

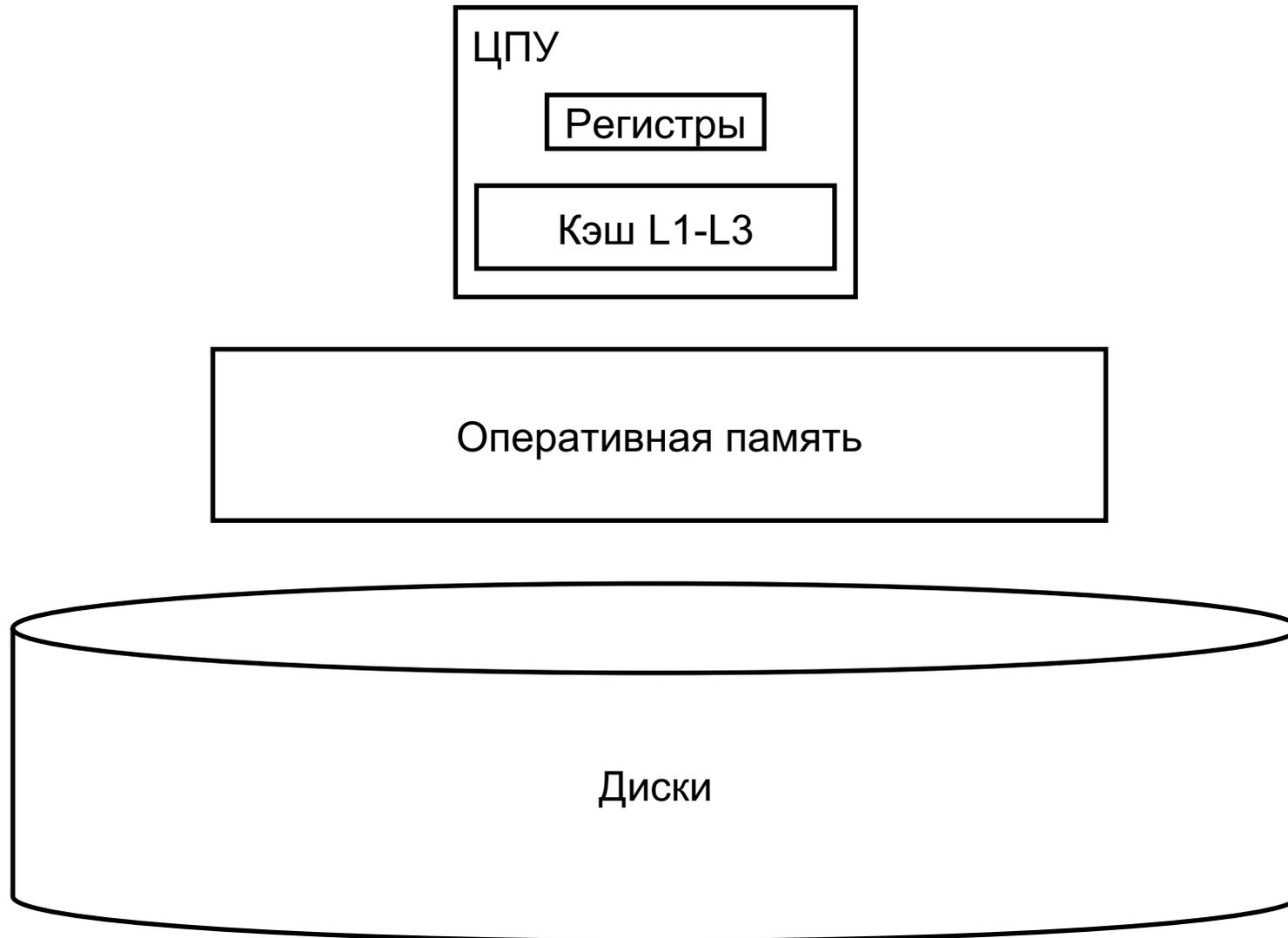
SAP HANA – универсальная среда обработки данных и разработки приложений

Антон Ноздрин

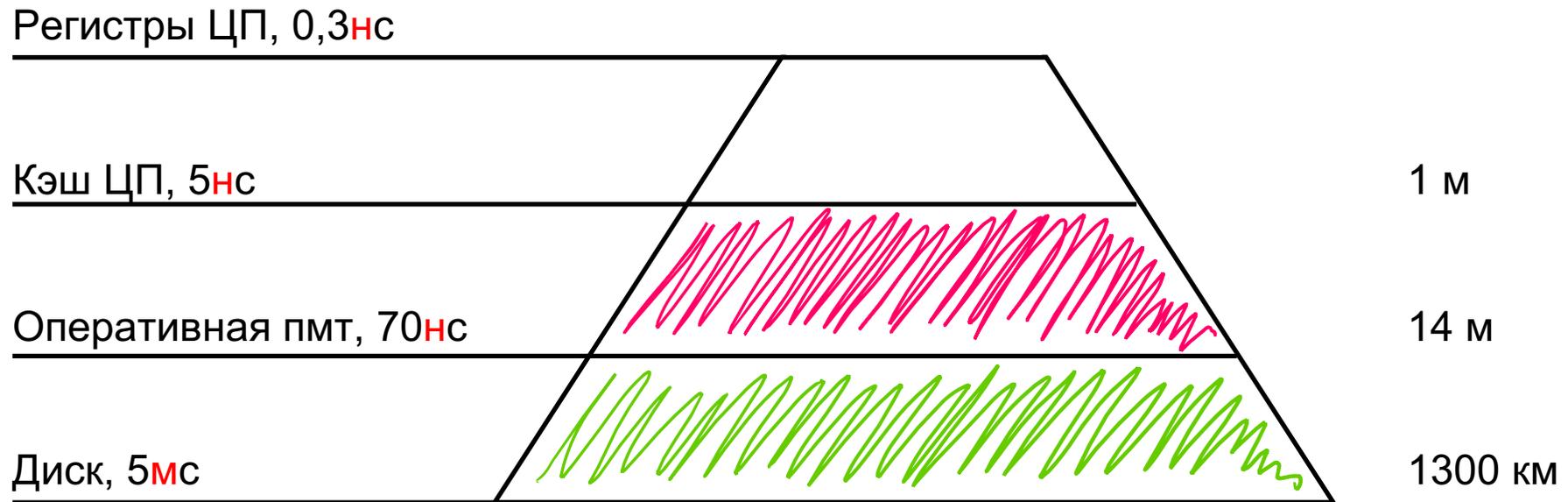
PUBLIC

Тезис 1: HANA – сверхбыстрая СУБД

Как устроена память ЭВМ



Иерархия памяти



Как работает кэш

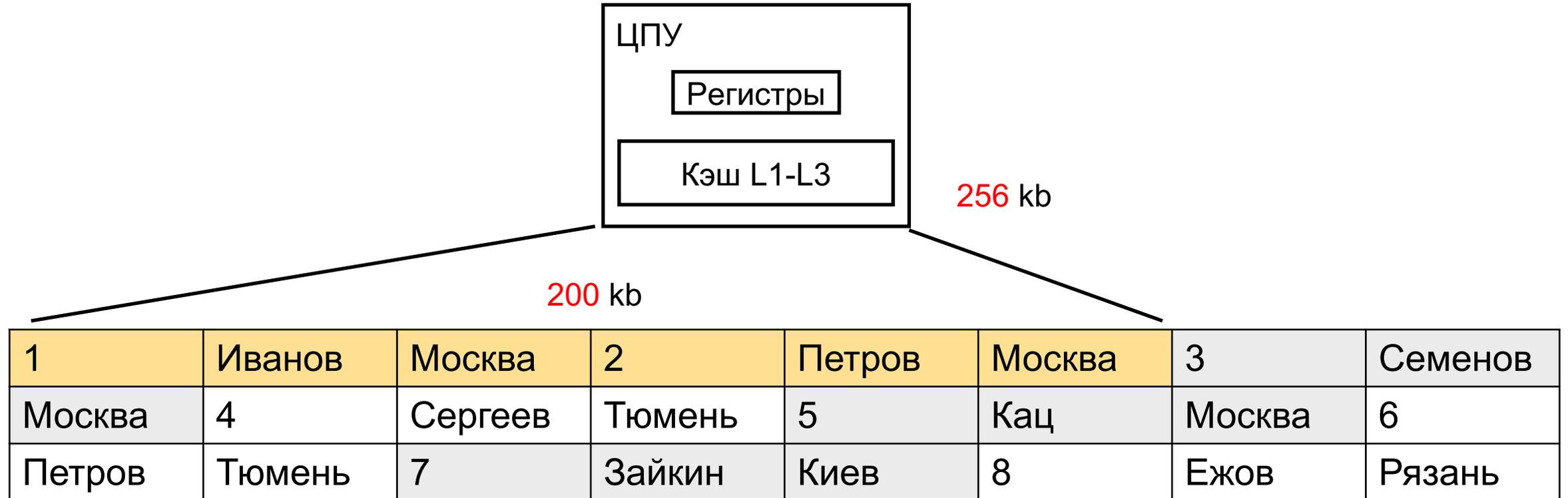
Найти человека с максимальной длиной фамилии



1	Иванов	Москва	2	Петров	Москва	3	Семенов
Москва	4	Сергеев	Тюмень	5	Кац	Москва	6
Петров	Тюмень	7	Зайкин	Киев	8	Ежов	Рязань

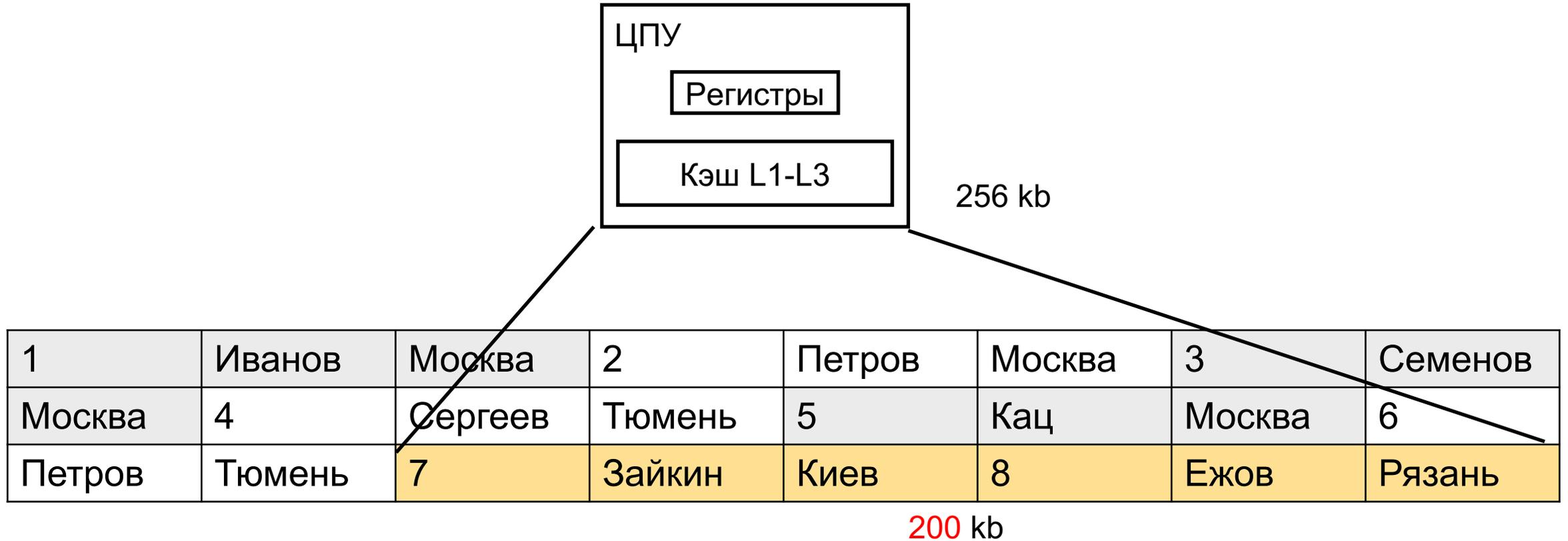
Как работает кэш

Найти человека с максимальной длиной фамилии



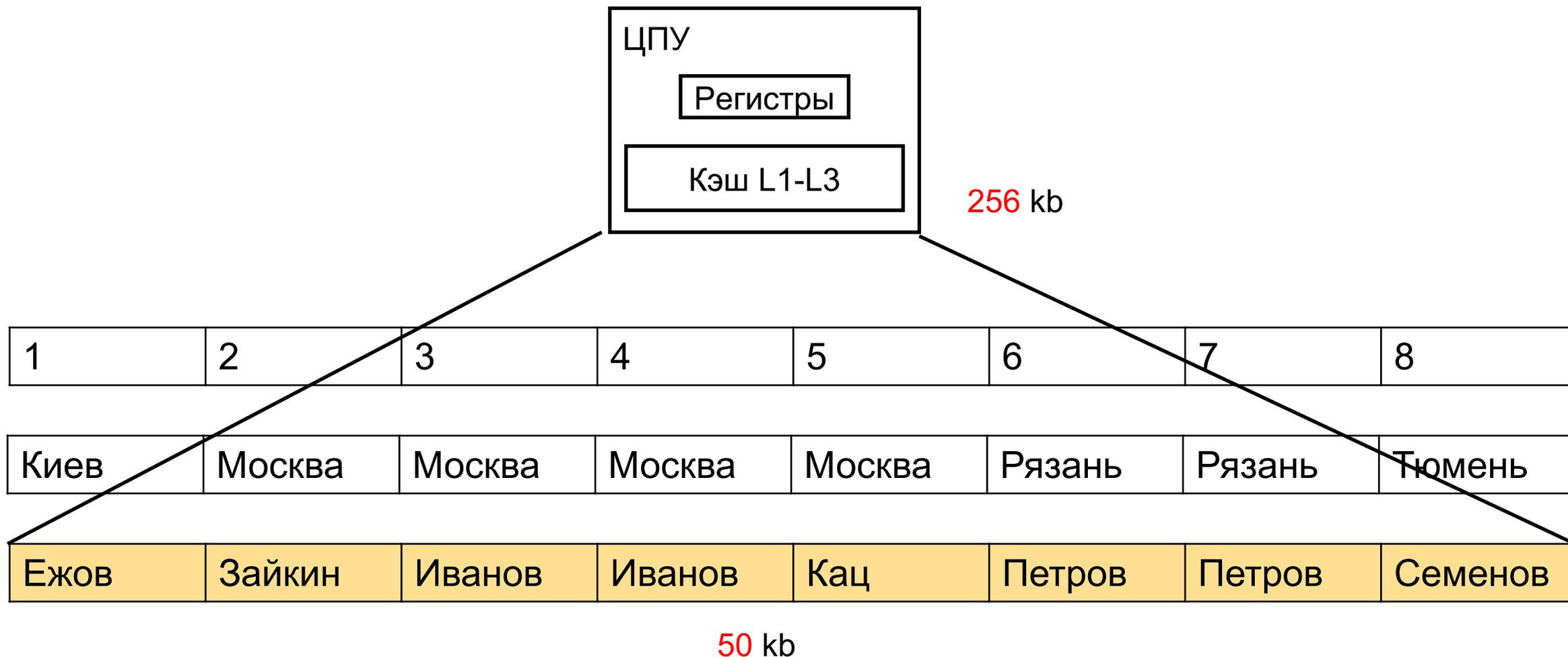
Как работает кэш

Найти человека с максимальной длиной фамилии



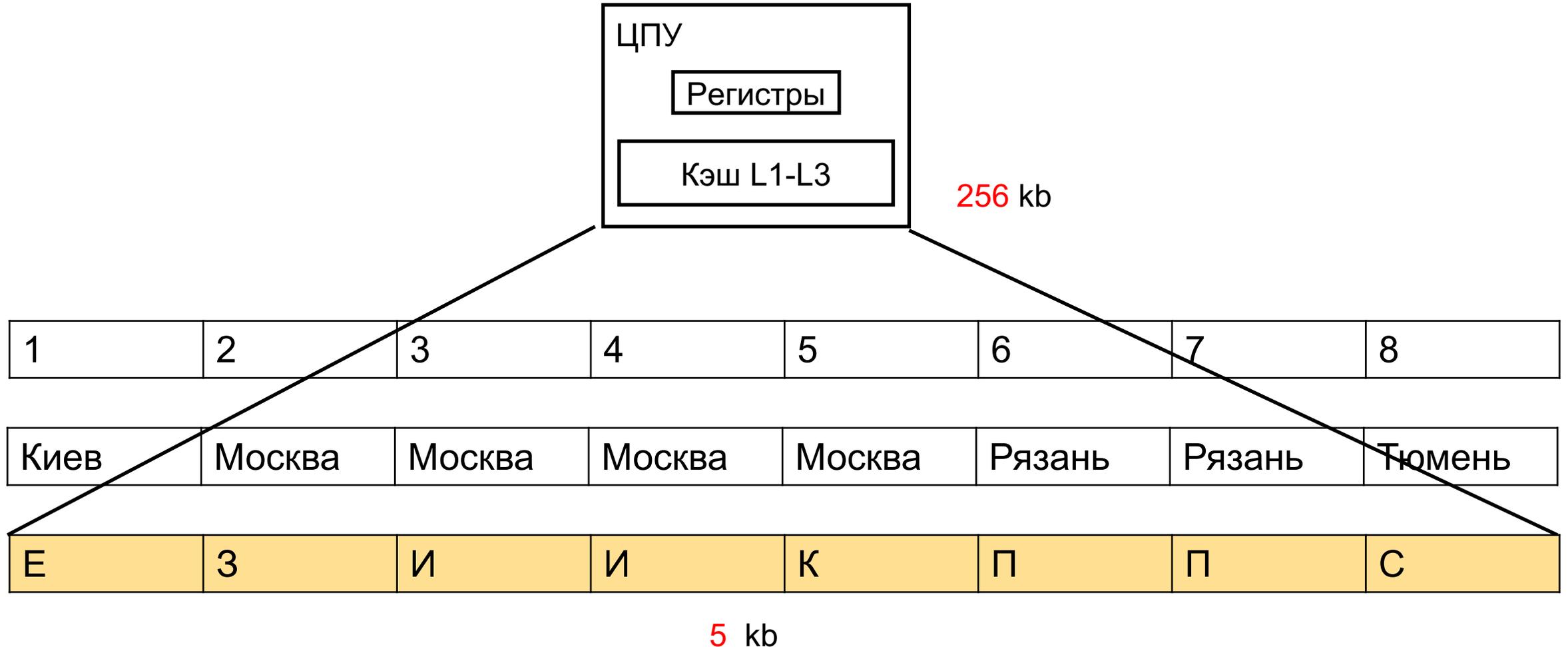
Как работает кэш

Найти человека с максимальной длиной фамилии



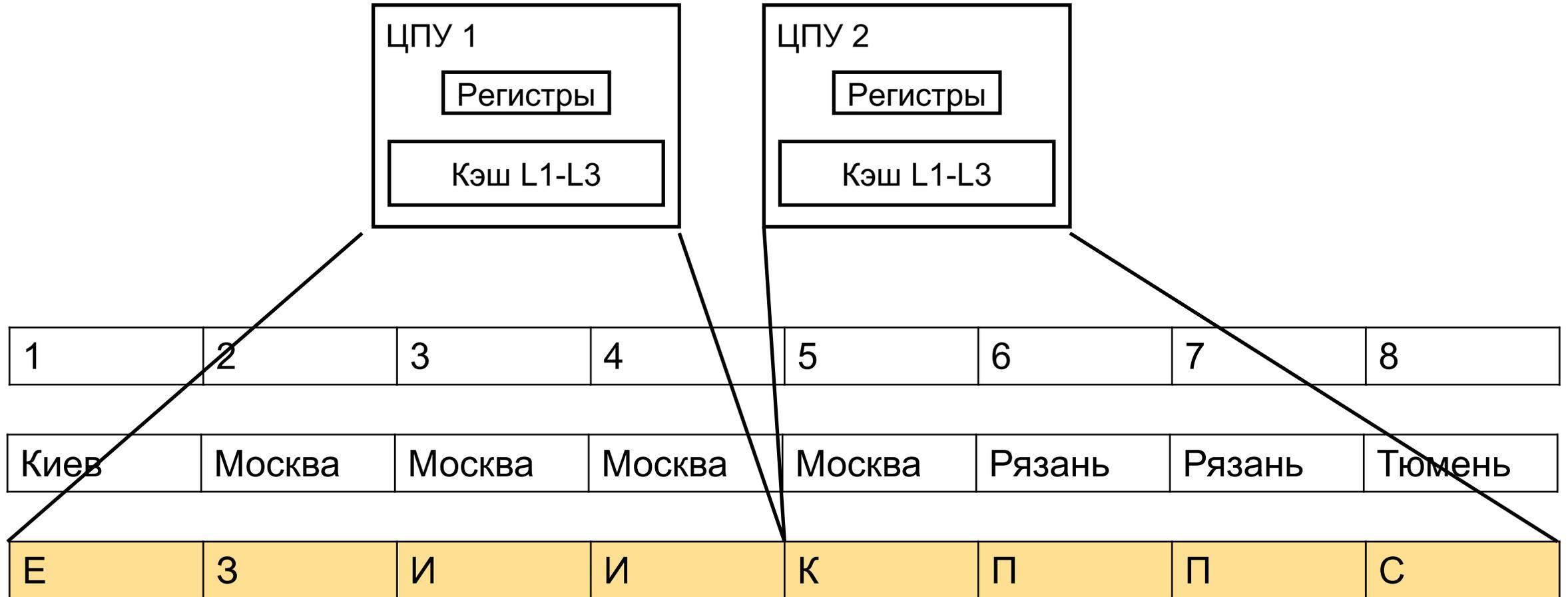
Как работает кэш

Найти человека с максимальной длиной фамилии



Особенности колоночного хранения

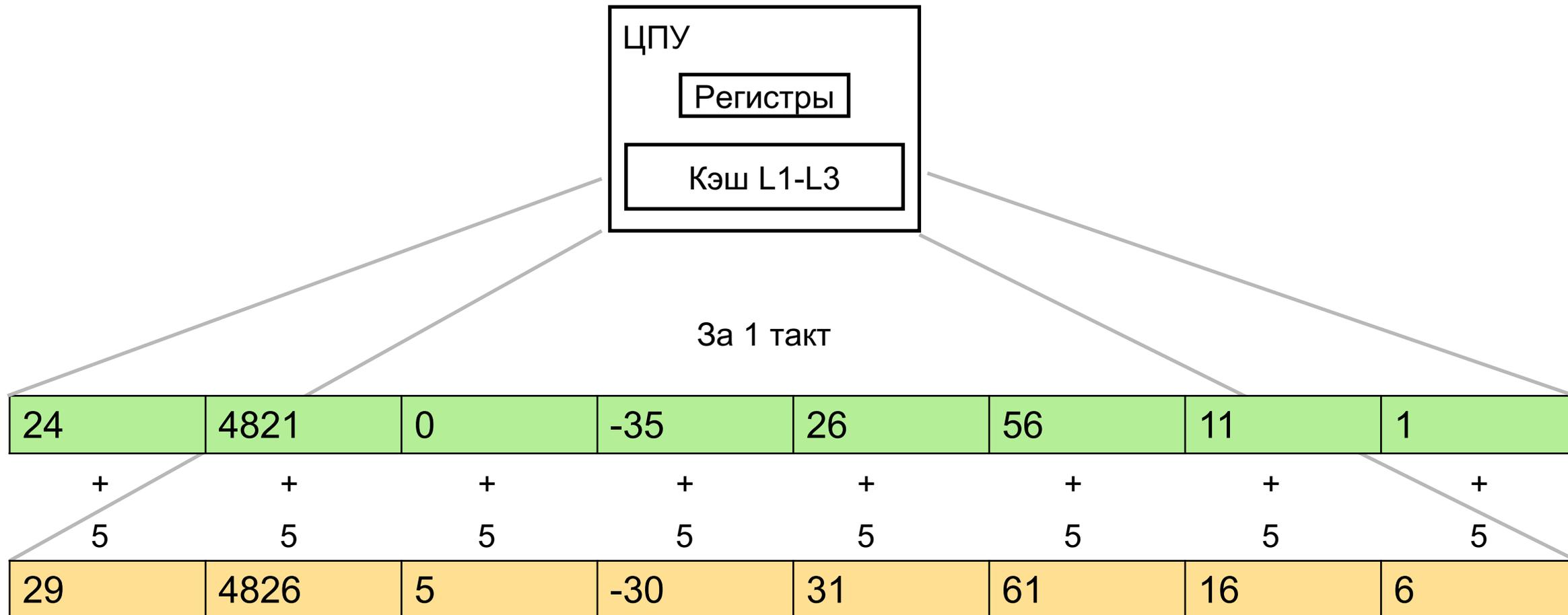
Найти человека с максимальной длиной фамилии



5 kb

Векторные вычисления

`select total_amount * 5`





Data Cafe?

- 3+ years, 250 Billion rows of detail POS
- No aggregates or pre-joins to get in the way of real time

Benefits?

- Queries analyze “ALL” of the data, full table scans, dynamic categorization on the fly, providing any insight possible
- Even during Black Friday load, data at most is 1 hour old (versus nightly on previous technologies)
- Queries come back in 2 seconds meeting Walmart’s “8 second rule”



Scale?

- Scale to thousands of users (1,500 named, 450 concurrent (20 reports/user)
- Tested to 36,000 completed queries per hour (9,000 reports/hour with 4 queries per report)
- Black Friday peak load, 3 million records per minute
- Scale-out with 16 2TB nodes (+HA)

TCO?

- Reduction in ETL complexity, direct population of base schemas
- Massive reduction in DBA tuning costs compared to previous technologies (Teradata and Greenplum)
- More efficient App Dev cycles with HANA XS engine and HTML5
- Data compressed from 46.5 TB to 5.7 TB

Но это было **не главное**

HANA умеет писать также хорошо как читать:

экстремальная скорость чтения и обработки данных (OLAP)

хорошая скорость записи данных (OLTP)

Изначально это и была цель создания SAP HANA

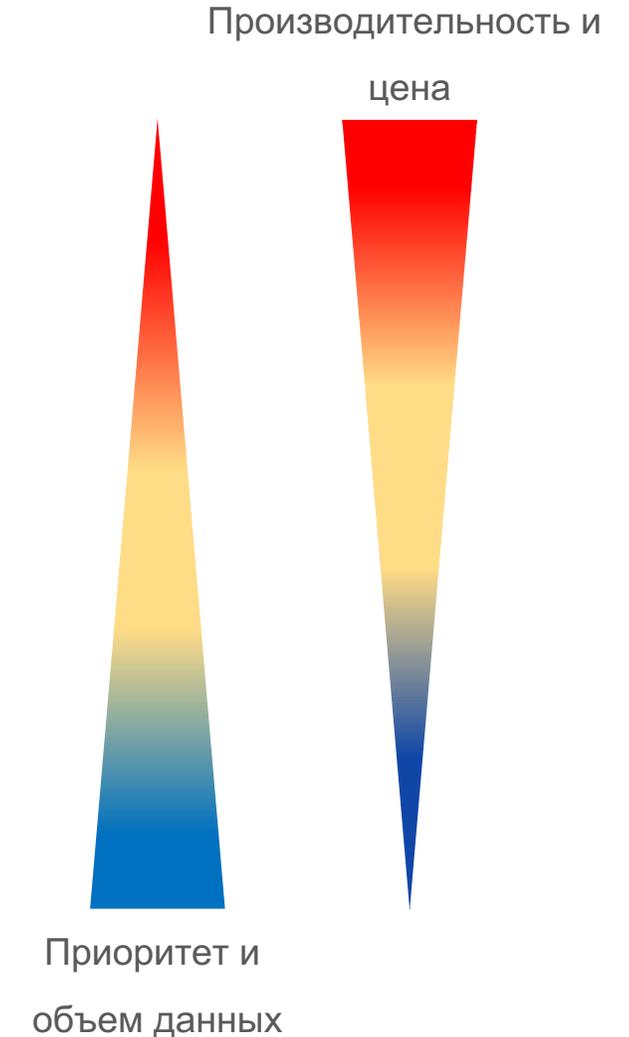
Что это все дает для SAP Digital Core

hana как супербыстрая СУБД

- Отсутствие индексов и агрегатов → упрощение модели данных
- Перенос VDM на уровень HANA → упрощение модели данных, перенос логики с Application Server (ABAP)
- Онлайн расчеты → не нужны «остатки»
- Онлайн расчеты → embedded analytics (не требует хранилища)

Тезис 2: HANA – EIM-платформа

Оптимизация хранения данных SAP BW/4HANA и SAP BW on HANA

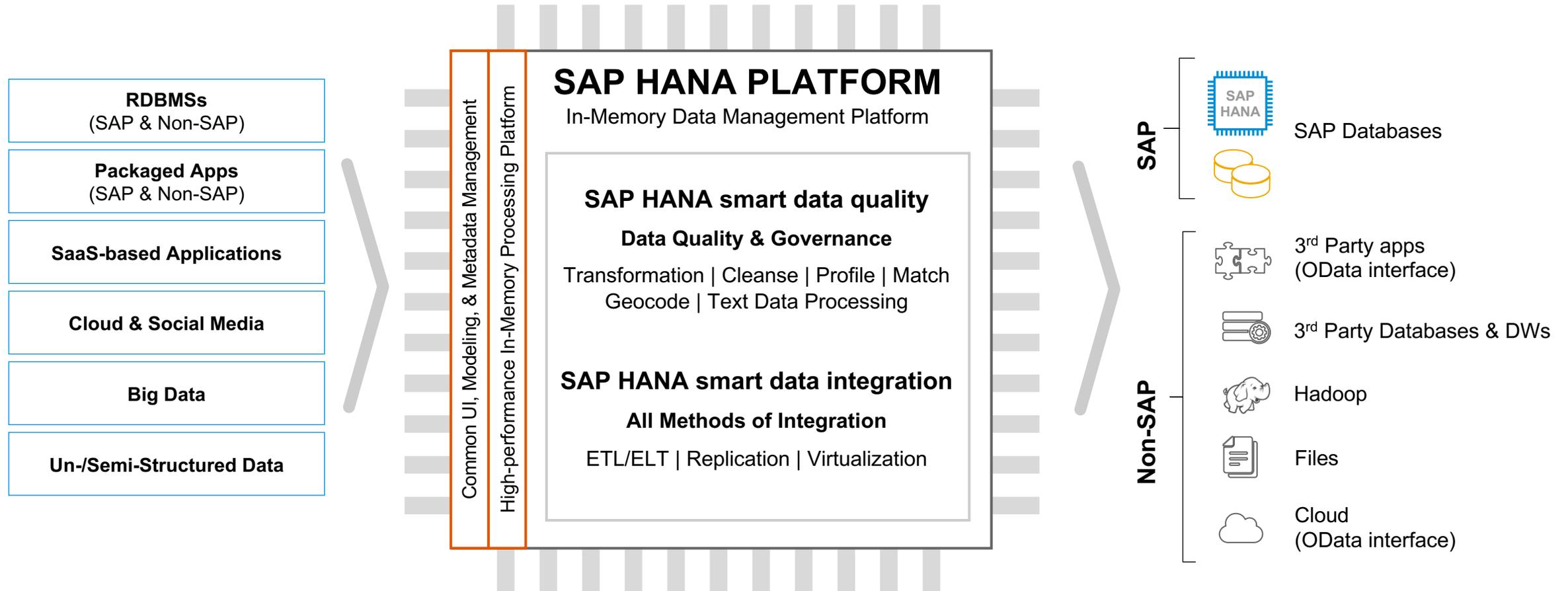


SAP HANA smart data integration and smart data quality

Любой источник

Унифицированные интеграция, доступ,
преобразование и управление качеством

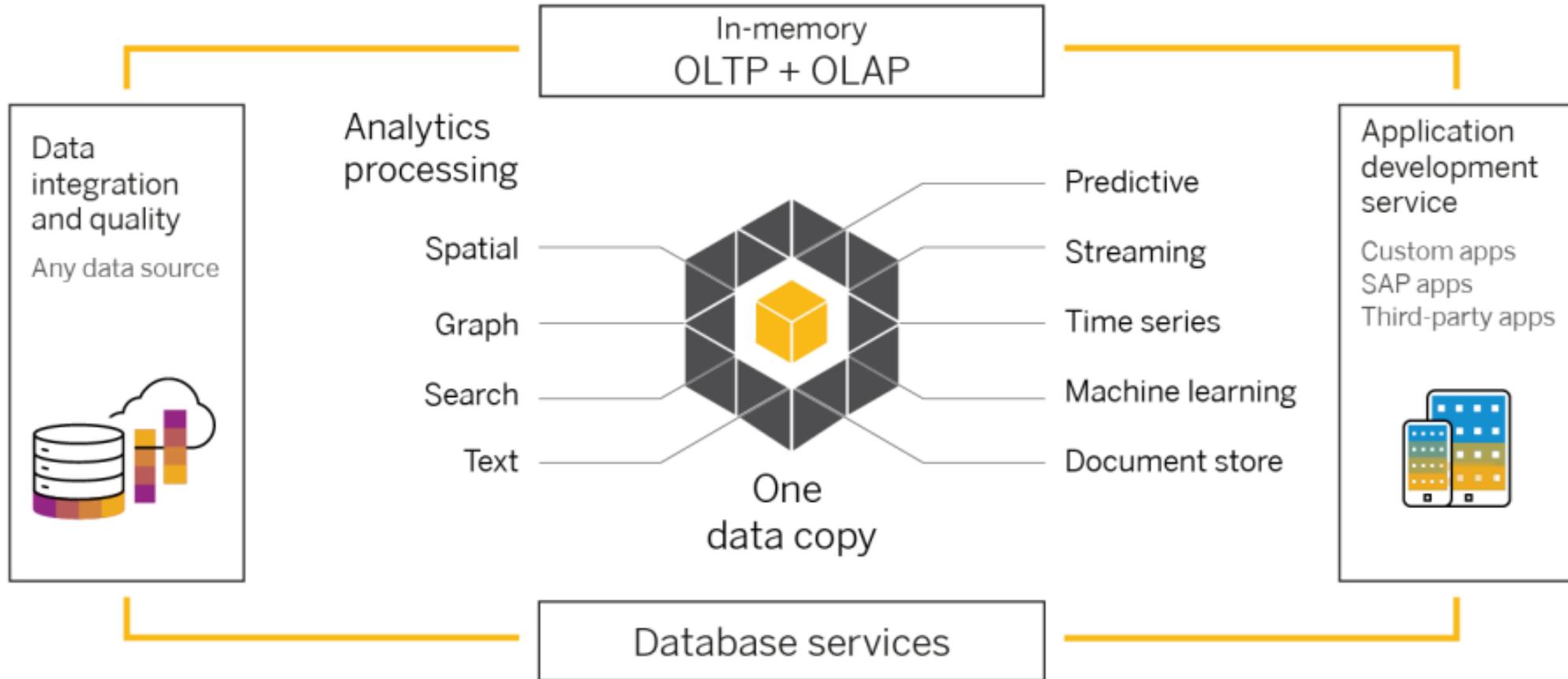
Любая цель



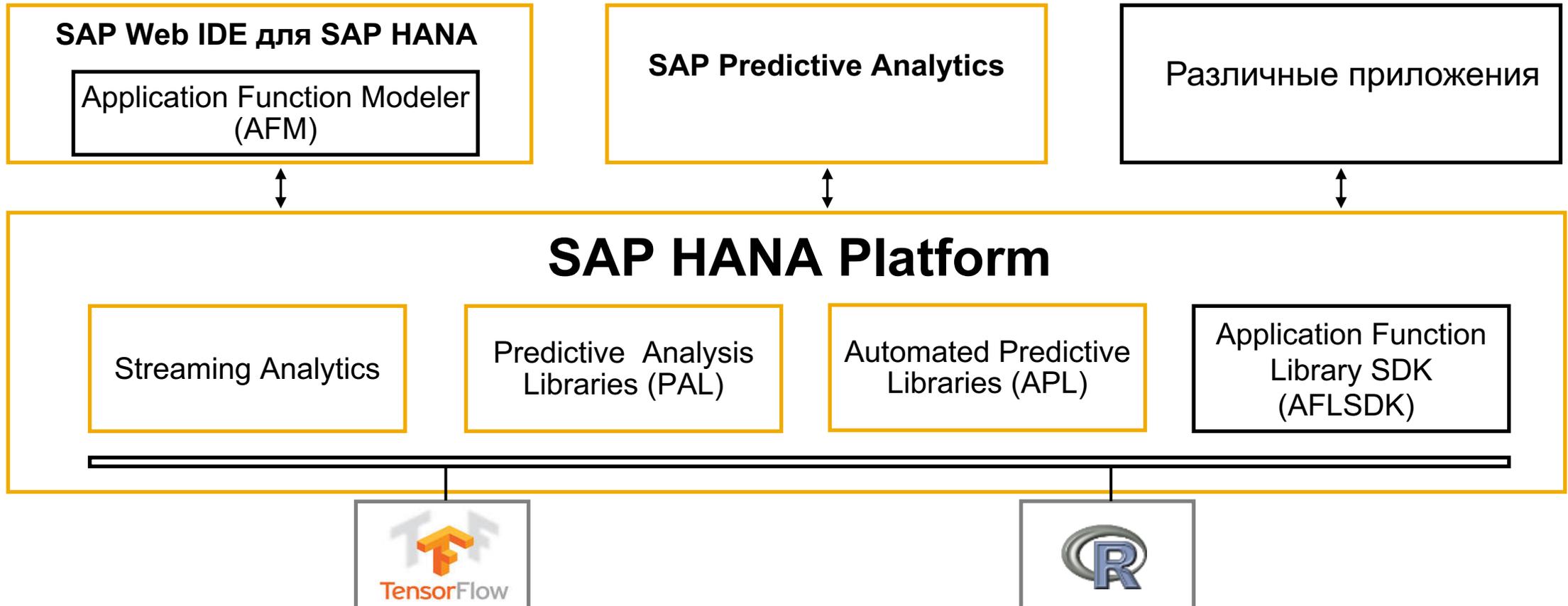
Тезис 3: HANA – набор уникальных сервисов

SAP HANA business data platform

On premise | Cloud | Hybrid



Машинное обучение и SAP HANA



Текстовый анализ



1. Извлечение значения
2. Преобразование в структурированные данные для анализа

Структурированная база данных



После структурирования может быть...

- Интегрирован
- Запрошен
- Анализирован
- Визуализирован
- Зарегистрирован

Извлечение ключевой информации из текстовых источников для распознавания бизнес-данных

Графы

Анализ пути/маршрута

- Построение кратчайшего маршрута с учетом ограничений
- Зависимости
 - Исследование денежного потока для прозрачности затрат или обнаружения мошенничества

Анализ связанности

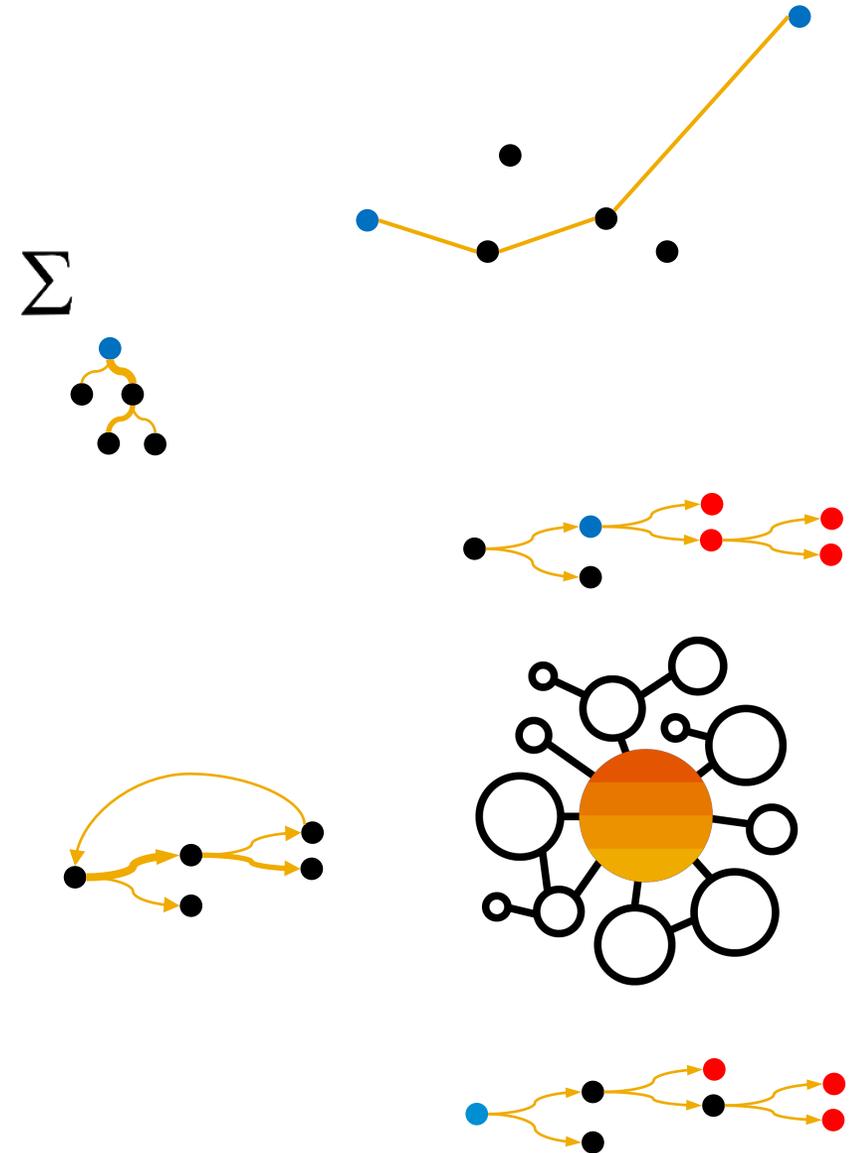
- Трассировка
 - Определить влияние обрыва в сети электроснабжения
- Место использования
 - Какие готовые продукты содержат специфическое сырье / полуфабрикаты
 - На каких клиентов окажет влияние дефектный материал

Анализ взаимосвязи

- Исследование «соседей»
 - Анализ социальных взаимосвязей, например, для определения всех субъектов, связанных с изучаемым субъектом
- Определение сообщества и анализ влияния
 - Выявление семейных групп для рекламы
 - Лидеры сообщества для более эффективной маркетинговой деятельности

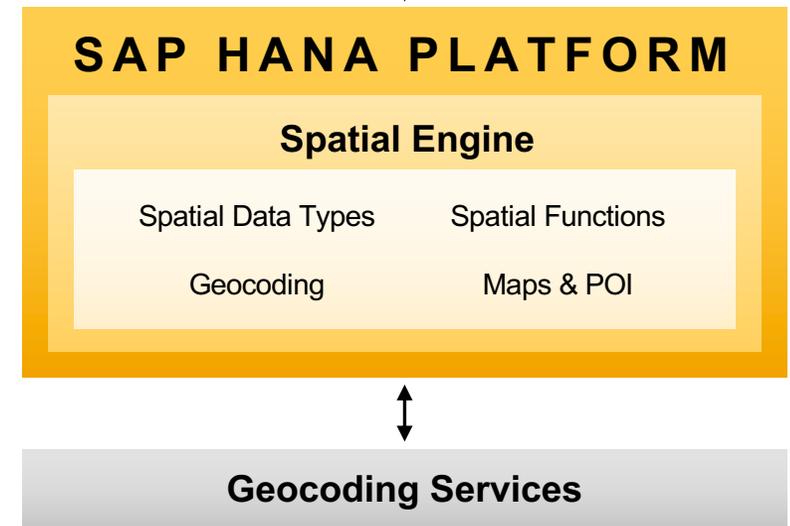
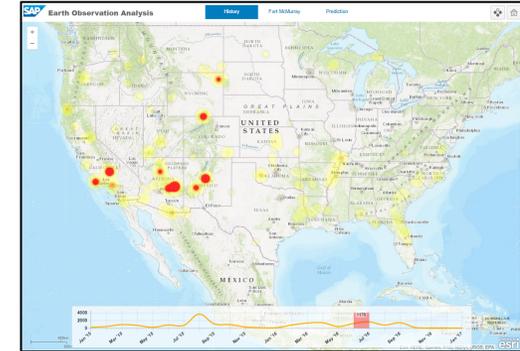
Обнаружение шаблонов

- Поиск шаблонов в больших графах – быстрое создание целевых групп

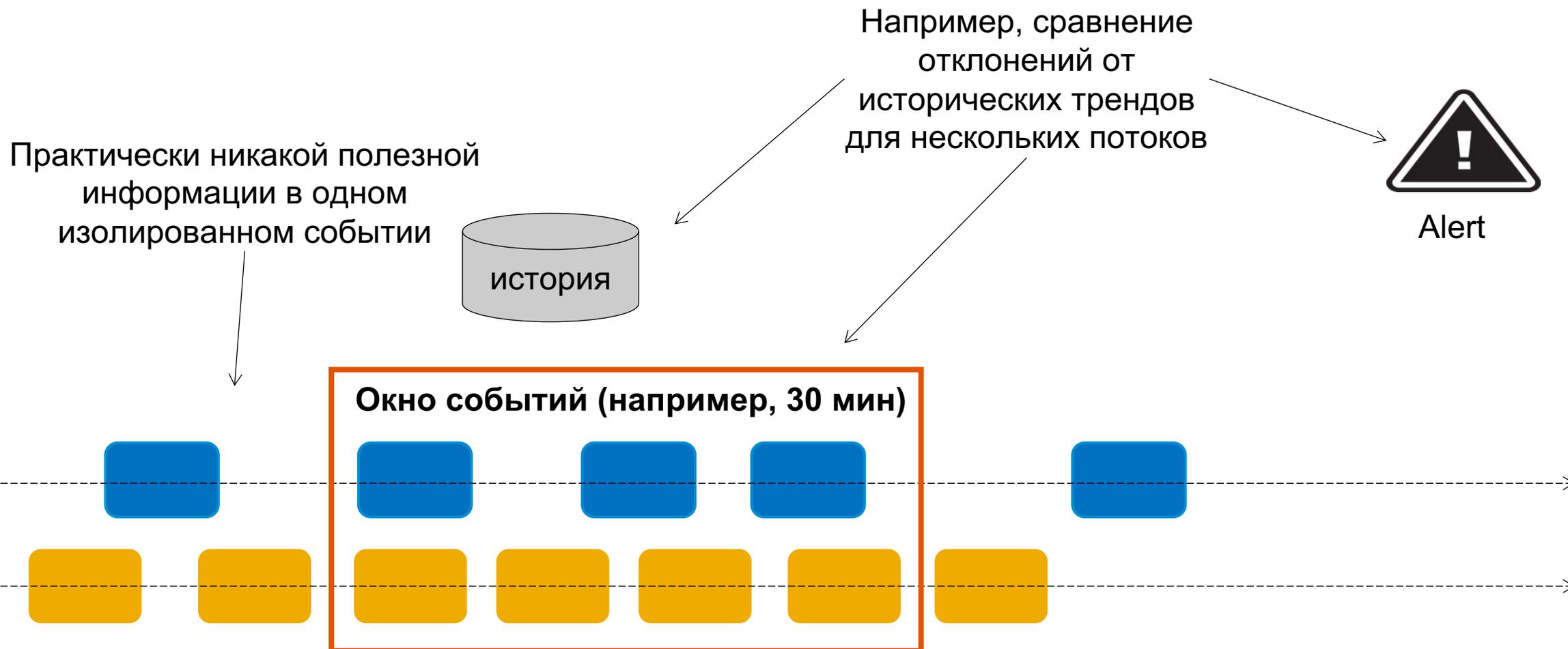


SAP HANA Spatial предоставляет хранение и обработку геоданных

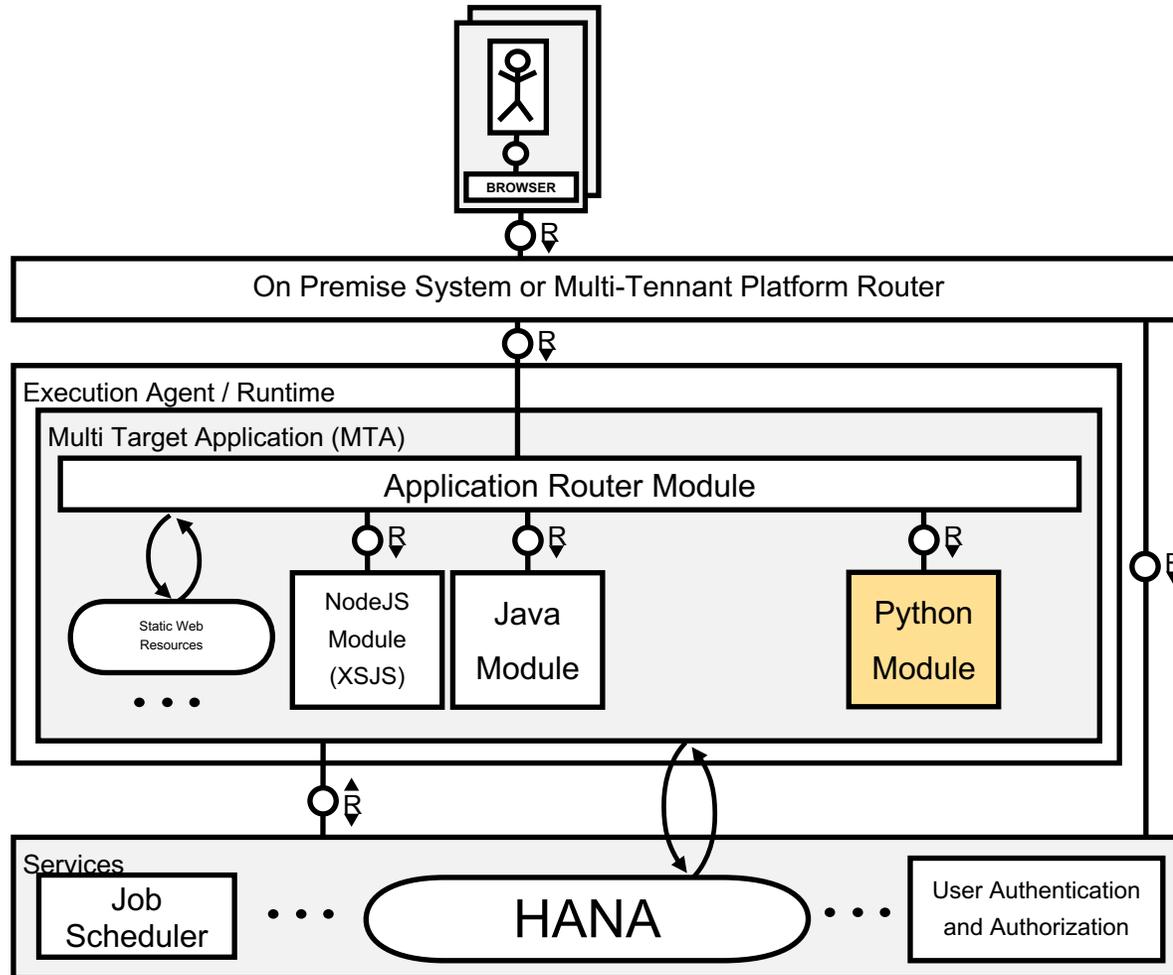
- 2D и 3D векторы – стандартные типы данных
- 50+ гео-функций и алгоритмов
- Прямое и обратное геокодирование
- Гео контент (GAB) и картографические сервисы
- Открытые стандарты (OGC, 1999 SQL/MM)
- SDK для создания своих алгоритмов
- Интеграция с потоковыми данными



Потоковые данные

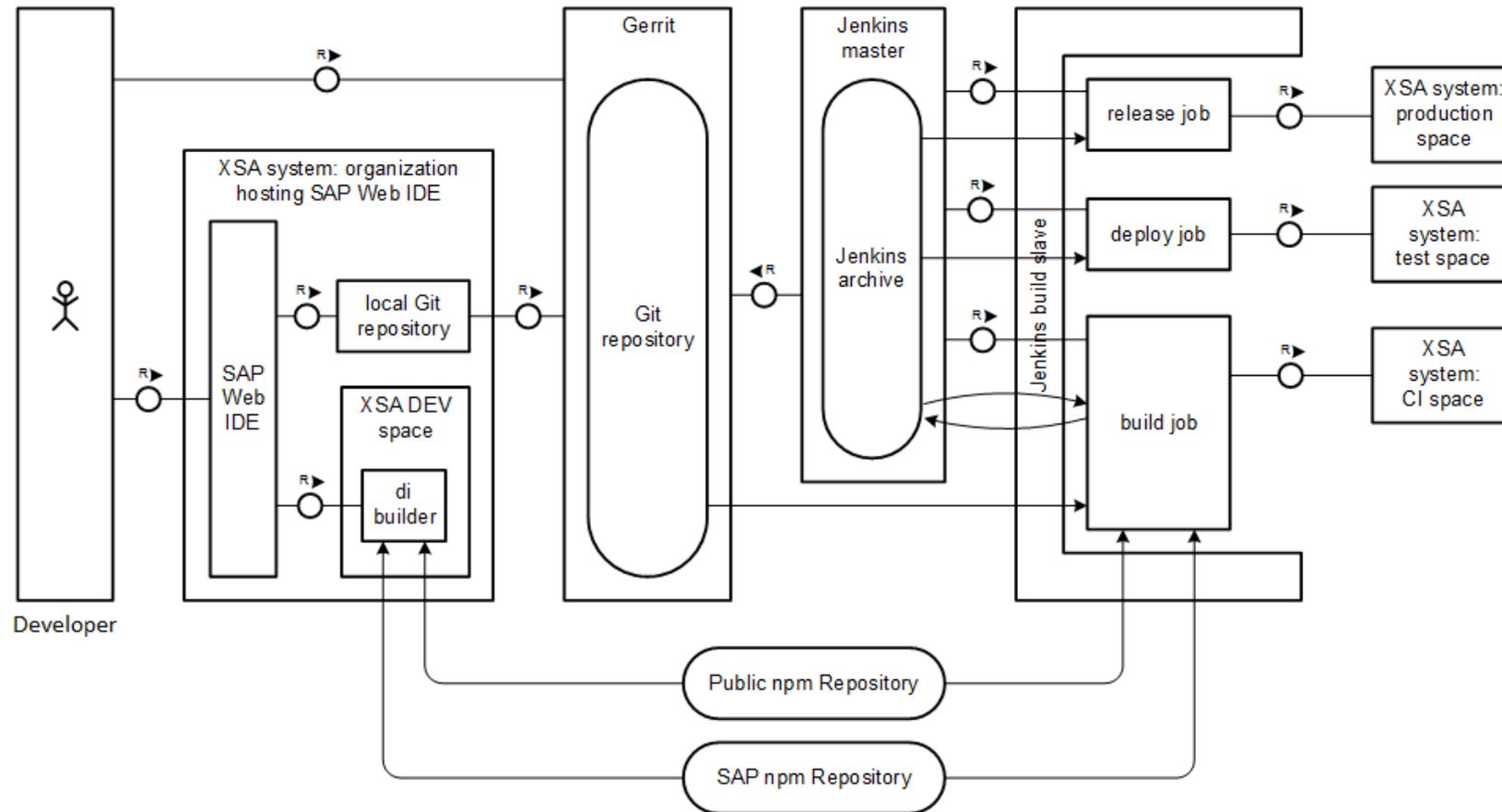


Тезис 4: HANA – универсальная платформа для разработки



- На базе Cloud Foundry (BYOL билдпаки)
- Близко к данным
- МТА-приложения
- Драйверы для некоторых языков (python, go, ruby и т.д.)

CI/CD* для HANA



*Continuous Integration/Continuous Delivery

Так что такое HANA?

o



[Hasso Plattner](#) on August 6, 2013

the hana team is maxdb+trex+ptime+sybase, but every phd student at the hpi in potsdam can tell you what is new in hana: delta-, main-, history-storage architecture, oltp without locks, parallel use of all available cores on a node for a single query, multi node operation, hana inside libraries for business functions, genomics, predictive analytics, business planning, text processing, etc. real time replication, multi tenancy, object aware caching etc in preparation. if you want to learn more i recommend my sigmod paper from 2009 or all the other papers, which followed by students of many universities.

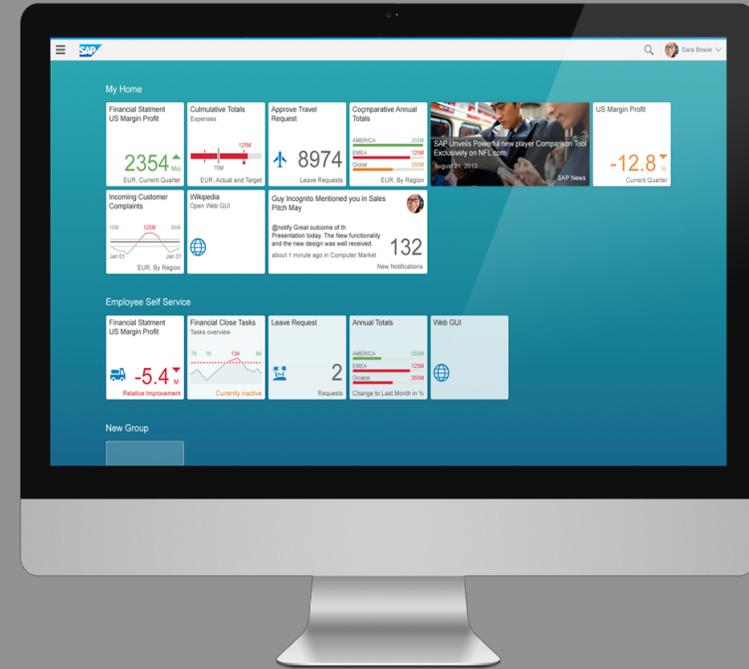
[Log in to Reply](#) | [Like \(1\)](#)

SAP HANA, Express Edition?

Бесплатная, доступная для скачивания, готовая к работе

SAP HANA, express edition -

- Модернизированная, упрощенная версия SAP HANA, оптимизированная под **быструю** и непрерывную разработку приложений;
- Предоставляется **бесплатно** для разработки (до 32Гб памяти)
- Доступна для скачивания на ноутбук, ПК, сервер, а также в облаке для опций **гибкой разработки и развертывания**.
- Предоставляется с **преднастроенной** виртуальной машиной, или **нативным** бинарным дистрибутивом Linux



Узнать больше: sap.com/developer/topics/sap-hana-express.html

Вопросы?

anton.nozdrin@sap.com