

DOI: <https://doi.org/10.15276/ict.03.2026.67>

УДК 004.62:004.853]:004.658.2

Збереження єдиного прогностичного контуру при еволюції телеметричного простору розподілених систем зберігання даних

Кравцов Сергій Вікторович¹⁾

ORCID: <https://orcid.org/0000-0003-2161-5111>; zelenij@gmail.com

Мартинюк Олександр Миколайович¹⁾

ORCID: <https://orcid.org/0000-0003-1461-2000>; martynyuk@op.edu.ua. Scopus Author ID: 60238390900

¹⁾ Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

АНОТАЦІЯ

У статті розглядається проблема збереження єдиного прогностичного контуру в умовах еволюції телеметричного простору програмно-визначених розподілених систем зберігання даних. Актуальність задачі зумовлена тим, що під час довготривалої експлуатації кластерів зазначених систем склад доступної телеметрії змінюється внаслідок оновлення версій платформи, появи нових діагностичних метрик, зміни складу та семантики повідомлень системних журналів, службових станів та складу обладнання; унаслідок цього побудова окремої моделі для кожної нової конфігурації призводить до фрагментації історичних даних і втрати наступності прогнозування. Запропоновано подання телеметричного простору як відкритої структури, що складається з інваріантного ядра ознак і змінних розширень. Інваріантне ядро виконує функцію шару сумісності між різними періодами експлуатації, тоді як змінні розширення відображають розвиток платформи, інфраструктури та доступної телеметрії. Для опису еволюції ознак введено типізовані відношення сумісності, що охоплюють збереження, перейменування, нормалізаційне відображення, агрегацію, декомпозицію та семантичну відповідність. Наукова новизна полягає у трактуванні інваріантного ядра не як механічного перетину однакових метрик різних періодів експлуатації, а як шару семантичної сумісності, в межах якого експлуатаційне значення ознак може зберігатися через допустимі перетворення. Це дає змогу підтримувати наступність прогностичної моделі навіть за зміни назв метрик, джерел телеметрії та форматів подання даних. Підхід проілюстровано на еволюції телеметричного простору Серф-кластера від Kraken до Reef. Показано, що зміна складу телеметрії може розглядатися як контрольоване розширення єдиного прогностичного контуру за умови збереження або відтворюваності інваріантного ядра. Якщо ця умова порушується, побудова окремої моделі стає обґрунтованою необхідною.

Ключові слова: телеметричний простір, прогнозування відмов, інваріантне ядро ознак, еволюція телеметрії, сумісність ознак, прогностичний контур

Для цитування: Кравцов С. В., Мартинюк О. М. «Збереження єдиного прогностичного контуру при еволюції телеметричного простору розподілених систем зберігання даних». *Інформатика. Культура. Техніка*. 2026; Том 3 № 2 (4):866–877. DOI: <https://doi.org/10.15276/ict.03.2026.67>

Актуальність. Програмно-визначені розподілені системи зберігання даних (Software Defined Storage – SDS) є одним із ключових інструментів побудови масштабованої інфраструктури для обробки та збереження великих обсягів даних. Їх розвиток пов'язаний із переходом від монолітних апаратних комплексів до гнучких програмних платформ, у яких функції розміщення, реплікації, відновлення та балансування даних реалізуються на рівні розподіленої логіки керування. Одним із репрезентативних прикладів такого підходу є Серф, архітектура якого базується на децентралізованому розміщенні даних, групах розміщення, OSD-компонентах та механізмах автоматичного відновлення [1], [2], [3], [4]. Однак саме ця гнучкість створює новий клас задач для систем прогнозування відмов. На відміну від традиційних систем зберігання, де основна увага приділялася стану окремих накопичувачів або контролерів, у середовищах програмно-визначених систем зберігання даних технічний стан системи визначається взаємодією багатьох рівнів: апаратних компонентів, мережних зв'язків, процесів реплікації, службових станів платформи, механізмів відновлення та ін. У суміжних дослідженнях хмарних і мікросервісних середовищ також підкреслюється, що аналіз аномалій та причин відмов потребує використання багатоджерельних сигналів, а не ізольованих метрик [5].

У попередній роботі автора [11] було розглянуто задачу формування мультимодального вектора ознак для прогнозування відмов розподілених систем зберігання даних. Було показано, що апаратні метрики, показники продуктивності, параметри реплікації та семантика системних журналів можуть бути інтегровані в єдиний вхідний простір моделі за умови спеціалізованої нормалізації гетерогенної телеметрії. У подальшій роботі [12] цей

підготовлений простір ознак було використано як основу ризик-орієнтованого прогнозування, де задача формалізується через оцінювання ймовірності критичної події на заданому часовому горизонті з урахуванням асиметричної вартості помилок.

Разом з тим обидва зазначені етапи не вичерпують проблему довготривалої експлуатації прогностичної системи. У реальних умовах інфраструктура програмно-визначеної системи зберігання даних не залишається сталою: оновлюються версії платформ, змінюється склад службових станів і журналів, розширюється набір діагностичних метрик, додаються нові рівні зберігання. Унаслідок цього телеметричний простір кластера програмно-визначеної системи зберігання даних еволюціонує насамперед через зміну самої структури програмно-визначеної системи та способів відображення її внутрішнього стану. Класичний підхід, за якого для кожної нової конфігурації будується окрема модель, призводить до фрагментації накопичених даних і втрати сумісності історичних спостережень. Описана проблема має практичний вимір і була актуалізована в процесі супроводу реального Serph-кластера, що експлуатується з 2017 року дотепер та в межах досліджуваного періоду пройшов еволюцію від Kraken (v11) до Reef (v18).

Аналіз останніх досліджень і публікацій. Архітектурні принципи програмно-визначених розподілених систем зберігання даних, зокрема Serph, детально описані у роботі Weil та ін. [1], а механізми груп розміщення, стани OSD-компонентів (Object Storage Daemon – демонів зберігання об'єктів) та діагностичні перевірки задокументовані в офіційній документації платформи [2], [3], [4]. Із цих джерел випливає, що надійність SDS визначається не станом окремих накопичувачів, а взаємодією багатьох службових шарів: OSD-компонентів, груп розміщення, механізмів реплікації, відновлення, балансування та внутрішньої діагностики. У суміжних дослідженнях з аналізу аномалій у хмарних і мікросервісних середовищах [5] також підкреслюється, що ефективна діагностика потребує багатоджерельних сигналів, а не ізольованих метрик.

Питання прогнозування відмов у великих системах зберігання розглядається у роботі Lu S. та ін. [10], де систематизовано метрики, методи та практичні складнощі побудови моделей у промислових умовах. Автори зазначають, що рідкість критичних подій і дисбаланс класів є ключовими обмеженнями під час навчання моделей, а накопичення достатньої кількості відмов потребує тривалих спостережень. Однак питання збереження єдиного прогностичного контуру при зміні конфігурацій інфраструктури, оновленні платформи або зміні складу доступної телеметрії не є центральним предметом цього дослідження.

У суміжних напрямках машинного навчання розроблено підходи до адаптації моделей у умовах дрейфу концепції [6], [7], потокових даних із дисбалансом класів [8] та еволюції простору ознак [9]. Ці роботи є важливими для розуміння загальної проблеми зміни статистичних властивостей даних у часі. Зокрема, Hou та ін. [9] розглядають сценарій, у якому нові ознаки з'являються вже після формування початкового простору спостереження, що концептуально близько до задачі цієї статті.

Разом з тим у задачах експлуатації SDS-платформ еволюція простору ознак має не лише статистичний, а й семантико-інженерний характер. Зміна версії платформи, формату лічильників або структури журналів не завжди означає появу принципово нової експлуатаційної властивості; часто вона відображає зміну способу подання вже відомого стану системи. З методологічної точки зору ця задача також близька до проблеми зіставлення схем даних, де необхідно встановлювати відповідність між елементами різних версій структури даних, навіть якщо вони відрізняються назвами, форматами або способом подання [16]. Тому для збереження прогностичного контуру недостатньо лише фіксувати факт появи або зникнення ознак, як це розглядається в загальних постановках еволюції простору ознак [9]. Необхідно формалізувати відношення між ознаками різних періодів експлуатації: збереження, перейменування, нормалізаційне відображення, агрегацію, декомпозицію та семантичну відповідність.

Наявні підходи до дрейфу даних, еволюції ознак та зіставлення схем даних [6], [7], [8], [9], [16] орієнтовані переважно на загальні схеми адаптації моделей або узгодження структур даних і не враховують специфіки SDS-телеметрії: залежності ознак від версії платформи,

зміни способів подання службових станів, семантики системних журналів та необхідності підтримувати інтерпретовану відповідність між історичними й поточними ознаками. У контексті довготривалої експлуатації Serph-кластера це означає, що ознака може змінити назву, джерело або спосіб подання, але залишатися релевантною для тієї самої прогностичної семантики. Саме тому потрібен не лише механізм додавання нових ознак, а й формальний опис умов, за яких ознаки різних періодів можуть вважатися сумісними.

У попередніх роботах автора розглянуто побудову мультимодального вектора ознак [11] та ризик-орієнтовану постановку прогнозування відмов [12]. Ці роботи формують базові елементи прогностичної системи: нормалізований простір входних ознак і постановку задачі прогнозування критичної події на заданому часовому горизонті. Однак вони не вирішують окрему задачу збереження єдиного прогностичного контуру при поступовій зміні складу телеметрії, появі нових лічильників, зміні форматів журналювання або модифікації службових станів платформи. Таким чином, невирішеною залишається задача формального опису того, за яких умов історичні та поточні ознаки SDS-телеметрії можуть вважатися сумісними в межах одного прогностичного контуру.

Мета дослідження. Метою цієї статті є формалізація методологічного підходу до збереження єдиного прогностичного контуру в умовах еволюції телеметричного простору розподілених систем зберігання даних шляхом виділення інваріантного ядра ознак, змінних розширень та визначення допустимих відношень сумісності між ознаками різних періодів експлуатації. Запропонований підхід має забезпечити використання історичних і поточних телеметричних спостережень у межах спільної прогностичної моделі без необхідності її повної перебудови при кожній зміні складу телеметрії. Наукова новизна полягає у трактуванні інваріантного ядра не як фіксованого набору тотожних метрик, а як шару семантичної сумісності ознак різних періодів експлуатації, що допускає збереження їх прогностичного значення через формалізовані відношення сумісності.

Методи. Формалізація відкритого телеметричного простору. У межах цієї статті попередньо сформований мультимодальний простір ознак [11] та ризик-орієнтована постановка прогнозування [12] розглядаються як базові елементи єдиного прогностичного контуру. Подальша задача полягає у формалізації такого подання телеметричного простору, яке допускає його еволюцію без втрати сумісності історичних і поточних спостережень.

Позначимо через $X(t)$ множину ознак, доступних для опису стану SDS-системи в момент або період часу (t). На відміну від статичної постановки, у даній роботі $X(t)$ розглядається як змінна структура, склад якої може еволюціонувати протягом життєвого циклу системи.

Для відображення цієї властивості телеметричний простір подається як відкрита структура, що складається з інваріантного ядра та змінних розширень:

$$X(t) = X_{core} \cup X_{ext}(t) \quad (1)$$

де X_{core} – стабільна частина простору ознак, що забезпечує наступність прогностичного контуру, а $X_{ext}(t)$ – змінна множина ознак, яка відображає поточний рівень розвитку SDS-платформи, інфраструктури та доступної телеметрії.

У такій постановці ядро підтримує сумісність, а змінні ознаки уточнюють поточний опис системи. Такий підхід узгоджується із загальною проблематикою еволюції простору ознак у суміжних напрямках машинного навчання [6], [7], [8], [9].

Інваріантне ядро ознак. Інваріантне ядро X_{core} формується з ознак, які мають стійке експлуатаційне значення для розподіленої системи зберігання даних незалежно від конкретної версії SDS-платформи, покоління обладнання або складу додаткових сервісів. До таких ознак належать узагальнені характеристики доступності вузлів, показники обслуговування запитів, агреговані метрики затримок, базові ознаки реплікації та відновлення, а також узагальнені характеристики службових журналів.

Методологічно X_{core} виконує функцію опорного шару прогностичної системи. Саме цей шар дозволяє зіставляти історичні та поточні спостереження, оскільки його компоненти

мають бути доступними або відтворюваними на різних етапах життєвого циклу системи. У первинному поданні інваріантне ядро може бути задане як перетин семантично узгоджених множин ознак, доступних або відтворюваних у ключових періодах експлуатації:

$$X_{core} = \bigcap_{t \in T} X(t) \quad (2)$$

де T – множина часових періодів, для яких накопичено телеметричні дані.

Таке визначення не означає, що інваріантне ядро має охоплювати більшу частину всіх доступних ознак. Його цінність полягає не в максимальній розмірності, а в стабільності, інтерпретованості та здатності підтримувати єдину логіку оцінювання стану системи. Надмірне включення до X_{core} платформозалежних або нестабільних параметрів знижує його роль як шару довготривалої сумісності.

З практичної точки зору ознака може бути віднесена до X_{core} , якщо вона відповідає трьом умовам: має зрозумілу експлуатаційну інтерпретацію, доступна або може бути приведена до узгодженого вигляду в різних періодах експлуатації, а також не залежить критично від окремої версії платформи чи вузької конфігурації інфраструктури. Ілюстрацію цього визначення на матеріалі реального Serph-кластера, що пройшов через вісім мажорних версій платформи, наведено в розділі результатів.

Змінні розширення ознак. Змінна частина $X_{ext}(t)$ відображає природну еволюцію SDS-інфраструктури. До неї можуть входити нові діагностичні метрики, нові службові стани, нові типи журнальних повідомлень, ознаки додаткових рівнів зберігання та обслуговування даних, характеристики географічно розподіленого трафіку, параметри маршрутизації запитів або нові показники роботи механізмів відновлення й балансування.

Формально розширення визначається як частина поточного простору ознак, що не входить до інваріантного ядра:

$$X_{ext}(t) = X(t) \setminus X_{core} \quad (3)$$

Принциповим є те, що $X_{ext}(t)$ не замінює ядро, а доповнює його. Нові ознаки можуть підвищувати деталізацію опису стану системи, уточнювати ризикові сценарії або відображати нові експлуатаційні процеси, однак вони не повинні руйнувати можливість порівняння з історичними спостереженнями. У цьому полягає відмінність відкритого телеметричного простору від підходу з побудовою окремих моделей для кожної нової конфігурації інфраструктури.

У межах запропонованої методології кожне спостереження стану системи в момент часу (t) може бути подане як комбінація ядра та доступних на цей момент розширень:

$$x(t) = ([x_{core}(t), x_{ext}(t)]) \quad (4)$$

Якщо частина розширених ознак відсутня в історичних даних, це не розглядається як втрата сумісності. Історичний контур зберігається через X_{core} , тоді як нові ознаки поступово включаються до аналізу як додатковий інформаційний шар. Така схема узгоджується із загальною ідеєю навчання в умовах еволюції ознак, де нові змінні можуть з'являтися вже після формування початкового простору спостереження [9].

Умова сумісності історичних і нових спостережень. Ключовою вимогою до відкритого телеметричного простору є збереження сумісності між спостереженнями, отриманими в різні періоди життєвого циклу SDS-системи.

Для цього необхідно, щоб будь-яке нове спостереження містило або дозволяло відновити компоненти інваріантного ядра:

$$\forall t \in T: X_{core} \subseteq X(t) \quad (5)$$

Ця умова означає, що незалежно від зміни платформи, оновлення службових шарів, додавання нових джерел телеметрії або зміни конфігурації інфраструктури базовий набір інтерпретованих ознак має залишатися доступним для оцінювання стану системи. Якщо ця

умова порушується, єдиний прогностичний контур втрачає основу для зіставлення історичних і поточних даних.

У практичному сенсі сумісність не вимагає повної однаковості всіх ознак у різні періоди експлуатації. Вона вимагає збереження спільного ядра, через яке історичні дані залишаються корисними для поточного аналізу. Це особливо важливо для задач прогнозування відмов, де рідкість критичних подій і дисбаланс класів ускладнюють накопичення достатньої навчальної вибірки [8], [10]. Порушення умови сумісності в таких умовах безпосередньо знижує статистичну основу прогнозування.

Отже, відкритий телеметричний простір має забезпечувати не механічну сталість усіх параметрів, а змістову наступність між різними станами інфраструктури. Саме ця наступність дозволяє розглядати систему прогнозування як довготривалий інструмент експлуатації, а не як набір локальних моделей, прив'язаних до окремих версій платформи або конфігурацій обладнання.

Принцип збереження прогностичного контуру. На основі введених визначень принцип збереження єдиного прогностичного контуру можна сформулювати так: еволюція телеметричного простору допускається лише в такій формі, за якої інваріантне ядро забезпечує сумісність історичних і поточних спостережень, а змінні розширення підвищують деталізацію опису без руйнування базової постановки задачі прогнозування.

Для кожного періоду експлуатації стан системи описується парою:

$$X(t) = (X_{core}, X_{ext}(t)), \quad (6)$$

а задача прогнозування зберігає загальний вигляд:

$$P(Y_{t+H}=1|X(t)) \quad (7)$$

де $Y_{t+H}=1$ означає настання критичної події на горизонті H .

Таким чином, зміна складу $X_{ext}(t)$ не порушує логіки прогнозування, а X_{core} забезпечує зв'язок із попередніми періодами експлуатації. Це створює основу для довготривалої експлуатації прогностичної системи без фрагментації накопичених даних.

Еволюцію схеми телеметричного простору, що є предметом цієї роботи, слід відрізнити від концептуального дрейфу, пов'язаного зі зміною статистичних властивостей даних та залежності між ознаками і прогнозованою подією. Методологію часової валідації без використання майбутніх даних, контроль концептуального дрейфу і дрейфу профілю ризику, а також умови рекалібрування та перенавчання прогностичної моделі детально розглянуто в попередній роботі автора [12]. У межах поточного дослідження розглядається інша задача - збереження сумісності телеметричного простору при зміні його структури.

Формалізація відношень між ознаками у часі. Для теоретичного обґрунтування механізму збереження інваріантного ядра в умовах еволюції телеметричного простору введено системно-множинне та предикатне подання ознак і відношень між ними.

Така постановка концептуально близька до задач зіставлення схем даних, у яких необхідно встановлювати відповідність між елементами різних структур подання інформації [16], однак у межах цієї роботи вона адаптується до телеметричного простору SDS та задачі збереження єдиного прогностичного контуру.

Нехай $F = \{f_i\}_{i=1}^N$ - генеральна множина потенційних телеметричних ознак SDS-системи. Кожна ознака $f_i \in F$ описується кортежем:

$$f_i = \langle id_i, name_i, kind_i, source_i, version_i, V_i, \psi_i \rangle, \quad (8)$$

де id_i - унікальний ідентифікатор ознаки; $name_i$ - системне ім'я або шлях метрики; $kind_i \in \{stat, dyn\}$ - роль ознаки за характером змінності; $source_i$ - джерело телеметрії; $version_i$ - версія платформи або схеми телеметрії; V_i - область допустимих значень; $\psi_i: T \rightarrow V_i$ - функція, що задає значення ознаки у часі.

Для періоду експлуатації (t) введемо предикат доступності:

$$A(f_i, t) = \begin{cases} 1 & \text{якщо ознака } f_i \text{ доступна або може бути відтворена в період } t \\ 0 & \text{інакше} \end{cases} \quad (9)$$

Тоді телеметричний простір періоду (t) визначається як:

$$X(t) = \{f_i \in F | A(f_i, t) = 1\} \quad (10)$$

За характером змінності ознаки поділяються на два класи. До статичних або умовно-статичних ознак належать конфігураційні, топологічні та архітектурні характеристики, семантика яких зберігається протягом розглянутого інтервалу експлуатації:

$$F_{stat} = \{f_i \in F | \forall t \in T: A(f_i, t) = 1 \wedge sem(f_i, t) = const\}. \quad (11)$$

До динамічних ознак належать метрики, лічильники та службові стани, значення яких змінюються в часі та формують часовий ряд стану системи:

$$F_{dyn} = \{f_i \in F | \exists t_1, t_2 \in T: \psi_i(t_1) \neq \psi_i(t_2)\}. \quad (12)$$

У прикладному наборі ознак, що використовується для опису стану SDS-вузла, можна виділити близько двадцяти базових ознак, з яких орієнтовно тринадцять мають статичний або умовно-статичний характер, а решта належить до динамічних. Таке розділення має не довідковий, а методологічний характер: статичні ознаки задають структурний контекст спостереження, тоді як динамічні ознаки відображають зміну поточного стану системи.

Еволюція телеметричного простору між двома періодами експлуатації t_a та t_b задається відношенням сумісності:

$$R_F \subseteq X(t_a) \times X(t_b) \quad (13)$$

Пара ознак $(f_i^{t_a}, f_j^{t_b}) \in R_F$, якщо існує оператор перетворення Γ_{ij} , що забезпечує відновлення або збереження релевантної для прогнозування семантики:

$$\Gamma_{ij}(f_j^{t_b}) \equiv_{sem} f_i^{t_a} \quad (14)$$

Тут $\equiv_{sem}$ означає семантичну еквівалентність у межах задачі прогнозування відмов, а не обов'язкову тотожність системних назв, форматів або джерел телеметрії. Типи допустимих відношень між ознаками наведено в Таблиці 1. Семантична еквівалентність встановлюється на підставі документованого призначення ознаки, її джерела, області допустимих значень та правила перетворення. Для тотожного відображення і перейменування відповідність має бути однозначною; для нормалізації, агрегації та декомпозиції вона підтверджується детермінованим правилом перетворення з урахуванням точності вихідних даних. Якщо для однієї ознаки існує декілька можливих відображень і жодне з них не може бути однозначно підтверджене документацією або правилом перетворення, така відповідність вважається невизначеною і не використовується для формування інваріантного ядра.

Позначимо множину допустимих типів відношень як:

$$T_{adm} = \{preserve, rename, normalize, aggregate, decompose, semantic\}. \quad (15)$$

Тоді умову збереження інваріантного ядра можна записати як:

$$\forall f_i \in X_{core}, \forall t \in T: \exists f_j \in X(t), \exists \Gamma_{ij}: \Gamma_{ij}(f_j, t) \equiv_{sem} f_i. \quad (16)$$

Умова (16) має бінарний характер: для кожного компонента X_{core} має існувати допустиме відношення сумісності та відповідний оператор Γ , що дозволяє отримати або відновити його узгоджене семантичне подання. Якщо хоча б для одного компонента інваріантного ядра таке відображення відсутнє, умова збереження ядра вважається невиконаною. Отже, перевірюваним критерієм структурної втрати є наявність невідновлюваних компонентів X_{core} для збереження єдиного прогностичного контуру їх кількість має дорівнювати нулю.

Водночас компоненти X_{core} не обов'язково повинні мати ідентичні системні імена або однакові джерела у сирій телеметрії різних періодів. За виконання умови (16) X_{core} слід розглядати не як механічний перетин однакових назв метрик, а як ядро сумісності, що забезпечує збереження єдиного прогностичного контуру.

Таблиця 1. Типи відношень між ознаками в еволюційному телеметричному просторі

Назва відношення	Формалізація / предикат	Оператор Γ	Статус для X_{core}
Збереження	$f^{t_b} \equiv sem f^{t_a}$	Тотожне відображення $\Gamma(f^{t_b}) = I(f^{t_b})$	Допустимо
Перейменування	$name(f^{t_b}) \neq name(f^{t_a}) \wedge sem(f^{t_b}) = sem(f^{t_a}) = sem(f^{t_a})$	Взаємно однозначне відображення ідентифікаторів	Допустимо
Нормалізаційне відображення	$\Gamma(f^{t_b}) = \alpha f^{t_b} + \beta$ або інше узгоджене масштабування	Лінійне або нелінійне приведення шкал до $\Gamma(f^{t_b}) \equiv sem f^{t_b}$	Допустимо
Агрегація	$\Gamma(f^{t_b}) = g(\{f_j^{t_b}\}_{j \in I})$	Інтегрування вузлових метрик до узагальненого виду f^{t_a}	Допустимо
Декомпозиція	$f^{t_a} \rightarrow \{f_1^{t_b}, f_2^{t_b}, \dots, f_k^{t_b}\}$	Проекція або лінійна комбінація нових ознак для відновлення f^{t_a}	Допустимо
Семантична відповідність	$P(f^{t_a}) \Leftrightarrow P(f^{t_b})$	Документоване або логічно обґрунтоване зіставлення через суміжні метрики	Допустимо через проміжний шар
Поява нової ознаки	$f^{t_b} \in X(t_b) \wedge \nexists f^{t_a} \in X(t_a): (f^{t_a}, f^{t_b}) \in R_F$	Додавання нового компонента	Належить до $X_{ext}(t)$
Втрата сумісності	$f^{t_a} \in X(t_a) \wedge \nexists f^{t_b} \in X(t_b): (f^{t_a}, f^{t_b}) \in R_F$	Відсутній	Недопустимо для ознак ядра

Джерело: розроблено авторами

Узагальнену UML-подібну структуру введених сутностей наведено на Рис. 1.

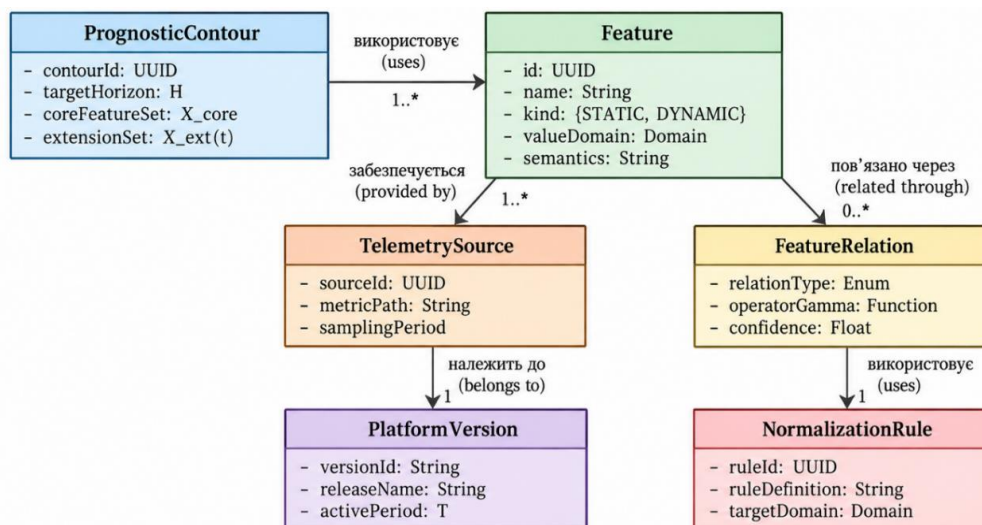


Рис. 1. UML-подібне подання реляційно-динамічної моделі ознак, джерел телеметрії, версій платформи, правил нормалізації та відношень сумісності в межах прогностичного контуру

Джерело: розроблено авторами

Результати. На основі формалізації, наведеної в методологічному розділі цієї статті, отримано три методологічні результати, що разом утворюють схему збереження єдиного прогностичного контуру при еволюції телеметричного простору SDS. Їх практичну інтерпретацію далі показано на прикладі еволюції телеметричного простору реального Serph-кластера.

Першим результатом є визначення ролі інваріантного ядра як шару історичної наступності. За наявності X_{core} дані, отримані на різних етапах життєвого циклу SDS-системи, можуть залишатися методологічно сумісними навіть тоді, коли склад додаткових телеметричних ознак змінюється. Це має особливе значення для задач прогнозування відмов, де накопичення достатньої кількості критичних подій є складним через їх рідкість, дисбаланс класів та тривалий цикл експлуатації інфраструктури [8], [10].

Другим результатом є відокремлення еволюції телеметрії від фрагментації прогностичних моделей. Поява нових службових станів, журнальних шаблонів, діагностичних метрик, рівнів зберігання або параметрів мережевої взаємодії не потребує автоматичного формування окремого прогностичного контуру. У межах запропонованої схеми такі зміни інтерпретуються як поява нових компонентів $X_{ext}(t)$, які уточнюють поточний опис системи, але не руйнують базову постановку оцінювання ризику, сформовану в попередній роботі [12].

Третім результатом є розмежування ознак за їх функціональною роллю в прогностичному контурі. Інваріантне ядро виконує роль стабільного шару сумісності, тоді як змінні розширення виконують роль адаптивного шару деталізації. Це розмежування наведено в Таблиці 2.

Таблиця 2. Функціональне розмежування інваріантного ядра та змінних розширень

Критерій	X_{core}	$X_{ext}(t)$
Основне призначення	Збереження історичної сумісності	Уточнення поточного стану системи
Роль у прогностичному контурі	Опорний шар	Адаптивний шар
Залежність від версії SDS-платформи	Мінімальна	Допустима
Наявність в історичних даних	Необхідна або відтворювана	Необов'язкова
Зміна в часі	Небажана	Очікувана
Методологічний ризик	Втрата сумісності при руйнуванні ядра	Перевантаження моделі нестабільними ознаками

Джерело: розроблено авторами

Зіставлення функціональних ролей у Таблиці 2 та формалізація відношень у Таблиці 1 показують, що цінність X_{core} визначається не кількістю ознак і не буквальним збігом назв лічильників або джерел телеметрії, а стабільністю експлуатаційної семантики, інтерпретованістю та здатністю підтримувати зв'язок між різними періодами експлуатації. Історична наступність може забезпечуватися сукупністю відношень перейменування, нормалізаційного відображення, агрегації, декомпозиції та семантичної відповідності. Така логіка узгоджується із загальною проблематикою зіставлення схем даних [16], але в цій роботі застосовується до часової еволюції SDS-телеметрії. Натомість $X_{ext}(t)$ може мати змінну розмірність, залежати від поточного стану розвитку інфраструктури та включати нові ознаки, які уточнюють опис системи без руйнування базового прогностичного контуру.

Зокрема, перехід від одних схем подання лічильників продуктивності до інших може бути формалізований як семантична відповідність із нормалізаційним або декомпозиційним відображенням, що дозволяє не відкидати раніше накопичені діагностичні дані.

Рис. 2 ілюструє принципову схему запропонованого підходу. У різні періоди експлуатації (t_1), (t_2), (t_3) склад змінних ознак $X_{ext}(t_i)$ може відрізнятися. Водночас наявність спільного інваріантного ядра X_{core} дозволяє формувати узгоджені простори $X(t_i)$, які не створюють окремих незалежних моделей, а залишаються частинами одного прогностичного контуру.

Аналіз офіційних приміток до випусків Serph дозволяє простежити еволюцію телеметричного простору протягом восьми мажорних версій – від Kraken (v11) до Reef (v18). Базові ознаки – стани OSD, деградація PG-груп, затримки операцій – присутні в усіх версіях і є кандидатами до X_{core} . Натомість кожне мажорне оновлення змінювало склад X_{ext} : у Luminous (v12) BlueStore став стандартним бекендом для нових OSD і додав власний пакет діагностичних лічильників [13]; показовим є приклад стабільної гілки Pacific (v16).

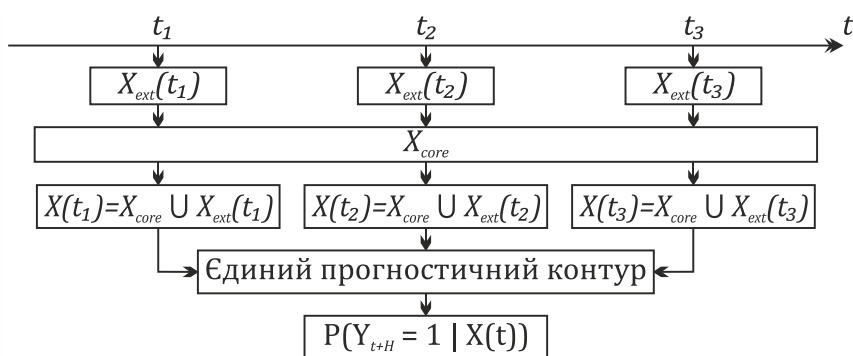


Рис. 2. Збереження єдиного прогностичного контуру при еволюції телеметричного простору SDS

Джерело: розроблено авторами

Зміну пріоритетів лічильників BlueStore, яка в основній гілці розкривала додатково 34 лічильники на OSD, у бекпорті було свідомо обмежено двома лічильниками, щоб не збільшувати навантаження телеметрії стабільного релізу. Це демонструє, що склад доступного телеметричного простору визначається не лише розвитком платформи, а й політикою супроводу конкретної гілки [14]. У Reef (v18) команди *perf dump* та *perf schema* оголошено застарілими на користь *counter dump* і *counter schema* з підтримкою *labeled counters* [15].

Практичний досвід експлуатації реального багатовузлового Серph-кластера з 2017 року по сьогодні ілюструє цю закономірність на практиці: у межах досліджуваного телеметричного архіву простір ознак еволюціонував від Kraken (v11) до Reef (v18), тоді як накопичені історичні спостереження залишалися цінними для побудови та підтримки прогностичної логіки. Поява нових ознак мала два сценарії: або повне переписування прогностичної логіки під нову конфігурацію телеметрії, або часткове розширення моделі шляхом додавання нових змінних без зміни базової постановки задачі. Запропонований підхід формалізує саме другий сценарій як методологічно пріоритетний - за умови збереження X_{core} як шару сумісності. Рис. 3 ілюструє цю еволюцію.

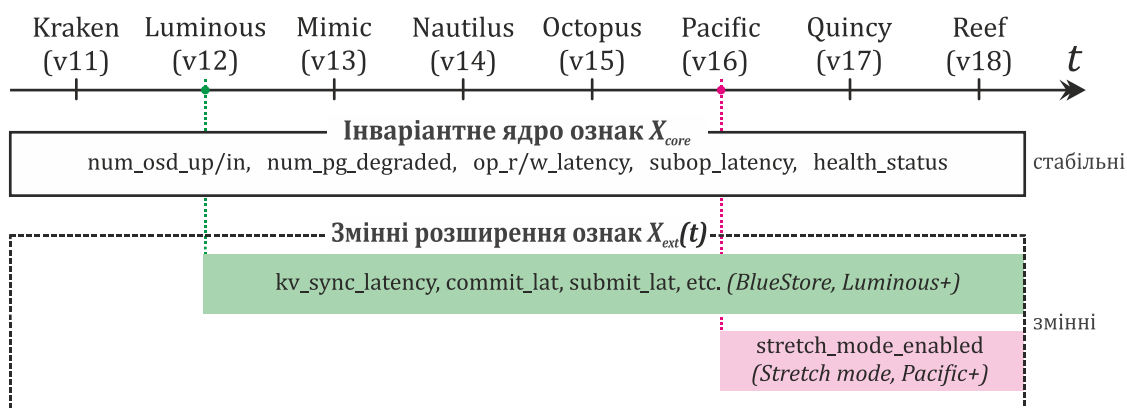


Рис. 3. Еволюція телеметричного простору Серph-кластера

Джерело: розроблено авторами

Наведені приклади мають ілюстративний характер і не претендують на повну каталогізацію міграції телеметричних ознак між усіма версіями Серph; метою їх використання є демонстрація застосування введених типів відношень сумісності до реальної еволюції телеметричного простору.

Водночас ця схема має чітку межу застосовності. Збереження єдиного прогностичного контуру можливе лише за умови, що інваріантне ядро залишається доступним або може бути відтворене в нових конфігураціях SDS-системи. Якщо така умова не виконується, побудова окремої моделі стає не методологічною помилкою, а вимушеним рішенням.

Для методологічного зіставлення підходу з побудовою окремих моделей та запропонованої схеми єдиного прогностичного контуру наведено в Таблицю 3.

Таблиця 3. Методологічне зіставлення окремих моделей і єдиного прогностичного контуру

Аспект	Окремі моделі для різних конфігурацій	Єдиний прогностичний контур
Поява нових ознак	Потребує нової або окремої моделі	Розглядається як розширення $X_{ext}(t)$
Використання історичних даних	Частково втрачає сумісність	Зберігається через X_{core}
Кількість прогностичних контурів	Декілька незалежних	Один еволюційний контур
Ризик фрагментації знань	Високий	Знижений
Зв'язок між періодами експлуатації	Обмежений	Підтримується через інваріантне ядро
Методологічний акцент	Модель під конфігурацію	Контур під життєвий цикл SDS

Джерело: розроблено авторами

Запропонована схема дозволяє зберігати єдиний прогностичний контур протягом життєвого циклу SDS-інфраструктури за умови збереження або відтворення інваріантного ядра ознак. Практична цінність підходу полягає у збереженні накопиченої історії спостережень як інформаційної основи для подальшого прогнозування, а не у її відкиданні при кожній зміні телеметричного простору.

Висновки. У статті формалізовано методологічний підхід до збереження єдиного прогностичного контуру в умовах еволюції телеметричного простору розподілених систем зберігання даних. Ключовим результатом є розмежування простору ознак на інваріантне ядро X_{core} та змінні розширення $X_{ext}(t)$, що дозволяє інтерпретувати зміну складу телеметрії як природне розширення поточного опису системи за умови збереження або відтворюваності інваріантного ядра, а не як безумовну підставу для побудови нової прогностичної моделі з нуля.

Розроблено формальний апарат опису телеметричного простору на основі множинного подання ознак, предиката доступності та типізованих відношень сумісності у часі. Ознаки розділено на статичні або умовно-статичні та динамічні, а для еволюції телеметрії визначено допустимі відношення: збереження, перейменування, нормалізаційне відображення, агрегація, декомпозиція та семантична відповідність. Показано, що збереження інваріантного ядра X_{core} можливе не лише за повної ідентичності метрик, а й за наявності операторів перетворення Γ , які відновлюють семантику опорного шару ознак і дозволяють зберігати сумісність історичних спостережень.

Встановлено, що збереження єдиного прогностичного контуру можливе за умови доступності або відтворюваності компонентів інваріантного ядра на всіх етапах життєвого циклу системи. Нові діагностичні метрики, службові стани або журнальні шаблони, які з'являються в процесі розвитку платформи, можуть бути включені до моделі як змінні розширення без порушення сумісності з історичними даними. Якщо хоча б для одного обов'язкового компонента інваріантного ядра умова сумісності не виконується і його неможливо відновити допустимим оператором Γ , умова збереження єдиного прогностичного контуру вважається порушеною, а побудова окремої або адаптованої моделі стає вимушеним рішенням.

Аналіз еволюції телеметричного простору в межах восьми мажорних версій Serp показав, що його склад визначається не лише розвитком платформи, а й політикою супроводу конкретної гілки. Підхід проілюстровано на матеріалі реального багатовузлового Serp-кластера, що експлуатується з 2017 року по сьогоднішній день та в межах досліджуваного періоду пройшов еволюцію від Kraken v11 до Reef v18. Наявність накопиченого телеметричного архіву та навченої моделі прогнозування відмов формує практичну основу для подальшої експериментальної апробації запропонованої схеми. Запропонований підхід забезпечує методологічну наступність між роботами серії та створює основу для довготривалої експлуатації прогностичної системи без фрагментації накопичених

телеметричних даних і без повного переосмислення прогностичної логіки при кожній зміні конфігурації інфраструктури.

Конфлікт інтересів. Автори заявляють про відсутність конфлікту інтересів.

Фінансування. Дослідження виконано без зовнішнього фінансування.

Доступність даних. Дослідження має теоретико-методологічний характер і спирається на досвід експлуатації реального багатовузлового Ceph-кластера з 2017 року. Первинна телеметрія та системні журнали виробничого кластера містять експлуатаційні дані, що не призначені для відкритого поширення, тому у відкритому доступі не публікуються. У статті наведено узагальнені приклади еволюції телеметричного простору, формалізовані типи відношень між ознаками та умови їх сумісності, що забезпечує відтворення методологічної логіки запропонованого підходу без доступу до первинного архіву. Для перевірки конкретних змін Ceph додатково використано відкриту документацію платформи.

Використання засобів штучного інтелекту. Засоби штучного інтелекту використовувалися для мовно-стилістичного редагування тексту (Copilot, Gemini, ChatGPT). Усі графічні матеріали, наукові положення, результати та висновки сформовані авторами самостійно.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Weil, S. A., Brandt, S. A., Miller, E. L., Long, D. D. E. & Maltzahn, C. “Ceph: A scalable, high-performance distributed file system”. *Proceedings of the 7th USENIX Symposium on Operating Systems Design and Implementation (OSDI '06)*. Seattle, WA, USA: USENIX Association. 2006. p. 307–320. – Available from: <https://www.usenix.org/conference/osdi-06/ceph-scalable-high-performance-distributed-file-system>. – [Accessed: 28.05.2026].
2. “Ceph documentation. Placement group states”. *Ceph Documentation*. – Available from: <https://docs.ceph.com/en/latest/rados/operations/pg-states/>. – [Accessed: 27.05.2026].
3. “Ceph documentation. OSD config reference”. *Ceph Documentation*. – Available from: <https://docs.ceph.com/en/latest/rados/configuration/osd-config-ref/>. – [Accessed: 27.05.2026].
4. “Ceph documentation. Health checks”. *Ceph Documentation*. – Available from: <https://docs.ceph.com/en/latest/rados/operations/health-checks/>. – [Accessed: 27.05.2026].
5. Soldani, J. & Brogi, A. “Anomaly detection and failure root cause analysis in (micro)service-based cloud applications: A survey”. *ACM Computing Surveys*. 2023; 55 (3): 59. DOI: <https://doi.org/10.1145/3501297>.
6. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. & Bouchachia, A. “A survey on concept drift adaptation”. *ACM Computing Surveys*. 2014; 46 (4): 44, <https://www.scopus.com/pages/publications/84901228061>. DOI: <https://doi.org/10.1145/2523813>.
7. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J. & Zhang, G. “Learning under concept drift: A review”. *IEEE Transactions on Knowledge and Data Engineering*. 2019; 31 (12): 2346–2363. DOI: <https://doi.org/10.1109/TKDE.2018.2876857>.
8. Hoens, T. R., Polikar, R. & Chawla, N. V. “Learning from streaming data with concept drift and imbalance: An overview”. *Progress in Artificial Intelligence*. 2012; 1 (1): 89–101, <https://www.scopus.com/pages/publications/85011285088>. DOI: <https://doi.org/10.1007/s13748-011-0008-0>
9. Hou, B.-J., Zhang, L. & Zhou, Z.-H. “Learning with feature evolvable streams”. *IEEE Transactions on Knowledge and Data Engineering*. 2021; 33 (6): 2602–2615. DOI: <https://doi.org/10.1109/TKDE.2019.2954090>.
10. Lu, S., et al. “Making disk failure predictions SMARTer!”. *18th USENIX Conference on File and Storage Technologies (FAST '20)*. 2020. p. 151–167. – Available from: <https://www.usenix.org/system/files/fast20-lu.pdf>. – [Accessed: 28.05.2026].
11. Кравцов, С. В. «Нормалізація гетерогенної телеметрії для побудови мультимодального вектора ознак у задачах прогнозування відмов розподілених систем зберігання даних». *Наука і техніка сьогодні*. 2026; 4 (58): 3575–3585. DOI: [https://doi.org/10.52058/2786-6025-2026-4\(58\)-3575-3585](https://doi.org/10.52058/2786-6025-2026-4(58)-3575-3585).
12. Кравцов, С. В. “Методологічні аспекти ризико-орієнтованого прогнозування відмов у розподілених системах зберігання даних”. *Наука і техніка сьогодні*. 2026; 5 (59): 4812–4826. DOI: [https://doi.org/10.52058/2786-6025-2026-5\(59\)-4812-4826](https://doi.org/10.52058/2786-6025-2026-5(59)-4812-4826).
13. “Ceph v12.2.0 Luminous Released”. *Ceph*. 2017. – Available from: <https://ceph.io/releases/v12-2-0-luminous-released>. – [Accessed: 27.05.2026].

14. “Pacific: os/bluestore: update perf counter priorities”. *Ceph, GitHub*. Pull Request #47095. 2022. – Available from: <https://github.com/ceph/ceph/pull/47095>. – [Accessed: 27.05.2026].

15. “Ceph v18.2.0 Reef Released”. *Ceph*. 2023. – Available from: <https://ceph.io/en/news/blog/2023/v18-2-0-reef-released>. – [Accessed: 27.05.2026].

16. Rahm, E. & Bernstein, P. A. “A survey of approaches to automatic schema matching”. *The VLDB Journal*. 2001; 10 (4): 334–350. DOI: <https://doi.org/10.1007/s007780100057>.

Отримано 03.09.2026

Отримано після перегляду 05.10.2026

Прийнято 12.10.2026

DOI: <https://doi.org/10.15276/ict.03.2026.67>

UDC 004.89:005.8:004.4

Maintaining a unified predictive framework under telemetry space evolution in distributed data storage systems

Serhii V. Kravtsov¹⁾

ORCID: <https://orcid.org/0000-0003-2161-5111>; zelenij@gmail.com

Oleksandr M. Martyniuk¹⁾

ORCID: <https://orcid.org/0000-0003-1461-2000>; martyniuk@op.edu.ua; Scopus Author ID: 60238390900

¹⁾ Odesa Polytechnic National University, 1 Shevchenko Ave., Odesa, 65044, Ukraine

ABSTRACT

The article addresses the problem of maintaining a unified predictive framework under the evolution of the telemetry space of software-defined distributed data storage systems. The relevance of the problem stems from the fact that, during long-term cluster operation, the available telemetry changes due to platform upgrades, the introduction of new diagnostic metrics, changes in the composition and semantics of system log messages and service states, and modifications to the hardware configuration. Consequently, constructing a separate model for each new configuration leads to fragmentation of historical data and loss of continuity in failure prediction. The telemetry space is represented as an open structure comprising an invariant feature core and variable extensions. The invariant core acts as a compatibility layer between different periods of system operation, whereas the variable extensions reflect the evolution of the platform, infrastructure, and available telemetry. To describe feature evolution, typed compatibility relations are introduced, including preservation, renaming, normalization mapping, aggregation, decomposition, and semantic correspondence. The scientific novelty lies in interpreting the invariant core not as the mechanical intersection of identical metrics across different operational periods, but as a layer of semantic compatibility in which the operational meaning of features can be preserved through admissible transformations. This approach makes it possible to maintain continuity of the predictive model even when metric names, telemetry sources, and data representation formats change. The proposed approach is illustrated using the evolution of the telemetry space of a Ceph cluster from Kraken to Reef. It is shown that changes in telemetry composition can be treated as controlled extensions of a unified predictive framework provided that the invariant core is preserved or can be reconstructed. If this condition is violated, constructing a separate predictive model becomes a justified necessity.

Keywords: Telemetry space; failure prediction; invariant feature core; telemetry evolution; feature compatibility; predictive framework

ІНФОРМАЦІЯ ПРО АВТОРІВ



Кравцов Сергій Вікторович - аспірант кафедри Комп'ютерних інтелектуальних систем та мереж. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Галузь досліджень: прогнозування відмов розподілених систем зберігання даних

Serhii V. Kravtsov - PhD student, Department of Computer Intelligent Systems and Networks. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ORCID: <https://orcid.org/0000-0003-2161-5111>; zelenij@gmail.com

Research field: Failure prediction in distributed data storage systems



Мартинюк Олександр Миколайович - кандидат технічних наук, доцент, завідувач кафедри Комп'ютерних інтелектуальних систем та мереж. Національний університет «Одеська політехніка», пр. Шевченка, 1. Одеса, 65044, Україна

Галузь досліджень: тестування та діагностика розподілених комп'ютерних систем і мереж, верифікація вбудованих систем

Oleksandr M. Martyniuk - PhD of Engineering Science, Associate Professor, Head of the Department of Computer Intelligent Systems and Networks. Odesa Polytechnic National University, 1, Shevchenko Ave. Odesa, 65044, Ukraine

ORCID: <https://orcid.org/0000-0003-1461-2000>; martyniuk@op.edu.ua. Scopus Author ID: 60238390900

Research field: Testing and diagnostics of distributed computer systems and networks, embedded systems verification