HDFS

Wiederholung: Übersicht HDFS

HDFS: Ziele und Annahmen

- Streaming Data Access
- Hardware Failure
- Large Data Sets
- Simple Coherency Model
- “Moving Computation is Cheaper than Moving Data”
- Portability Across Heterogeneous Hardware and Software Platforms

HDFS - Eigenschaften

- *data locality*
- *large block size*
- *data replication*
- *fault tolerance* gegenüber Software und Hardware Fehler

Name Nodes

- Single Master
- „Volume Manager“ von HDFS
  - In-Memory mapping der Files zu den Blocks
  - d.h. für jeden Block:
    - auf welchen Nodes ist dieser verfügbar
  - Clients fragen Name Node ab, wenn sie *File System Operations* ausführen wollen, dann direkter Datentransfer zwischen Client und Datanode
- Eigentlichen Daten laufen nie über NameNodes
- Single Point of Failure - außer in “HDFS High-Availability”

Data Nodes

- Verantwortlich für das Speichern der File Blocks
- Pipeline Write möglich
- Kommunikation mit Name Node

HDFS Client

- Anwendung, die eine File System Aktivität ausführen will, wie File schreiben, löschen etc.
- Anfrage an Name Node um Informationen zu erhalten
- direkte Interaktion mit Data Nodes, z.B. für Datei-Lesen und -Schreiben

File System Namespace

- Hierarchisch, wie in anderen Dateisystemen
- Bisher: Keine Hard- oder Softlinks
- Verantwortlich: Name Node

Data Replication - Ziele

- Reliability (Ausfallsicherheit)
- Availability (Verfügbarkeit)
- Performance (Performanz)

Data Replication

- Files bestehen aus Block-Sequenzen
- Blocks werden repliziert
- Replication Factor: Anzahl der Kopien eines Files
- Files sind write-ones
- Name Nodes entscheiden über die Replikation der Blocks
- Name Nodes erhalten Heartbeats und Blockreports von jedem Data Node

Secondary NameNode

- kein wirklicher zweiter NameNode
- sondern zur Entlastung des NameNodes
  - NameNode speicher alle Metadaten in *FSImage.* Dieses wird beim Starten in den Hauptspeicher geladen und beim Herunterfahen persistiert.
  - Um den NameNode zu Entlasten alle Änderungen im FSImage festzuhalten, Erzeugt der seondary NameNode eine Edit-Log des *FSImage*

Dateisystemschnittstellen

- FUSE (Filesystem im Userspace)
  - zum Mounten von Dateisystemen – normalerweise im Systemmodus
  - HDSF Implementierung
- WebHDFS: REST-basierter Zugriff (HTTP)
- Java und C-API (libhdfs)
- Thrift für Zugriff mit anderen Programmiersprachen

Rechtemanagement

Quotas

Dateisystem Shell

Admin Shell

Literatur

- Tom White, „Hadoop The Definite Guide“, third edition, 2012, O'Reilly