

Hard Disk Drive SMART Data Analysis using Apache Spark

ECE 590 – Big Data Technologies-Spring 2021

Team - 18

Shamili M Tetali
George Mason University
Fairfax, Virginia, 22030
stetali@gmu.edu

Matiyas Maru
George Mason University
Fairfax, Virginia, 22030
Mmaru@gmu.edu

Muhammed Hassan
George Mason University
Fairfax, Virginia, 22030
Mhassa9@gmu.edu

Abstract

The enormous growth of data acquisition use cases in recent years has indeed demanded more storage resources, primarily hard disk drives (HDD) where data resides. Based on the organization's IT design choices these physical storage components could either be housed at Cloud storage or on-premises datacenters. Data availability from any of these storage implementations heavily depends on reliability characteristics of the underlying infrastructure. Apparently, of all the hardware components, hard disks experience high failure rates and in several cases its quite uncertain to accurately predict failures ahead of the impact, thereby losing service credibility and negatively influences financial prospects.

The scope of this project is primarily based on selected research papers[2][3][4] and data collected from Backblaze datasets. In this paper, we will leverage Big Data analytic concepts and characterize data, based on Hard Disks inherent feature called Self-Monitoring Analysis and Reporting Technology (SMART) attributes. Essentially, for the purpose of this study we obtained historical datasets from Backblaze[1] and considered SMART parameters that indicate abnormal hard disk behavior. Data analysis is performed by using Apache Spark implementation and finally established an analytical framework that will help to understand and determine co-relation between parameters (like temperature, capacity, manufacturer etc).

Introduction

As demand for IT Infrastructure intensifies, deployments surged at an unprecedented rate and massively expanding at global scale. Reliability and availability being important factors, service providers and engineers focuses their attention to maintaining the ecosystem and aims to delivering uninterrupted service. Therefore, fault prevention requires advanced data analysis methods to early detection of failures

analysis. Proactive detection and identification of abnormalities is the key to meeting the standards. Disks are among the most frequently failed components and observing abnormalities to predicting the impending failure of hard disks in the field can help systems at large organizations, datacenters and other crucial places to take corrective actions before the failure to avoid loss of data and performance degradation.

Data reliability, availability and low ownership costs are basic expectations from data users and it's essential for engineers to understanding the nature of failures and come up with predictive analysis mechanisms. With major advancements in machine learning, data mining etc, it has become possible to devise data analytic methods to proactively determine disk failure rates and preemptively solve large scale manufacturing concerns, thereby reducing the TCO and costs associated to HDD manufacturing and return merchandize processes.

The primary goal of this study is to implement Big Data analytics on HDD Smart attribute datasets and to understanding the relationship between various parameters that could negatively impact disk drive performance and premature failures. A significant size of 52 million records are considered towards the scope of this project and utilize Apache Spark as the analytics engine. The study will also demonstrate Big Data analytic Framework capabilities for various reporting and how we could derive value added solutions for Hard Disk Drive predictive failures.

Data Flow Architecture

Fig(1) shows the schematic diagram of our analysis using Apache Spark. The data is collected from Backblaze, stored in our Local Machine. These are further analyzed using Apache Zeppelin notebook which triggers the spark jobs computation and visualize the results.

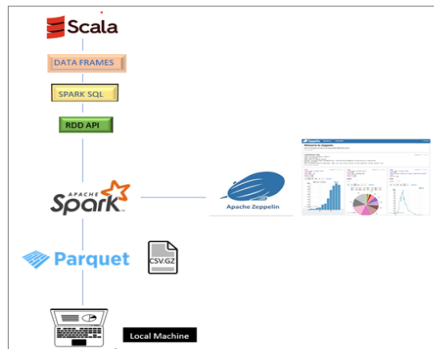


Fig (1) : Data Flow Architecture

For Analysis visualization we used open source software as listed below.

1. Apache Spark computation engine: version-2.4.7
2. Apache Zeppelin web notebook for querying and visualization: version-0.8.2
3. SCALA programming language : version-2.11.12

Implementation

Data Collection:

The source of Data is collected from Backblaze which is primarily a Cloud Storage and Backup company[link1]. hard drive dataset. For the purpose of this project, we reviewed data files for year 2020 (Q1-Q4) that consisted 52286398 records (**approx. 52 million**)

Further breakdown of Data files specifics:

Total Files Size	16.37GB
Files Count	370 files
Drive Count	162,299
Drive Failures	1,302
Drive Days	51.2M
Drive Capacity Range	240GB – 18TB

The first row of each file contains the column names, the remaining rows are the actual data. The columns are as follows:

Date – The date of the file in yyyy-mm-dd format.
Serial Number – Mfg. assigned serial number of the drive.
Model – Mfg. assigned model number of the drive.
Capacity – The drive capacity in bytes.

Failure – Contains a “0” if the drive is OK. Contains a “1” if this is the last day the drive was operational before failing.

The remaining columns are Smart Attributes associated to Normalized and RAW values that ranges from 1 – 255 and each value signify a drive operating characteristic as reported by the drive built in smart function.

Hard Disk SMART Data Analysis:

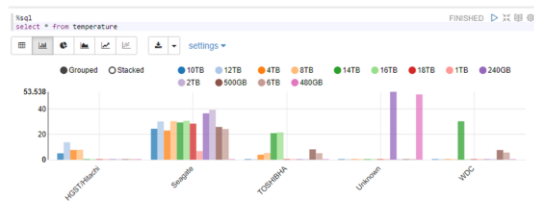
The data consists of daily snapshots of the SMART statistics and a failure label for all operational hard drives in a data center in 2020. SMART stats are meant to be indicators of drive reliability and should, in theory, provide good input features to a predictive model of drive failure. The first step is to get all data in csv format arrangement and combine each month data and further archived those files for efficiency purposes. Data is converted to parquet format which allows compression schemes to be specified on a per-column basis. The data is then read via Spark and preview of most used column data is shown below in table[1].

Datafiles in consists of both “normalized” and “raw” columns for each of the SMART attribute. “Normalized” columns are ignored for the purpose of this analysis as standardized RAW data would suffice the need and provide detailed information required to determining drive statistics. Manufacturer-specific normalizations would be applicable to partial datasets due to transformations and therefore creates discrepancies. Hence, for the scope of this analysis we considered denormalized Raw values to establishing a model across all manufacturers drives.

		model	capacity	failure	smart_1_raw	smart_1_normalized	smart_2_raw	smart_2_normalized	smart_3_raw	smart_3_normalized	smart_4_raw	smart_4_normalized	smart_5_raw	smart_5_normalized	smart_6_raw	smart_6_normalized	smart_7_raw	smart_7_normalized	smart_8_raw	smart_8_normalized	smart_9_raw	smart_9_normalized	smart_10_raw	smart_10_normalized	smart_11_raw	smart_11_normalized	smart_12_raw	smart_12_normalized	smart_13_raw	smart_13_normalized	smart_14_raw	smart_14_normalized	smart_15_raw	smart_15_normalized	smart_16_raw	smart_16_normalized	smart_17_raw	smart_17_normalized	smart_18_raw	smart_18_normalized	smart_19_raw	smart_19_normalized	smart_20_raw	smart_20_normalized	smart_21_raw	smart_21_normalized	smart_22_raw	smart_22_normalized	smart_23_raw	smart_23_normalized	smart_24_raw	smart_24_normalized	smart_25_raw	smart_25_normalized	smart_26_raw	smart_26_normalized	smart_27_raw	smart_27_normalized	smart_28_raw	smart_28_normalized	smart_29_raw	smart_29_normalized	smart_30_raw	smart_30_normalized	smart_31_raw	smart_31_normalized	smart_32_raw	smart_32_normalized	smart_33_raw	smart_33_normalized	smart_34_raw	smart_34_normalized	smart_35_raw	smart_35_normalized	smart_36_raw	smart_36_normalized	smart_37_raw	smart_37_normalized	smart_38_raw	smart_38_normalized	smart_39_raw	smart_39_normalized	smart_40_raw	smart_40_normalized	smart_41_raw	smart_41_normalized	smart_42_raw	smart_42_normalized	smart_43_raw	smart_43_normalized	smart_44_raw	smart_44_normalized	smart_45_raw	smart_45_normalized	smart_46_raw	smart_46_normalized	smart_47_raw	smart_47_normalized	smart_48_raw	smart_48_normalized	smart_49_raw	smart_49_normalized	smart_50_raw	smart_50_normalized	smart_51_raw	smart_51_normalized	smart_52_raw	smart_52_normalized	smart_53_raw	smart_53_normalized	smart_54_raw	smart_54_normalized	smart_55_raw	smart_55_normalized	smart_56_raw	smart_56_normalized	smart_57_raw	smart_57_normalized	smart_58_raw	smart_58_normalized	smart_59_raw	smart_59_normalized	smart_60_raw	smart_60_normalized	smart_61_raw	smart_61_normalized	smart_62_raw	smart_62_normalized	smart_63_raw	smart_63_normalized	smart_64_raw	smart_64_normalized	smart_65_raw	smart_65_normalized	smart_66_raw	smart_66_normalized	smart_67_raw	smart_67_normalized	smart_68_raw	smart_68_normalized	smart_69_raw	smart_69_normalized	smart_70_raw	smart_70_normalized	smart_71_raw	smart_71_normalized	smart_72_raw	smart_72_normalized	smart_73_raw	smart_73_normalized	smart_74_raw	smart_74_normalized	smart_75_raw	smart_75_normalized	smart_76_raw	smart_76_normalized	smart_77_raw	smart_77_normalized	smart_78_raw	smart_78_normalized	smart_79_raw	smart_79_normalized	smart_80_raw	smart_80_normalized	smart_81_raw	smart_81_normalized	smart_82_raw	smart_82_normalized	smart_83_raw	smart_83_normalized	smart_84_raw	smart_84_normalized	smart_85_raw	smart_85_normalized	smart_86_raw	smart_86_normalized	smart_87_raw	smart_87_normalized	smart_88_raw	smart_88_normalized	smart_89_raw	smart_89_normalized	smart_90_raw	smart_90_normalized	smart_91_raw	smart_91_normalized	smart_92_raw	smart_92_normalized	smart_93_raw	smart_93_normalized	smart_94_raw	smart_94_normalized	smart_95_raw	smart_95_normalized	smart_96_raw	smart_96_normalized	smart_97_raw	smart_97_normalized	smart_98_raw	smart_98_normalized	smart_99_raw	smart_99_normalized	smart_100_raw	smart_100_normalized	smart_101_raw	smart_101_normalized	smart_102_raw	smart_102_normalized	smart_103_raw	smart_103_normalized	smart_104_raw	smart_104_normalized	smart_105_raw	smart_105_normalized	smart_106_raw	smart_106_normalized	smart_107_raw	smart_107_normalized	smart_108_raw	smart_108_normalized	smart_109_raw	smart_109_normalized	smart_110_raw	smart_110_normalized	smart_111_raw	smart_111_normalized	smart_112_raw	smart_112_normalized	smart_113_raw	smart_113_normalized	smart_114_raw	smart_114_normalized	smart_115_raw	smart_115_normalized	smart_116_raw	smart_116_normalized	smart_117_raw	smart_117_normalized	smart_118_raw	smart_118_normalized	smart_119_raw	smart_119_normalized	smart_120_raw	smart_120_normalized	smart_121_raw	smart_121_normalized	smart_122_raw	smart_122_normalized	smart_123_raw	smart_123_normalized	smart_124_raw	smart_124_normalized	smart_125_raw	smart_125_normalized	smart_126_raw	smart_126_normalized	smart_127_raw	smart_127_normalized	smart_128_raw	smart_128_normalized	smart_129_raw	smart_129_normalized	smart_130_raw	smart_130_normalized	smart_131_raw	smart_131_normalized	smart_132_raw	smart_132_normalized	smart_133_raw	smart_133_normalized	smart_134_raw	smart_134_normalized	smart_135_raw	smart_135_normalized	smart_136_raw	smart_136_normalized	smart_137_raw	smart_137_normalized	smart_138_raw	smart_138_normalized	smart_139_raw	smart_139_normalized	smart_140_raw	smart_140_normalized	smart_141_raw	smart_141_normalized	smart_142_raw	smart_142_normalized	smart_143_raw	smart_143_normalized	smart_144_raw	smart_144_normalized	smart_145_raw	smart_145_normalized	smart_146_raw	smart_146_normalized	smart_147_raw	smart_147_normalized	smart_148_raw	smart_148_normalized	smart_149_raw	smart_149_normalized	smart_150_raw	smart_150_normalized	smart_151_raw	smart_151_normalized	smart_152_raw	smart_152_normalized	smart_153_raw	smart_153_normalized	smart_154_raw	smart_154_normalized	smart_155_raw	smart_155_normalized	smart_156_raw	smart_156_normalized	smart_157_raw	smart_157_normalized	smart_158_raw	smart_158_normalized	smart_159_raw	smart_159_normalized	smart_160_raw	smart_160_normalized	smart_161_raw	smart_161_normalized	smart_162_raw	smart_162_normalized	smart_163_raw	smart_163_normalized	smart_164_raw	smart_164_normalized	smart_165_raw	smart_165_normalized	smart_166_raw	smart_166_normalized	smart_167_raw	smart_167_normalized	smart_168_raw	smart_168_normalized	smart_169_raw	smart_169_normalized	smart_170_raw	smart_170_normalized	smart_171_raw	smart_171_normalized	smart_172_raw	smart_172_normalized	smart_173_raw	smart_173_normalized	smart_174_raw	smart_174_normalized	smart_175_raw	smart_175_normalized	smart_176_raw	smart_176_normalized	smart_177_raw	smart_177_normalized	smart_178_raw	smart_178_normalized	smart_179_raw	smart_179_normalized	smart_180_raw	smart_180_normalized	smart_181_raw	smart_181_normalized	smart_182_raw	smart_182_normalized	smart_183_raw	smart_183_normalized	smart_184_raw	smart_184_normalized	smart_185_raw	smart_185_normalized	smart_186_raw	smart_186_normalized	smart_187_raw	smart_187_normalized	smart_188_raw	smart_188_normalized	smart_189_raw	smart_189_normalized	smart_190_raw	smart_190_normalized	smart_191_raw	smart_191_normalized	smart_192_raw	smart_192_normalized	smart_193_raw	smart_193_normalized	smart_194_raw	smart_194_normalized	smart_195_raw	smart_195_normalized	smart_196_raw	smart_196_normalized	smart_197_raw	smart_197_normalized	smart_198_raw	smart_198_normalized	smart_199_raw	smart_199_normalized	smart_200_raw	smart_200_normalized	smart_201_raw	smart_201_normalized	smart_202_raw	smart_202_normalized	smart_203_raw	smart_203_normalized	smart_204_raw	smart_204_normalized	smart_205_raw	smart_205_normalized	smart_206_raw	smart_206_normalized	smart_207_raw	smart_207_normalized	smart_208_raw	smart_208_normalized	smart_209_raw	smart_209_normalized	smart_210_raw	smart_210_normalized	smart_211_raw	smart_211_normalized	smart_212_raw	smart_212_normalized	smart_213_raw	smart_213_normalized	smart_214_raw	smart_214_normalized	smart_215_raw	smart_215_normalized	smart_216_raw	smart_216_normalized	smart_217_raw	smart_217_normalized	smart_218_raw	smart_218_normalized	smart_219_raw	smart_219_normalized	smart_220_raw	smart_220_normalized	smart_221_raw	smart_221_normalized	smart_222_raw	smart_222_normalized	smart_223_raw	smart_223_normalized	smart_224_raw	smart_224_normalized	smart_225_raw	smart_225_normalized	smart_226_raw	smart_226_normalized	smart_227_raw	smart_227_normalized	smart_228_raw	smart_228_normalized	smart_229_raw	smart_229_normalized	smart_230_raw	smart_230_normalized	smart_231_raw	smart_231_normalized	smart_232_raw	smart_232_normalized	smart_233_raw	smart_233_normalized	smart_234_raw	smart_234_normalized	smart_235_raw	smart_235_normalized	smart_236_raw	smart_236_normalized	smart_237_raw	smart_237_normalized	smart_238_raw	smart_238_normalized	smart_239_raw	smart_239_normalized	smart_240_raw	smart_240_normalized	smart_241_raw	smart_241_normalized	smart_242_raw	smart_242_normalized	smart_243_raw	smart_243_normalized	smart_244_raw	smart_244_normalized	smart_245_raw	smart_245_normalized	smart_246_raw	smart_246_normalized	smart_247_raw	smart_247_normalized	smart_248_raw	smart_248_normalized	smart_249_raw	smart_249_normalized	smart_250_raw	smart_250_normalized	smart_251_raw	smart_251_normalized	smart_252_raw	smart_252_normalized	smart_253_raw	smart_253_normalized	smart_254_raw	smart_254_normalized	smart_255_raw	smart_255_normalized	smart_256_raw	smart_256_normalized	smart_257_raw	smart_257_normalized	smart_258_raw	smart_258_normalized	smart_259_raw	smart_259_normalized	smart_260_raw	smart_260_normalized	smart_261_raw	smart_261_normalized	smart_262_raw	smart_262_normalized	smart_263_raw	smart_263_normalized	smart_264_raw	smart_264_normalized	smart_265_raw	smart_265_normalized	smart_266_raw	smart_266_normalized	smart_267_raw	smart_267_normalized	smart_268_raw	smart_268_normalized	smart_269_raw	smart_269_normalized	smart_270_raw	smart_270_normalized	smart_271_raw	smart_271_normalized	smart_272_raw	smart_272_normalized	smart_273_raw	smart_273_normalized	smart_274_raw	smart_274_normalized	smart_275_raw	smart_275_normalized	smart_276_raw	smart_276_normalized	smart_277_raw	smart_277_normalized	smart_278_raw	smart_278_normalized	smart_279_raw	smart_279_normalized	smart_280_raw	smart_280_normalized	smart_281_raw	smart_281_normalized	smart_282_raw	smart_282_normalized	smart_283_raw	smart_283_normalized	smart_284_raw	smart_284_normalized	smart_285_raw	smart_285_normalized	smart_286_raw	smart_286_normalized	smart_287_raw	smart_287_normalized	smart_288_raw	smart_288_normalized	smart_289_raw	smart_289_normalized	smart_290_raw	smart_290_normalized	smart_291_raw	smart_291_normalized	smart_292_raw	smart_292_normalized	smart_293_raw	smart_293_normalized	smart_294_raw	smart_294_normalized	smart_295_raw	smart_295_normalized	smart_296_raw	smart_296_normalized	smart_297_raw	smart_297_normalized	smart_298_raw	smart_298_normalized	smart_299_raw	smart_299_normalized	smart_300_raw	smart_300_normalized	smart_301_raw	smart_301_normalized	smart_302_raw	smart_302_normalized	smart_303_raw	smart_303_normalized	smart_304_raw	smart_304_normalized	smart_305_raw	smart_305_normalized	smart_306_raw	smart_306_normalized	smart_307_raw	smart_307_normalized	smart_308_raw	smart_308_normalized	smart_309_raw	smart_309_normalized	smart_310_raw	smart_310_normalized	smart_311_raw	smart_311_normalized	smart_312_raw	smart_312_normalized	smart_313_raw	smart_313_normalized	smart_314_raw	smart_314_normalized	smart_315_raw	smart_315_normalized	smart_316_raw	smart_316_normalized	smart_317_raw	smart_317_normalized	smart_318_raw	smart_318_normalized	smart_319_raw	smart_319_normalized	smart_320_raw	smart_320_normalized	smart_321_raw	smart_321_normalized	smart_322_raw	smart_322_normalized	smart_323_raw	smart_323_normalized	smart_324_raw	smart_324_normalized	smart_325_raw	smart_325_normalized	smart_326_raw	smart_326_normalized	smart_327_raw	smart_327_normalized	smart_328_raw	smart_328_normalized	smart_329_raw	smart_329_normalized	smart_330_raw	smart_330_normalized	smart_331_raw	smart_331_normalized	smart_332_raw	smart_332_normalized	smart_333_raw	smart_333_normalized	smart_334_raw	smart_334_normalized	smart_335_raw	smart_335_normalized	smart_336_raw	smart_336_normalized	smart_337_raw	smart_337_normalized	smart_338_raw	smart_338_normalized	smart_339_raw	smart_339_normalized	smart_340_raw	smart_340_normalized	smart_341_raw	smart_341_normalized	smart_342_raw	smart_342_normalized	smart_343_raw	smart_343_normalized	smart_344_raw	smart_344_normalized	smart_345_raw	smart_345_normalized	smart_346_raw	smart_346_normalized	smart_347_raw	smart_347_normalized	smart_348_raw	smart_348_normalized	smart_349_raw	smart_349_normalized	smart_350_raw	smart_350_normalized	smart_351_raw	smart_351_normalized	smart_352_raw	smart_352_normalized	smart_353_raw	smart_353_normalized	smart_354_raw	smart_354_normalized	smart_355_raw	smart_355_normalized	smart_356_raw	smart_356_normalized	smart_357_raw	smart_357_normalized	smart_358_raw	smart_358_normalized	smart_359_raw	smart_359_normalized	smart_360_raw	smart_360_normalized	smart_361_raw	smart_361_normalized	smart_362_raw	smart_362_normalized	smart_363_raw	smart_363_normalized	smart_364_raw	smart_364_normalized	smart_365_raw	smart_365_normalized	smart_366_raw	smart_366_normalized	smart_367_raw	smart_367_normalized	smart_368_raw	smart_368_normalized	smart_369_raw	smart_369_normalized	smart_370_raw	smart_370_normalized	smart_371_raw	smart_371_normalized	smart_372_raw	smart_372_normalized	smart_373_raw	smart_373_normalized	smart_374_raw	smart_374_normalized	smart_375_raw	smart_375_normalized	smart_376_raw	smart_376_normalized	smart_377_raw	smart_377_normalized	smart_378_raw	smart_378_normalized	smart_379_raw	smart_379_normalized	smart_380_raw	smart_380_normalized	smart_381_raw	smart_381_normalized	smart_382_raw	smart_382_normalized	smart_383_raw	smart_383_normalized	smart_384_raw	smart_384_normalized	smart_385_raw	smart_385_normalized	smart_386_raw	smart_386_normalized	smart_387_raw	smart_387_normalized	smart_388_raw	smart_388_normalized	smart_389_raw	smart_389_normalized	smart_390_raw	smart_390_normalized	smart_391_raw	smart_391_normalized	smart_392_raw	smart_392_normalized	smart_393_raw	smart_393_normalized	smart_394_raw	smart_394_normalized	smart_395_raw	smart_395_normalized	smart_396_raw	smart_396_normalized	smart_397_raw	smart_397_normalized	smart_398_raw	smart_398_normalized	smart_399_raw	smart_399_normalized	smart_400_raw	smart_400_normalized	smart_401_raw	smart_401_normalized	smart_402_raw	smart_402_normalized	smart_403_raw	smart_403_normalized	smart_404_raw	smart_404_normalized	smart_405_raw	smart_405_normalized	smart_406_raw	smart_406_normalized	smart_407_raw	smart_407_normalized	smart_408_raw	smart_408_normalized	smart_409_raw	smart_409_normalized	smart_410_raw	smart_410_normalized	smart_411_raw	smart_411_normalized	smart_412_raw	smart_412_normalized	smart_413_raw	smart_413_normalized	smart_414_raw	smart_414_normalized	smart_415_raw	smart_415_normalized	smart_416_raw	smart_416_normalized	smart_417_raw	smart_417_normalized	smart_418_raw	smart_418_normalized	smart_419_raw	smart_419_normalized	smart_420_raw	smart_420_normalized	smart_421_raw	smart_421_normalized	smart_422_raw	smart_422_normalized	smart_423_raw	smart_423_normalized	smart_424_raw	smart_424_normalized	smart_425_raw	smart_425_normalized	smart_426_raw	smart_426_normalized	smart_427_raw	smart_427_normalized	smart_428_raw	smart_428_normalized	smart_429_raw	smart_429_normalized	smart_430_raw	smart_430_normalized	smart_431_raw	smart_431_normalized	smart_432_raw	smart_432_normalized	smart_433_raw	smart_433_normalized	smart_434_raw	smart_434_normalized	smart_435_raw	smart_435_normalized	smart_436_raw	smart_436_normalized	smart_437_raw	smart_437_normalized	smart_438_raw	smart_438_normalized	smart_439_raw	smart_439_normalized	smart_440_raw	smart_440_normalized	smart_441_raw	smart_441_normalized	smart_442_raw	smart_442_normalized	smart_443_raw	smart_443_normalized	smart_444_raw	smart_444_normalized	smart_445_raw	smart_445_normalized	smart_446_raw	smart_446_normalized	smart_447_raw	smart_447_normalized	smart_448_raw	smart_448_normalized	smart_449_raw	smart_449_normalized	smart_450_raw	smart_450_normalized	smart_451_raw	smart_451_normalized	smart_452_raw	smart_452_normalized	smart_453_raw	smart_453_normalized	smart_454_raw	smart_454_normalized	smart_455_raw	smart_455_normalized	smart_456_raw	smart_456_normalized	smart_457_raw	smart_457_normalized	smart_458_raw	smart_458_normalized	smart_459_raw	smart_459_normalized	smart_460_raw	smart_460_normalized	smart_461_raw	smart_461_normalized	smart_462_raw	smart_462_normalized	smart_463_raw	smart_463_normalized	smart_464_raw	smart_464_normalized	smart_465_raw	smart_465_normalized	smart_466_raw	smart_466_normalized	smart_467_raw	smart_467_normalized	smart_468_raw	smart_468_normalized	smart_469_raw	smart_469_normalized	smart_470_raw	smart_470_normalized	smart_471_raw	smart_471_normalized	smart_472_raw	smart_472_normalized	smart_473_raw	smart_473_normalized	smart_474_raw	smart_474_normalized	smart_475_raw	smart_475_normalized	smart_476_raw	smart_476_normalized	smart_477_raw	smart_477_normalized	smart_478_raw	smart_478_normalized	smart_479_raw	smart_479_normalized	smart_480_raw	smart_480_normalized	smart_481_raw	smart_481_normalized	smart_482_raw	smart_482_normalized	smart_483_raw	smart_483_normalized	smart_484_raw	smart_484_normalized	smart_485_raw	smart_485_normalized	smart_486_raw	smart_486_normalized	smart_487_raw	smart_487_normalized	smart_488_raw	smart_488_normalized	smart_489_raw	smart_489_normalized	smart_490_raw	smart_490_normalized	smart_491_raw	smart_491_normalized	smart_492_raw	smart_492_normalized	smart_493_raw	smart_493_normalized	smart_494_raw	smart_494_normalized	smart_495_raw	smart_495_normalized	smart_496_raw	smart_496_normalized	smart_497_raw	smart_497_normalized	smart_498_raw	smart_498_normalized	smart_499_raw	smart_499_normalized	smart_500_raw	smart_500_normalized	smart_501_raw	smart_501_normalized	smart_502_raw	smart_502_normalized	smart_503_raw	smart_503_normalized	smart_504_raw	smart_504_normalized	smart_505_raw	smart_505_normalized	smart_506_raw	smart_506_normalized	smart_507_raw	smart_507_normalized	smart_508_raw	smart_508_normalized	smart_509_raw	smart_509_normalized	smart_510_raw	smart_510_normalized	smart_511_raw	smart_511_normalized	smart_512_raw	smart_512_normalized	smart_513_raw	smart_513_normalized	smart_514_raw	smart_514_normalized	smart_515_raw	smart_515_normalized	smart_516_raw	smart_516_normalized	smart_517_raw	smart_517_normalized	smart_518_raw	smart_518_normalized	smart_519_raw	smart_519_normalized	smart_520_raw	smart_520_normalized	smart_521_raw	smart_521_normalized	smart_522_raw	smart_522_normalized	smart_523_raw	smart_523_normalized	smart_524_raw	smart_524_normalized	smart_525_raw	smart_525_normalized	smart_526_raw	smart_526_normalized	smart_527_raw	smart_527_normalized	smart_528_raw	smart_528_normalized	smart_529_raw	smart_529_normalized	smart_530_raw	smart_530_normalized	smart_531_raw	smart_531_normalized	smart_532_raw	smart_532_normalized	smart_533_raw	smart_533_normalized	smart_534_raw	smart_534_normalized	smart_535_raw	smart_535_normalized	smart_536_raw	smart_536_normalized	smart_537_raw	smart_537_normalized	smart_538_raw	smart_538_normalized	smart_539_raw	smart_539_normalized	smart_540_raw	smart_540_normalized	smart_541_raw	smart_541_normalized	smart_542_raw	smart_542_normalized	smart_543_raw	smart_543_normalized	smart_544_raw	smart_544_normalized	smart_545_raw	smart_545_normalized	smart_546_raw	smart_546_normalized	smart_547_raw	smart_547_normalized	smart_548_raw	smart_548_normalized	smart_549_raw	smart_549_normalized	smart_550_raw	smart_550_normalized	smart_551_raw	smart_551_normalized	smart_552_raw	smart_552_normalized	smart_553_raw	smart_553_normalized	smart_554_raw	smart_554_normalized	smart_555_raw	smart_555_normalized	smart_556_raw	smart_556_normalized	smart_557_raw	smart_557_normalized	smart_558_raw	smart_558_normalized	smart_559_raw	smart_559_normalized	smart_560_raw	smart_560_normalized	smart_561_raw	smart_561_normalized	smart_562_raw	smart_562_normalized	smart_563_raw	smart_563_normalized	smart_564_raw	smart_564_normalized	smart_565_raw	smart_565_normalized	smart_566_raw	smart_566_normalized	smart_567_raw	smart_567_normalized	smart_568_raw	smart_568_normalized	smart_569_raw	smart_569_normalized	smart_570_raw	smart_570_normalized	smart_571_raw	smart_571_normalized	smart_572_raw</
--	--	-------	----------	---------	-------------	--------------------	-------------	--------------------	-------------	--------------------	-------------	--------------------	-------------	--------------------	-------------	--------------------	-------------	--------------------	-------------	--------------------	-------------	--------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	--------------	---------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	---------------	----------------------	-----------------

Analysis of Temperature on hard drives

We analyzed the Smart_194_raw value which represent the temperature of the drive for various manufacturers. Through the visualization below, the distribution of drive temperatures for our four most popular drives. Depicted in bar graph below Fig(2), all drives were well operated between 0° (or 5°) to 60° as specified by manufacturers and within threshold. However, typical operating temperature range of Hard Disk drives could potentially vary by specific disk design characteristics.



RESULTS & DISCUSSIONS

Computing Annual Failure Rate:

Considering a given group of drives (i.e. model, manufacturer, etc.) an attempt is made to compute the AFR for a period of observation as follows:

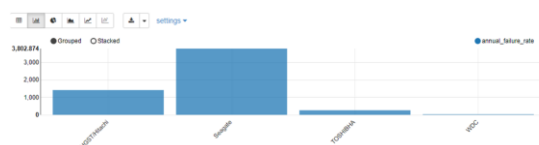
$$AFR = (\text{Drive Failures} / (\text{Drive Days} / 366)) * 100$$

where:

Drive Failures are the number of drives that failed during the period of observation.

Drive Days is number of days all of the observed drives were operational during the period of observation.

There are 366 days in 2020, obviously in non-leap years we would consider 365.



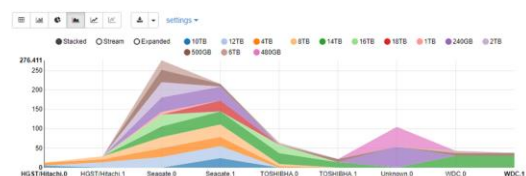
manufacturer	model	capacity_bytes	drivedays	failures	temperature	annual_failure_rate
HGST/Hitachi	HGST HD55C4040ALE630	4TB	9276	1	0	3.545666
HGST/Hitachi	HGST HM55C4040ALE640	4TB	1083641	8	19.5	0.2702
HGST/Hitachi	HGST HM55C4040BLE640	4TB	4662611	34	200.1470588	0.266889
HGST/Hitachi	HGST HUH721212ALE600	12TB	820272	7	11.42857143	0.312335
HGST/Hitachi	HGST HUH721212ALE604	12TB	275036	9	24.11111111	1.197661
HGST/Hitachi	HGST HUH721212ALN604	12TB	3968303	50	3.84	0.461154
HGST/Hitachi	HGST HUH728080ALE600	8TB	371930	3	13.33333333	0.295217
Seagate	ST10000NM0086	10TB	110451	7	28.57142857	2.319581
Seagate	ST12000NM0007	12TB	2314237	66	28.3030303	1.0438
Seagate	ST12000NM0008	12TB	1740779	49	32.85714286	1.030228
Seagate	ST12000NM001G	12TB	610091	12	33.83333333	0.719893
Seagate	ST14000NM001G	14TB	431057	13	32.53846154	1.103798
Seagate	ST18000NM000J	18TB	5491	2	30	13.330905
Seagate	ST4000DM000	4TB	1745899	83	23.79518072	1.739963
Seagate	ST500LM012 HN	500GB	170910	34	5.352941176	7.281025
Seagate	ST8000DM002	8TB	900304	28	33.28571429	1.138282
Seagate	ST8000NM0055	8TB	1325815	47	36.9787234	1.297466
Seagate	Seagate SSD	240GB	9902	1	38	3.696223
TOSHIBA	TOSHIBA MG07ACA14TA	14TB	4100116	102	12.56862745	0.910511
TOSHIBA	TOSHIBA MQ01ABF050	500GB	143891	96	7.625	24.418483
TOSHIBA	TOSHIBA MQ01ABF050M	500GB	145672	32	10.3125	8.03998
WDC	WDC WD5000LPCX	500GB	19476	1	0	1.879236
WDC	WDC WD5000LPVX	500GB	73535	10	0	4.977222
WDC	WDC WUH721414ALE6L4	14TB	226848	1	40	0.161342

Table : Annual Failure Rate based on 2020 stats

After performing a comparative analysis, it's determined that on an average Seagate Drives shows highest rate of Annual failure and then followed by Hitachi as shown in Fig(3) and the table above.

Effect of Disk Temp on Predictive Failure Abnormalities

Overall, there is no direct correlation between operating temperature and failure rates. However, from the observation, Seagate drives models from 2020 are generating more heat while experiencing predictive failure abnormalities and therefore disrupts thermal profiles in Datacenter Environments. This is an important metric as to understanding HDD behavior, which in this case Seagate generate elevated heat levels and contribute to thermal profile variations in storage deployments and thereby increasing Operating expenses. Refer Fig(4)



Relationship between Predictive failure and 5 main SMART Parameters:

SMART readings presented by Hard Disks can be out-of-bounds, noisy, or inaccurate based on disk design and operating conditions, and therefore sometimes have quite a bit of missing data. One step to address these problems is to filter out columns where a lot of the entries are null.

The 5 Key SMART attributes for predictive errors:

SMART 5 - Reallocated_Sector_Count.
SMART 187 - Reported_Uncorrectable_Errors.
SMART 188 - Command_Timeout.
SMART 197 - Current_Pending_Sector_Count.
SMART 198 - Offline_Uncorrectable

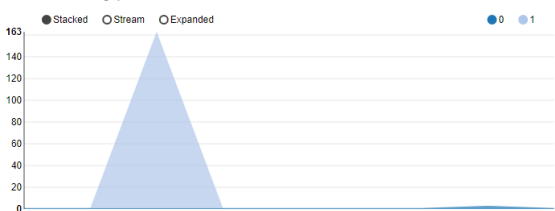
As stated by Backblaze, When RAW value for one of these five attributes is greater than zero, we have a clear indication to consider the value and investigate. Below graphs are to see whether there is a true relation between these SMART parameters and the failure.

As per the observation, except for SMART 188, the definitions described above from Backblaze holds true for every other SMART value.

SMART 5



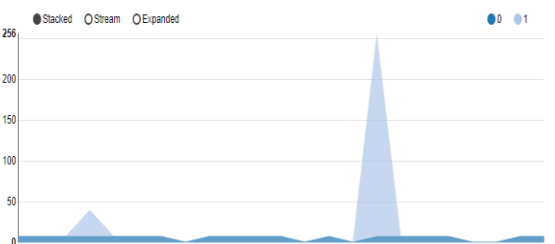
SMART 187



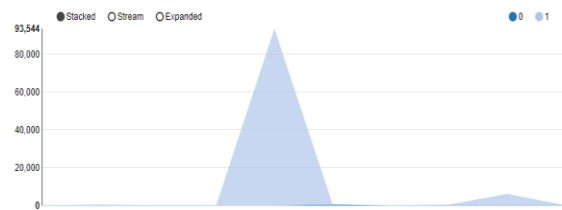
SMART 188



SMART 197



SMART 198



Considering the scope for next level analysis we have implemented linear regression model with these smart parameters as features and failure column as label. However, the coefficients are sometimes visible as blank values and thus we cannot perform our analysis further to display predictions. Therefore, in the future scope we would like to investigate which regressions would best suit to predict the drive failure using machine learning training models.

Conclusion:

Proactive monitoring and management capabilities are the first steps towards improving current standards and modelling systems for gaining Infrastructure Insights. As summarized in each section, various methods and approaches are taken to demonstrate the powerful capabilities of Big Data concepts that included Spark and other supporting utilities. In this study, Hard Disks SMARTs were analyzed for measuring Hard Disk Drive Quality that essentially adds value to improving the availability and reliability of underlying storage. Primarily based upon engineering analysis, we considered the characteristics of each query, the type of data and size of data set were the main factors to optimally demonstrating the capabilities of Big Data Analytics and estimated the failure rates, therefore plotting information to minimizing the scope of impact to infrastructure. The methodologies described can provide proactive and valuable Insights to HDD Manufacturers and Customer Deployments. The paper also establishes relationship between various characteristics of the Hard Disks to understanding the operational overheads like thermals and discusses its impact.

Future Scope:

In this paper, we proposed and established a solution that can be further extended to developing Telemetry & Transform Management system with advanced machine learning techniques. Applying deep data analytics to telemetry data to potentially enable Self Managing, Self-Healing and Self Optimizing features and improving the availability and performance of Storage infrastructure.

References:

1. <https://www.backblaze.com/b2/hard-drive-test-data.html>
2. D. D. Mishra, S. Pathan and C. Murthy, "Apache Spark Based Analytics of Squid Proxy Logs," 2018 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), 2018, pp. 1-6, doi: 10.1109/ANTS.2018.8710044.
3. Singh, Archana & Mittal, Mamta & Kapoor, Namita. (2018). Data Processing Framework Using Apache and Spark Technologies in Big Data. 10.1007/978-981-13-0550-4_5.
4. Botezatu, Mirela & Giurgiu, Ioana & Bogojeska, Jasmina & Wiesmann, Dorothea. (2016). Predicting Disk Replacement towards Reliable Data Centers. 39-48. 10.1145/2939672.2939699.
5. <https://renovacloud.com/an-introduction-to-and-evaluation-of-apache-spark-for-big-data-architectures/?lang=en>