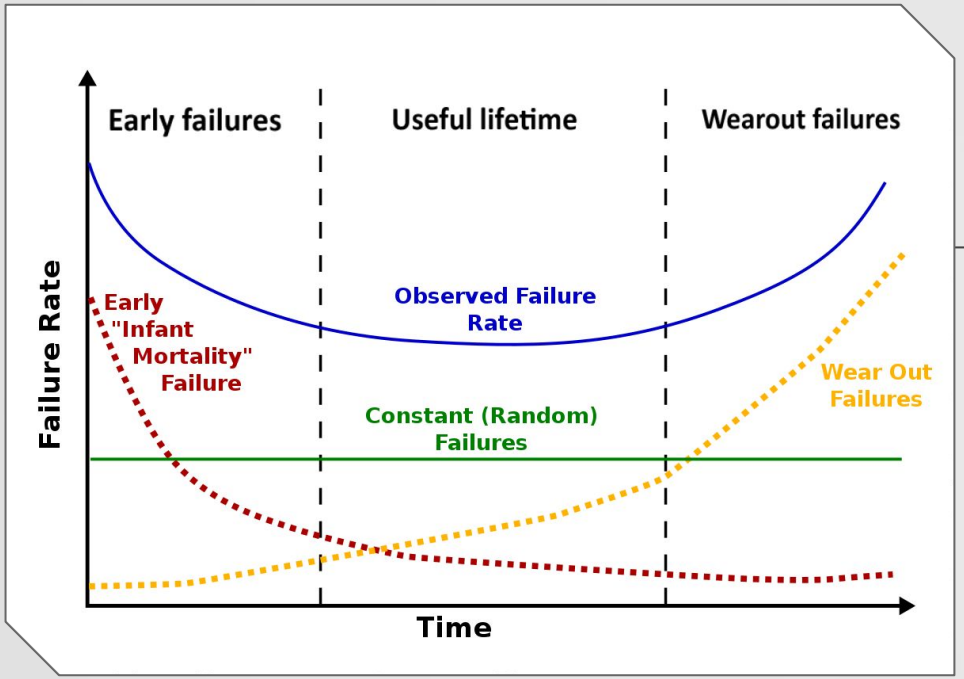
Methods of Statistical Analysis in Server Hardware Failure Prediction

Albert Dzakhoev
Nikita Balashov
Elbrus Olisaev

Project Motivation

Relevance: This project addresses the need for faster replacement of faulty equipment, minimizing delays that would otherwise occur while waiting for parts to be shipped. Additionally, when post-warranty service is our responsibility, having spare parts readily available will further expedite the repair process.

Project Objective: Development of a software tool based on statistical analysis methods to predict server equipment failures. This will enable better planning and stock management, reducing dependence on external suppliers.



Bathtub curve

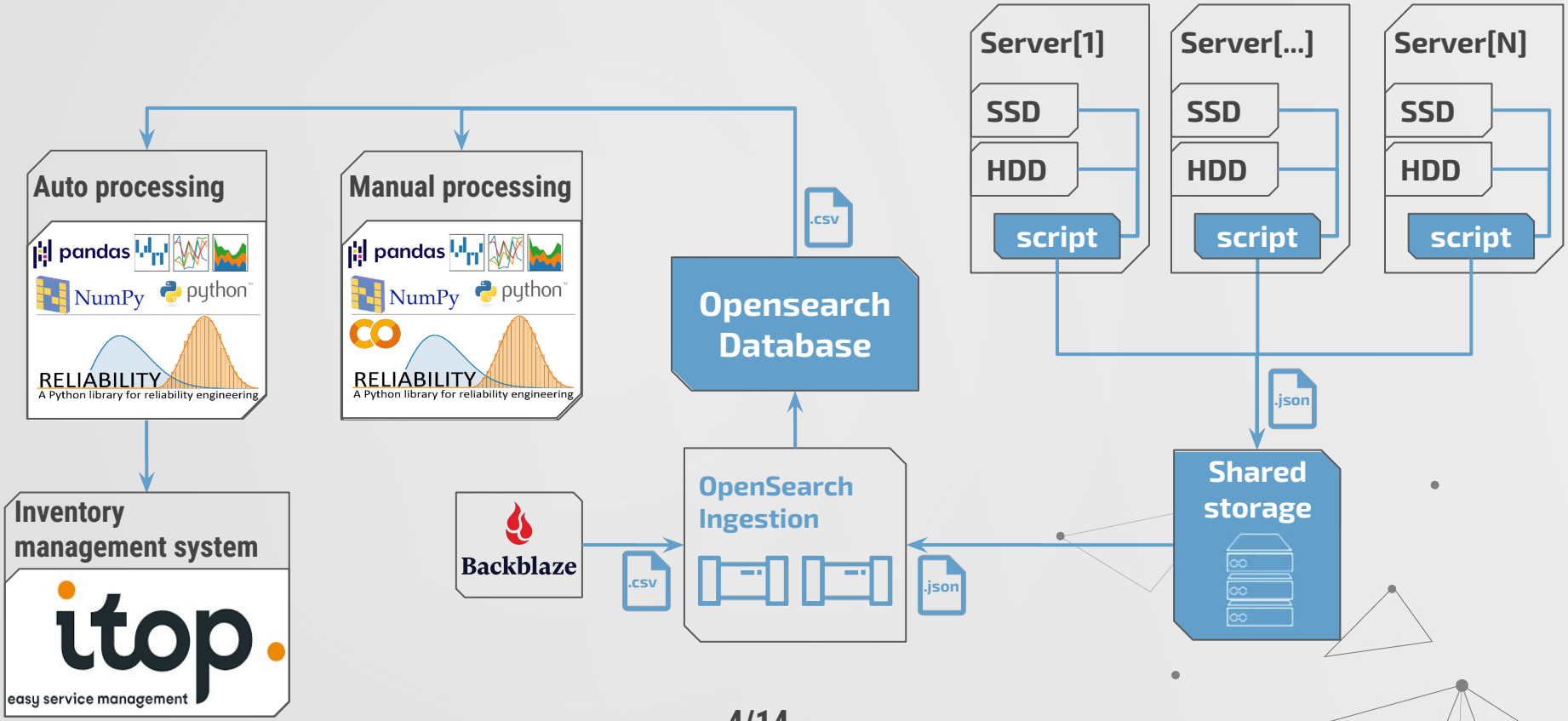
The bathtub curve is a graph that represents the change in the observed failure rate of components over time. It consists of three main stages: **Early failures** with a high failure frequency, **Useful lifetime** with a nearly constant failure frequency due to random events, and **Wearout failures** with an increasing failure frequency due to wear.

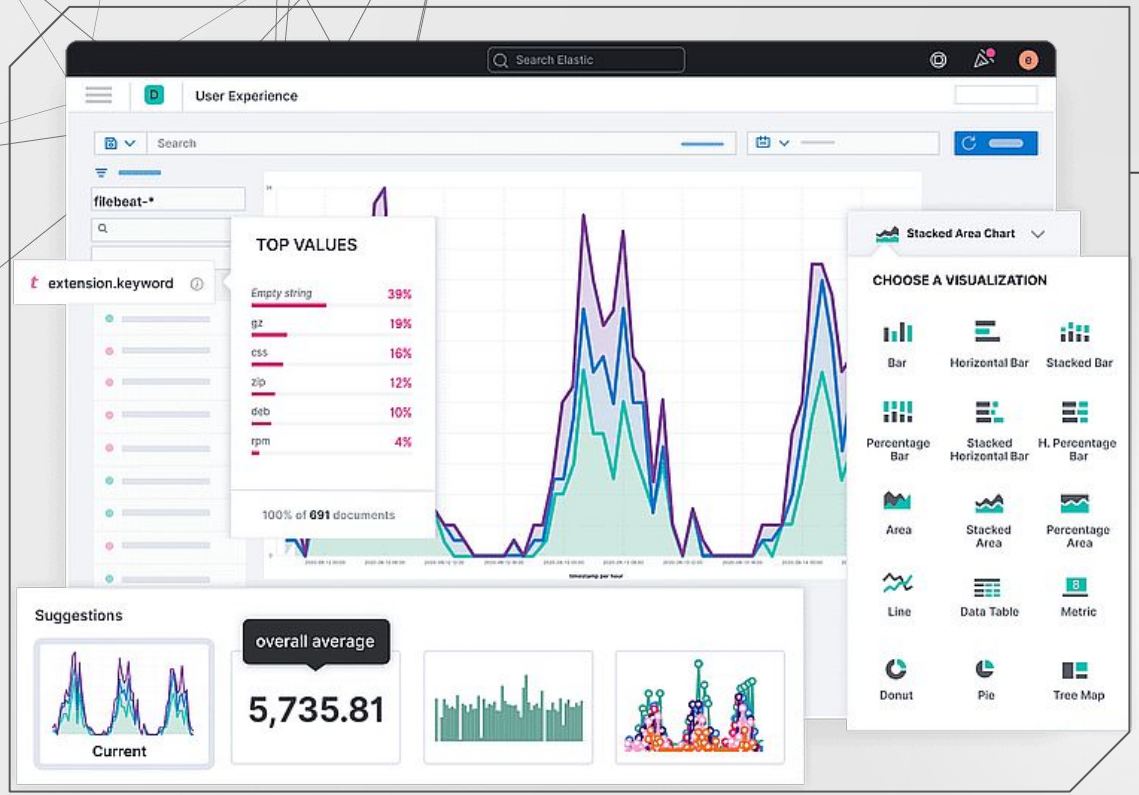
```
=== START OF INFORMATION SECTION ===
Vendor:               SEAGATE
Product:              ST18000NM004J
Revision:             E004
Compliance:           SPC-5
User Capacity:        18,000,207,937,536 bytes [18.0 TB]
Logical block size:   512 bytes
Physical block size:  4096 bytes
LU is fully provisioned
Rotation Rate:        7200 rpm
Form Factor:          3.5 inches
Logical Unit id:      0x5000c500dad5185f
Serial number:        ZR5ETNBY0000W332TVTX
Device type:          disk
Transport protocol:   SAS (SPL-3)
Local Time is:        Mon Oct 23 12:35:22 2023 MSK
SMART support is:     Available - device has SMART capability.
SMART support is:     Enabled
Temperature Warning:  Enabled
Read Cache is:        Enabled
Writeback Cache is:   Enabled

=== START OF READ SMART DATA SECTION ===
SMART Health Status: OK
```

S.M.A.R.T. REPORTS

S.M.A.R.T. (Self-Monitoring, Analysis, and Reporting Technology) - is a technology used in hard drives to monitor and assess their condition. **SMART**-reports contain data on the hard drive's status, such as temperature, the number of read/write errors, the number of reboots, and so on.

System Architecture

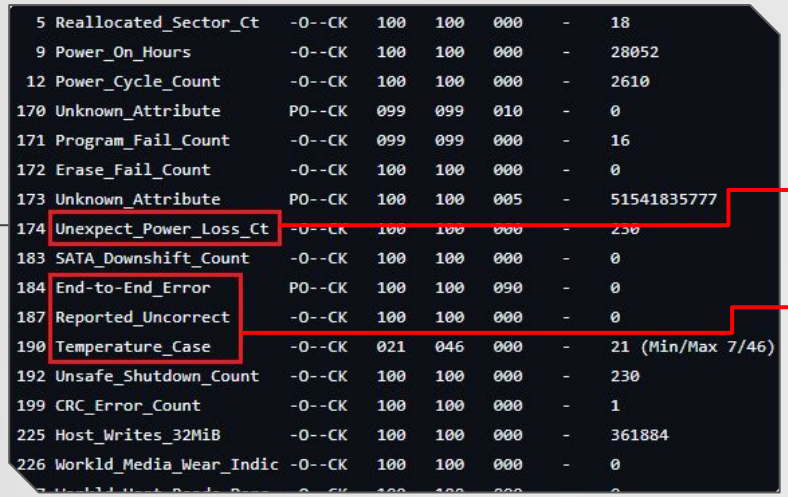




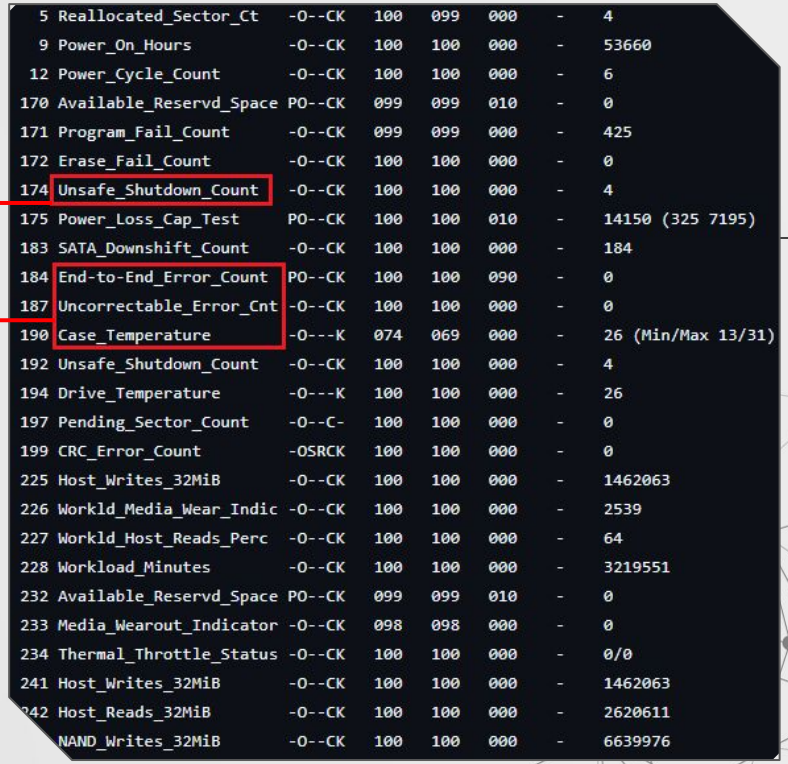
OPENSEARCH STACK

A platform for data storage and retrieval that provides tools for visualization, monitoring, and managing large volumes of information in real-time. It provides the ability to use pipelines for preliminary data processing, which allows us to standardize and transform data before indexing.

Reports Differences



```
  5 Reallocated_Sector_Ct   -O--CK   100   100   000    -    18
  9 Power_On_Hours          -O--CK   100   100   000    -    28052
 12 Power_Cycle_Count       -O--CK   100   100   000    -    2610
170 Unknown_Attribute       PO--CK   099   099   010    -    0
171 Program_Fail_Count      -O--CK   099   099   000    -    16
172 Erase_Fail_Count        -O--CK   100   100   000    -    0
173 Unknown_Attribute       PO--CK   100   100   005    -    51541835777
174 Unexpect_Power_Loss_Ct  -O--CK   100   100   000    -    230
183 SATA_Downshift_Count    -O--CK   100   100   000    -    0
184 End-to-End_Error        PO--CK   100   100   090    -    0
187 Reported_Uncorrect      -O--CK   100   100   000    -    0
190 Temperature_Case        -O--CK   021   046   000    -    21 (Min/Max 7/46)
192 Unsafe_Shutdown_Count   -O--CK   100   100   000    -    230
199 CRC_Error_Count         -O--CK   100   100   000    -    1
225 Host_Writes_32MiB       -O--CK   100   100   000    -    361884
226 Workld_Media_Wear_Indic -O--CK   100   100   000    -    0
```



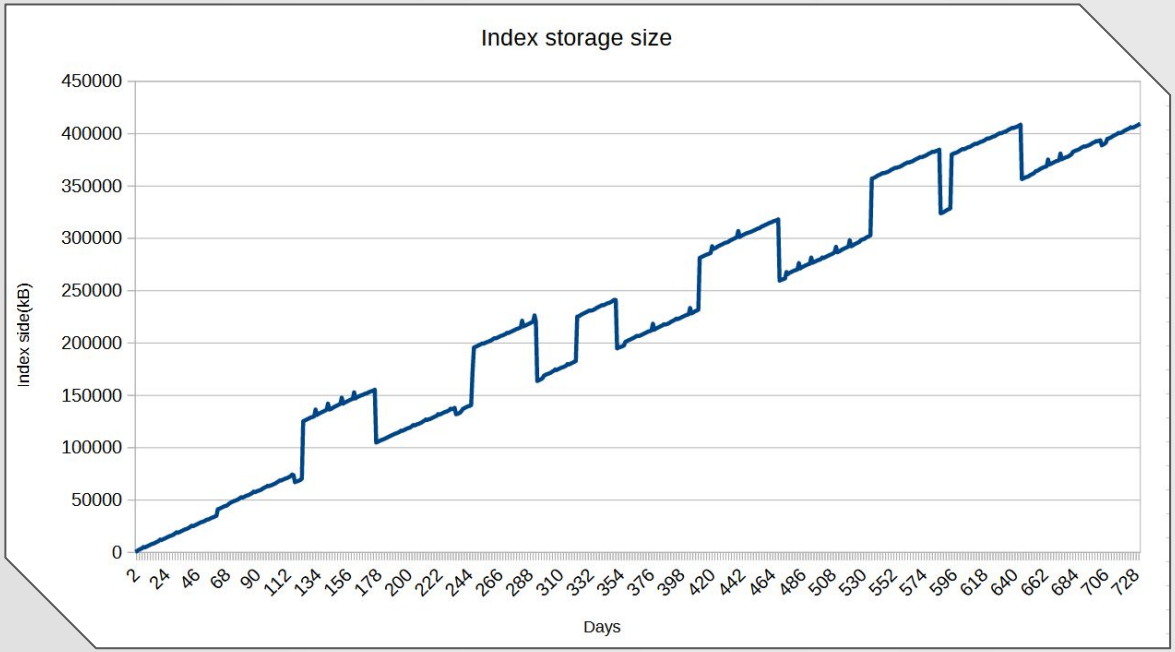
```
  5 Reallocated_Sector_Ct   -O--CK   100   099   000    -    4
  9 Power_On_Hours          -O--CK   100   100   000    -    53660
 12 Power_Cycle_Count       -O--CK   100   100   000    -    6
170 Available_Reservd_Space PO--CK   099   099   010    -    0
171 Program_Fail_Count      -O--CK   099   099   000    -    425
172 Erase_Fail_Count        -O--CK   100   100   000    -    0
174 Unsafe_Shutdown_Count   -O--CK   100   100   000    -    4
175 Power_Loss_Cap_Test     PO--CK   100   100   010    -    14150 (325 7195)
183 SATA_Downshift_Count    -O--CK   100   100   000    -    184
184 End-to-End_Error_Count  PO--CK   100   100   090    -    0
187 Uncorrectable_Error_Cnt -O--CK   100   100   000    -    0
190 Case_Temperature        -O---K   074   069   000    -    26 (Min/Max 13/31)
192 Unsafe_Shutdown_Count   -O--CK   100   100   000    -    4
194 Drive_Temperature       -O---K   100   100   000    -    26
197 Pending_Sector_Count    -O--C-   100   100   000    -    0
199 CRC_Error_Count         -OSRCK   100   100   000    -    0
225 Host_Writes_32MiB       -O--CK   100   100   000    -    1462063
226 Workld_Media_Wear_Indic -O--CK   100   100   000    -    2539
227 Workld_Host_Reads_Perc  -O--CK   100   100   000    -    64
228 Workload_Minutes        -O--CK   100   100   000    -    3219551
232 Available_Reservd_Space PO--CK   099   099   010    -    0
233 Media_Wearout_Indicator -O--CK   098   098   000    -    0
234 Thermal_Throttle_Status -O--CK   100   100   000    -    0/0
241 Host_Writes_32MiB       -O--CK   100   100   000    -    1462063
242 Host_Reads_32MiB        -O--CK   100   100   000    -    2620611
    NAND_Writes_32MiB       -O--CK   100   100   000    -    6639976
```

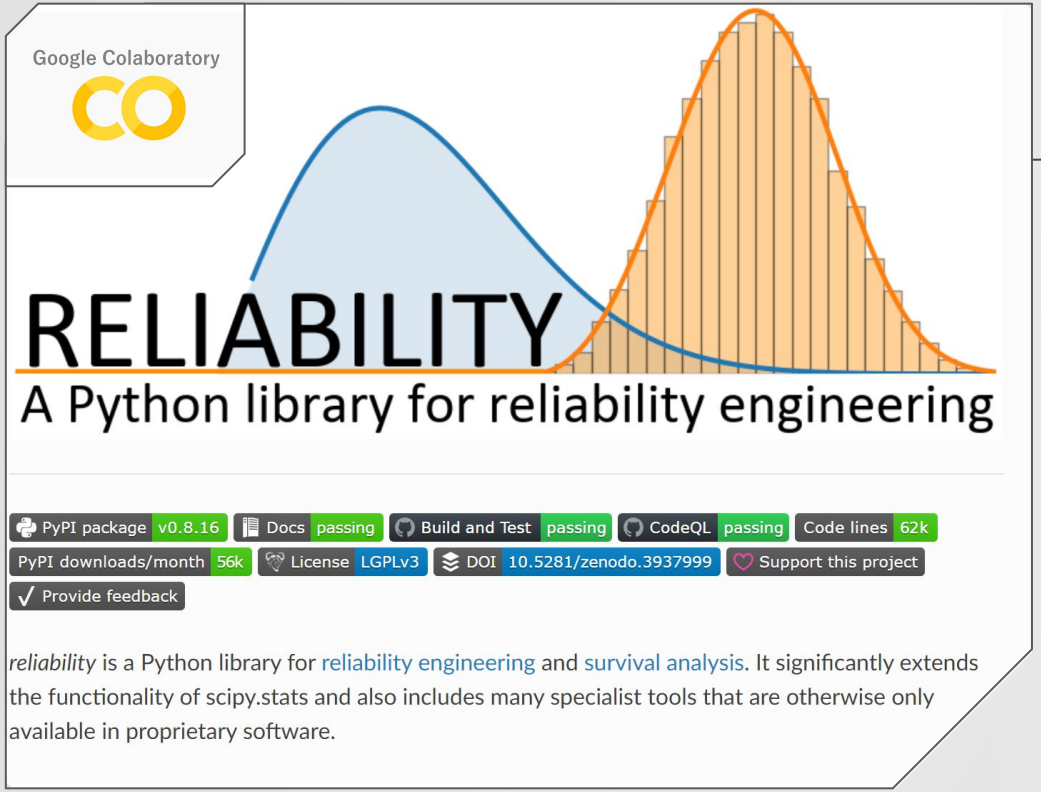
S.M.A.R.T. reports for **Intel S3520 SSDs** and **Intel 545s**

Parameter names **174**, **184**, **187**, and **190** vary between the two models, even though both are Intel SSDs. Utilizing pipelines simplifies subsequent analysis by eliminating duplication and enhancing data consistency.

Required Space Estimation

The index database size increase remains linear over two years of data accumulation





RELIABILITY

A library that implements various methods for calculating and analyzing reliability data, including the Weibull distribution.

Backblaze

An American company specializing in providing online data storage and backup services. They publish detailed statistics on the performance of thousands of hard drives used in their data centers.

Drive Stats 2023 Snapshot

Drive Count	Drive Failures	Drive Days	Annualized Failure Rate
269,756	16,962	424,474,539	1.46%

Hard Drive Reliability - Annualized Hard Drive Failure Rates

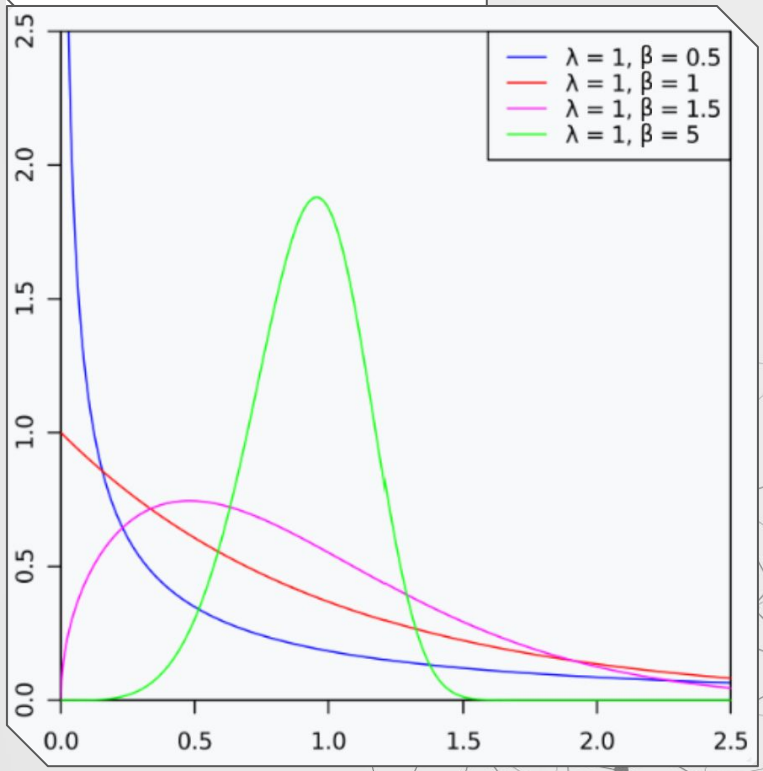
Weibull distribution

The **Weibull distribution** is used in reliability analysis, for example, to calculate the mean time to failure (MTTF) of a device.

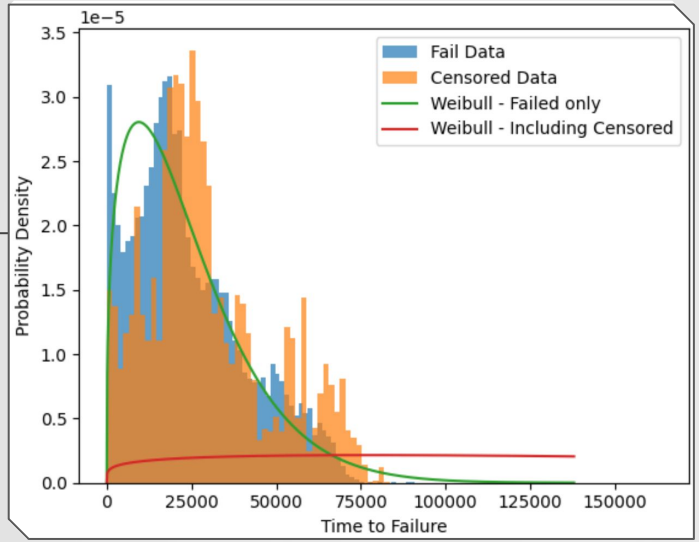
Applications:

- Modeling the time between events.
- Survival analysis.

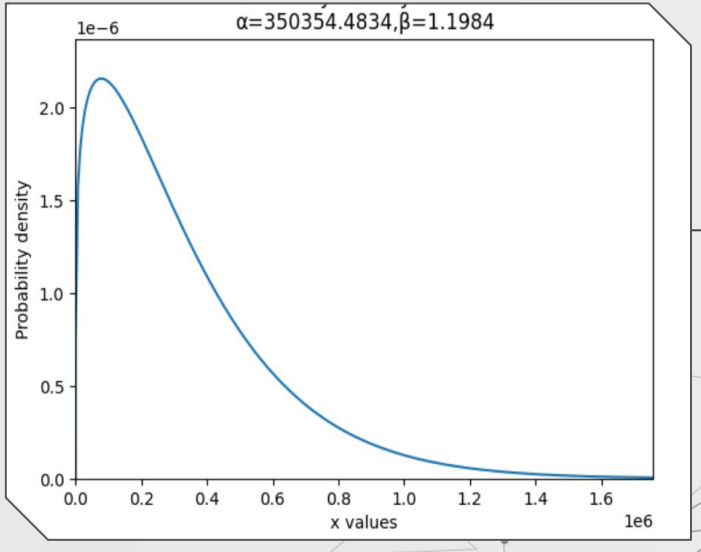
$$f_X(x) = \begin{cases} \frac{\beta}{\lambda}\left(\frac{x}{\lambda}\right)^{\beta-1} e^{-\left(\frac{x}{\lambda}\right)^{\beta}}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$



Probability Density Function



Histogram of the Data Used



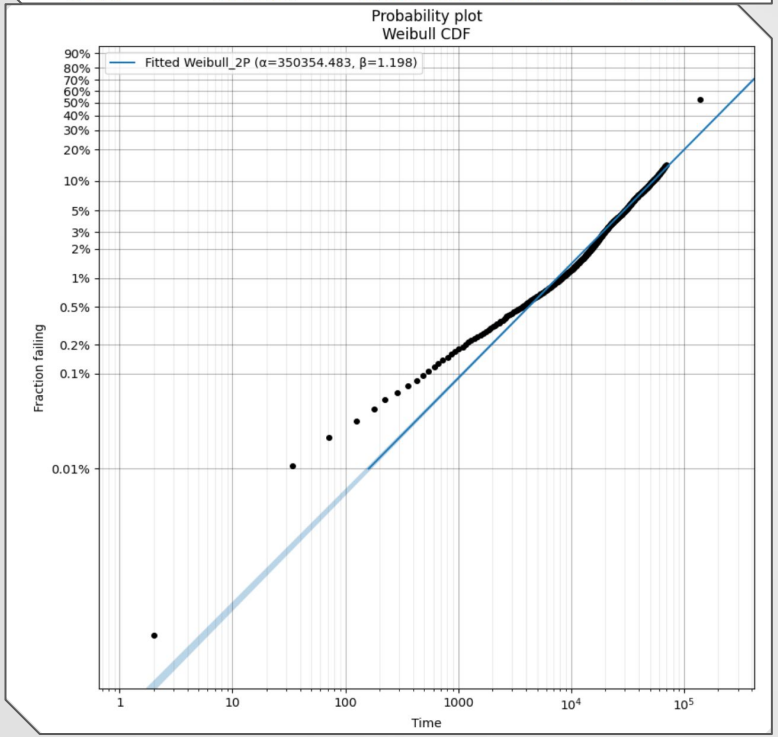
Graph of the Probability Density Function (PDF)

- Fail Data - operating time of failed disks
- Censored Data – operating time of disks still in operation

The Probability Density Function (**PDF**) is used to describe the likelihood of a component failing at a specific moment in time or over a specific time period.

Maximum Likelihood Estimation

```
Results from Fit Weibull 2P (95% CI):
Analysis method: Maximum Likelihood Estimation (MLE)
Optimizer: TNC
Failures / Right censored: 20897/370412 (94.65972% right censored)

Parameter  Point Estimate  Standard Error  Lower CI  Upper CI
    Alpha          350354         4995.48    340699    360284
     Beta         1.19836      0.00702269   1.18468   1.21221
```



Weibull distribution analysis conducted using Maximum Likelihood Estimation (**MLE**). We use the coefficients obtained from processing these data for future calculations. The plot indicates the likelihood of failure over time for the dataset and provides a visual confirmation of the Weibull model fit.

Data Overview:

- Total samples: 370412
- Failures: 20897
- Right censored: 94.65972%, meaning that most data points did not fail within the observed period

Analysis Example

These are the forecasts based on **Backblaze** data for the approximate parameters of the computing infrastructure at **JINR**.

```
Probability of failure and expected number of failures by specific times:
Time 10000:
  Probability of failure of one disk by time 10000: 1.3998%
  Expected number of failures out of 4000 disks by time 10000: 56.0

Time 20000:
  Probability of failure of one disk by time 20000: 3.1831%
  Expected number of failures out of 4000 disks by time 20000: 127.3

Time 30000:
  Probability of failure of one disk by time 30000: 5.1229%
  Expected number of failures out of 4000 disks by time 30000: 204.9

Approximate number of components that will fail in the time interval between the first and second year [8760, 17520]: 61.08
```

Current Results and Future Plans

Ready-to-implement solutions

Research and development solutions

Future plans