fallacy is IoT applications are low bandwidth. Depending on the sensor type and required sampling rates, bandwidth from an individual end point may in fact be low. Aggregating sensors at the gateway, then multiplying the effect across multiple clusters each with their own gateway, can change the situation quickly. An infrastructure unable to keep up with the sum total of data flowing into it risks soft failure. Even brief periods of congestion can result in dropped data, missed events, and faulty analytics results.

How should IoT infrastructure design prevent this, while still meeting all the other requirements? IT and OT teams are turning more and more to cloud implementations. By distributing computer resources, platforms share loads and connectivity. This provides resilience to single-point hard failures and reduces bottlenecks due to processing power or network traffic. Achieving scale means adding “instances”, classes of server platforms for particular tasks.

The most popular public cloud platform today is Amazon Elastic Compute Cloud (EC2). Amazon EC2 provides a wide-ranging set of instance choices, with different platforms and availability-based pricing. Platforms range from compute-intensive to instances with GPUs and transactional platforms with SSD storage. Amazon provides a streaming data service, Kinesis, for big data processing.
Public cloud resources like EC2 can be helpful in presenting back- end analytics results, or archiving IoT data, or providing users an overview of what kind of data is coming from where. While these are a breakthrough in a web-user context, they can fall short in a mission-critical IoT context. For high performance IoT applications with many sensors, faster data rates, and split-second decision- making, public clouds leave many concerns unaddressed:

- Determinism: Most public clouds offer few assurances of latency, or quality of service. They handle web interaction measured in seconds, with a person in the loop. On the IoT, latency of more than a few milliseconds can throw a carefully orchestrated process out of control. In normal big data analytics tasks, real time might mean hourly or daily. IoT analytics are often looking for sub-second exceptions.
- Symmetry: Unless a truly dedicated instance is purchased, most public cloud instances are on virtualized servers. When the

application scales, a second instance could be on another processor on the same server – or on an entirely different machine hundreds or thousands of miles away, with different latency. A class of instance may provide the same type of resource, but what else is running on that virtual server?

- Trusted execution: IoT applications likely take steps to authenticate sensors and gateways to prevent intrusion or hacking. A public cloud instance can introduce a random server into the mix, weakening the trust chain considerably. Even reputable cloud vendors can be subject to spoofing, DDoS, or other attacks targeting instances that are loosely trusted.
- Data security: Where, exactly, is the data in the cloud? For example, to deal with compliance issues, Amazon introduced AWS GovCloud – an isolated, quasi-public region with specialized security and monitoring tools. Public instances often lack control over the physical location where data is stored. Is data encrypted, and how, and is it protected end-to-end or just at the point of storage?
- Protocols, ports, and programming environments: The basic transport layer is usually TCP/IP or UDP/IP, but other protocols are of interest. Unless a fully dedicated network interface exists on a public cloud instance, supporting middleware can become challenging if port numbers are controlled or unavailable. Installing environments such as Java may be non-trivial, and unsupported by the cloud vendor.

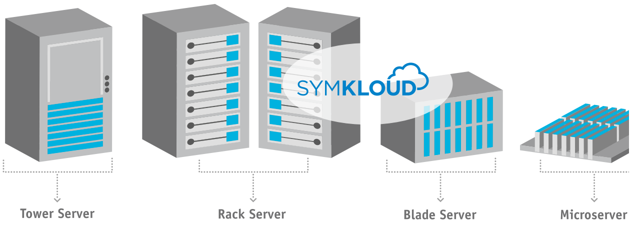
For most IoT applications of significant size, the best answer is a scalable hybrid cloud. This coordinates dedicated private servers handling asset management, real-time streaming, traffic switching, processing and analytics, security, and trusted execution with public cloud resources for presentation and data archiving.

INSIDE CONVERGED MODULAR SERVERS

A new architecture is driving the scalable IoT hybrid cloud: the converged modular server. Compared to a traditional server using a scaled-up processor such as the Intel Xeon processor E5 with up to 18 cores, converged modular servers are reaching into the manycore realm to deliver denser, more flexible compute resources.

Rather than splitting a large workload across several cores, the converged server approach is best for many threads of execution – exactly the right model for IoT infrastructure. By using lower- power multicore processors, modular servers can reduce overall power consumption and space while enabling scale-out performance increases within and between platforms.

Innovation in Server Form Factors



Many of these modular servers feature the Intel Xeon processor E3 family, designed specifically for lighter, web-scale workloads.

Fabric interconnect is common in blade-based servers. Converged modular server architecture also connects multiple processors via a fast, low latency, configurable fabric within a single enclosure. Integrated fabric enables low-latency load balancing between processor cores. An Ethernet-based platform can drop into an existing enterprise network, distributing and shaping traffic.

Redundant, integrated IP switching provides more control over latency and quality of service, and guarantees each node its allocated bandwidth. Traffic shaping also segregates management traffic from data, providing continuous network integrity under all loading conditions.

Also borrowed from blade servers is the idea of hot swappable modules. Instead of a unified design, many converged modular servers are implemented in a compact rackmount chassis, with self- contained modular server subsystems each having their own fabric connection, cooling, and monitoring. Adding or replacing toolless subsystem modules with the platform powered on makes for easier maintenance and upgrades.

An advantage some modular servers have in power consumption is dynamic idle, managing power on a per-server basis. Careful task partitioning means processors can sleep, or even completely power down, until needed. For example, if a modular server were dedicated to asset management or analytics, it could be idled if there were no pending tasks - without affecting incoming IoT data streams handled by other servers in the same platform.

Converged modular servers bring greater rackspace density, better scalability, and lower power consumption. For example, the Kontron SYMKLOUD Series currently packs up to 36 Intel Xeon processor E3 cores or up to 72 Intel Core i7 processor cores in a 2U enclosure – leading to perhaps 756 or 1512 cores in a 42U rack. Near term, new processors will push core counts over 100 per 2U. This framework introduces many new possibilities for partitioning a server optimized for IoT infrastructure use.

HANDLING IOT WORKLOADS

At the sensor tier, calculating data rates for a given sensor and sampling frequency is straightforward. Most IoT applications scale out by adding sensors. This can be more sensors of the same type at a location, new sensors at a new geographic location, or new sensor types for capturing different variables.

Gateways present more complexity at the second tier. They can scale out as new geographic locations are added, supporting entirely new sensor clusters. As sensors multiply on an existing gateway, they may surpass its processing capability, creating a scale-up need for a larger platform. Adding sensors of varying protocols requires more protocol conversion. More pre-processing or localized sensor fusion algorithms also increases workload.

Another factor is the role of advanced middleware, such as publish- subscribe. Physical connections for sensors are usually in hub or mesh topology, aggregated into a gateway. Virtual connections between sensors, gateways, and infrastructure may be more complex. Publish- subscribe networks such as DDS and MQTT are suited for event- driven models, where a sensor publishes readings to any number of interested subscribers anywhere on the network. Pub-sub models usually feature guaranteed message delivery and low latency.

Aggregated IoT traffic might suggest a scale-up server approach for the infrastructure, with bigger platforms to handle more streaming from gateways. Converged modular servers offer a unique approach to scale. Dozens of processor cores with each added platform create more architectural possibilities for dealing with varied workloads.

- Rather than an incidental task run occasionally, a core dedicated to provisioning, management, and security can constantly check on the integrity and health of the IoT network.
- Dedicated cores can service a gateway, processing of incoming data.
- Even finer granularity is possible with cores dedica- ted to individual sensor data streams.
- Mixing core types, including CPU, GPU, and DSP, enables more effective signal processing.

Any of these steps could help with real-time responsiveness to data or events on the IoT network. The most intriguing possibility for the infrastructure tier is dedicating clusters of cores for real- time, predictive analytics. Instead of simply acting on events as they unfold and analyzing data later, predictive analytics can look for trends and patterns that suggest a future event. For instance, instead of scheduled maintenance on running equipment, condition monitoring may suggest the likelihood of an upcoming failure, or determine that all is well and postponing maintenance until necessary.

Big data platforms also enter the picture. Hadoop is an open source scalable framework for distributed data processing and storage. Its key benefit in an IoT context is data can be stored flexibly, either in a single very large file, or in many smaller files – and processed with system-wide visibility. It scales linearly with processing nodes, and with a low-latency fabric interconnect Hadoop shines for non-CPU intensive tasks.

Hadoop is also adept at dealing with both structured and unstructured data. Real-time IoT data streams are often structured data, a flow of sensor readings and timestamps. Enterprise data such as ads triggered by beacons, item purchase history, warranty information, maintenance diagrams, mapping data, photos and video, user manuals, and more is often unstructured. Combining the two enables rich analysis providing breakthrough insights and delivery of information tied to the context of events.

Microsoft has evaluated Hadoop scale-up versus scale-out for IT infrastructure. Their findings are that most actual jobs are sub tera- scale, instead of peta-scale, and scale-up performance becomes a competitive option. One conclusion is more snugly-coupled cores in a single server perform better, as opposed to the same number cores distributed loosely in clusters across a cloud. What they describe as a scale-up server could in fact be a modular server featuring up to 72 cores – such as the Kontron SYMKLOUD Series – with Ethernet fabric interconnect.

For an IoT infrastructure, a hybrid cloud built around converged modular servers helps scaling in all directions.

FOR MORE INSIGHT:
KONTRON IOT MICROSITE | KONTRON SYMKLOUD MICROSITE

Amazon Elastic Compute Cloud, http://aws.amazon.com/ec2/
"Why Hybrid Cloud Makes Sense", Ari Banerjee, Light Reading, October 30, 2014
OMG Data Distribution Service Portal, http://portals.omg.org/dds/

REDRAWING BOUNDARIES

A hybrid cloud approach to IoT infrastructure featuring modular servers also helps free application developers from many other boundaries.

- On a private modular server, virtualization environments are fully controlled – everything running on the platform is known. This opens the potential for network functions virtualization (NFV), reducing both capital expenditures and lifecycle costs as appliances are consolidated.
- Developers can select and deploy the right software approach. Operating systems can mix and match between real-time, Linux, or Windows. Installing and configuring middleware such as OSGi, DDS, MQTT, or other services is unconstrained on a private platform in the hybrid cloud.
- Manycore processing and fabric interconnect mean the ability to ingest and process more IoT data and provide faster and more detailed analytics. This can lead to improvements or cost reduction in existing services, or incremental revenue from new service offerings.
- The highest potential of the IoT lies in monetizing analytics and data, and creating powerful new business models. With a hybrid cloud in place, data assets can be developed and shared or fully protected. Ability to grow and adapt the infrastructure quickly means improved competitiveness.

All these benefits apply not only to OEMs developing IoT applications, but also to cloud service providers (CSPs) looking to extend their existing M2M offerings or break into IoT services. A hybrid cloud running on converged modular servers offers better scalability, improved rackspace density, reduced power consumption and cooling requirements, and flexibility to target IoT application needs more precisely.

HYPERCONNECTING THE IOT

The deployment of converged modular servers in hybrid clouds is just part of Kontron's overall strategy in hyperconnecting the IoT. With unique experience in both industrial computing and telecom/data center solutions, Kontron teams have experience with all the aspects needed for end-to-end success in IoT applications.

Kontron is a Premier Member of the Intel Internet of Things Solutions Alliance. This means Kontron's customers have access to the latest Intel technology combined with extended lifecycle support for embedded applications.

KONTRON SYMKLOUD MS2900 AND MS2910

Hybrid cloud IoT infrastructure demands a platform that delivers quality of service for years under any conditions. Kontron's SYMKLOUD Series brings converged modular servers onto the IoT, with unique support for IT/OT integration and mission-critical applications.

KEY FEATURES OF THE SYMKLOUD SERIES:

- 2U height, 21" (533.4mm) depth
- Up to nine Intel® Xeon E3-1265 Lv2 quad-core processors – up to a total of 36 cores
- Up to 18 Intel Core i7-4860EQ GT3e Iris Pro processors - up to a total of 72 cores
- Up to nine Intel Core i7-4860EQ processors combined with PCIe expansion slot
- Ethernet fabric interconnect, GigE (MS2900) or 10GigE (MS2910)
- Integrated, redundant L4 to L7 switching
- Up to 2 load balancer subsystems
- All modules hot-swappable
- Up to 13.5 TB storage (HDD or SSD)

The ruggedized SYMKLOUD Series borrows design philosophy from NEBS telecom-grade infrastructure. It is able to endure higher temperatures, and more shock and vibration.

Packing 36 cores in a 2U rackmount is remarkable, but depth is also critical. A 21" chassis allows SYMKLOUD into both data centers and applications such as transportation and industrial environments.

IT-style platform management features include 1-click updates, BMC with advanced options, SNMP and IPMI, and remote management for diagnostics and provisioning.

The SYMKLOUD Series has 5 to 7 year lifecycle support, reducing capex, development, and maintenance costs compared to typical servers.

About Kontron

Kontron, a global leader in embedded computing technology and trusted advisor in IoT, works closely with its customers, allowing them to focus on their core competencies by offering a complete and integrated portfolio of hardware, software and services designed to help them make the most of their applications.

With a significant percentage of employees in research and development, Kontron creates many of the standards that drive the world's embedded computing platforms; bringing to life numerous technologies and applications that touch millions of lives. The result is an accelerated time-to-market, reduced total-cost-of-ownership, product longevity and the best possible overall application with leading-edge, highest reliability embedded technology.

Kontron is a listed company. Its shares are traded in the Prime Standard segment of the Frankfurt Stock Exchange and on other exchanges under the symbol "KBC". For more information, please visit: **www.kontron.com**

CORPORATE OFFICES

EUROPE, MIDDLE EAST
& AFRICA

Lise-Meitner-Str. 3-5
86156 Augsburg
Germany
Tel.: +49 821 4086-0
Fax: +49 821 4086-111
info@kontron.com

NORTH AMERICA

14118 Stowe Drive
Poway, CA 92064-7147
USA
Tel.: +1 888 294 4558
Fax: +1 858 677 0898
info@us.kontron.com

ASIA PACIFIC

1~2F, 10 Building, No. 8 Liangshuihe 2nd Street,
Economical & Techonological Development Zone,
Beijing, 100176, P.R. China
Tel.: +86 10 63751188
Fax: +86 10 83682438
info@kontron.cn