THREE STEPS TO ANALYTICS-READINESS

Gaining insights from analytics requires careful preparation and alignment of organization, data and processes across the business.
When building an analytics strategy, ensure you address three key steps:

STEP 1
BUILD AN ORGANIZATIONAL FOUNDATION FOR ANALYTICS

Work with stakeholders at all levels to ensure universal commitment, shared vision, and clarity on roles and responsibilities.

Collaborate to identify the analytics use cases of highest value to the business.

Plan expected outputs into existing workflows so they can be easily shared and used.

BUSINESS
Valid business case and KPIs

ORGANIZATIONAL FOUNDATION FOR ANALYTICS

PEOPLE
Business-wide stakeholders with clear roles and responsibilities

PROCESS
Workflows and methodologies to share and use insights

STEP 2
MAP THE DATA PIPELINE

INGEST
Extract data, detecting what has changed, and move it into a system where it can be stored and analyzed.

PREPARE
Data comes in different structures, sizes, and speeds. Storage should align with cost, access, volume and performance requirements.

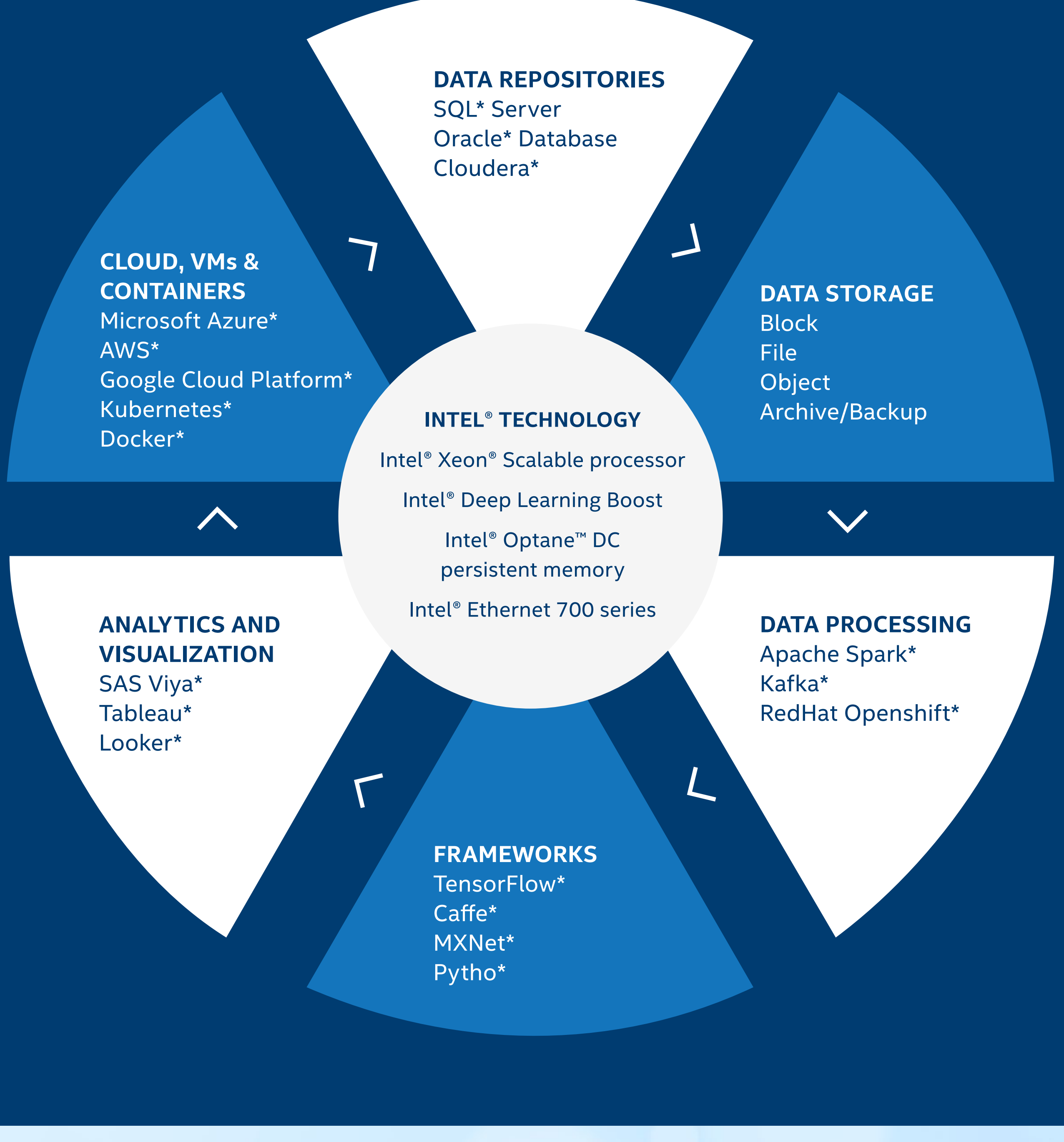
ANALYZE
Data is cleansed, normalized, processed and stored in analytical systems to allow for ad-hoc querying and exploration.

ACT
Perform in-depth exploration and visualization to pull actionable insights from the data.

Ingest			Prepare		Analyze		Act		
Data creation	Data transmission	Data ingestion	Data integration	Data staging	Data clean, normalize	Models and experiment	Tuning and distillation	Visualize and deploy	Archival

STEP 3
PRODUCTIONALIZE ANALYTICS

Make sure you have the tools in place to support analytics workloads **today and in the future.**



The latest Intel® technology is optimized to deliver breakthrough analytics performance. **Intel® Optane™ DC persistent memory** is delivered with **2nd generation Intel® Xeon® Scalable processors.** Compared to previous, non-persistent memory technologies, it:

DELIVERS UP TO 8X PERFORMANCE IMPROVEMENT FOR QUERIES[1]

The new 2nd generation Intel® Xeon® Scalable processors with Intel® Deep Learning Boost accelerates AI inference up to **30x[2]**

Learn more about preparing your organization to meet its analytics ambitions by reading our eGuide *From Data to Insights*

Read the eGuide

Software and workloads used in performance tests may have been optimized for performance only on Intel microprocessors.

Performance tests, such as SYSmark and MobileMark, are measured using specific computer systems, components, software, operations and functions. Any change to any of those factors may cause the results to vary. You should consult other information and performance tests to assist you in fully evaluating your contemplated purchases, including the performance of that product when combined with other products. For more complete information, visit http://www.intel.com/benchmarks.

[1] 8x improvement in queries result based on testing by Intel on 1 November 2018. Baseline configuration: Platform: S2600WF (Wolf Pass); number of nodes: 1; number of sockets: 2; CPU: Intel® Xeon® Platinum 8280L CPU @ 2.70 GHz; Cores/socket, threads/socket: 28 cores/socket, 2 threads/socket; ucode: (microcode: 0x400000a); HT: Enabled; Turbo: Off; BIOS version: SE5C620.86B.0D.01.0134.100420181737; BKC version: WW06'19; FW version: N/A; System DDR Mem Config slots/cap/run-speed: DDR Mem: 24 / 32GB / 2666 MT/s; System Intel Optane DC persistent memory Config: slots / cap / run-speed: N/A; Total Memory/Node (DDR, Intel DC Optane perst. mem.): 768GB DDR; Storage – boot: SATA SSD 500GB; Storage - application drives: HDD (ST1000NX0313) * 8; NIC: 10-Gigabit SFI/SFP+ Network Connection; Software: Spark Hadoop; OS: Fedora release 29 (Twenty Nine); Kernal: Linux-4.18.8-100.fc27.x86_64-x86_64-with-fedora-27-Twenty_Seven BIOS: SE5C620.86B .0D.01.0299.122420180146; Mitigation log attached: 1,2,3,3a,4, L1TF; Intel Optane DC persistent memory mode: N/A; Run Method: Run 9 I/O intensive queries together in a few iterations; Iterations and result choice: 3 iterations and choose the execution time for second or third iteration; Dataset size: 2TB/3TB; Workload & version: Decision Support I/O intensive queries; Compiler: gcc (GCC) 8.3.1 20190223 (Red Hat 8.3.1-2), JDK 1.8.0_201; Libraries: Memkind; Other software: Spark + Hadoop. New configuration: Platform: S2600WF (Wolf Pass); number of nodes: 1; number of sockets: 2; CPU: Intel® Xeon® Platinum 8280L CPU @ 2.70 GHz; Cores/socket, threads/socket: 28 cores/socket, 2 threads/socket; ucode: (microcode: 0x400000a); HT: Enabled; Turbo: Off; BIOS version: SE5C620.86B.0D.01.0134.100420181737; BKC version: WW06'19; FW version: N/A; System DDR Mem Config slots/cap/run-speed: DDR Mem: 12 / 16GB / 2666 MT/s; System Intel Optane DC persistent memory Config: slots / cap / run-speed: 8 / 128GB / 2666 MT/s; Total Memory/Node (DDR, Intel DC Optane perst. mem.): 192GB DDR + 1TB DCPMM; Storage – boot: SATA SSD 500GB; Storage - application drives: HDD (ST1000NX0313) * 8; NIC: 10-Gigabit SFI/SFP+ Network Connection; Software: Spark Hadoop; OS: Fedora release 29 (Twenty Nine); Kernal: Linux-4.18.8-100.fc27.x86_64-x86_64-with-fedora-27-Twenty_Seven BIOS: SE5C620.86B.0D.01.0299.122420180146; Mitigation log attached: 1,2,3,3a,4, L1TF; Intel Optane DC persistent memory mode: App Direct; Run Method: Run 9 I/O intensive queries together in a few iterations; Iterations and result choice: 3 iterations and choose the execution time for second or third iteration; Dataset size: 2TB/3TB; Workload & version: Decision Support I/O intensive queries; Compiler: gcc (GCC) 8.3.1 20190223 (Red Hat 8.3.1-2), JDK 1.8.0_201; Libraries: Memkind; Other software: Spark + Hadoop.

[2] 30x inference throughput improvement on Intel® Xeon® Platinum 9282 processor with Intel® DL Boost: Tested by Intel as of 2/26/2019. Platform: Dragon rock 2 socket Intel® Xeon® Platinum 9282 (56 cores per socket), HT ON, turbo ON, Total Memory 768 Gb (24 slots/ 32 Gb/ 2933 MHz), BIOS:SE5C620.86B.0D.01.0241.112020180249, CentOS 7 Kernel 3.10.0-957.5.1.el7.x86_64, Deep Learning Framework: Intel® Optimization for Caffe* version: https://github.com/intel/caffe d554cbf1, ICC 2019.2.187, MKL DNN version: v0.17 (commit hash: 830a10059a018cd2634d94195140cf2d8790a75a), model: https://github.com/intel/caffe/blob/master/models/intel_optimized_models/int8/resnet50_int8_full_conv.prototxt, BS=64, No datalayer syntheticData:3x224x224, 56 instance/2 socket, Datatype: INT8 vs. Tested by Intel as of July 11, 2017: 2S Intel® Xeon® Platinum 8180 CPU @ 2.50 GHz (28 cores), HT disabled, turbo disabled, scaling governor set to "performance" via intel_pstate driver, 384 Gb DDR4-2666 ECC RAM. CentOS Linux* release 7.3.1611 (Core), Linux kernel 3.10.0-514.10.2.el7.x86_64. SSD: Intel® SSD DC S3700 Series (800 Gb, 2.5in SATA 6 Gb/s, 25nm, MLC). Performance measured with: Environment variables: KMP_AFFINITY='granularity=fine, compact', OMP_NUM_THREADS=56, CPU Freq set with cpupower frequency-set -d 2.5G -u 3.8G -g performance. Caffe: (http://github.com/intel/caffe/), revision f96b759f71b2281835f690af267158b82b150b5c. Inference measured with "caffe time --forward_only" command, training measured with "caffe time" command. For "ConvNet" topologies, synthetic data set was used. For other topologies, data was stored on local storage and cached in memory before training. Topology specs from https://github.com/intel/caffe/tree/master/models/intel_optimized_models (ResNet-50). Intel C++ compiler ver. 17.0.2 20170213, Intel® MKL small libraries version 2018.0.20170425. Caffe run with "numactl -l".

Cost reduction scenarios described are intended as examples of how a given Intel- based product, in the specified circumstances and configurations, may affect future costs and provide cost savings. Circumstances will vary. Intel does not guarantee any costs or cost reduction.

Intel's compilers may or may not optimize to the same degree for non-Intel microprocessors for optimizations that are not unique to Intel microprocessors. These optimizations include SSE2, SSE3, and SSSE3 instruction sets and other optimizations. Intel does not guarantee the availability, functionality, or effectiveness of any optimization on microprocessors not manufactured by Intel. Microprocessor-dependent optimizations in this product are intended for use with Intel microprocessors. Certain optimizations not specific to Intel microarchitecture are reserved for Intel microprocessors. Please refer to the applicable product User and Reference Guides for more information regarding the specific instruction sets covered by this notice.

Intel does not control or audit third-party data. You should review this content, consult other sources, and confirm whether referenced data are accurate.

0719/KR/CAT/PDF

340925-001EN