



Урок 20. HW4K8S

Михаил Цветков
инженер
Intel



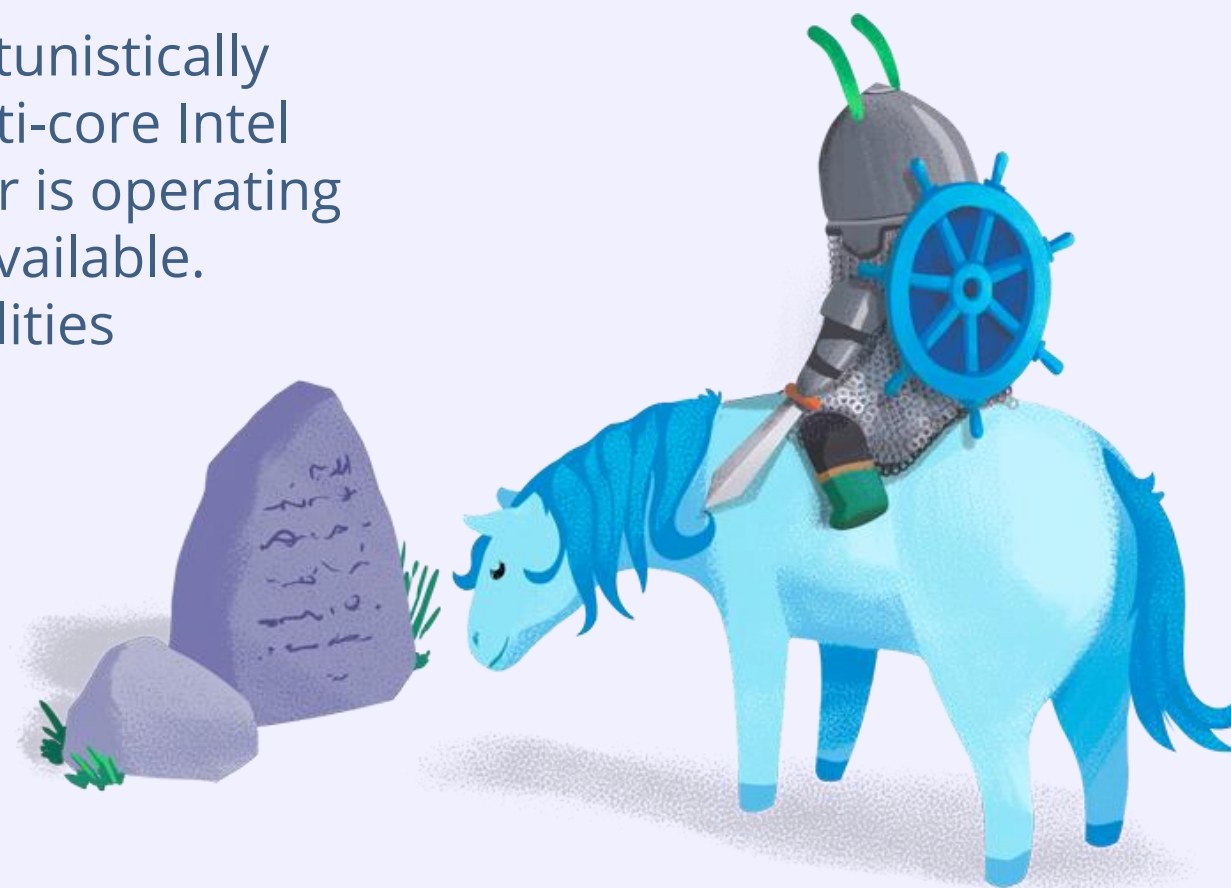
Михаил Цветков

Технический директор Intel в России
Инженер

- участвовал в разработке микроархитектур Intel
- многократно и успешно применял микроархитектуры Intel на практике
- видел, как строят датацентры и знает, как они работают
- любит об этом рассказывать
- характер нордический. женат.

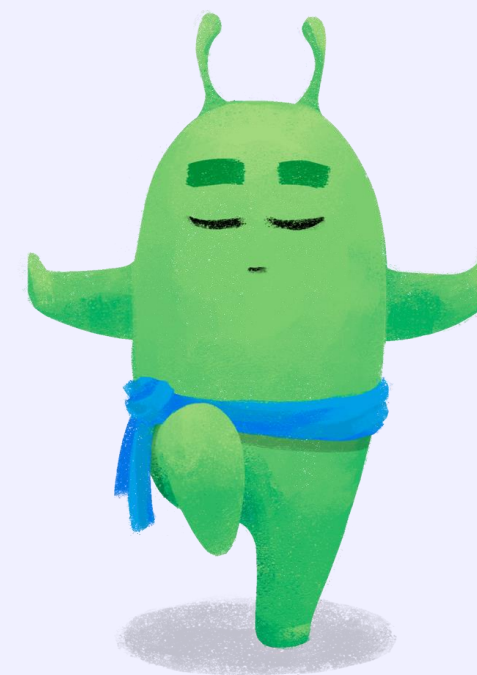
Notices and Disclaimers

- Performance varies by use, configuration and other factors. Learn more at www.Intel.com/PerformanceIndex
- Performance results are based on testing as of dates shown in configurations and may not reflect all publicly available updates. See backup for configuration details.
- No product or component can be absolutely secure.
- Includes the effect of Intel Thermal Velocity Boost, a feature that opportunistically and automatically increases clock frequency above single-core and multi-core Intel Turbo Boost Technology frequencies based on how much the processor is operating below its maximum temperature and whether turbo power budget is available. The frequency gain and duration is dependent on the workload, capabilities of the processor and the processor cooling solution.
- Code names are used by Intel to identify products, technologies, or services that are in development and not publicly available. These are not "commercial" names and not intended to function as trademarks.
- © Intel Corporation. Intel, the Intel logo, and other Intel marks are trademarks of Intel Corporation or its subsidiaries. Other names and brands may be claimed as the property of others.



Для кого

- Для тех, кто разрабатывает и зарабатывает на микросервисах
- Для тех, кто только собирается и выбирает платформу для K8S
- Для всех, кто интересуется современным железом



План

1

Теория и История

2

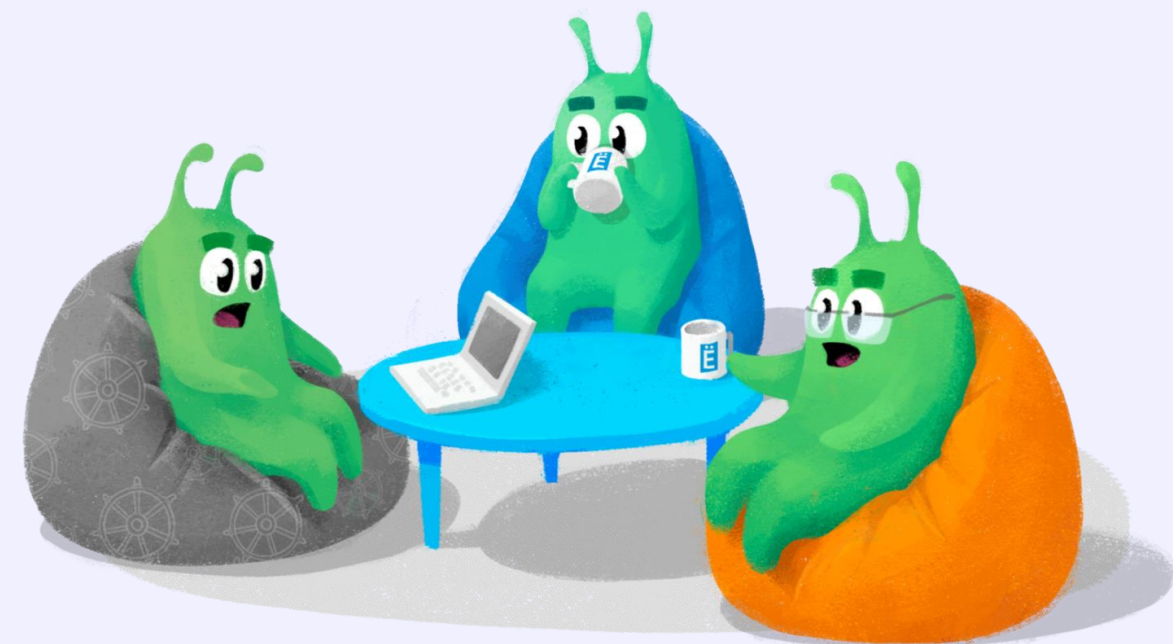
Особенности микросервисов

3

Серверные платформы для K8S

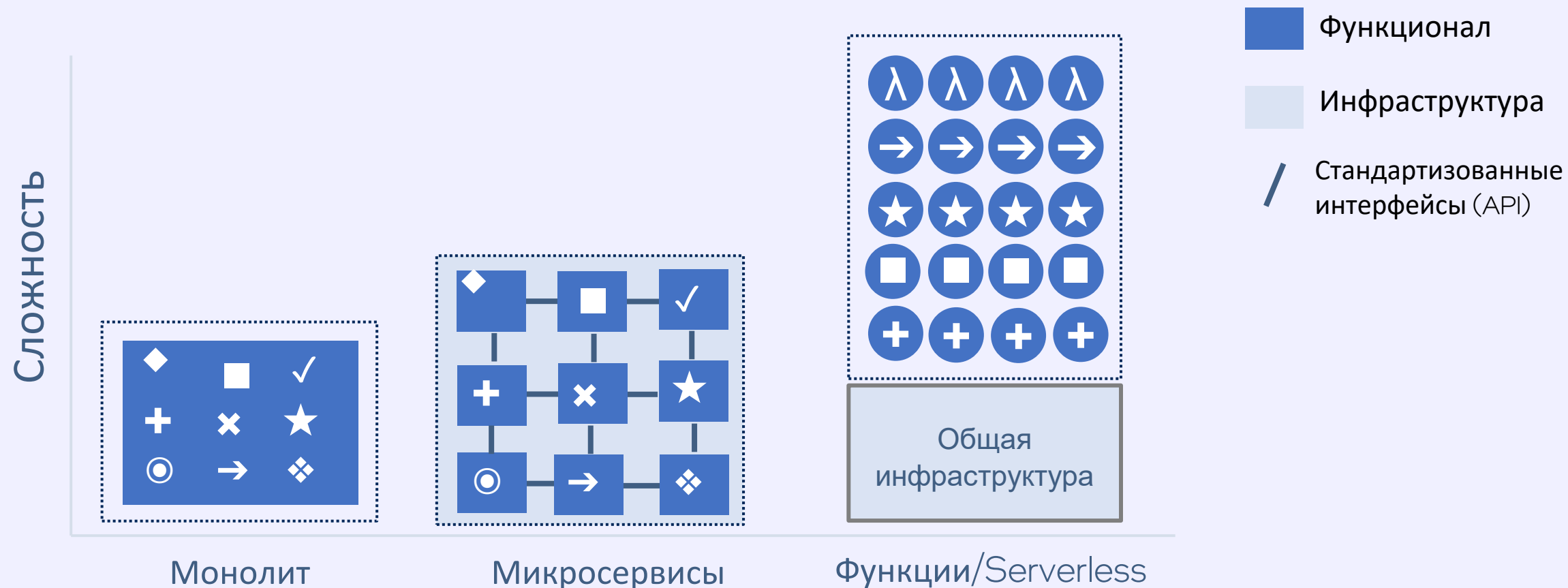
4

Инфраструктура датацентров



Почему микросервисы

- “Эксперты предсказывают, что в 2022 году 90% всех приложений в датацентрах/облаках будет разрабатываться на микросервисной архитектуре”*
- 83% всех новых облачных SaaS приложений используют микросервисы**



Переход на микросервисы



* <https://www.charterglobal.com/five-microservices-trends-in-2020/>

**Source: Intel Microservices Insights Study, June 2021

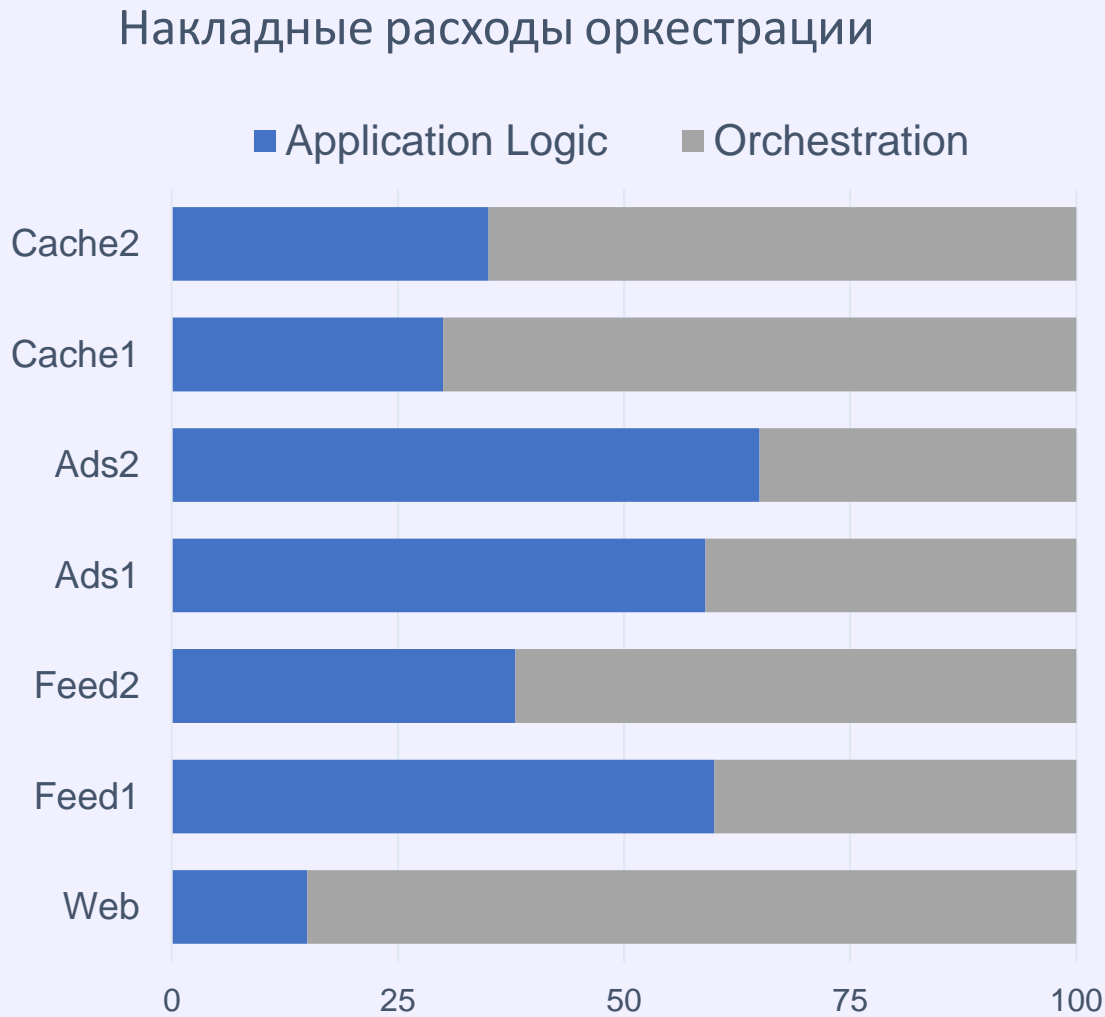
*** Container Infrastructure Software Market Assessment: x86 Containers Forecast, 2018–2023; IDC 2020

Немного истории

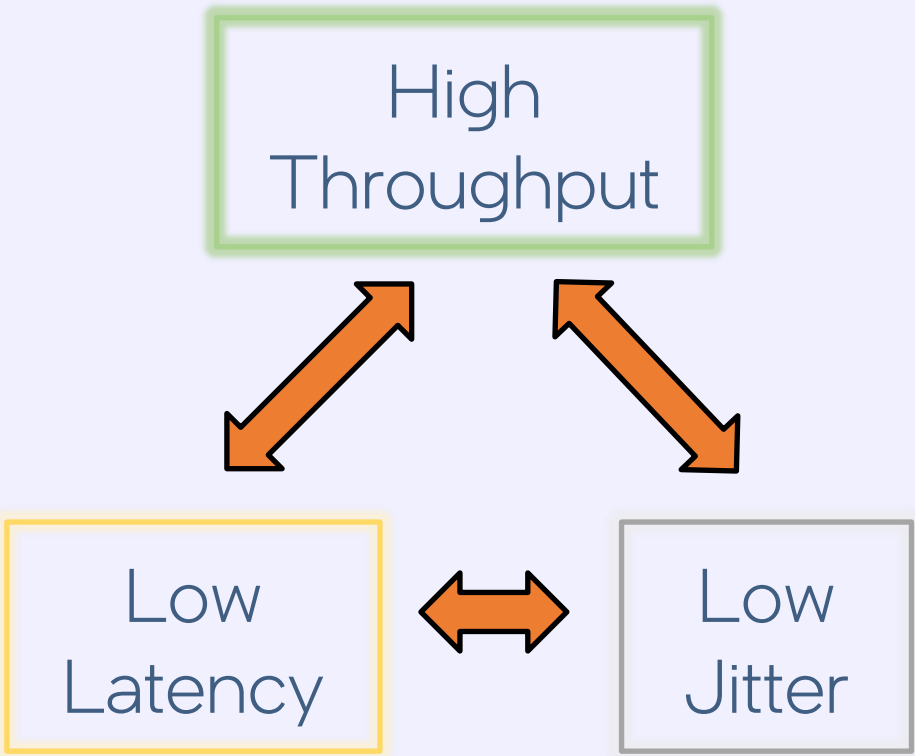
- 2005 – **Intel virtualization (VT-x)**
 - 2008 – Intel Core, Nehalem uArch
 - 2013 – Docker
 - 2014 – Kubernetes
 - 2021 - **Intel Software Guard Extensions (Intel SGX) 2.0**
-
- Полная Аппаратная Виртуализация
IaaS, конвергенция сети и хранения на IA
- Программная Виртуализация на уровне OS
SaaS, микросервисная декомпозиция в облаках

Особенности микросервисов

Facebook production microservices**



Требования к железу



** From Accelerometer: Understanding Acceleration Opportunities for Data Center Overheads at Hyperscale, Akshitha Sriraman, Abhishek Dhanotia. Facebook. 2020

Особенности микросервисов

Быстродействие
Макс задержки

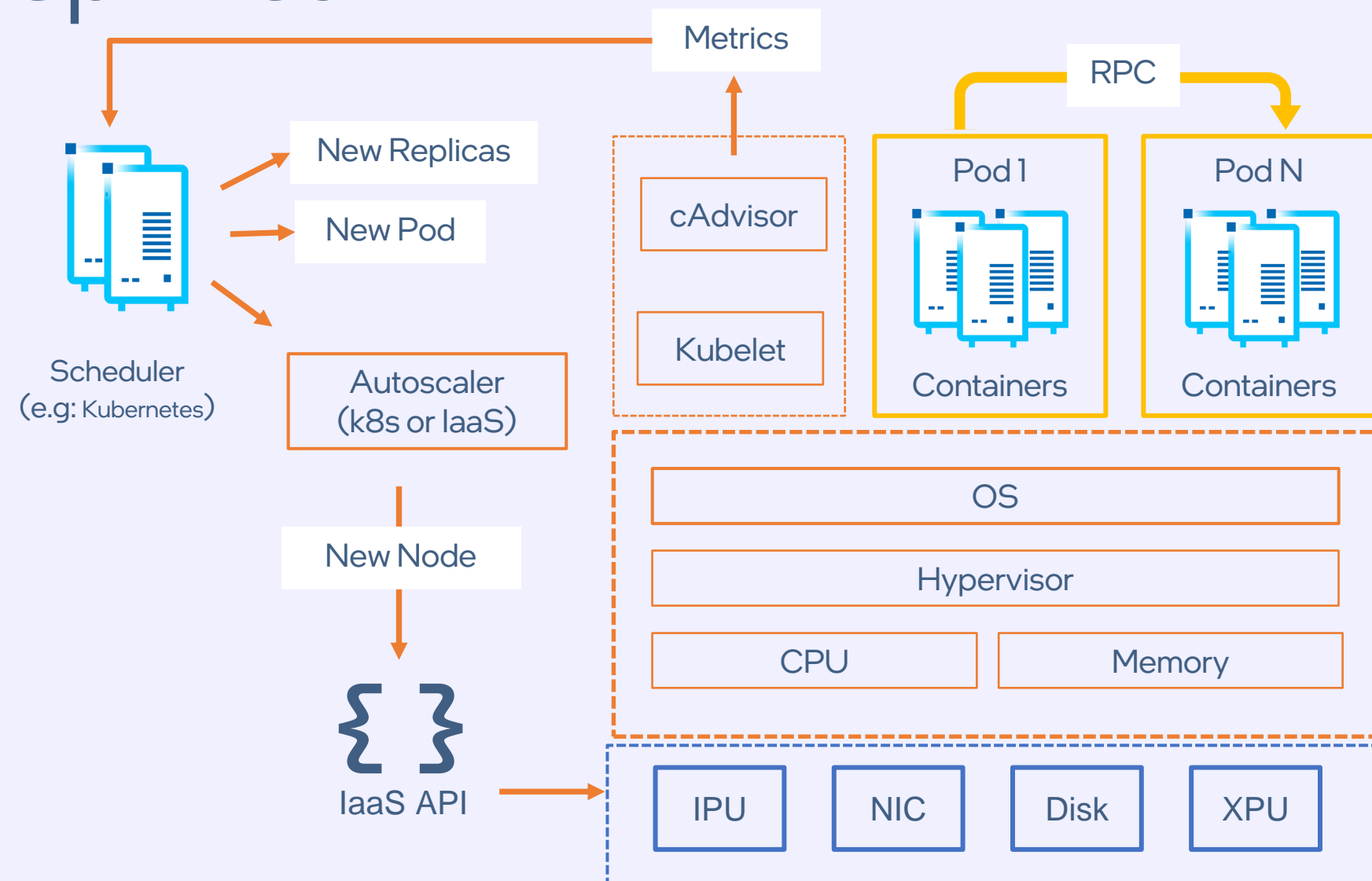
SLO на макс задержку при масштабировании на миллионы реплик
Текущее решение: резервирование ресурсов (overprovisioning)

Наблюдаемость

Высокая динамика, переходные процессы – что и где происходит
Текущее решение : Traces, Metrics and Logs for service health and SLO trends

Расходы на инфраструктуру

Load Balancing, Orchestration, communication, resource contention
Current Solutions: service mesh, policies for scheduler, extensive tuning and optimization



“Rethinking search architecture to deal with tail latencies, overhead increasing non-linearly with traffic growth, it’s a million-dollar pain point”: Large SaaS Provider

“We’re constantly adding metrics to understand usage, ensure rightsizing and prevent resource starvation” : Large US B2C Cloud-Native SaaS Company

“Resource contention can cause CPU/MEM limits to be hit, which sends traffic to remaining pods, continuing the failure cycle”: Large B2C SaaS Company

“memset takes up high single digit percent of cpu for a large-scale service, that’s a lot of cpu cycles for writing zero’s”: US Cloud Provider

*Microservices Trends and Challenges:

Google : <https://static.googleusercontent.com/media/research.google.com/en//pubs/archive/44271.pdf>

Twitter: https://blog.twitter.com/engineering/en_us/topics/infrastructure/2017/the-infrastructure-behind-twitter-scale.html

Netflix: <https://medium.com/swlh/a-design-analysis-of-cloud-based-microservices-architecture-at-netflix-98836b2da45f>

Facebook: <https://research.fb.com/wp-content/uploads/2019/05/SoftSKU-Optimizing-Server-Architectures-for-Microservice-Diversity-@Scale.pdf>

Процессоры Intel для датацентров и облачных приложений

2019

Cascade Lake

Intel DL Boost (VNNI)
Intel® Optane™ Persistent Memory

2020

Cooper Lake

Intel DL Boost (BFLOAT16)

2021

Ice Lake

Intel Software Guard Extensions
(Intel® SGX)

2022

Next Gen Xeon

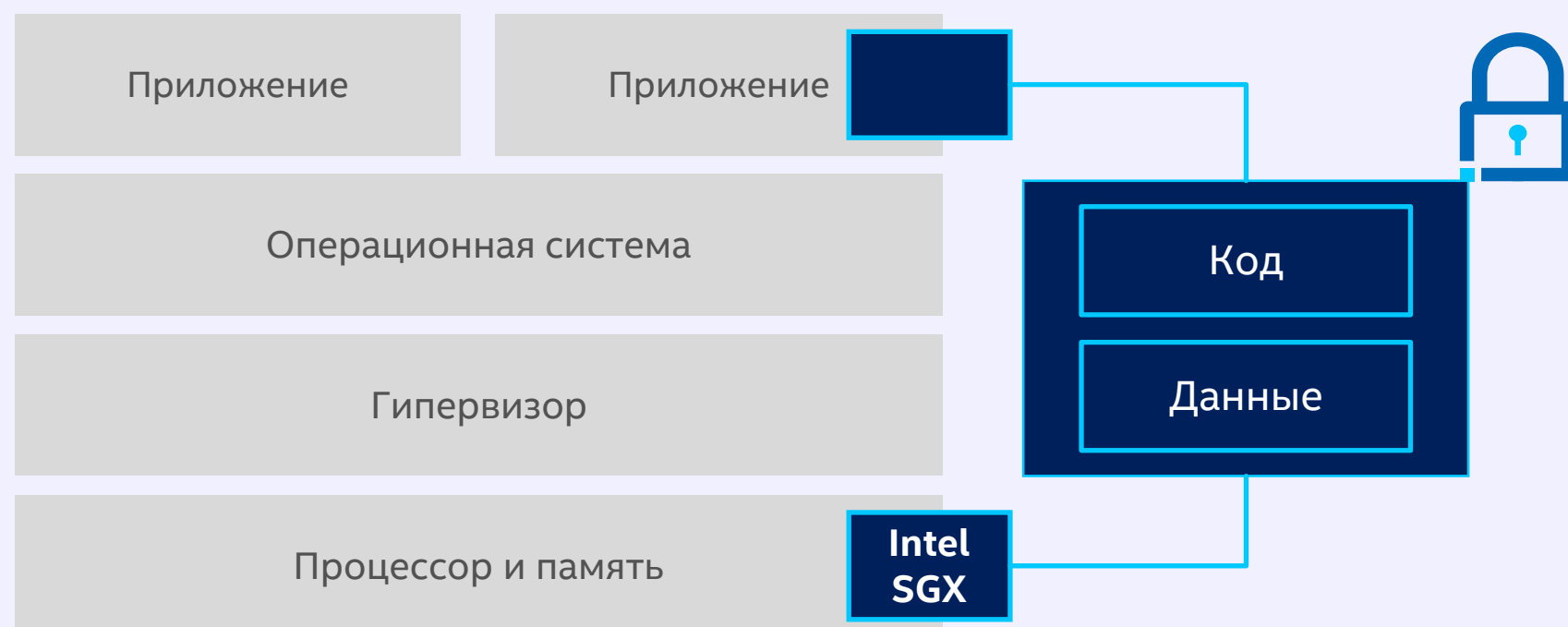
Sapphire Rapids

Intel Data Stream Accelerator (DSA)
Accelerator Interfacing Architecture (AiA)
Intel Dynamic Load Balancer (DLB)
Intel QuickAssist Technology (QAT)
Intel Shared Virtual Memory (SVM)
Intel Scalable IO Virtualization (SIOV)

Масштабируемые процессоры Intel® Xeon® 3-го поколения Ice Lake – Intel Software Guard Extensions (Intel SGX)

Конфиденциальность для важных сегментов данных без ущерба для производительности

Огромные анклавы теперь поддерживают требования современных рабочих нагрузок (до 1 ТБ памяти)

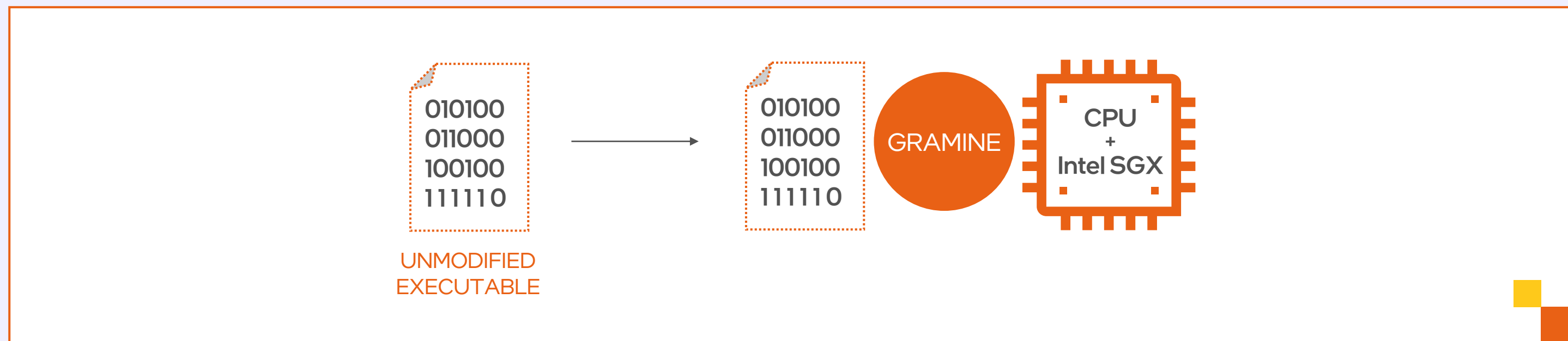
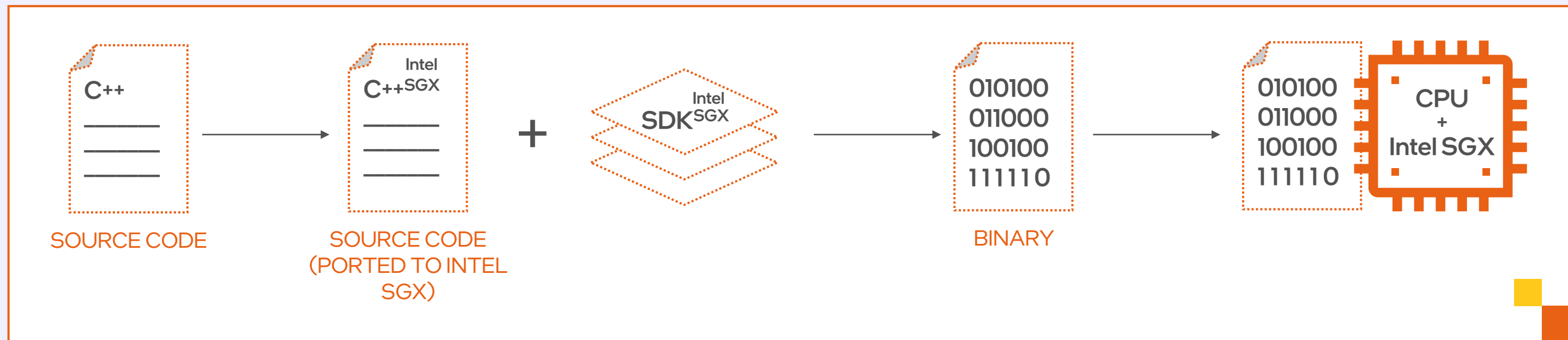


Intel SGX в K8S



<https://github.com/scontain/hello-world-kubernetes>

Gramine Project: Library OS для SGX контейнеров



Масштабируемые процессоры Intel® Xeon® 3-го поколения Ice Lake – **новые инструкции**

Cryptography

- Big-Number Arithmetic (AVX-512 Integer IFMA)
- Vector AES and Vector Carry-less Multiply Instructions
- Galois Field New Instructions (GFNI)
- SHA-NI

Compression/Decompression and Special SIMD

- Bit Algebra
- VBMI – Vector Bit Manipulation Instruction

New SIMD ISA Utilizing AVX512 on ICX

Vector **CLMUL**

Vector **AES**

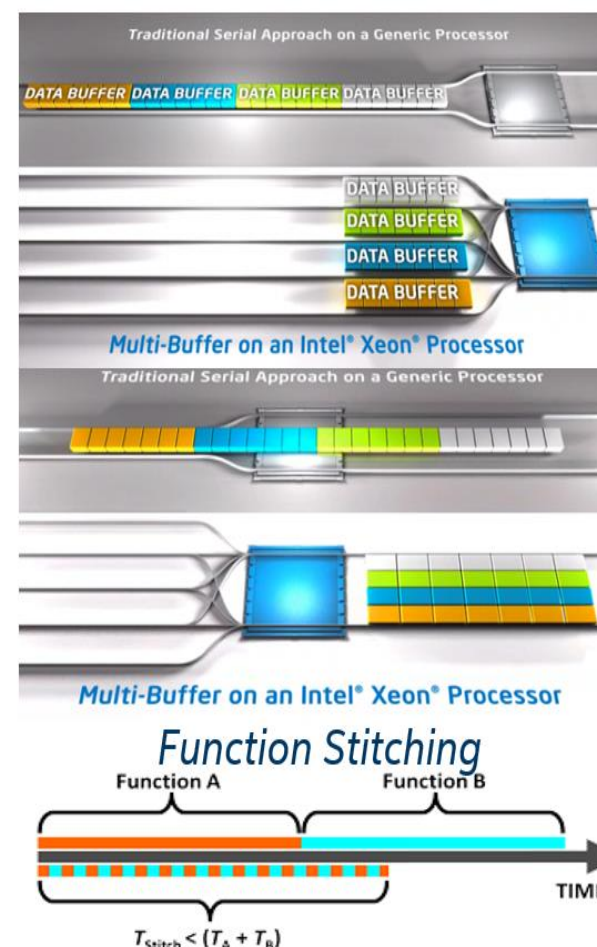
VPMADD52

SHA Extensions

GFNI

Software / Algorithms

Multi-Buffer



Ice Lake vs. Cascade Lake Per Core Performance

ECDHE x25519 **4.12X**

RSA Sign 2048 **5.63X**

ECDHE p256 **2.73X**

AES-CTR **3.84X**

AES-CMAC **3.78X**

AES-XTS **3.5X**

AES-GCM **3.34x**

ECDSA Sign p256 **1.9X**

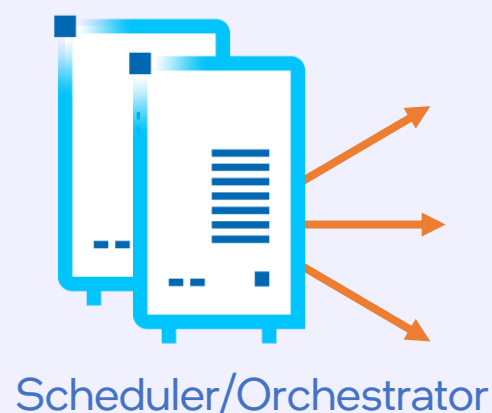
CRC **2.3X**

ZUC **1.5X**

Performance varies by use, configuration and other factors. Configurations see appendix [1,2,3]

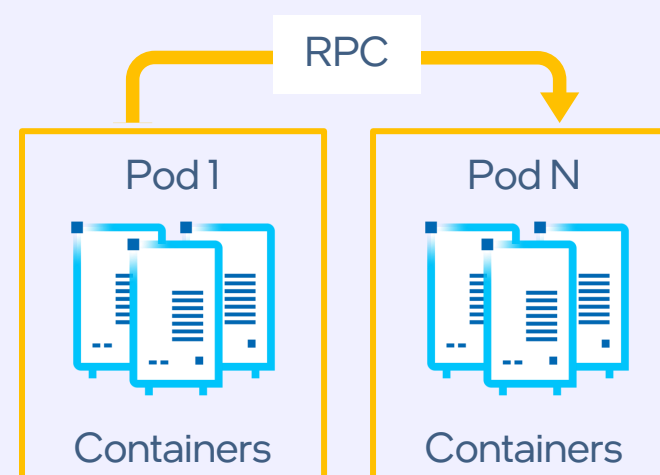
Процессоры Intel® Xeon® следующего поколения Sapphire Rapids – акселератор для микросервисов

Эффективность планировки



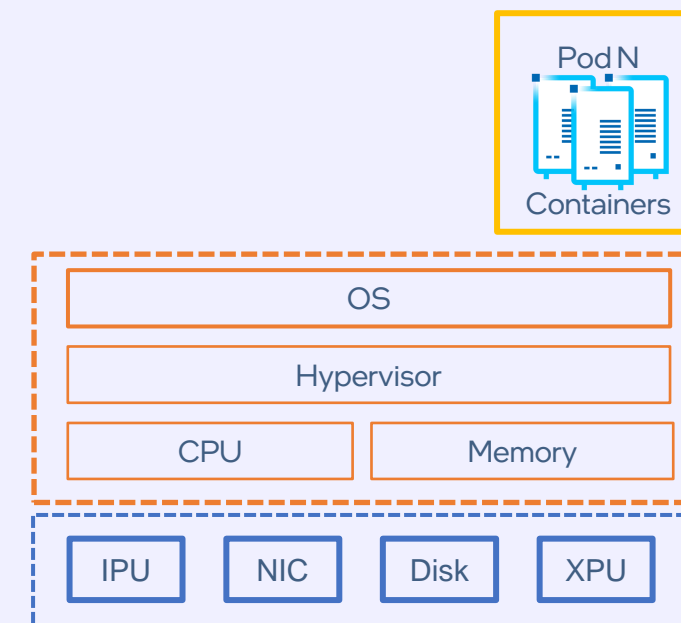
- Device/Feature Awareness (Kubernetes Enhancements)
- Faster container spin-up / tear-down (DSA)
- Enhanced Observability (New PMU capabilities)

Быстрее коммуникации



- Low Latency Comms (QAT, DLB)
- Efficient IPU Offload
- IPU 'brownfield' use cases: QoS, Ops (provisioning, caching, audits, etc.), Telemetry

Производительность



- Runtime Language Optimization (iTLB, iCache, Vectorization, ISA's)
- Optimized Software Stack (kernel, glibc, other perf libs)
- Platform optimizations (SVM, RDT, etc.)

Процессоры Intel® Xeon® следующего поколения Sapphire Rapids – Микросервисы

Goal

Enable higher throughput while meeting latency requirements and reducing infrastructure overhead for execution, monitoring, and orchestration thousands of microservices

Improved Performance and Quality of Service

Runtime Languages – lower latency for Runtime Languages
AiA ISA's – efficient worker threads, signaling, and synch.

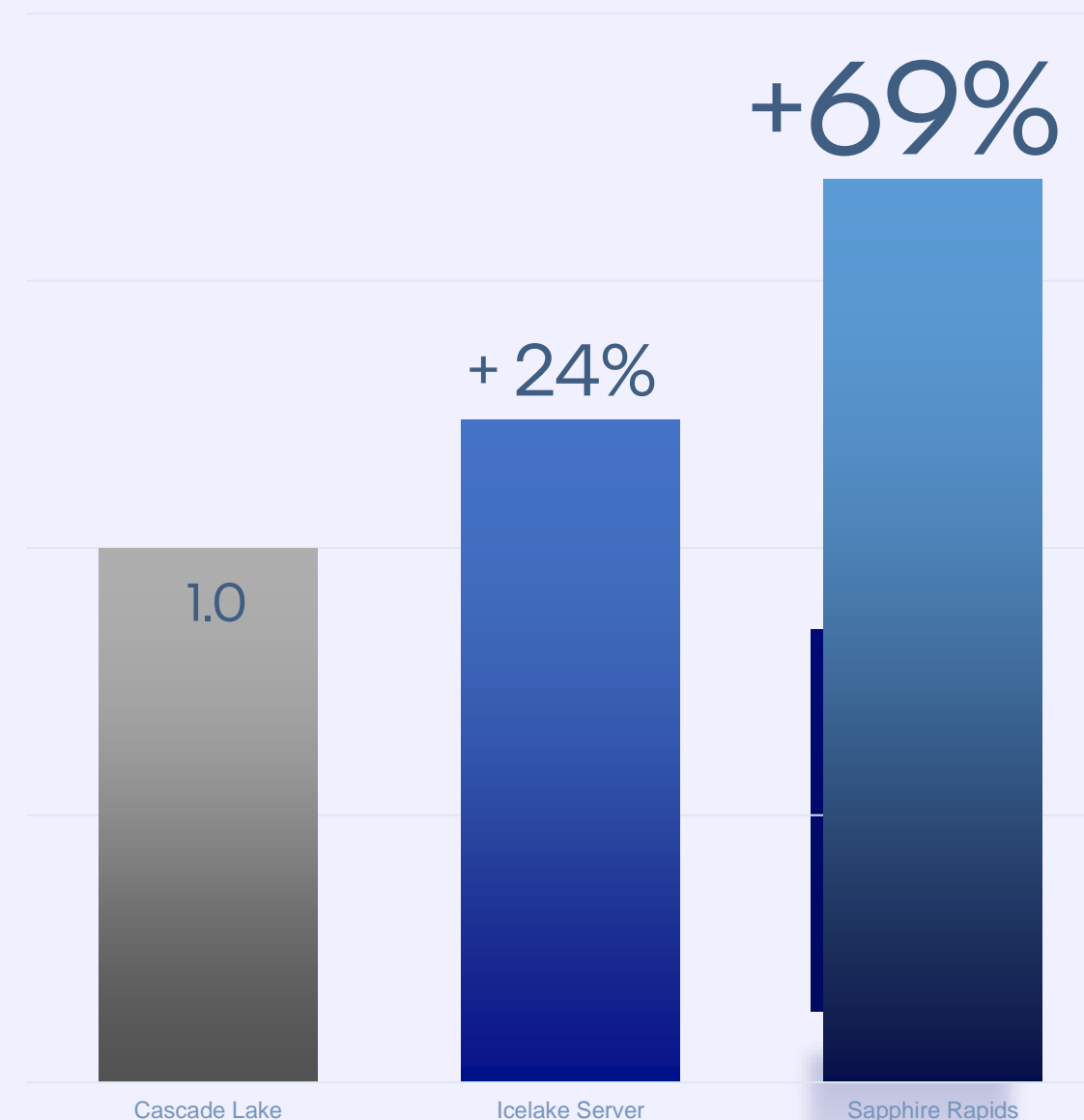
Reduced Infrastructure Overhead

Kubernetes – enhanced for scaling, placement, and policies
Advanced Telemetry – easier analysis & optimization

Better Distributed Communication

Improved latency of Remote procedure calls and service-mesh
QAT, DSA etc. – optimized networking and data movement

Throughput per Core under Latency SLA of p99 <30ms



Results have been estimated or simulated based on testing on pre-production hardware and software. For workloads and configurations visit www.intel.com/InnovationEventClaims. Results may vary

Инфраструктура датацентров: сеть и хранение

Передавать быстрее



Intel® Ethernet E810-2CQDA2

До 200GbE на один PCIe 4.0 слот для высоконагруженных приложений

Хранить больше



Intel® Optane™ SSD P5800X

Быстрейший SSD на планете



Intel® Optane™ Persistent Memory 200 series

До 6TB памяти на сокет + постоянное хранилище данных



Intel® SSD D5-P5316

Первый PCIe 4.0 144-слойный QLC 3D NAND делает возможным 1PB хранения в 1U корпусе

Обрабатывать все



Процессоры

Intel® Xeon® Scalable третьего поколения

Самый быстрый серверный процессор Intel со встроенными решениями для ИИ и безопасности

Intel® Agilex™ FPGA

Передовое решение по производительности FPGA логики и энергоэффективности

Оптимизированные решения



>500
Партнерских Решений

Container Bare Metal Reference Architecture Guide

HARDWARE & OPEN-SOURCE SOFTWARE

K8s NETWORKING

PACKET PROCESSING

RESOURCE MANAGEMENT

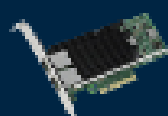
OBSERVABILITY



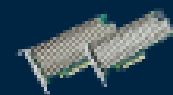
Intel FPGA



Intel® Optane™



Intel® Ethernet
Controller



Intel® QuickAssist
Technology

INSTALLATION PLAYBOOK

ANSIBLE SCRIPTS



OPERATORS



HELM CHARTS



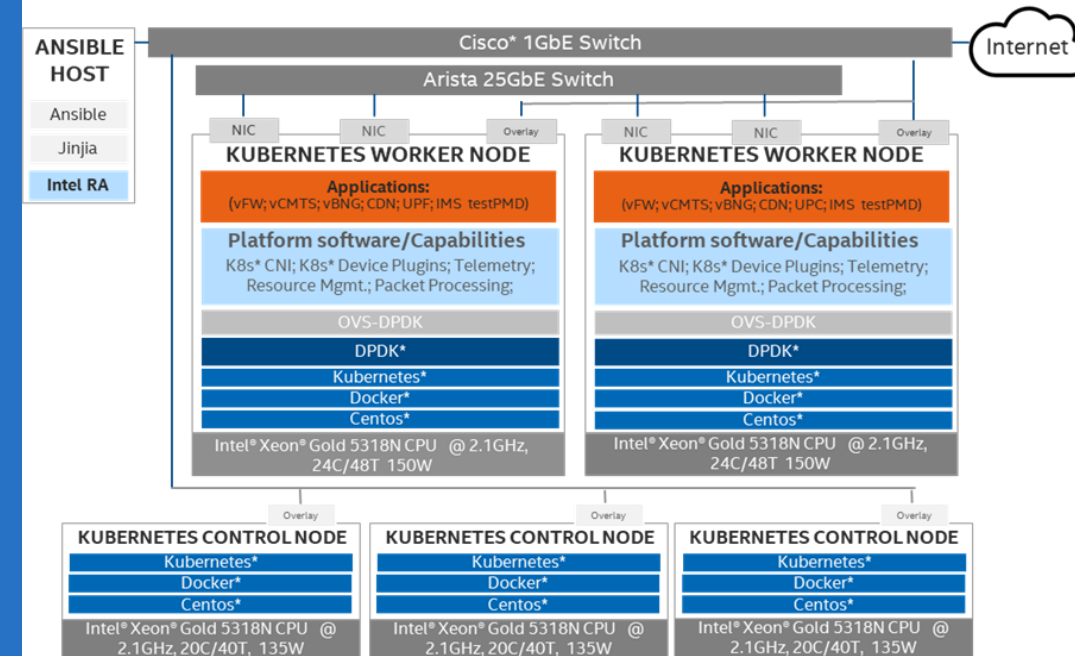
KUBESPRAY



Optimized Setup
<30 min

INTEL REFERENCE ARCHITECTURES

Kubernetes* Cluster



CONFIGURATION PROFILES PER NETWORK LOCATION & WORKLOADS

ON
PREMISES
EDGE

VISUAL CLOUD:
ADI; SMTc

REMOTE
CENTRAL
OFFICE

5G-UPF; CMTS;
BNG; CDN

REGIONAL
DATA
CENTER

5G-UPF; CMTS;
BNG; CDN



Intel: Container Experience Kits



Container Reference Architecture

The container bare metal reference architecture release represents...



Kubernetes Networking

Kubernetes support for network functionality required in NFV use cases.



Acceleration

Open platforms provide the scalability and flexibility needed for almost any network usage....



Resource Management

Kubernetes match the server capabilities to workload requirements and...



Telemetry

Ensure platform and container metrics and events are accessible through industry standard...



Training and Demo Links

Watch the following training and demo videos to understand the newly developed...



Cable Use Case

Intel optimizes virtualized, cable modem termination system (vCMTS) architecture...

Intel: Container Experience Kits



<https://networkbuilders.intel.com/intel-technologies/container-experience-kits>

СЛЕРМ при поддержке

 VK Cloud Solutions

+

 intel.

СЛЕПМ при поддержке



VK Cloud Solutions



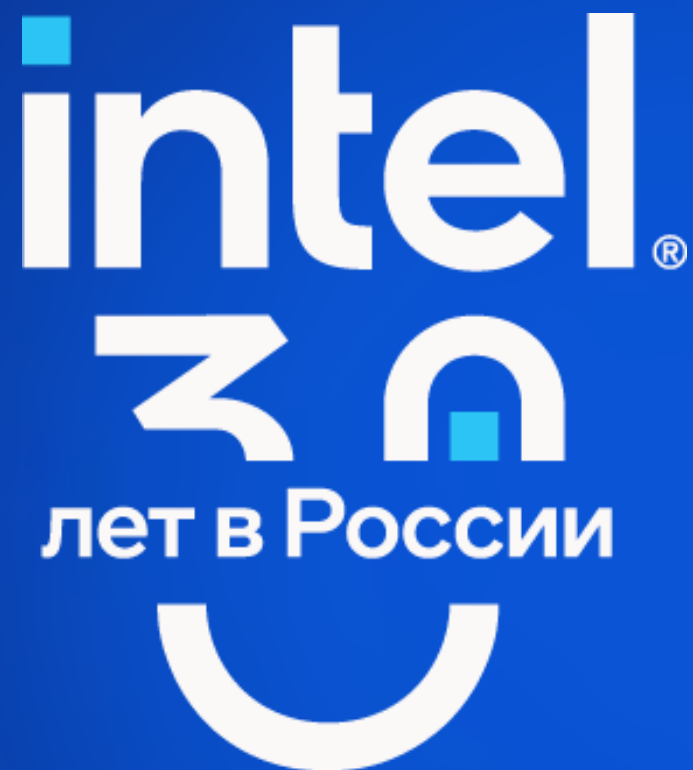
intel.



Спасибо!

Appendix

1. **3.34x higher IPsec AES-GCM performance, 3.78x higher IPsec AES-CMAC performance, 3.84x higher IPsec AES-CTR performance, 1.5x higher IPsec ZUC performance:** 8380: 1-node, 2x Intel(R) Xeon(R) Platinum 8380 CPU on M50CYP2SB2U with 512 GB (16 slots/ 32GB/ 3200) total DDR4 memory, ucode 0x8d055260, HT On, Turbo Off, Ubuntu 20.04.2 LTS, 5.4.0-66-generic, 1x Intel 1.8TB SSD OS Drive, intel-ipsec-mb v0.55, gcc 9.3.0, Glibc 2.31, test by Intel on 3/17/2021. 8280M: 1-node, 2x Intel(R) Xeon(R) Platinum 8280M CPU on S2600WFT with 384 GB (12 slots/ 32GB/ 2933) total DDR4 memory, ucode 0x4003003, HT On, Turbo Off, Ubuntu 20.04.2 LTS, 5.4.0-66-generic, 1x Intel 1.8TB SSD OS Drive, intel-ipsec-mb v0.55, gcc 9.3.0, Glibc 2.31, test by Intel on 3/8/2021.
2. **3.5x higher ISA-L AES-XTS performance, 2.30x higher ISA-L CRC performance:** ISA-L: 8380: 1-node, 2x Intel® Xeon® Platinum 8380 Processor, 40 cores HT On Turbo OFF Total Memory 512 GB (16 slots/ 32GB/ 3200 MHz), Data protection (Reed Solomon EC (10+4)), Data integrity (CRC64), Hashing (Multibuffer MD5), Data encryption (AES-XTS 128 Expanded Key), Data Compression (Level 3 Compression (Calgary Corpus)), BIOS: SE5C6200.86B.3021.D40.2103160200 (ucode: 0x8d05a260), Ubuntu 20.04.2, 5.4.0-67-generic, gcc 9.3.0 compiler, yasm 1.3.0, nasm 2.14.02, isal 2.30, isal_crypto 2.23, OpenSSL 1.1.1.i, zlib 1.2.11, Test by Intel as of 03/19/2021. 8280: 1-node, 2x Intel® Xeon® Platinum 8280 Processor, 28 cores HT On Turbo OFF Total Memory 384 GB (12 slots/ 32GB/ 2933 MHz), BIOS: SE5C620.86B.02.01.0013.121520200651 (ucode:0x4003003), Ubuntu 20.04.2, 5.4.0-67-generic,, gcc 9.3.0 compiler, yasm 1.3.0, nasm 2.14.02, isal 2.30, isal_crypto 2.23, OpenSSL 1.1.1.i, zlib 1.2.11 Test by Intel as of 2/9/2021. Performance measured on single core.
3. **5.63x higher OpenSSL RSA Sign 2048 performance, 1.90x higher OpenSSL ECDSA Sign p256 performance, 4.12x higher OpenSSL ECDHE x25519 performance, 2.73x higher OpenSSL ECDHE p256 performance,** 8280M: 1-node, 2x Intel(R) Xeon(R) Platinum 8280M CPU on S2600WFT with 384 GB (12 slots/ 32GB/ 2933) total DDR4 memory, ucode 0x5003003, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x INTEL_SSDSC2KG01, OpenSSL 1.1.1j, GCC 9.3.0, test by Intel on 3/5/2021. 8380: 1-node, 2x Intel(R) Xeon(R) Platinum 8380 CPU on M50CYP2SB2U with 512 GB (16 slots/ 32GB/ 3200) total DDR4 memory, ucode 0xd000270, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x INTEL_SSDSC2KG01, OpenSSL 1.1.1j, GCC 9.3.0, QAT Engine v0.6.4, test by Intel on 3/24/2021. 8380: 1-node, 2x Intel(R) Xeon(R) Platinum 8380 CPU on M50CYP2SB2U with 512 GB (16 slots/ 32GB/ 3200) total DDR4 memory, ucode 0xd000270, HT On, Turbo Off, Ubuntu 20.04.1 LTS, 5.4.0-65-generic, 1x INTEL_SSDSC2KG01, OpenSSL 1.1.1j, GCC 9.3.0, QAT Engine v0.6.5, test by Intel on 3/24/2021.
4. 20% IPC improvement: 3rd Gen Xeon Scalable processor: 1-node, 2x 28-core 3rd Gen Intel Xeon Scalable processor, Wilson City platform, 512GB (16 slots / 32GB / 3200) total DDR4 memory, HT on, ucode=x270, RHEL 8.0, Kernel Version 4.18.0-80.el8.x86_64, test by Intel on 3/30/2021. 2nd Gen Intel Xeon Scalable processor: 1-node, 2x 28-core 2nd Gen Intel Xeon Scalable processor, Neon City platform, 384GB (12 slots / 32GB / 2933) total DDR4 memory, HT on, ucode=x2f00, RHEL 8.0, Kernel Version 4.18.0-80.el8.x86_64, test by Intel on 3/30/2021. SPECrate2017_int_base (est). Tests at equal frequency, equal uncore frequency, equal compiler.
5. Intel® Optane™ Persistent Memory 200 Series, Average 32% more memory bandwidth: Based on testing by Intel as of April 27, 2020 (Baseline) and March 23, 2021 (New). Baseline configuration: 1-node, 1x Intel® Xeon® Platinum 8280L processor (28 cores at 2.7 GHz) on Neon City with a single Intel® Optane™ PMem module configuration (6 x 32 GB DRAM; 1 x {128 GB, 256 GB, 512 GB} Intel® Optane™ PMem module), ucode rev: 04002F00 running Fedora 29 kernel 5.1.18-200.fc29.x86_64 and Intel Memory Latency Checker (Intel MLC) version 3.8 with App Direct Mode. New Configuration: 1-node, 1x pre-production 3rd Gen Intel® Xeon® Scalable processor (38 cores at 2.0 GHz) on Wilson City with a single Intel® Optane™ PMem module configuration (8 x 32 GB DRAM; 1 x {128 GB, 256 GB, 512 GB} Intel® Optane™ PMem module), ucode rev: 8d000270 running RHEL 8.1 kernel 4.18.0-147.el8.x86_64 and Intel MLC version 3.9 with App Direct Mode.



Мы ищем разработчиков:

Experienced:

- DevOps Engineer (Integration, Computer Vision)
- Software Validation Engineer (DevOps, Computer Vision)
- Infrastructure and DevOps Engineer
- Cloud SW Development Engineer (Computer Vision)

Interns (для студентов очной формы):

- DevOps intern (Parallel Runtimes Engineering team)
- Infrastructure and DevOps Intern