Medium + Eng Blog

Dara Store

• Presto at Pinterest

https://medium.com/pinterest-engineering/presto-at-pinterest-a8bda7515e52

Data Lake

• Data lake at Robinhood

https://robinhood.engineering/data-lake-at-robinhood-3e9cdf963368

• Data warehouse vs data lake

https://medium.com/xplenty-blog/etl-vs-elt-or-data-warehouse-vs-data-lake-44a6a008de25

https://medium.com/readwrite/data-lake-vs-data-warehouse-which-is-the-best-data-architecture-ebda7192f99f

Data Streaming

https://medium.com/@lewisdgavin/what-is-real-time-data-37331ff91704

https://medium.com/pinterest-engineering/scalable-and-reliable-data-ingestion-at-pinterest-b921c2ee8754

Streaming Data Platform

https://medium.com/pinterest-engineering/unified-flink-source-at-pinterest-streaming-data-processing-c9d4e89f2ed6

https://netflixtechblog.com/keystone-real-time-stream-processing-platform-a3ee651812a

Spark

https://medium.com/explore-artificial-intelligence/introduction-to-core-spark-concepts-1e4ec879b9af

https://medium.com/explore-artificial-intelligence/sparks-structured-api-f3a5cef32955

Pinot(distributed, online analytical processing system)

• Pinot at Uber

https://eng.uber.com/operating-apache-pinot/

• Distributed, Online Analytical Processing (OLAP) system

• Use cases: dashboards, analytical applications, near real-time exploration

• Pinot platform at Uber: Ingestion / Storage / Query

• Ingestion from Kafka (together with FlinkSQL)

• Ingestion from Offline source (hive query -> pinot segments)

• Storage components: Pinot Cluster + Kafka + HDFS

• Pinot is multi-tenant

• Presto is also integrated with Pinot

Use case at Uber: Uber Eats restaurant manager: <https://eng.uber.com/restaurant-manager/>

A good example to use both real-time analytics & batch analytics: <https://eng.uber.com/analyzing-customer-user-experience/>

System Design

https://medium.com/swlh/vending-machine-design-a-state-design-pattern-approach-5b7e1a026cd2

Data Platform examples

https://medium.com/netflix-techblog/delta-a-data-synchronization-and-enrichment-platform-e82c36a79aee

https://medium.com/pinterest-engineering/building-a-real-time-anomaly-detection-system-for-time-series-at-pinterest-a833e6856ddd