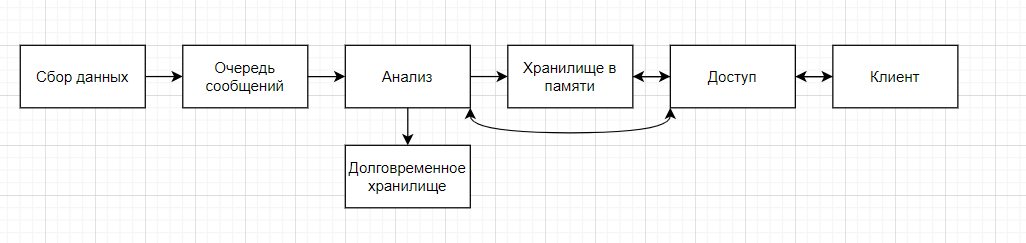
Бизнес-процесс

Потоковая обработка позволяет строить системы “мягкого реального времени”. Общая архитектура потоковой системы приведена на диаграмме:



На рынке существует немало технологий для обработки потоков. Из популярных продуктов с открытым исходным кодом можно отметить:

* Spark Streaming
* Storm
* Flink
* Samza

Все перечисленные проекты являются проектами фонда Apache.

Слой приёма ("заглатывания") сообщений обеспечивает поступление, буферизацию, предобработку и фильтрацию сообщений. Обычно, имеет ограничения в части длительности хранения (часы, дни) и ограниченный набор функций доступа к данным без возможности произвольного доступа. Можно использовать Kafka, Flume, Rabbit MQ, Active MQ, Pravega, Ignite.

Слой долговременного хранения предназначен для хранения данных после окончания обработки. Может быть использован для аналитики, пакетно-ориентированной обработки данных. Обеспечивает гарантированное хранение. Принципиально не предназначен для обслуживания запросов длительностью порядка секунды. Можно использовать HDFS, Cassandra, HBASE, KUDU, Druid, Elasticsearch.

Слой потоковой обработки использует сообщения, накопленных Ingestion-слоем. Обеспечивает обработку в режиме реального времени с гарантией сохранения данных.

Аномалии обычно относят к одному из трёх типов:

* Точечные аномалии
* Контекстные аномалии
* Коллективные аномалии

**Точечные аномалии**

возникают в ситуации, когда отдельный экземпляр данных может рассматриваться как безусловно аномальный по отношению к остальным.

**Контекстные аномалии**

наблюдаются, если экземпляр аномальный в определённом контексте, или при выполнении определённого условия (поэтому также называются условными).

**Коллективные аномалии**

возникают, когда последовательность связанных экземпляров данных (например, участок временного ряда) является аномальной по отношению к остальным данных.

Отдельный экземпляр может не являться отклонением, однако совместное появление таких экземпляров будет коллективной аномалией.

Аномалии можно хранить с особой меткой и далее выполнять поиск и анализировать уже эту метку.