

# Customer “Pain Points”

## 1 **The Downside of Becoming a Data Company**

Companies that have pivoted to AI and analytics in order to monetize their data are experiencing accelerating data growth rates

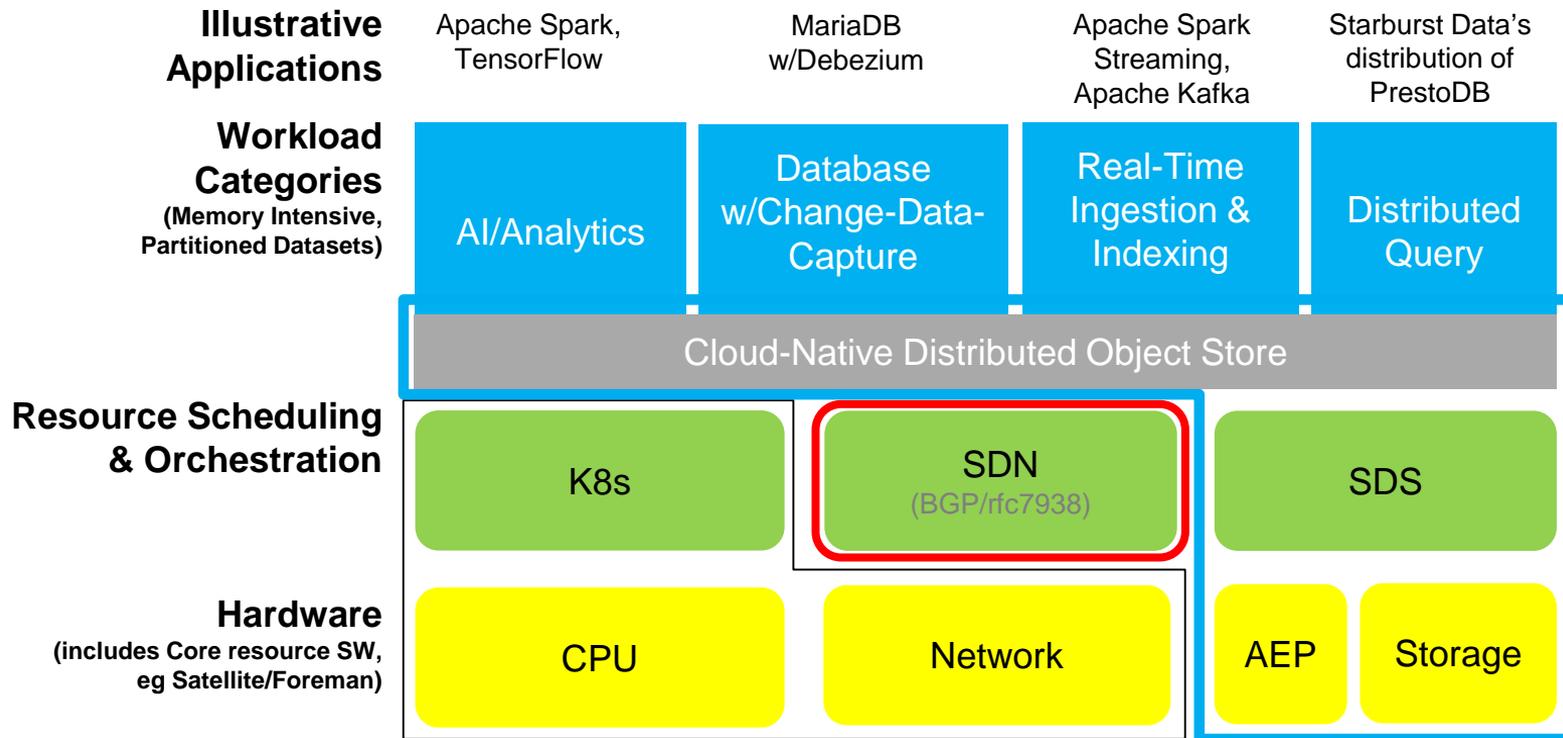
## 2 **Disaggregating Storage**

This data growth challenge makes disaggregating storage from compute attractive because the company can scale their storage capacity to match their data growth, independent of compute.

## 3 **Cloud-Like, Open-Source, and AI/Analytics-Ready**

Many of these companies are using Hadoop\* and HDFS\*, but still find it difficult to share data. They expressed a desire to move to something more cloud-like that readily supports their AI/analytics investments.

# DMP Workload Overview



# Analytics on Intel's Data Management Platform

Accommodating Data Growth via Disaggregated Storage while Keeping Compute Close to the Data

## Workload Description

Companies are aggressively adopting cloud-like data management techniques to insure their data scientists and analysts are as productive as possible. This centers around insuring their data pipelines have access to the requisite data during execution.

## Solution: an 'on-prem' equivalent to AWS's s3+Elastic MapReduce (EMR) offerings

Hardware-based appliance integrated with Kubernetes platform. These host a Spark/Tensorflow offering collocated with S3-like object storage.

## Customer Value

- Accommodate Data Growth
- Pipelines Execute Close to Data
- Space Efficiency

## Solution Overview

The solution exploits Intel's 3D Xpoint™ technology on the cluster's compute servers and Intel's QLC-based, 3D NAND on the storage servers. The combined performance/capacity advantage of these products are exposed via the combination of:

- A cluster-wide volume manager with support for NVMe-over-Fabric to enable physical Storage Disaggregation.
- An s3-compatible storage service that supports the requirements of AI/Analytics.
- Use of the S3A Hadoop client, at least in the case of Spark.

