

Cloudera Data Platform: Designing Edge to AI Applications on Hybrid Data Cloud

Introduction

- Team activity: Team Introductions

Technology Review

- HDFS
- HBase
- Kudu
- Map Reduce
- Spark, including SparkSQL and SparkML
- Hive
- Impala
- Relational Database Management Systems
- Spark streaming
- Apache Flume
- Apache NiFi
- Apache Kafka

Workshop Application Use Cases

- Oz Metropolitan
- Architectural questions
- Team activity: Review Metroz Use Cases and Logical Architecture

Application Vertical Slice

- Definition
- Minimizing risk of an unsound architecture
- Selecting a vertical slice
- Team activity: Metroz Vertical Slice

Application Processing

- Real time, near real time processing
- Batch processing

- Data access patterns
- Delivery and processing guarantees
- Data consistency and ACID transactions
- Stream processing guarantees
- Machine Learning pipelines
- Team activity: Metroz Processing

Application Data

- Three V's of Big Data
- Data Lifecycle
- Data Formats
- Transforming Data
- Team activity: Metroz Data Requirements

Scalable Applications

- Scale up, scale out, scale to X
- Determining if an application will scale
- Poll: scalable airport terminal designs
- Spark scalability and parallel processing
- Scalable storage engines: HDFS, Ozone, Kafka and Kudu
- Team activity: Scaling Metroz

Fault-Tolerant Distributed Systems

- Principles
- Transparency
- Hardware vs. Software redundancy

- Tolerating disasters
- Stateless functional fault tolerance
- Stateful fault tolerance
- Replication and group consistency
- Application tolerance for failures

- Team activity: Failures in Metroz

Security and Privacy

- Principles
- Security Architecture
- Knox Security Architecture
- Ranger Security Architecture
- Setting security policies with Ranger
- Threat Analysis
- Team activity: Securing Metroz

Deployment

- Cluster sizing and evolution
- On-premise vs. Cloud
- Edge computing
- Team activity: Deploying Metroz

Software Architecture

- Architecture artifacts
- Team activity: Metroz Physical Architecture

Potential CDP Solutions

- Review of Uber and Lyft Big data platforms
- Review of Metroz CDP solution architectures